

ECONOMIC RESEARCH REPORTS

Cooperative vs. Non-Cooperative Truels: Little Agreement, But Does that matter?

by

**Walter Bossert, Steven J. Brams,
&
D. Marc Kilgour**

RR#: 2000-19

October 2000



C.V. Starr Center for Applied Economics

Department of Economics

Faculty of Arts and Science

New York University

269 Mercer Street, 3rd Floor

New York, New York 10003-6687

Cooperative vs. Non-Cooperative Truels: Little Agreement, But Does That Matter?*

WALTER BOSSERT

Département de Sciences Economiques and C.R.D.E., Université de Montréal
C.P. 6128, succursale Centre-ville, Montréal QC H3C 3J7, Canada

STEVEN J. BRAMS

Department of Politics, New York University
New York, NY 10003, USA

D. MARC KILGOUR

Department of Mathematics, Wilfrid Laurier University
Waterloo, ON N2L 3C5, Canada

This version: October 2000. This copy: October 26, 2000.

* Walter Bossert and D. Marc Kilgour gratefully acknowledge the support of the Social Sciences and Humanities Research Council of Canada, and Steven J. Brams the support of the C.V. Starr Center for Applied Economics at New York University.

Cooperative vs. Non-Cooperative Truels: Little Agreement, But Does That Matter?

Abstract

It is well-known that non-cooperative and cooperative game theory may yield different solutions to games. These differences are particularly dramatic in the case of truels, or three-person duels, in which the players may fire sequentially or simultaneously, and the games may be one-round or n -round. Thus, it is never a Nash equilibrium for all players to hold their fire in any of these games, whereas in simultaneous one-round and n -round truels such cooperation, wherein everybody survives, is in both the α -core and β -core. On the other hand, both cores may be empty, indicating a lack of stability, when the unique Nash equilibrium is one survivor.

Conditions under which each approach seems most applicable are discussed. Although it might be desirable to subsume the two approaches within a unified framework, such unification seems unlikely since the two approaches are grounded in fundamentally different notions of stability.

JEL Classification: C71, C72, D74

Keywords: Cooperative game, non-cooperative game, truel, Nash Equilibrium, core

1 Introduction

There has long been a tension in game theory between cooperative and non-cooperative analysis. Although five-sixths of von Neumann and Morgenstern (1953) is devoted to cooperative game theory, it is non-cooperative game theory, pioneered by Nash (1951) and epitomized by the Nash equilibrium, that has held sway for at least a generation, largely due to the influence of economists and the strategic models they have developed in microeconomics, industrial organization, and other fields.

In recent years, cooperative models have made somewhat of a comeback, in part because Nash equilibria have not proved satisfactory as a solution concept in some fields. To be sure, numerous refinements have been made in this concept, the most important being subgame perfection that we, in fact, will use here, but there has not been a consensus on which refinements are most useful, and in what strategic situations. Other notions of equilibrium, less myopic than Nash and grounded in theories that propose very different conceptions of a game and its rules of play (e.g., Greenberg, 1990; Brams, 1994), have proved useful but have not met with widespread acceptance.

One advantage of cooperative theory is that it does not require a detailed specification of the moves of players, the sequence in which they are made, who knows what about whom and when, and so on. This descriptive detail is frequently missing in situations that we might want to model. But even when it is available, conclusions that can be derived about optimal play are often highly game-specific and non-robust: small changes in the strategy set, sequencing, or information conditions can lead to entirely different outcomes.

This is not to say, however, that non-cooperative theorists have failed to provide quite general answers to certain strategic questions. Indeed, in this paper we will describe a class of three-person games and draw general conclusions about non-cooperative behavior in them—for example, that players will never refrain, entirely, from shooting at their opponents. On the other hand, from the cooperative perspective this conclusion is not supported, because there are games in this class in which it is rational for all the players to exercise self-restraint.

One purpose of this paper is to highlight differences between the solutions that the cooperative and non-cooperative approaches yield in three-person games. The cooperative approach, which we operationalize in terms of the α -core and the β -core, makes self-restraint a rational choice by showing that players can, on occasion, guarantee themselves more—or cannot be prevented from getting more—by agreeing not to shoot, *provided this*

agreement is enforceable. But it is precisely this presumption of enforceability that the non-cooperative approach questions, rendering shooting rational in *all* the games.

In fact, even when the cooperative and non-cooperative approaches both prescribe shooting, the outcomes may be different. Moreover, *within* the cooperative approach, similar solution concepts can yield different outcomes.

For the “grand unification” theorists of game theory, this divergence may be unsettling if not dismaying. But we take a more sanguine view, invoking an analogy from physics. Just as the particle and wave interpretations of light have not proved entirely reconcilable, there is no good reason to expect a complete reconciliation of the cooperative and non-cooperative approaches in game theory.

The two interpretations of light have, nevertheless, proved useful in explaining different observable physical phenomena. Likewise, the cooperative and non-cooperative approaches in game theory offer insight into possibly different outcomes to expect in games when they are viewed from the two different perspectives.

Just as most physicists have lived quite well with the different interpretations of light, we should not despair at this lack of intellectual unity in game theory. Perhaps someday there will be an overarching theory that subsumes the cooperative and non-cooperative approaches within a single compelling framework. One possibility is Greenberg’s (1990) theory of social situations. However, we believe a good pragmatic stance to adopt is that of asking which approach works best in what situations. We will offer some thoughts on this subject in the concluding section.

We next describe the games we will analyze, after which we will compare the solutions each approach prescribes. These games are among the simplest 3-person (and therefore n -person) games one can imagine in the sense that (1) the players have only three qualitatively different choices and (2) they are the same choices for each player. On the other hand, simple as these games are, they give rise to a rich menu of outcomes from both the cooperative and non-cooperative perspectives, rendering comparisons between the two perspectives challenging.

The games we analyze, called *truels*, extend duels, or two-person shooting matches, to three players. The shooting goes in rounds, whereby in any round each surviving player can shoot at another survivor or fire into the air (i.e., not fire in that round). The game ends when either (i) a pre-specified number of rounds has been played or (ii) there is at most one survivor.

Several versions of truels have been discussed in the literature: see Kilgour and Brams (1997) for a survey. In this paper, we analyze *sequential* truels, in which players fire one

at a time in a prespecified order that is fixed throughout the game, and *simultaneous* truels, in which all surviving players fire simultaneously in each round (but possibly into the air). Other variations found in the literature include truels with random firing order, truels that end probabilistically, and truels in which firing into the air is prohibited.

