Beyond indifferent players: On the existence of Prisoners Dilemmas in games with amicable and adversarial preferences

Andrew W. Horowitz, Renato Galvão Flôres Junior

Dezembro de 2004
Os artigos publicados são de inteira responsabilidade de seus autores. As opiniões neles emitidas não exprimem, necessariamente, o ponto de vista da Fundação Getúlio Vargas.
Beyond Indifferent Players:

On the Existence of Prisoners Dilemmas in games with amicable and adversarial preferences

(October 2004)

JEL C7, D6, H4
Keywords: Prisoners Dilemmas, Game Theory, Non-Cooperative Games

By

Andrew W. Horowitz* and Renato G. Flôres Jr**

Department of Economics and EPGE
Sam M. Walton College of Business and Fundação Getulio Vargas
University of Arkansas, and Rio de Janeiro
Fayetteville, AR 72701 and Brasil

Abstract: Why don’t agents cooperate when they both stand to gain? This question ranks among the most fundamental in the social sciences. Explanations abound. Among the most compelling are various configurations of the prisoner’s dilemma (PD), or public goods problem. Payoffs in PD’s are specified in one of two ways: as primitive cardinal payoffs or as ordinal final utility. However, as final utility is objectively unobservable, only the primitive payoff games are ever observed. This paper explores mappings from primitive payoff to utility payoff games and demonstrates that though an observable game is a PD there are broad classes of utility functions for which there exists no associated utility PD. In particular we show that even small amounts of either altruism or enmity may disrupt the mapping from primitive payoff to utility PD. We then examine some implications of these results.

* Horowitz@walton.uark.edu; phone: (1)-479-575-6228
** rflores@fgv.br; phone (55)-21-25595909.

We thank Arilton Teixeira, Aloisio Pessoa de Araújo, Amy Farmer, Cary Deck, Joel Carron, and seminar participants at Fundação Getulio Vargas (Rio de Janeiro), IOB – University of Antwerp, ISEG – Technical University of Lisbon, Ibmec-Rio, Walton College of Business, and Federal University of Rio Grande de Sul. The usual disclaimers apply.
1. Introduction

Prisoner dilemmas (PDs) have been employed across the social and business sciences, philosophy, and biology as prime examples of the tension between individual and collective rationality.¹ They constitute powerful illustrations of the gains foregone when strategic structure precludes cooperation as an equilibrium strategy.

The payoffs in PD’s have two forms. First, they may be cardinal observable payoffs (e.g., years in prison, nuclear warheads, or advertising budgets). We refer to such games as Primitive Prisoner’s Dilemmas (PPDs). Alternatively, payoffs may be specified as final utility, which is inherently unobservable. We refer to these games as Utility Prisoner’s Dilemmas (UPDs). In either case there is an implicit mapping between observable payoff and final utility that has received scant attention in the literature. Though this neglect may be innocuous for some mappings we show that when a player has amicable or adversarial inclination towards the other player there are broad classes of utility functions for which it is impossible for a PPD to map into a UPD. We identify classes of utility functions under which games that are not prisoners dilemmas in observable payoffs are, in fact, prisoner’s dilemmas in the unobserved utility game.

Why our focus on amicable and adversarial preferences? First, there exists a large body of experimental evidence (see Fehr and Gachter 2000 for a survey) that casts doubt on the indifference of players with regard to the payoffs of other players. We will demonstrate that only in the case of truly indifferent players will a game that is a PD in

¹ A nice survey of economic applications of the PD can be found in Rapoport (1987). In political science, Brams’ (1994) “Theory of Moves” provides a novel analysis of the Prisoner’s Dilemma (PD) and argues that mutual cooperation will typically emerge.
observable payoffs necessarily be a PD in the unobserved utility game. In fact, the body of experimental evidence cited above indicates unambiguously that pure neutrality towards the welfare of the other players is the exception, rather the rule. Beyond the experimental literature, the potential for altruism in strategic environments has long been recognized. For example, strategic frameworks are frequently employed to model intra-household interactions (see Browning and Chiaporri 1998). Moreover, intra-household and kin altruism is implied by evolutionary biology.

Adversarial relationships, in the sense of competition, arise in virtually all economic environments. However in the typical strategic setting adversarial incentives are inherent in the payoff-structure rather than embodied in preferences. Thus, the incentive to adopt a particular strategy is typically governed by own payoff maximization rather than explicit consideration of rivals’ payoff. In contrast, we consider strategic behavior when a player’s utility is decreasing in the other player’s cardinal payoff. Such preferences may correspond to conventional notions envy or malice. These terms, “envy” and “malice,” have precise economic meanings (see Hammond 1987 and Brennan 1973), and though this literature addresses issues tangentially related to this paper, it never addresses the implications of such utility mappings on the existence of the PD.

For those who remain skeptical of amicability or enmity in preferences per-se, there exists an alternative motivation that is also entirely consistent with the model and results. Namely, if the observable payoff of one player yields an externality (in utils) to others, the analysis is identical. Formally, these externalities would create a wedge between the

\[ \text{In zero sum games these objectives would be equivalent. But as noted, our analysis does not concern zero-sum games.} \]
observable payoff game and the unobservable “welfare” game that is formally equivalent to either amicable or adversarial preferences. Such an interpretation opens a plethora of applications in economics as well as political science.

The remainder of the paper is organized as follows. Section 2 introduces notation and definitions necessary to analyze PDs with neutral, amicable, and adversarial players. Section 3 presents our most general existence results and specific congruence results for amicable, adversarial, indifferent, and asymmetric players. A Cobb-Douglas example is also provided in Section 3. Section 4 summarizes and concludes.