Among the truels discussed by Kilgour and Brams (1997) are those in which players are perfect shots—that is, a player’s target is killed with certainty. In such truels, players’ preferences depend only on the *outcome* of the truel, or who survives.

Outcomes are evaluated lexicographically, in the order given below, according to three goals:

1. A player most prefers to survive.
2. The fewer the opponents who survive, the better the outcome for a player.
3. If exactly one opponent survives, a player prefers one opponent to the other.

The third goal is implemented by identifying, for each player, which of its opponents is its *antagonist*. Its other opponent, which it would rather see survive, is its *non-antagonist*.

After defining terms and fixing notation in section 2, we derive in section 3 the subgame-perfect and other Nash equilibria for sequential and simultaneous truels with n rounds, where $1 \leq n < \infty$. In section 4, we turn to the cooperative analysis, giving the α - and β -cores of the truels. Finally, we compare the results of the cooperative and non-cooperative analyses in section 5, drawing conclusions about the appropriateness of each kind of analysis in different situations.

2 Truels: Definitions and Notation

The set of players of a truel is $N = \{1, 2, 3\}$, and 2^N is the set of all non-empty subsets of N . For $S \in 2^N$, \mathbb{R}^S denotes an $|S|$ -dimensional Euclidean space with the coordinates of points indexed by the members of S .

We use $\Gamma^q(n)$, and $\Gamma^m(n)$ to denote n -round sequential and simultaneous truels, respectively. Sequential truels are assumed to have the fixed firing order $\langle 1, 2, 3, 1, 2, \dots \rangle$. Let $t \in \{q, m\}$. The set of strategies for player $i \in N$ in game $\Gamma^t(n)$ is $\Sigma_i^t(n)$. For example, the set of strategies of player 1 in a one-round simultaneous truel is $\Sigma_1^m(1) = \{\text{shoot at player 2, shoot at player 3, shoot into the air}\}$. (These strategies for player 1 will be denoted $1 \rightarrow 2$, $1 \rightarrow 3$, and $1 \uparrow$, respectively.) The set of all strategy profiles in game Γ^t

is $\Sigma^t = \Sigma_1^t \times \Sigma_2^t \times \Sigma_3^t$. For $S \in 2^N$, let $\Sigma_S^t = \times_{i \in S} \Sigma_i^t$. Dependence on n will be suppressed when no confusion can arise.

An outcome of a truel is a set of survivors, that is, a subset of N . We denote outcomes by $[\emptyset]$, $[1]$, $[1, 2]$, etc., to distinguish them from subsets of players, or coalitions, which we denote using braces instead of brackets. Thus, the set of all conceivable outcomes is

$$X = \{[\emptyset], [1], [2], [3], [1, 2], [1, 3], [2, 3], [1, 2, 3]\}.$$

The outcome function for the game $\Gamma^t(n)$ is given by $f_n^t: \Sigma^t(n) \rightarrow X$. Clearly, $f_1^q(\Sigma^q(1)) = X \setminus \{[\emptyset], [1]\}$, $f_n^q(\Sigma^q(n)) = X \setminus \{[\emptyset]\}$ if $n > 1$, and $f_n^m(\Sigma^m(n)) = X$ for all n .

To complete the specifications of these truels, we must define players' preferences. For each player $i \in N$, let $a(i) \in N \setminus \{i\}$ denote i 's *antagonist*, the opponent i dislikes more than its *non-antagonist* opponent, $na(i) \in N$. Note that $N = \{i, a(i), na(i)\}$.

As indicated in section 1, i prefers any outcome wherein it survives to any outcome wherein it does not (goal 1). Comparing two outcomes wherein i survives at both or does not survive at both, i prefers the one wherein fewer of its opponents survive (goal 2). If exactly one opponent survives at both outcomes (and i survives, or does not survive, at both), i prefers the outcome wherein its non-antagonist, $na(i)$, survives to the outcome where its antagonist, $a(i)$, survives (goal 3). Therefore, player i 's antisymmetric preference ordering \succ_i on X is given by

$$[i] \succ_i [i, na(i)] \succ_i [i, a(i)] \succ_i [1, 2, 3] \succ_i [\emptyset] \succ_i [na(i)] \succ_i [a(i)] \succ_i [na(i), a(i)].$$

Although we work in a purely ordinal framework, it will be convenient later to represent individual preferences using utility. For all $i \in N$, define, without loss of generality, $U_i: X \rightarrow \mathbb{R}$ by letting

$$\begin{aligned} U_i([i]) &= 8, U_i([i, na(i)]) = 7, U_i([i, a(i)]) = 6, U_i([1, 2, 3]) = 5, \\ U_i([\emptyset]) &= 4, U_i([na(i)]) = 3, U_i([a(i)]) = 2, U_i([na(i), a(i)]) = 1. \end{aligned}$$

3 Non-Cooperative Analysis: Equilibria

In most of the literature (see, for example, Brams and Kilgour, 1998, 2000, Kilgour, 1972, 1975, 1978), truels are analyzed as non-cooperative games. The non-cooperative solution concepts we use for truels are the following: For sequential truels, we identify subgame-perfect equilibria, because these are games of perfect information. For the $n = 1$ round simultaneous truel, we identify the Nash equilibria as the non-cooperative solutions,

because these are strategic- (or normal-) form games. For simultaneous truels with $n > 1$ rounds, we identify the Nash equilibria that are subgame-perfect in the sense that the behavior called for in any subsequent round is an equilibrium in the induced subgame.

One-Round Sequential Truel Kilgour and Brams (1997) show that there is a unique subgame-perfect equilibrium in the one-round sequential truel, with outcome

$$E_1^q = \begin{cases} [na(1)] & \text{if } a(2) = a(3) = 1 \\ [1, 2] & \text{otherwise} \end{cases}$$

The strategies at equilibrium can be described as follows:

1. If $a(3) = 2$ and all three players survive until 3's turn (because 1 and 2 do not fire at another player), 3 will eliminate 2 producing outcome $[1, 3]$. But 2 prefers $[1, 2]$ to this outcome and can achieve it by firing earlier at 3, in which case 1 can do no better than fire into the air.
2. If $a(3) = 1$ and $a(2) = 3$, 2 prefers to implement outcome $[1, 2]$ by eliminating 3, rather than $[2, 3]$ by firing into the air. Again, 1 will fire into the air.
3. If $a(2) = a(3) = 1$, 3 will implement $[2, 3]$ if both 1 and 2 fire into the air; 2 prefers this outcome to either $[1, 2]$ or $[3]$, which it can achieve by firing at 3 or 1, respectively. Thus 1, faced with the choice of $[a(1)]$, $[na(1)]$, or $[2, 3]$, prefers $[na(1)]$, so it eliminates $a(1)$, and then $na(1)$ eliminates 1.

Thus, in the one-round sequential truel, *there is either one survivor or two*.

n -Round Sequential Truel, $n > 1$ Kilgour and Brams (1997) show that the unique subgame-perfect equilibrium of the sequential truel with $n > 1$ rounds yields

$$E_n^q = \begin{cases} [3] & \text{if } a(1) = 2, a(2) = 3, a(3) = 1, \text{ and } n \text{ even} \\ [1] & \text{if } a(1) = 2, a(2) = 3, a(3) = 1, \text{ and } n \text{ odd} \\ [2] & \text{if } a(1) = 3, a(3) = 2, a(2) = 1 \\ [i] & \text{if for some } i \in N, na(j) = i \text{ for all } j \in N \setminus \{i\} \end{cases}$$

A player is a *common non-antagonist* if and only if it is the non-antagonist of each of its two opponents. There can be at most one common non-antagonist. The sole survivor of the n -round sequential truel, when $n > 1$, is the common non-antagonist, if there is one. Otherwise, the identity of the unique surviving player depends on the particular cyclic pattern of antagonisms, and sometimes on the parity of n .

For descriptions of the equilibrium strategies, see Kilgour and Brams (1997) or Kilgour (1975). Whatever the length of the truel, all firing occurs in the first or second round, and *there is exactly one survivor*.

One-Round Simultaneous Truel The one-round simultaneous truel is a strategic-form game whose game matrix is shown below. Note that the three 3×3 matrices correspond to player 3's strategies of shoot at 1, shoot at 2, and shoot into the air, respectively. Within each matrix, the rows correspond to player 1's strategies, and the columns to player 2's strategies.

		<u>3 → 1</u>						
		2 → 3	2 → 1	2 ↑				
1 → 2		[∅]*	[3]	[3]				
1 → 3		[2]	[2]	[2]* ²				
1 ↑		[2]	[2, 3]	[2, 3]				

		<u>3 → 2</u>						
		2 → 3	2 → 1	2 ↑				
1 → 2		[1]	[3]	[1, 3]				
1 → 3		[1]	[∅]*	[1]				
1 ↑		[1]* ¹	[3]	[1, 3]				

		<u>3 ↑</u>						
		2 → 3	2 → 1	2 ↑				
1 → 2		[1]	[3]* ³	[1, 3]				
1 → 3		[1, 2]	[2]	[1, 2]				
1 ↑		[1, 2]	[2, 3]	[1, 2, 3]				

A Nash equilibrium is a strategy profile, associated with a cell, such that no player can obtain a preferred outcome by unilaterally changing its strategy. The Nash equilibria of the one-round simultaneous truel are indicated by asterisks in the game matrix, interpreted as follows:

- * Always an equilibrium.
- *¹ An equilibrium if and only if 1 is common non-antagonist.
- *² An equilibrium if and only if 2 is common non-antagonist.
- *³ An equilibrium if and only if 3 is common non-antagonist.

Notice that, if there is no common non-antagonist, then there are no survivors in equilibrium, and the strategies are one of the cyclic firing patterns, $1 \rightarrow 2 \rightarrow 3 \rightarrow 1$ or $1 \rightarrow 3 \rightarrow 2 \rightarrow 1$. These cyclic firing patterns continue to be equilibria if there is a common non-antagonist, but another equilibrium emerges as well: the common non-antagonist fires into the air, and the other two players eliminate each other. Thus, in the one-round simultaneous truel, *there is either one survivor or none*.

n -Round Simultaneous Truel, $n > 1$ To analyze the n -round simultaneous truel, recall our assumption that, should all three players survive the first round, their strategies in the resulting subgame must be subgame-perfect equilibria in the $(n - 1)$ -round simultaneous truel. To represent strategies and outcomes in any round up to the last, the game matrix for the one-round simultaneous truel must be modified in two ways.

First, when there are exactly two survivors of the k th round prior to the end, those survivors play a $(k - 1)$ -round simultaneous duel. It is easy to verify that the strategy of firing at the opponent in every round is dominant for a player in such a duel so that, provided $k \geq 2$, the outcome of the $(k - 1)$ -round simultaneous duel is $[\emptyset]$. (In fact, such a duel is a Prisoners' Dilemma.) Thus, each of the six outcomes with two survivors in the matrix of the one-round simultaneous truel must be replaced by the outcome $[\emptyset]$. Second, if all players fire into the air in the k th round prior to the end, the outcome is an equilibrium outcome of the $(k - 1)$ -round simultaneous truel, which we denote for now as " x ."

		<u>3 → 1</u>			<u>3 → 2</u>			<u>3 ↑</u>		
		2 → 3	2 → 1	2 ↑	2 → 3	2 → 1	2 ↑	2 → 3	2 → 1	2 ↑
1 → 2		$[\emptyset]^*$	[3]	[3]	[1]	[3]	$[\emptyset]$	[1]	[3]	\emptyset
1 → 3		[2]	[2]	[2]	[1]	$[\emptyset]^*$	[1]	$[\emptyset]$	[2]	$[\emptyset]$
1 ↑		[2]	$[\emptyset]$	$[\emptyset]$	[1]	[3]	$[\emptyset]$	$[\emptyset]$	$[\emptyset]$	x

Observe that the two cyclic firing patterns resulting in $[\emptyset]$ continue to be equilibria in the first round of any n -round simultaneous truel. It is easy to verify that none of the other outcomes, with the possible exception of x , can possibly be equilibria.

Assume that $n = 2$, and recall that the equilibria of the one-round simultaneous truel have as outcome either $x = [\emptyset]$ or $x = [i]$, where $i \in N$ is the common non-antagonist. It is easy to see that $x = [i]$ cannot correspond to an equilibrium, because any player other than i prefers $[\emptyset]$; moreover, this player can achieve $[\emptyset]$ by firing at either opponent in the first round. The only possibility for an equilibrium that involves all players firing into the air in the first round is $x = [\emptyset]$, but the players can achieve $[\emptyset]$ in equilibrium by firing in the first round.