2. Notation and Definitions

The Game

Consider a two-player game and call the players A and B and their cardinal (observable) payoffs $\alpha$ and $\beta$ respectively. Each player has two strategies. Denote the players’ strategy sets and strategy choice as respectively: $S^p = \{1, 2\}$ and $s^p$ for $p = A, B$. So the joint strategy space has four elements and denote the associated observable primitive (cardinal) payoff vectors as $\pi_{ij} = [\alpha_{ij}, \beta_{ij}]$ where $i = s^A$ and $j = s^B$ with the payoff space denoted as $\Pi \subset \mathbb{R}^2$. Let $r^p(s)$ denote the best response of player $p$ to strategy $s$ by the other player. Without loss of generality, payoffs are non-negative and when the clarity constraint permits we suppress the subscripts on $\alpha$ and $\beta$. The one-stage

---

3 Yet another motivation is a game where joint strategies map into two-good payoffs with one of the primitive payoffs is a “good” and other is a “bad” for one player, while the second player has reverse preferences towards the payoffs. For example, we can imagine roommates who have contradictory preferences towards classical and rock music. For one roommate classical is a good and rock is a bad, while the reverse holds for the other roommate. Joint strategies yield quantities of both goods, and it is easy to construct a PD (i.e., Pareto Inferior equilibrium) in this environment.
game defined by the above triplet $\Gamma = [P, S, \Pi]$ will be called the primitive game. All its elements are observable and fully known by both players.

A primitive prisoner’s dilemma (PPD) occurs when the Nash Equilibrium of the primitive game yields a payoff ($\pi$) that is vector dominated by some non-equilibrium payoff. Without loss of generality let $s_p = 1$ for $(p = A, B)$ be the strategies that map to the vector dominated primitive payoff and $s_p = 2$ for $(p = A, B)$ the strategies that map to the vector dominant payoff. Using the notation introduced above the payoff vectors are: $\pi_{22} > \pi_{11}$, where a vector inequality indicates vector dominance.

Each player has unobservable preferences over the primitive payoff space that are complete, transitive, and reflexive. In a slight (but innocuous) abuse of notation that yields considerable notational economy we denote the unobservable utility functions as: $A(\alpha, \beta), B(\alpha, \beta)$. Let $U_{ij} = [A(\pi_{ij}), B(\pi_{ij})]$ be the vector of final utility payoffs when player $A$ plays strategy $i$ and $B$ plays strategy $j$ (where $i$ may equal $j$). The functions $A(\pi_{ij})$ and $B(\pi_{ij})$ may map non-monotonically from own-primitive-payoff ($\alpha$ for $A$, $\beta$ for $B$) to final own-utility due to either amicable or adversarial preferences. For a given $U$ every Primitive Game maps to an associated Utility Game ($UG$) and we define the associated $UG$ as $V(\Gamma) = [P, S, U(\Pi)]$. If $U$ does not order payoffs as in the observable primitive game, $V$ will be a weakly better predictor of players’ strategic behavior than $\Gamma$. For expositional convenience we will assume henceforth that the utility functions are

4 For ease of exposition we consider Prisoners Dilemmas where the equilibrium is strictly inferior for both players. Naturally, the definition Pareto inferior would allow only one player to be worse off, while all other players might be indifferent. Focusing on strict PDs considerably streamlines the paper. However, it is critical to note that versions of all propositions and results can be obtained with the weaker PD definition – though at a considerable cost in tedium.
differentiable. Extension to well-behaved non-differentiable utility functions is straightforward for virtually the entire analysis.

A number of indifference curves will have special significance in our analysis and we employ the following notation: \( A_{ij} = \{ \pi \in \Pi: \pi \sim \pi_{ij} \} \) for player \( A \), while analogously \( B_{ij} \) denotes player \( B \)’s indifference set with \( \pi_{ij} \) where \( i, j = 1, 2 \). So \( A_{11} \) is the set of all joint payoffs that \( A \) finds indifferent to \( \pi_{11} \). We use strong versions of the upper and lower contour sets of \( \pi_{ij} \) for player \( p \), defining them respectively as follows: \( UCS^p_{ij} = \{ \pi \in \Pi: \pi \succ \pi_{ij} \} \) for player \( p \), \( LCS^p_{ij} = \{ \pi \in \Pi: \pi \prec \pi_{ij} \} \) for player \( p \). Again, all propositions hold with weak forms of the upper and lower contour sets, though the exposition is more tedious. The required modification of the proofs with weak contour sets is indicated subsequently.

**Payoff Space Partitions**

The following payoff space partitions are central to our analysis. We will subsequently provide graphically illustrations of these sets for amicable, adversarial, and indifferent players. Note that all sets are subsets of the primitive joint payoff space.

1. **Superior Set (S)**
   \[
   S = UCS^A_{11} \cap UCS^B_{11}
   \]

2. **Far Set (F)**
   \[
   F = UCS^A_{22} \cap UCS^B_{22}
   \]

3. **Central Set (C)**
   \[
   C = S \cap LCS^A_{22} \cap LCS^B_{22}
   \]
(4). Dominant Set of player p \( (D^p) \)

\[ D^p = UCS^p_{22} \cap LCS^{\neg p}_{11}, \]

where \( \neg p \) indicates player “not p.”