The situation with $n > 2$ is similar. We conclude that, if $n > 1$, the n -round simultaneous truel *has no survivors*. This is the unique equilibrium outcome. While the most plausible equilibria are those involving a cyclic firing pattern in the first round, this may

not occur in the absence of communication and coordination, a point we will return to in the concluding section.

In summary, whatever the nature of the truel—sequential or simultaneous, one-round or n -round—at least one player shoots another in some round. Consequently, it is never the case that all players survive in equilibrium.

4 Cooperative Analysis: Cores

The core is probably the best-known cooperative game solution concept. It is usually defined in terms of von Neumann–Morgenstern utility functions, which we will use on occasion in the subsequent analysis. However, only ordinal information about the players' preferences for the possible outcomes is essential in what follows.

In games without transferable utility, several procedures for finding characteristic functions, each leading to a different notion of the core, have been suggested; see Aumann (1961, 1967) and Aumann and Peleg (1960). Following this literature, we find the α - and β -characteristic functions and then use them to find α -cores and β -cores for sequential and simultaneous truels. It turns out that the α -core and the β -core coincide for sequential truels, but they may be different for simultaneous truels.

The characteristic function $v: 2^N \rightarrow \cup_{S \in 2^N} 2^{\mathbb{R}^S}$ of a game assigns a nonempty set of payoff vectors to each coalition S such that $v(S) \in 2^{\mathbb{R}^S}$ for all $S \in 2^N$. Loosely speaking, the α -characteristic function, $v_\alpha(\cdot)$, is the set of payoff vectors that a coalition can guarantee its members, whereas the β -characteristic function, $v_\beta(\cdot)$, is the set of payoff vectors that the complement of the coalition cannot prevent the members of the coalition from receiving.

Let $t \in \{g, m\}$. For $u \in \mathbb{R}^N$, it is clear that $u \in v_\alpha^t(N) = v_\beta^t(N)$ if and only if there exists $\sigma^t \in \Sigma^t$ such that $U_i(f^t(\sigma^t)) \geq u_i$ for all $i \in N$.

Similarly, for a coalition $S \in 2^N$ and a payoff vector $u \in \mathbb{R}^S$, $u \in v_\alpha^t(S)$ if and only if there exists $\sigma_S^t \in \Sigma_S^t$ such that $U_i(f^t(\sigma_S^t, \sigma_{N \setminus S}^t)) \geq u_i$ for all $i \in S$ and for all $\sigma_{N \setminus S}^t \in \Sigma_{N \setminus S}^t$. Analogously, $u \in v_\beta^t(S)$ if and only if, for each $\sigma_{N \setminus S}^t \in \Sigma_{N \setminus S}^t$, there exists $\sigma_S^t \in \Sigma_S^t$ such that $U_i(f^t(\sigma_S^t, \sigma_{N \setminus S}^t)) \geq u_i$ for all $i \in S$.

An outcome $x \in f^t(\Sigma^t)$ is in the α -core of the game Γ^t if and only if there exists no (S, u) , with $S \in 2^N$ and $u \in v_\alpha^t(S)$, such that $u_i > U_i(x)$ for all $i \in S$. Analogously, $x \in f^t(\Sigma^t)$ is in the β -core of Γ^t if and only if there exists no (S, u) with $S \in 2^N$ and $u \in v_\beta^t(S)$ such that $u_i > U_i(x)$ for all $i \in S$. We say that a coalition $S \in 2^N$ α -blocks (respectively, β -blocks) an outcome $x \in f^t(\Sigma^t)$ if and only if there exists $u \in v_\alpha^t(S)$ (respectively, $v_\beta^t(S)$) such that $u_i > U_i(x)$ for all $i \in S$.

(respectively, $u \in v_\beta^t(S)$) such that $u_i > U_i(x)$ for all $i \in S$.

For $t \in \{q, m\}$ and an n -round truel, the α -core and the β -core of $\Gamma^t(n)$ are denoted by $C_\alpha^t(n)$ and $C_\beta^t(n)$, respectively. Again, the dependence on the number of rounds is suppressed when no ambiguities can arise. Note that our results would be unaffected by the “stronger” core definitions that replace “ $u_i > U_i(x)$ for all $i \in S$ ” by “ $u_i \geq U_i(x)$ for all $i \in S$ with at least one strict inequality.” For $t \in \{q, m\}$, it is clear that $v_\alpha^t(S) \subseteq v_\beta^t(S)$ for all $S \in 2^N$. Consequently,

$$C_\beta^q \subseteq C_\alpha^q \text{ and } C_\beta^m \subseteq C_\alpha^m. \quad (1)$$

One-Round Sequential Truel Theorem 1 shows that the one-round sequential truel, $\Gamma^q(1)$, has a unique core outcome, which is identical for the two core definitions.

Theorem 1 $C_\alpha^q(1) = C_\beta^q(1) = \{[1, na(1)]\}$.

Proof: By (1), it is sufficient to show that

- (i) $[1, na(1)]$ belongs to $C_\beta^q(1)$;
- (ii) no other outcome in $f_1^q(\Sigma^q(1))$ is in $C_\alpha^q(1)$.

(i) First, note that no coalition containing player 1 will block $[1, na(1)]$, because this is player 1’s most preferred outcome in $f_1^q(\Sigma^q(1))$. To show that the coalition $\{2\}$ cannot β -block $[1, na(1)]$, consider the strategy pair $(\sigma_1^q, \sigma_3^q) = (1 \rightarrow 2, 3 \uparrow) \in \Sigma_{\{1,3\}}^q(1)$ in which player 1 shoots player 2 and player 3 shoots in the air. The outcome is $[1, 3]$, the worst possible outcome for player 2, independent of player 2’s strategy choice. Hence, if the complement of $\{2\}$ chooses $(1 \rightarrow 2, 3 \uparrow)$, there is no strategy for player 2 that produces an outcome it prefers to $[1, na(1)]$. Thus $\{2\}$ cannot block $[1, na(1)]$ according to the β -core definition. That $\{3\}$ cannot block this outcome is shown analogously.