(5). Central Set of player p \( (C^p) \)

\[ C^p = UCS^p_{11} \cap LCS^{\neg p}_{11} \cap LCS^p_{22}. \]

(6). Far Set of player p \( (F^p) \)

\[ F^p = UCS^p_{22} \cap LCS^{\neg p}_{22} \cap LCS^p_{11}. \]

**Payoff Partitions When Both Players are Indifferent**

Since players’ subjective amicable, adversarial, or indifferent attitude towards one another are not directly observable the standard assumption is one of indifference – that is, each player’s strategy choices are governed by their own cardinal payoffs alone. Of course, it is also possible that such indifference is in fact a player’s true preference towards others. Letting subscripts denote partials the indifferent player’s preferences are:

\[ A_\alpha > 0, \quad A_\beta = 0, \quad B_\beta > 0, \quad B_\alpha = 0, \]

and indifference curves are linear in the joint-payoff space. Figure 1 below illustrates the payoff-space partition for indifferent players. These sets have different topology for amicable or adversarial players and we will rigorously characterize the relationship between them under the various preferences in the following section.
Payoff Partitions When Both Players are Amicable

We say player A is amicable at \( \pi \), if \( A_\alpha(\pi) > 0 \) and \( A_\beta(\pi) > 0 \) and a globally amicable if the inequalities hold at all \( \pi \). A similar definition applies for player B. An extreme form of amicability is altruism. Player A is an altruist at \( \pi \) if and only if \( \frac{\partial \ln A(\pi)}{\partial \ln \beta} > \frac{\partial \ln A(\pi)}{\partial \ln \alpha} \) and a global altruist if the condition holds at all \( \pi \). When comparing preferences \( A^o(\pi) \) and \( A^*(\pi) \) we say that \( A^o \) is more amicable than \( A^* \) at \( \pi \) if \( \frac{-A^o_\alpha}{A^o_\beta} > \frac{-A^*_\alpha}{A^*_\beta} \). Given our definitions an amicable player’s indifference curves of are downward sloping in the joint payoff-space. Figure 2 illustrates a payoff-space partition for amicable players, with indifferent players’ partitions indicated by the dashed lines.
The juxtaposition of Figure 2 and 1 provides a striking illustration of the distortions of the payoff-partitions vis-à-vis the indifferent player. We will demonstrate that this non-congruence has critical implication for the interpretation and existence of PDs in the unobserved utility game.

**Figure 2**
The Payoff Partition – Amicable Players

The juxtaposition Figure 2 and 1 provides a striking illustration of the distortions of the payoff-partitions vis-à-vis the indifferent player. We will demonstrate that this non-congruence has critical implication for the interpretation and existence of PDs in the unobserved utility game.

*Payoff Partitions When Both Players are Adversarial*

We say player A is an *adversary* (or has enmity) at \( \pi \) if \( A_\alpha(\pi) > 0 \) and \( A_\beta(\pi) < 0 \) and a *global adversary* if the inequalities hold at all \( \pi \). Indifference curves of a player with enmity are upward sloping (with finite slope) in the joint payoff space. We say that
player A has strong enmity for the other player at $\pi$ if:

$$\left| \frac{\partial \ln A(\pi)}{\partial \ln \beta} \right| > \frac{\partial \ln A(\pi)}{\partial \ln \alpha},$$

and that preferences $A^o$ display less enmity than $A^*$ at $\pi$ if $-A^o_{\alpha}/A^o_{\beta} > -A^*_{\alpha}/A^*_{\beta}$ at $\pi$.

Figure 3 below illustrates a payoff-space partition for adversaries, with indifferent players’ partitions again in the background.

![Figure 3](image-url)

**Figure 3**

The Payoff Partition -- Adversaries

As in the case of amicable preferences, Figure 3 reveals dramatic “distortions” in the payoff-space partitions – though they are markedly different. Note that for these adversaries, as opposed to indifferent and amicable players, the Far Sets ($F, F^A, F^B$) are now bounded. Also note that the Central Set ($C$) has remained bounded in all scenarios.

We now move to consideration of the existence of PD under these various preferences.
3. Prisoners Dilemmas under Alternative Preferences

3.1 Necessary and Sufficient Conditions for Prisoners Dilemmas in the Utility Game

We are now in a position to connect the existence of PDs with the payoff-space partitions. We begin by defining two forms of PDs: Strong and Weak.

- **Strong Prisoners Dilemma (SPD)**
  A game is a SPD if the strategy yielding the Pareto Inferior payoff is a dominant strategy for both players.

- **Weak Prisoners Dilemma (WPD)**
  A game is a WPD if the strategy yielding the Pareto Inferior payoff is a dominant strategy for only one player.

The following Propositions provide the necessary and sufficient conditions for the various forms of PDs, for multiple equilibrium, and for no equilibrium in our framework. We note that Propositions 1-5 hold for any types of attitudes between the players: amicable, adversarial, and indifferent.

**Proposition 1.** Given any $\pi_{11}$ and $\pi_{22} \in S$, $\pi_{ij} \in D^p$ for $i,j = 1, 2$ $i \neq j$ and $p = A, B$ are necessary and sufficient conditions for the unique Nash Equilibrium to be a SPD.

**Proof:** Sufficient: First consider player A’s best responses. Given the above conditions: $\pi_{12} \in UCS^A_{22}$ and $\pi_{21} \in LCS^A_{11}$, therefore $s^A = 1$ is a dominant strategy for A. An analogous argument holds for B. Necessary: Again first consider player A. Suppose the conditions Proposition 1 are not satisfied. If $\pi_{12} \not\in D^A$ then either $\pi_{12} \not\in UCS^A_{22}$ or $\pi_{12} \not\in LCS^B_{11}$. If $\pi_{12} \not\in UCS^A_{22}$, $\pi_{12} \in LCS^A_{22}$ (recall our strong definitions of UCS and LCS) and $r^A(2) = 2$ so that $s^A = 1$ is no longer a dominant strategy. If $\pi_{12} \not\in LCS^B_{11}$, $\pi_{12} \in UCS^B_{11}$ so that $r^B(1) = 2$ and $s^B = 1$ is no longer a dominant strategy. A similar argument holds for $\pi_{21} \not\in D^B$.
Corollary 1: A necessary condition for the existence of a SPD is that $B_{11} \cap A_{22} \neq \emptyset$ and $B_{22} \cap A_{11} \neq \emptyset$.

Proof: Immediate given Theorem 1. If $B_{11} \cap A_{22} = \emptyset$, $D_v^A$ is empty.

Proposition 2: Given any $\pi_{11}$ and a $\pi_{22} \in S_v$, a sufficient condition for a unique Nash Equilibrium which is a WPD is: $\pi_{ij} \in D^p$ and $\pi_{ji} \in C^{-p}$ for $i, j = 1, 2$ where $i \neq j$.

Proof: First consider the case where $\pi_{12} \in D^A$ and $\pi_{21} \in C^B$. Then $\pi_{12} \in UCS_{22}^A$ and $\pi_{21} \in LCS_{11}^A$, therefore $s^A = 1$ is a dominant strategy for $A$. For player $B$, $\pi_{12} \in LCS_{11}^B$ so $r^B(1) = 1$ and $\pi_{21} \in LCS_{22}^B$ so $r^B(2) = 2$. So $B$ has no dominant strategy and $s = \{1,1\}$ is the unique Nash Equilibrium. An analogous argument holds for the case of $\pi_{21} \in D^B$ and $\pi_{12} \in C^A$, in which case player $B$ is the player with the dominant strategy.

Proposition 3: Given any $\pi_{11}$ and a $\pi_{22} \in S_v$ if $\pi_{ij} \in D^p$ and $\pi_{ji} \in C^i \neq j$, $i, j = 1, 2$, then $(j, i)$ is the unique Nash equilibrium of the game.

Proof: First consider the case where $\pi_{12} \in D^A$ and $\pi_{21} \in F^B$. Then $\pi_{12} \in UCS_{22}^A$ and $\pi_{21} \in UCS_{11}^A$, so $r^A(2) = 1$ and $r^A(1) = 2$. For player $B$, $\pi_{12} \in LCS_{11}^B$ so $r^B(1) = 1$ and $\pi_{21} \in UCS_{22}^B$ so $r^B(2) = 1$. So $B$ has a dominant strategy and $s = \{2,1\}$ is the unique Nash Equilibrium – which is not a PD. An analogous argument holds for the case of $\pi_{21} \in D^B$ and $\pi_{12} \in F^A$, in which case player $A$ has the dominant strategy.

Proposition 4: Given any $\pi_{11}$ and a $\pi_{22} \in S_v$ if $\pi_{ij} \in D^p$ and $\pi_{ji} \in C i \neq j$, $i, j = 1, 2$ then the game has no Nash equilibrium.
Proof: First consider the case where \( \pi_{12} \in D^A \) and \( \pi_{21} \in C \). Then \( \pi_{12} \in UCS^A_{32} \) and \( \pi_{21} \in UCS^A_{11} \), so \( r^A(2) = 1 \) and \( r^A(1) = 2 \). For player B, \( \pi_{12} \in LCS^B_{11} \) so \( r^B(1) = 1 \) and \( \pi_{21} \in UCS^B_{11} \) so \( r^B(2) = 1 \). So there is no Nash Equilibrium. An analogous argument holds for the case of \( \pi_{21} \in D^B \) and \( \pi_{12} \in C \).

Proposition 5: Given any \( \pi_{11} \) and a \( \pi_{22} \in S_v \), both \( s = \{1, 1\} \) and \( s' = \{2, 2\} \) are equilibrium if: \( \pi_{12} \in C^A \) and \( \pi_{21} \in C^B \).

Proof: \( \pi_{12} \in LCS^A_{32} \) so \( r^A(2) = 2 \) and \( \pi_{21} \in LCS^A_{11} \) so \( r^A(1) = 1 \). For player B \( \pi_{12} \in LCS^B_{11} \) so \( r^B(1) = 1 \) and \( \pi_{21} \in LCS^B_{32} \) so \( r^B(1) = 1 \) and \( r^B(2) = 2 \).

Discussion

Propositions 1 through 5 make clear that it is the membership of the \( \pi_{ij} \) payoffs in the various partitions of the joint-payoff space that determine the nature of the equilibrium, or lack thereof. Of critical relevance to the existence of Prisoner’s Dilemmas is the membership of at least one of the \( \pi_{ij} \) payoffs in a player’s Dominant Set. The Figures of Section 2 suggest that under amicable or adversarial preferences the Players’ Dominant Sets contract and expand respectively. An immediate implication is that though the observable payoff structure of a game suggests a Prisoners Dilemma equilibrium, unobserved amicable or adversarial attitudes of the players may transform the utility game to one with a different equilibrium. The mechanism of this transformation is the “migration” of \( \pi_{ij} \) payoffs between Payoff-space Partitions as we move from the primitive game, with the implied indifference of players, to a utility game with amicable or adversarial preferences. Only in the case of truly indifferent player can we be certain that the Dominant Sets in the observable game and utility games are congruent.
3.2 Payoff-Space Partition Congruence with non-Indifferent Players

The Figures of Section 2 were merely suggestive of the types of Payoff-space Set transformation that may occur when non-indifferent players are present. We now formally characterize these transformations. Note that in the prior and proceeding analysis, multiple-crossing of an indifference curve of an amicable player with a particular indifference curve of another amicable player would complicate the analysis. To keep the paper of manageable length we focus on single crossing indifference curves of amicable players and note that the results would be modified in fairly obvious ways in the presence of multiple crossing curves. To facilitate presentation of the next results we introduce the following additional notation: for each set of the payoff-space partition let the subscripts \( \Gamma \) or \( V \) indicate respectively the primitive or utility game partition. For example, \( D^p_{\Gamma} \) is player \( p \)’s Dominant Set in observable payoffs while \( D^p_{V} \) is player \( p \)’s Dominant Set in the utility game.