Finally, consider the coalition $\{2, 3\}$. Let $\sigma_1^q = (1 \rightarrow na(1)) \in \Sigma_1^q(1)$ be the strategy in which player 1 shoots $na(1)$. Independent of the strategy choice of the coalition $\{2, 3\}$, the resulting outcome is less preferred than $[1, na(1)]$ for player $na(1)$; thus, coalition $\{2, 3\}$ cannot block either. We have shown that $[1, na(1)]$ is in $C_\beta^q(1)$ and, therefore, in $C_\alpha^q(1)$.

(ii) The outcome $[2]$ is not in the α -core because it is blocked by coalition $\{1, 3\}$. To see this, let $(\sigma_1^q, \sigma_3^q) = (1 \rightarrow 2, 3 \uparrow) \in \Sigma_{\{1,3\}}^q(1)$. For any strategy choice of player 2, the resulting outcome is $[1, 3]$, which is preferred to $[2]$ by both coalition members. That $[3]$ cannot be in the α -core is shown analogously.

To show that the outcome $[1, a(1)]$ is blocked by the coalition $\{1, na(1)\}$, consider a strategy pair $(\sigma_1^q, \sigma_{na(1)}^q) = (1 \rightarrow a(1), na(1) \uparrow) \in \Sigma_{\{1, na(1)\}}^q(1)$, where player 1 shoots player $a(1)$ and player $na(1)$ shoots into the air. The resulting outcome is $[1, na(1)]$ for all strategy choices of player $a(1)$. Because this outcome is preferred to $[1, a(1)]$ by players 1 and $na(1)$, $[1, a(1)]$ cannot be in the α -core.

Also, the outcome $[2, 3]$ is blocked by the coalition $\{1\}$, as can be seen by choosing $\sigma_1^q = (1 \rightarrow 2) \in \Sigma_1^q(1)$, i.e., player 1 shoots player 2. Because player 1 can guarantee itself an outcome preferred to $[2, 3]$ (its least preferred outcome) independent of $\{2, 3\}$'s strategy choice, outcome $[2, 3]$ is not in the α -core.

Finally, the outcome $[1, 2, 3]$, in which all players survive, can be blocked by the coalition $\{1, 2\}$. Choose $(\sigma_1^q, \sigma_2^q) = (1 \rightarrow 3, 2 \uparrow) \in \Sigma_{\{1, 2\}}^q(1)$ such that player 1 shoots player 3 and player 2 shoots in the air. The outcome is $[1, 2]$, independent of player 3's strategy choice. Because $[1, 2]$ is preferred to $[1, 2, 3]$ by both players 1 and 2, the members of the coalition $\{1, 2\}$ can guarantee themselves a more-preferred outcome, so $[1, 2, 3]$ is not in the α -core. ■

n -Round Sequential Truel, $n > 1$ Strategies are more complex in the n -round than in the one-round sequential truel because players can, for example, make their actions contingent on the history of play. As well, outcome $[1]$ can arise when there are two or more rounds and it is another possible core element. Nonetheless, much of the analysis of outcomes other than $[1, na(1)]$ and $[1]$ in Theorem 1 applies here as well, in part because the consequences of many different strategy choices can be summarized analogously in one-round and n -round sequential truels. For example, outcomes not in the core of $\Gamma^q(n)$ when $n = 1$ are also not in the core when $n > 1$.

The primary issues to be addressed in determining cores in the n -round sequential truel are (i) whether $[1, na(1)]$ survives in the α -core or the β -core of $\Gamma^q(n)$, and (ii) whether $[1]$ enters the α -core or β -core of $\Gamma^q(n)$. It turns out that the answer to (i) is “no,” and the answer to (ii) depends on the assignment of antagonists to players 2 and 3. If player 1 is the antagonist of both its opponents, the α -core and the β -core are empty, but if player 1 is the non-antagonist of at least one other player, the α -core and the β -core coincide, and contain only the element $[1]$, as shown in the following theorem.

Theorem 2 (a) $C_\alpha^q(n) = C_\beta^q(n) = \emptyset$ if and only if $n \geq 2$ and $a(2) = a(3) = 1$;

(b) $C_\alpha^q(n) = C_\beta^q(n) = \{[1]\}$ if and only if $n \geq 2$ and either $na(2) = 1$ or $na(3) = 1$.

Proof: We first prove that no outcome other than $[1]$ can be in the α -core of $\Gamma^q(n)$. By (1), this implies that no such outcome can be in the β -core either.

The outcome $[2]$ is not in the α -core because it is blocked by coalition $\{1, 3\}$. Let $(\sigma_1^q, \sigma_3^q) \in \Sigma_{\{1,3\}}^q(n)$ be such that player 1 shoots player 2 ($1 \rightarrow 2$) in round 1 and no further shots are fired in any round. For any strategy choice of player 2, the resulting outcome is $[1, 3]$, which is preferred to $[2]$ by both coalition members. That $[3]$ cannot be in the core is shown analogously.

The outcome $[1, a(1)]$ is blocked by the coalition $\{1, na(1)\}$. Consider a strategy pair $(\sigma_1^q, \sigma_{na(1)}^q) \in \Sigma_{\{1,na(1)\}}^q(n)$ such that player 1 shoots player $a(1)$ in the first round and no further shots are fired. The resulting outcome is $[1, na(1)]$ for all strategy choices of player $a(1)$. Because this outcome is preferred to $[1, a(1)]$ by players 1 and $na(1)$, $[1, a(1)]$ cannot be in the α -core.

The outcome $[1, na(1)]$ is blocked by the coalition $\{1, a(1)\}$. Consider a strategy pair $(\sigma_1^q, \sigma_{a(1)}^q) \in \Sigma_{\{1,a(1)\}}^q(n)$ such that player 1 shoots player $na(1)$ in round 1, player $a(1)$ shoots in the air in round 1, and player 1 shoots player $a(1)$ in round 2. The resulting outcome is $[1]$ for all strategy choices of player $na(1)$. Because this outcome is preferred to $[1, na(1)]$ by players 1 and $a(1)$, $[1, na(1)]$ cannot be in the α -core.

The outcome $[2, 3]$ is blocked by the coalition $\{1\}$. By choosing $\sigma_1^q \in \Sigma_1^q(n)$ such that player 1 shoots player 2 in round 1 and shoots in the air subsequently, player 1 can guarantee itself a higher utility than in outcome $[2, 3]$ (its worst possible outcome), independent of the strategy choice of $\{2, 3\}$. Therefore, $[2, 3]$ is not in the α -core.