**Both Players are Adversaries**

*Proposition 6.* With adversarial preferences a player’s Primitive Dominant Set is a strict sub-set of their Utility Dominant Set: \( D^p_{\Gamma} \subset D^p_{V} \).

*Proof.* Consider the point \( \pi' = (\alpha_{22}, \beta_{11}) \). With adversarial preferences \( \pi' \in UCS^A_{22} \) and \( \pi' \in LCS^B_{11} \), so \( \pi' \in D^A_{\Gamma} \). With regard to the primitive game \( \pi' \notin D^A_{\gamma} \), since \( \pi' \in A_{22} \) and \( \pi' \in B_{11} \). A similar argument holds for player \( B \). To see that every element of \( D^B_{\gamma} \) must be an element of \( D^B_{\psi} \) simply note that because of the finite upward slope of indifference curves with adversarial preferences \( \forall \pi \in D^A_{\gamma}, \pi \in UCS^A_{22} \) and \( \pi \in LCS^B_{11} \) :
Note that even if the upper and lower contour sets were defined weakly we could find a point in an open ball centered on \( \pi' \) that is an element of \( D'_V \) but not \( D'_\Gamma \). This general argument holds for all subsequent propositions, and will not be repeated.

**Proposition 7.** If both players have adversarial preferences the Utility Central Set is a strict sub-set of the Primitive Central Set: \( C_V \subset C_\Gamma \)

**Proof:** \( C_\Gamma \) is the quadrilateral defined by \( \{ \pi | \pi_{11} < \pi < \pi_{22} \} \). Given that \( B_{11} \) and \( A_{22} \) have finite positive slopes and pass through \( \pi_{11} \) and \( \pi_{22} \) respectively, they must intersect in the interior of \( C_\Gamma \) since \( B_{11} \) cannot intersect \( B_{22} \) which also passes through \( \pi_{22} \). Likewise for \( A_{11} \) and \( B_{22} \). Therefore \( C_V \subset C_\Gamma \).

**Proposition 8.** With adversarial preferences the Superior Set of the utility game is a strict sub-set of the primitive Superior Set: \( S_V \subset S_\Gamma \).

**Proof.** \( S_\Gamma \) is the quadrant defined by \( \pi > \pi_{11} \). Since adversarial indifference curves have finite positive upward slope \( A_{11} \) and \( B_{11} \) are contained in \( S_\Gamma \) for all \( \pi > \pi_{11} \). The intersection of the upper contour sets for \( \pi > \pi_{11} \) must therefore be empty or contained in \( S_\Gamma \). So every element of \( S_V \) must also be an element of \( S_\Gamma \). To see that not every element of \( S_\Gamma \) is an element of \( S_V \) let \( B'_e(\pi') \) be a closed ball of radius \( e \) centered on \( \pi' \in A_{11} \) for some \( \pi' > \pi_{11} \), where we choose \( e \) such that \( B'_e(\pi') \subset S_\Gamma \). By the Jordan Curve Theorem the indifference curve through \( \pi' \) divides the ball into two distinct domains, one a subset of \( UCS_{11}^A \) and the other a subset of \( LCS_{11}^A \) where by definition if \( \pi'' \in B'_e(\pi') \) and \( \pi'' \in LCS_{11}^A \), \( \pi'' \notin S_V \).
We can make the following stronger characterization of $S_V$ when both players are both strong adversaries.

**Proposition 9.** $S_V$ is either empty or bounded if players are strong global adversaries.

**Proof:** With $\alpha$ on the ordinate and $\beta$ the abscissa, as in Figure 3, A’s and B’s indifference curves are respectively convex and concave with positive finite slope.

(i). If the indifference curves are tangent at $\pi_{11}$ the intersection of the upper contour sets is empty.

(ii). If the slope of $A_{11}$ exceeds that of $B_{11}$ at $\pi_{11}$ the intersection of the upper contour sets must lie to the southwest of $\pi_{11}$ and is contained in the bounded set: \{ $\pi$ : $\alpha < \alpha_{11}$, $\beta < \beta_{11}$ \}.

(iii.) If the slope of $B_{11}$ exceeds that of $A_{11}$ at $\pi_{11}$ the indifference curves must intersect again (since $B_{11}$ is strictly concave and $A_{11}$ strictly convex). Call this intersection $\pi'$. In this case the contour sets intersection must lie to the southwest of $\pi'$: in the bounded set $\{ \pi : \alpha < \alpha'$ and $\beta < \beta' \}.$.

**Corollary to Proposition 9.** If players are strong adversaries $F^V$ is either empty or bounded.

**Proof:** Simply repeat the above proof substituting $\pi_{22}$ for $\pi_{11}$, $A_{22}$ for $A_{11}$, and $B_{22}$ for $B_{11}$.

The following very strong proposition is the principal non-existence results of our analysis.

**Proposition 10.** If players are strong adversaries at $\pi_{11}$ a game which is PD in observable payoffs can never be a PD in the utility game.