Finally, the outcome $[1, 2, 3]$ can be blocked by the coalition $\{1, 2\}$. If $(\sigma_1^q, \sigma_2^q) \in \Sigma_{\{1,2\}}^q(n)$ is chosen such that player 1 shoots player 3 in round 1 and no further shots are fired, the outcome will be $[1, 2]$, independent of player 3's strategy choice. Because $[1, 2]$ is preferred to $[1, 2, 3]$ by both players 1 and 2, the coalition members can guarantee themselves higher payoffs. Thus, $[1, 2, 3]$ is not in the α -core.

To complete the proof of part (a), suppose $a(2) = a(3) = 1$. Note that $U_2([1]) = U_3([1]) = 2$. Choose $(\sigma_2^q, \sigma_3^q) \in \Sigma_{\{2,3\}}^q(n)$ such that player 2 shoots player 1 in round 1 if player 2 is alive at its turn, and player 3 shoots player 1 in round 1 if players 1 and 3 are both still alive at player 3's turn. After round 1, these players always shoot in the air. Given these strategies, the worst possible outcome for player 2 is $[3]$, and the worst possible outcome for player 3 is $[2]$. But $U_2([3]) = U_3([2]) = 3$, so the members of the coalition $\{2, 3\}$ have guaranteed themselves the payoff vector $(3, 3)$, which strictly dominates the payoffs $(2, 2)$ achieved at $[1]$. Therefore, $[1]$ is not in the α -core and thus not in the β -core.

To prove part (b), suppose there exists $j \in \{2, 3\}$ such that $na(j) = 1$, and consider the outcome $[1]$. Clearly, no coalition containing player 1 can block this outcome. To

see that coalition $\{2\}$ cannot block the outcome $[1]$, let $(\sigma_1^q, \sigma_3^q) \in \Sigma_{\{1,3\}}^q(n)$ be such that player 1 shoots player 2 in round 1, and no further shots are fired. Independent of player 2's strategy choice, the outcome is $[1, 3]$, which is worse than $[1]$ for player 2. Therefore, $\{2\}$ cannot achieve a higher payoff, given the strategy choice of its complement $\{1, 3\}$. That $\{3\}$ cannot block $[1]$ is shown analogously.

Finally, consider coalition $\{2, 3\}$. Let $\sigma_1^q \in \Sigma_1^q(n)$ be such that player 1 shoots player j in round 1 and does not shoot in any other round. Given player 1's strategy and the fact that $na(j) = 1$, the best possible outcome for player j is $[1]$. It follows that coalition $\{2, 3\}$ cannot block $[1]$, completing the proof. ■

Simultaneous Truel Unlike sequential truels, the α -core and the β -core of simultaneous truels may not coincide. However, the two cores do not depend on the length of the truel. As shown in the next theorem, the β -core of any simultaneous truel consists of the grand coalition only. Moreover, the α -core always contains the grand coalition and, in addition, the common non-antagonist (if there is one).

Theorem 3 (a) $C_\alpha^m = \{[1, 2, 3]\}$ if and only if, for each $i \in N$, there exists $j \in N \setminus \{i\}$ such that $a(j) = i$;
 (b) For all $i \in N$, $C_\alpha^m = \{[1, 2, 3], [i]\}$ if and only if $na(j) = i$ for all $j \in N \setminus \{i\}$;
 (c) $C_\beta^m = \{[1, 2, 3]\}$.

Proof: First, we show that outcomes containing zero or two players cannot be in the α -core. By (1), these outcomes are not in the β -core, either.

The outcome $[\emptyset]$, in which nobody survives, is blocked by the grand coalition, N . If $\sigma^m \in \Sigma^m$ is such that no shots are fired, the resulting outcome is $[1, 2, 3]$, which is preferred to $[\emptyset]$ by all players.

Now consider an outcome $[i, j]$ with $i \neq j$. Let $\{k\} = N \setminus \{i, j\}$. The coalition $\{k\}$ can choose a strategy $\sigma_k^m \in \Sigma_k^m$ such that k shoots player i in round 1 and does not fire again under any circumstances. For any strategy choice of $\{i, j\}$, this guarantees an outcome that is preferred to $[i, j]$ by k ; thus $[i, j]$ cannot be in the α -core.

Next we prove that $[1, 2, 3]$ is in the β -core and, hence, in the α -core. Clearly, N cannot block $[1, 2, 3]$, because any improvement for one player would make at least one opponent worse off. To see that a singleton $\{i\}$ cannot block $[1, 2, 3]$, let i 's opponents be j and k , and consider a strategy pair $(\sigma_j^m, \sigma_k^m) \in \Sigma_{\{j,k\}}^m$ in which j shoots i in round 1 and does not fire in any subsequent round, and k never fires. For any strategy choice of

player i , i will not survive and, consequently, the resulting outcome is worse than $[1, 2, 3]$. Hence, $\{i\}$ cannot block $[1, 2, 3]$.

Finally, consider a coalition $\{i, j\}$ with $i \neq j$. Let $\{k\} = N \setminus \{i, j\}$, and consider a strategy $\sigma_k^m \in \Sigma_k^m$ such that k shoots i in round 1 and never fires again. Again, i cannot survive, so the resulting outcome is worse than $[1, 2, 3]$ for player i for any strategy choice of $\{i, j\}$. Therefore, $\{i, j\}$ cannot block $[1, 2, 3]$, and $[1, 2, 3]$ must be in the β -core; by (1), it is in the α -core also.

To complete the proof of (a), let $i \in N$, and suppose there exists $j \in N \setminus \{i\}$ such that $a(j) = i$. We show that $\{i\}$ cannot be in the α -core. Consider the coalition $\{j\}$, and let $\sigma_j^m \in \Sigma_j^m$ be such that j shoots i in round 1 and fires no further shots. Independent of the strategy choices of the remaining two players, the worst possible outcome for j is $[na(j)]$ (i.e., only its non-antagonist survives), which j prefers to $[i]$ because $a(j) = i$. Therefore, j can guarantee itself a utility of at least 3, which exceeds $U_j([i]) = 2$. Hence, $[i]$ is not in the α -core and therefore not in the β -core.

Now suppose that $i \in N$ is a common non-antagonist, i.e., if j and k denote i 's opponents ($j \neq k$), then $na(j) = na(k) = i$. Of course, there is at most one common non-antagonist.

To complete the proof of (b), we show that $[i]$ is in the α -core. First, it is clear that no coalition containing i can block $[i]$, i 's most-preferred outcome. The greatest payoff that j can guarantee for itself is $U_j([i]) = 3$ by firing at k in the first round. No higher payoff can be guaranteed because i and k could, for example, choose a strategy pair $(\sigma_i^m, \sigma_k^m) \in \Sigma_{\{i,k\}}^m$ such that i and k shoot at j in the first round and fire no further shots. Therefore, $\{j\}$ cannot block $[i]$ according to the α -core definition.

Now, to obtain a contradiction, suppose there exists $(\sigma_j^m, \sigma_k^m) \in \Sigma_{\{j,k\}}^m$ such that both

$$U_j(f^m(\sigma_j^m, \sigma_k^m, \sigma_i^m)) > 3 = U_j([i]) \quad (2)$$

$$U_k(f^m(\sigma_j^m, \sigma_k^m, \sigma_i^m)) > 3 = U_k([i]) \quad (3)$$

are true for all $\sigma_i^m \in \Sigma_i^m$. We distinguish three cases, according to the properties of (σ_j^m, σ_k^m) :

- (i) j 's first shot is aimed at i , or k 's first shot is aimed at i ;
- (ii) j 's first shot is aimed at k , or k 's first shot is aimed at j ;
- (iii) neither j nor k fires any shots.

(i) Suppose that j 's first shot is aimed at i . Let σ_i^m be such that i shoots j in round 1 and never shoots again. Then k survives and j does not, making $U_j([k]) = 2$ the greatest

possible payoff for j , contradicting (2). The argument is similar if k 's first shot is aimed at i , leading to a contradiction of (3).

(ii) Rename j and k if necessary so that j 's first shot is aimed at k . This case would be covered by (i) if k 's first shot were aimed at i . Therefore, k 's first shot is at j , or k does not shoot. There are two subcases:

(ii.1) If k fires its first shot at j , and it fires in the same round as j or an earlier round, let $\sigma_i^m \in \Sigma_i^m$ be such that i fires its first and only shot at k in the same round in which k fires its first shot at j . The resulting outcome is $[i]$, contradicting (2) and (3).

(ii.2) If k fires its first shot at j but in a later round than j 's first shot, or k does not fire at all, let $\sigma_i^m \in \Sigma_i^m$ be such that i shoots j in round 1 and k in round 2. Again, outcome $[i]$ results, contradicting (2) and (3).

(iii) Let $\sigma_i^m \in \Sigma_i^m$ be such that i shoots j in round 1. The resulting outcome is $[i, k]$, contradicting (2).

This completes the proof of (b).

To complete the proof of (c), we show that the common non-antagonist, $[i]$, is not in the β -core of Γ^m . For any $\sigma_i^m \in \Sigma_i^m$, we provide a strategy pair $(\sigma_j^m, \sigma_k^m) \in \Sigma_{\{j,k\}}^m$ such that $U_j(f^m(\sigma_j^m, \sigma_k^m, \sigma_i^m)) > 3 = U_j([i])$ and $U_k(f^m(\sigma_j^m, \sigma_k^m, \sigma_i^m)) > 3 = U_k([i])$, which demonstrates that $\{j, k\}$ blocks $[i]$ from the β -core.

Let $\sigma_i^m \in \Sigma_i^m$. We distinguish two cases:

- (i) i shoots j or k in round 1;
- (ii) i does not fire in round 1.

(i) If i shoots j in round 1, let $(\sigma_j^m, \sigma_k^m) \in \Sigma_{\{j,k\}}^m$ be such that j shoots k in round 1 and k shoots i in round 1. The resulting outcome is $[\emptyset]$, which is preferred to $[i]$ by both j and k . The argument is analogous if i shoots k in round 1.

(ii) Let $(\sigma_j^m, \sigma_k^m) \in \Sigma_{\{j,k\}}^m$ be such that j shoots i in round 1 and no further shots are fired. The resulting outcome is $[j, k]$, which is preferred to $[i]$ by both j and k . ■

5 Cooperative and Non-Cooperative Truels

Taking equilibria as representative outcomes of non-cooperative analysis, and core outcomes as representative of cooperative analysis, we compare these two views of truels. Table 1 summarizes the numbers of survivors under each form of analysis for each version of the truel.

Table 1: Numbers of Survivors in Truels

Truel Type	Rounds	Non-Cooperative	Cooperative	
		Equilibrium	α -Core	β -Core
Sequential	1	1 or 2	2	2
Sequential	$n > 1$	1	— or 1	— or 1
Simultaneous	1	0 or 1	1 or 3	3
Simultaneous	$n > 1$	0	1 or 3	3

Observe that there are one or two numbers in all cells, except when the cores may be empty, denoted “—,” which can occur in two instances of the sequential truel with $n > 1$.

Cooperative game theory generally produces more survivors than non-cooperative game theory. In fact, this is not the case only in the $n > 1$ sequential truel, which has the same number of survivors whether the analysis is cooperative or non-cooperative, provided the core is non-empty. If the core is empty, the cooperative analysis suggests a lack of stability rather than just fewer survivors.

Table 2 is a more detailed summary of the results reported above. It shows all outcomes that are supported by equilibria, and all core outcomes, in both the sequential and simultaneous truels. In Table 2, *CNA* means “Common Non-Antagonist,” $Ant = (i, j, k)$ means $a(1) = i$, $a(2) = j$, and $a(3) = k$, and $Ant = (\cdot, 1, 1)$ means either $a(1) = 2$, $a(2) = 1$, and $a(3) = 1$ or $a(1) = 3$, $a(2) = 1$, and $a(3) = 1$. Note that player 1 is CNA if and only if $Ant = (\cdot, 3, 2)$, player 2 is CNA if and only if $Ant = (3, \cdot, 1)$, player 3 is CNA if and only if $Ant = (2, 1, \cdot)$, and “no CNA” means either $Ant = (2, 3, 1)$ or $Ant = (3, 1, 2)$.

Place Table 2 here

Table 2 shows that it is generally beneficial to be a common non-antagonist. There is only one instance when the existence of a common antagonist (player 1 in the $n > 1$ round sequential truel) makes for an empty core, so we might say that nobody benefits in this case.