**Proof:** Suppose in contradiction to Proposition 10 a game which is a PD in cardinal payoffs is also a PD in the utility game. Then $\pi_{22} \in S_V$, and $A(\pi_{22}) > A(\pi_{11})$ and $B(\pi_{11}) < B(\pi_{22})$. As both preference functions are continuously differentiable, there exists an open ball $B_\delta(\pi_{11})$, and a $\pi' = (\alpha_{11} + d\alpha, \beta_{11} + d\beta) \in B_\delta(\pi_{11})$, with $d\alpha$, $d\beta > 0$, such that $A_\alpha(\pi_{11}) d\alpha + A_\beta(\pi_{11}) d\beta > 0$ and $B_\alpha(\pi_{11}) d\alpha + B_\beta(\pi_{11}) d\beta > 0$. Taking into account
the signs of the partial derivatives these inequalities yield: \( A_\alpha(\pi_{11}) / A_\beta(\pi_{11}) < B_\alpha(\pi_{11}) / B_\beta(\pi_{11}) \). By the definition of strong adversaries at \( \pi_{11} \), however, \( A_\alpha(\pi_{11}) / A_\beta(\pi_{11}) > B_\alpha(\pi_{11}) / B_\beta(\pi_{11}) \), a contradiction. ∴

Proposition 10 extends an important implication of Proposition 9. That is, when players are global strong adversaries a primitive Prisoner’s Dilemma can never be a UPD. It also provides a dramatic example of a more general result which holds for all forms of adversarial preferences. Namely, adversarial behaviour reduces from an infinite to a finite (Lebesgue) measure the set of primitive payoffs that could possibly be associated with a utility prisoner’s dilemma. Moreover, in the strong adversary case of Proposition 10 the bounded set \( F^V \) will never include \( \pi_{22} \). Since all prisoners dilemmas (primitive or utility) require a Pareto dominant payoff, the non-existence of the utility prisoners dilemma follows. This non-congruence of the observable primitive game and the inherently unobservable utility game has profound implications for the interpretation of a wide range of economic applications – including the public goods problem.

Both Players are Amicable

Proposition 11. If both players are amicable the Dominant Sets of the utility game (if they exist) are strict sub-sets of their primitive game Dominant Sets: \( D_\Gamma^V \subset D_\Gamma^I \).

Proof. Since both \( A_{22} \) and \( B_{11} \) have negative finite slope their intersection must occur in \( D_\Gamma^A \), if at all. Thus every element of \( D_\Gamma^A \) is also an element of \( D_\Gamma^I \). By Corollary 1 \( A_{22} \) and \( B_{11} \) must intersect for \( D_\Gamma^A \) to be non-empty. If it occurs call the intersection point \( \pi' \in D_\Gamma^A \). Now consider an \( \epsilon > 0 \) such that the closed ball \( B_\epsilon(\pi') \subset D_\Gamma^A \), and \( A_{11} \)
partitions $B^c_r(\pi^*)$ into distinct domains one of which contains elements of $LCS^A_{22}$, which are not members of $D^A_V$ but are elements of $D^A_\Gamma$. \\ 

**Proposition 12.** If both players have amicable preferences the Primitive Central Set is a strict sub-set of the Utility Central Set: $C_\Gamma \subset C_V$. 

**Proof:** $C_\Gamma$ is the quadrilateral defined by $\{ \pi | \pi_{11} < \pi < \pi_{22} \}$. Given that $B_{11}$ and $A_{22}$ have finite negative slopes and pass through $\pi_{11}$ and $\pi_{22}$ respectively, they must intersect in the interior of $D_\Gamma$, if at all. Likewise for $A_{11}$ and $B_{22}$. Therefore $C_\Gamma \subset C_V$. 

**Proposition 13.** With amicable preferences the Superior Set of the primitive game is a strict sub-set of the Superior Set of the utility game: $S_\Gamma \subset S_V$. 

**Proof.** Immediate. With amicable preferences the indifference curves $A_{11}$ and $B_{22}$ are support functions for $S_\Gamma$. 

Figure 4 below provides a summary of the principal results of this sub-section.

---

**Figure 4**  
Summary of Payoff Partition Congruence  
Players are:

<table>
<thead>
<tr>
<th>Indifferent</th>
<th>Adversaries</th>
<th>Amicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D^p_\Gamma = D^p_V$</td>
<td>$D^p_\Gamma \subset D^p_V$</td>
<td>$D^p_V \subset D^p_\Gamma$</td>
</tr>
<tr>
<td>$S_\Gamma = S_V$</td>
<td>$S_V \subset S_\Gamma$</td>
<td>$S_\Gamma \subset S_V$</td>
</tr>
<tr>
<td>$C_\Gamma = C_V$</td>
<td>$C_V \subset C_\Gamma$</td>
<td>$C_\Gamma \subset C_V$</td>
</tr>
</tbody>
</table>

PP UP  PP UP  PP UPD
Note that the likelihood of a PPD having an associated UPD is reduced in different ways for adversaries and amicable players. In the case of adversaries the likelihood that the Pareto Superior payoff in the primitive game is also Pareto Superior in the utility game is reduced. For amicable players the likelihood that the “off diagonal” payoff provides defection incentive from the Pareto Superior payoff is reduced.

**Asymmetric Attitudes and the Central Set as a Measure of Amicability**

Asymmetric attitudes across players generate a rich set of possibilities. In this sub-section we briefly explore congruence properties of the Central Set when players have asymmetric attitudes. We focus on the Central Set since its boundedness properties and sensitivity to alternative preferences render it a good measure of general amicability.

*Proposition 14.* Suppose player B is neutral. If player A is amicable \( C_{\Gamma} \subset C_{V} \) while if player A is an adversary: \( C_{V} \subset C_{\Gamma} \).

*Proof.* Immediate given slight adaptations of Propositions 7 and 12.

As attitudes move from indifference to amicability, the defection incentive that gives rise to the PPD is attenuated and the potential for no-equilibrium increases. By Propositions 7 and 12 the Central Set expansion comes at the expense of the Dominant Sets. The following result, which combines amicable and adversarial players, is also immediate:

*Proposition 15.* Suppose player A is amicable and B adversarial. Ceteris paribus,

i). the larger the ratio \(-B_{\alpha} / B_{\beta}\) on the interval \([\alpha_{11}, \alpha_{22}]\), the smaller \( C_{V} \).

ii). the smaller the ratio \(-A_{\alpha} / A_{\beta}\) on the interval \([\alpha_{11}, \alpha_{22}]\), the smaller \( C_{V} \).

*Proof.* Again immediate given slight modifications to Propositions 7 and 12.
Together Propositions 7, 12, 14, and 15 suggest that the area of the Central Set is an intriguing metric of the aggregate “friendliness” of the players. Recall that the Central Set is always bounded when both players are indifferent or adversarial. As preferences move from adversarial to indifference to amicability the area of Central Set increases monotonically. When at least one player is amicable, the Central Set is no longer necessarily bounded, though it remains bounded under many well behaved utility functions and its area increases uniformly with increasing amicability as defined earlier. The Cobb-Douglas example of the following section will further illustrate this property.

3.3. A Cobb-Douglas Example

We begin with amicable preferences and to simplify exposition express B’s preferences in terms of primitive payoffs “a” for player A and “b” for own primitive payoff. The utility functions are then:

\[
A (\alpha, \beta) = \alpha^x \beta^{1-x} \\
B (a, b) = a^{1-y} b^y,
\]

where \(x\) and \(y\) are non-negative. Note that if \(x \in (0,1)\), Player A is amicable, and is an altruist if \(x \in (0,1/2)\).

By Corollary 1, \(A_{22}\) and \(B_{11}\) must intersect for \(D^A\) to exist and membership of a \(\pi_{12}\) in \(D^A\) is a necessary condition for the existence of a SPD. Let \(\bar{B}_{11}\) and \(\bar{A}_{22}\) be the utility levels associated with indifference curves \(B_{11}\) and \(A_{22}\). For an arbitrary \(\alpha = a, b \neq \beta\) in these indifference sets except for a point of intersection. Considering an arbitrary \(\alpha = a\), substituting \(\alpha\) for “a” in (7), defining \(\frac{\beta}{b} \big|_{a=a} R(\alpha; x, y)\), and rearranging (7) yields:
(8) \[ R(\alpha; x, y) = \left( \frac{A_{22}^y}{B_{11}^1} \right)^{\alpha^{1-(x+y)}}. \]

**Proposition 16.** The Dominant Sets are non-empty if either: i). both players are amicable but not altruists; ii). one is an altruist and one merely amicable, with \( x+y > 1 \).

**Proof:** For either i) or ii) above at \( \alpha = \alpha_{11}, \beta > b \) so \( R(\alpha_{11}; x, y) > 1 \). Moreover if either i) or ii) are satisfied \( x + y > 1 \) and \( \lim_{\alpha \to \infty} R(\alpha; x, y) = 0 \) so the indifference curves cross at some \( \alpha \), with the intersection obtained by solving \( R(\alpha; x, y) = 1 \).

**Proposition 17.** The Dominant Sets are empty if either: i). both players are altruists; ii). one is an altruist and the other merely amicable, with \( x+y < 1 \); iii). players are amicable with \( x+y = 1 \).

**Proof:** Reasoning similar to the Proof of (16) implies for i) and ii) the curves never intercept at \( \alpha > \alpha_{11} \). If \( x+y = 1 \), the \( \beta \) is independent of \( \alpha \), and \( \lim_{\alpha \to \infty} R(\alpha; x, y) = \lim_{\alpha \to 0} R(\alpha; x, y) = \frac{A_{22}^y}{B_{11}^1} \), a constant. Thus the curves are either parallel or coincide, but never cross.

These Propositions indicate that when both players are amicable and at least one is sufficiently *altruistic*, a utility Prisoner’s Dilemma will *never occur*. With these explicit utility functions we can also compute the “gains and losses” in the Dominant Set from different attitudes. For example, for simplicity letting \( x=y \) we can derive a lower bound for the reduction in the dominant sets when both players are amicable. Using (8) the intersection occurs at: \( \alpha^* = (\frac{A_{22}}{2})^{x/1-2x} (\frac{B_{11}}{1})^{1-x/1-2x} \). Recalling that \( A_{22} \) and \( B_{22} \) pass through \( \pi_{22} \) one can also write:
\[ \alpha_{22} = \left( \frac{A_{22}}{x^{1/2x}} \right) \frac{B_{22}}{1-2x}^{1-x/1.2x}, \text{ so that} \]
\[ \alpha^* / \alpha_{22} = \left( \frac{B_{11}}{B_{22}} \right)^{1-x/1-2x} > 1. \]

Therefore player A’s utility dominant set relative to primitive dominant set, is reduced by at least the area: \( \beta_{11} \cdot \alpha_{22} \left[ \left( \frac{B_{22}}{B_{11}} \right)^{1-x/2x-1} - 1 \right] \). To this it must be added the area beyond point \( \alpha^* \) and below the indifference curve \( A_{22} \). Computing this integral yields:

\[ \left( \frac{A_{22}}{1-x/2x-1} \right) \cdot \alpha^*^{1-2x/1-x} = \left( \frac{1-x}{2x-1} \right) \left( \frac{B_{11}}{B_{22}} \right)^{1-x/2x-1} \]

In spite of curve \( B_{11} \) going to zero, Player’s A utility dominant set (which remains unbounded) does not have finite measure. A more precise bound may be obtained by computing the area outside the UDS between \( B_{11} \) and the vertical line passing through \( \pi_{11} \) until an \( \alpha' \geq \alpha^* \). We shall not pursue this computation here. The above results suggest the following proposition.