Note the special features of truels with only one round. In the sequential truel, player 1 is first to fire and can affect who survives (while not always surviving itself); player 3 is last to fire and is often eliminated simply because it is the one remaining threat to one of its opponents. In the one-round simultaneous truel, on the other hand, a common non-antagonist is guaranteed survival, though not always alone, in both the cooperative and the non-cooperative analyses.

6 Conclusions

Non-cooperative analysis may be thought of as more demanding, because it insists that players enforce an outcome on their own—specifically, that an outcome is stable because no player has an incentive unilaterally to depart from it. Although this notion of stability would appear to make non-cooperative analysis more hard-nosed and realistic, it is, paradoxically, less realistic if an equilibrium outcome that the analysis prescribes cannot be implemented unless the players can communicate and coordinate their choices (which is usually *not* assumed in non-cooperative game theory).

To be sure, in the case of *sequential* truels, this is less of a problem, because the order of choice is fixed. Given this order and the players' knowledge of each other's antagonisms, we showed that if player 1 is the *common* non-antagonist, its survival in equilibrium is guaranteed. Other players may also be equilibrium survivors in a sequential truel, depending on the pattern of antagonisms and the parity of the number of rounds.

For *simultaneous* truels, on the other hand, the situation is far more ambiguous. Except for one-round truels, in which a common non-antagonist can survive in equilibrium, there will be no survivors in equilibrium in either the one-round or the n -round truels.

But the equilibrium prediction of no survivors in n -round simultaneous truels will not be realized unless the players follow a cyclic pattern of shooting. If they do not, and, say, players 1 and 2 both shoot player 3, and player 3 shoots player 2, then player 1 will be the sole survivor. While this is not an equilibrium outcome because player 2 could have done better shooting player 1, it cannot always be achieved in the absence of communication and coordination, whether the simultaneous truel is one round or n rounds.

To try to circumvent this problem, assume that there is communication, and the players agree to a cyclic pattern of shooting in the first round, leading to no survivors in a simultaneous truel. Is this scenario more realistic than the players, realizing that they all can do better by not shooting in the first round, agreeing to fire in the air? Indeed, three survivors, which is a predicted outcome of both the α -core and β -core, seems eminently reasonable, except for one thing: each player has an incentive to defect in round 1, or beginning in round 1 if the truel has n rounds.

At the same time, there is also a *disincentive* to defect in the n -round truel:

1. If only one player defects, the two surviving players will shoot each other in a duel in the next round, so the defector will do no better than if it had stuck to the agreed-upon cyclic shooting pattern in round 1.

2. If two players defect and shoot each other, they will actually be worse off because there will be one survivor; if they do not shoot each other but both shoot the third player, there will again be a duel in the next round and no survivors. Only if both defectors shoot different players will one survive, but then the defector who does not survive is worse off than if all players had initially abstained from shooting each other.
3. If all three players defect but do not shoot different opponents, there will again be one survivor, which may or may not be oneself (and therefore better or worse than no survivors).

Thus, the core predictions of three survivors in n -round simultaneous truels is not so implausible—even from a non-cooperative perspective—depending on how many survivors there are, and who they are, if one or more players defects in the first round.

Curiously, the core analysis is, in a sense, more demanding than the equilibrium analysis for n -round *sequential* truels. For these truels, each approach predicts one survivor, though not necessarily the same player, except in one case: when both players 2 and 3 consider player 1 their common antagonist. In this case, the usual advantage that player 1 enjoys from going first is nullified by the other players' mutual antagonism toward it, rendering no coalition stable in the cooperative analysis (i.e., the core is empty).

We conclude that both Nash equilibria and the α -cores and β -cores of truels make reasonable predictions when they do not overlap: there are plausible circumstances under which each prediction might be fulfilled. Generally speaking, the non-cooperative analysis predicts fewer, and never more, survivors than the core analysis. However, its predictions will not be readily achieved if there is no communication or coordination by the players, especially in the case of simultaneous truels, rendering these predictions dubious.

When communication and coordination are possible, the cores become more compelling as solutions, even in the absence of an enforceable agreement, as we illustrated above. The prediction of a non-empty core is also sensible in an n -round sequential truel in which the normally favored player 1 is the antagonist of both its opponents.

The lack of agreement of the two different kinds of solution concepts does not, we think, signal a crisis in game theory. Cooperative and non-cooperative approaches have both played an important role in the development of the theory. The question is how they might best be combined to produce a coherent and rounded analysis of different strategic situations, and their probable outcomes, that highlights when each kind of outcome is likely to arise.

References

- AUMANN, R.J. (1961), “The core of a cooperative game without side payments,” *Transactions of the American Mathematical Society* **98**, 539–552.
- AUMANN, R.J. (1967), “A survey of cooperative games without side payments,” in SHUBIK, M. (ed.), *Essays in Mathematical Economics in Honor of Oskar Morgenstern*, Princeton: Princeton University Press, 3–27.
- AUMANN, R.J. AND PELEG, B. (1960), “Von Neumann-Morgenstern solutions to cooperative games without side payments,” *Bulletin of the American Mathematical Society* **66**, 173–179.
- BRAMS, S.J. (1994), *Theory of Moves*, New York: Cambridge University Press.
- BRAMS, S.J. AND KILGOUR, D.M. (1998), “Backward induction is not robust: The parity problem and the uncertainty problem,” *Theory and Decision* **45**, 263–289.
- BRAMS, S.J. AND KILGOUR, D.M. (2000), “Games That End in a Bang or a Whimper,” Preprint, Department of Politics, New York University.
- GREENBERG, J. (1990), *The Theory of Social Situations: An Alternative Game-Theoretic Approach*, Cambridge, UK: Cambridge University Press.
- KILGOUR, D.M. (1972), “The simultaneous truel,” *International Journal of Game Theory* **1**, 229–242.
- KILGOUR, D.M. (1975), “The sequential truel,” *International Journal of Game Theory* **4**, 151–174.
- KILGOUR, D.M. (1978), “Equilibrium points of infinite sequential truels,” *International Journal of Game Theory* **6**, 167–180.
- KILGOUR, D.M. AND BRAMS, S.J. (1997), “The truel,” *Mathematics Magazine* **70**, 315–326.
- NASH, J. (1951), “Non-cooperative Games,” *Annals of Mathematics* **54**, 286–295.
- VON NEUMANN, J. AND MORGENSTERN, O. (1953), *Theory of Games and Economic Behavior*, 3rd edition, Princeton: Princeton University Press.