**Proposition 18.** If \( x=y \) and \( 2x > 1 \), the lower bound to the reduction in Player’s A utility dominant set is increasing with the ratio \( r_B = B_{22} / B_{11} \) whenever \( r_B > \left( \frac{A_{22}}{\beta_{11}} \right)^{1-x/2x} \) and decreasing if the reverse occurs.

Proof. The increasing result is immediate. For the decreasing it suffices to compute the derivative, w.r.t. \( r_B \), of the combined area.

The main importance of Proposition 18 is to emphasize that it is the normalized utility values at points \( \pi_{11} \) and \( \pi_{22} \) that are crucial in determining the distortions in the relevant PD sets.
Adversarial Preferences

For adversaries additional flexibility is obtained if we re-specify the Cobb-Douglas utility functions as follows:

\[ A (\alpha, \beta) = \alpha^x \beta^y \]
\[ B (\alpha, \beta) = a^w b^z, \]

with all exponents non-negative and \( x \) and \( z \) less than 1. Note that if \( y > x \) and \( w > z \) both players have strong global enmity. In this case all indifference curves emanate from origin and have a single-crossing property. If \( x > y \) and \( w < z \), the concavities of indifference curves are reversed, though all indifference curves still emanate from the origin. Finally, if \( x > y \) and \( w > z \), or \( x < y \) and \( w < z \), multiple crossing of indifference curves are possible and a set of complex possibilities arise.

Now return to the strong global enmity case \( (y > x \text{ and } w > z) \). Proposition 9 states that \( F^V \) is either empty or bounded. It follows that the area of \( F \) can be obtained.

**Proposition 18.** With strong global enmity the area of \( F^V \) shrinks from an infinite (Lebesgue-measure) value to \( \alpha^{22} \beta^{22} \frac{(wy - zx)(w+z)(x+y)}{22}. \)

**Proof:** The \( A_{22} \) and \( B_{22} \) curves intercept at \((0, 0)\) and \((\alpha_{22}, \beta_{22})\). With \( \alpha \) on the ordinate \( A_{22} \) is “below” \( B_{22} \) on the interval \((0; \alpha_{22})\) and we have the function: \( F(\alpha) = (\overline{A}_{22})^{-1/y} \alpha^{x/y} - (\overline{B}_{22})^{1/z} \alpha^{w/z} \). Integrating this function on the interval and recalling that: \( (\overline{A}_{22})^{-1/y} \alpha_{22}^{x/y} = \beta_{22} = (\overline{B}_{22})^{1/z} \alpha_{22}^{w/z} \), one arrives at the result. Notice that the relationship between the exponents ensures that \( wy - zx > 0 \).
4. Summary and Conclusion

Prisoner’s dilemmas provide a fundamental paradigm of the tension between individual and collective rationality. Analysis of their structure and operation has provided insight into issues ranging from the public goods problem to arms races. Yet the predictive power of the paradigm depends critically on implicit assumptions on the nature of the mapping from observable primitive payoff to unobservable final utility. When unobservable final utility depends only on own-primitive-payoff the equilibrium of a primitive-payoff-game and the associated utility-games are identical. Under this circumstance the specific properties of the unobservable utility function are immaterial for predictions of strategy choice and a primitive game with a PD equilibrium is a perfect proxy for the unobservable final utility game. However, when linkages exist between the primitive payoff of one player and the utility of another, PD equilibrium in the observable game may not correspond to equilibrium in the utility game. Moreover, as discussed previously, a large body of experimental evidence is generally inconsistent with pure indifference of players to the payoffs of others.

This paper explores the implications of two types of linkages between the players’ final utility and the other player’s primitive payoff: adversarial and amicable preferences. We demonstrate that such non-indifference generates specific non-congruencies of the “primitive-dominant-set” and “utility-dominant-sets,” which has the consequence of mapping apparent PD’s into other (non-PD) equilibrium. On the other hand, utility PDs may arise in games that do not exhibit PD structure in primitive payoffs.

To appreciate the implications of this non-congruence consider a standard two-person PD in observable payoffs. Both players are apparently better off through
cooperation than competition, though the temptation of defection precludes cooperation as a non-cooperative equilibrium. It would therefore seem that players have incentive to create institutions that can support the Pareto-superior payoffs. Indeed, we observe many situations where institutions supporting the Pareto Superior outcome are created and the gains from cooperation can be realized. However, we also observe many situations where such institutions do not emerge and where all players appear to reap inefficiently low returns. This paper proposes a new explanation for such phenomenon. Namely, that the joint-strategy Pareto superior utility-payoffs do not in fact exist. The Kyoto Protocol, and the “2003 WTO Cancun Meeting”, could be examples of this phenomenon. On the flip side, equilibrium that appears Pareto optimal in primitive payoffs may in fact be PDs in utility payoffs. Ongoing research applies this theory to the successes and failures of a variety of trade and political institutions.
References


Últimos Ensaios Econômicos da EPGE


