

# The Topology of Conflict and Co-operation

**Sam Perlo-Freeman**

School of Economics,  
Bristol Business School  
University of the West of England, Bristol  
Email: Samuel.Perlo-Freeman@uwe.ac.uk

**May 2006**

## Abstract

The class of simultaneous 2x2 pure-strategy ordinal games (which include well-known games such as Prisoner's Dilemma, Chicken and Stag Hunt) have received considerable attention, including complete classification schemes by amongst others Rapoport & Guyer (1978) and Robinson & Goforth (2005). This paper focuses on a particularly pertinent subset of these games, described as the 'Co-operate-Defect' (C-D) games, which are characterised by each player having a dominant preference for a particular strategy by the *other* player. These games are therefore relevant in a number of contexts, including arms race games and collective action problems. The C-D games may be efficiently classified by assigning each player one of six distinct types, a classification that cannot be naturally extended to the full class of 2x2 games. The six types and the resulting game forms are analysed, and the subclass of CD games are identified within a topological structure for the 2x2 games devised by Robinson & Goforth (2005).

## 1. Introduction

The humble 2x2 game has been the subject of a great deal of study in the social sciences. They have been applied in particular to the phenomenon of an 'arms race' between two hostile powers, and more generally, they are used as a tool of analysis of conflictual situations. Study of these games typically starts (and sometimes ends) with the notorious 'Prisoner's Dilemma', although other game forms have also received attention. While obviously representing a monumental simplification of complex situations, they can be perhaps seen as a simple, but remarkably fruitful metaphor for understanding what is going on.

A few attempts have been made to classify all of the possible 2x2 *ordinal* games, where only the preference ordering of each player between the four possible outcomes is considered. A *Taxonomy of the 2x2 Games*, by Rapoport & Guyer (1978) was an early example. More recently, David Robinson and David Goforth (2005) in *The Topology of the 2x2 Games*, go a step further by underpinning their classification of the 2x2 games with a topological structure, by defining two games as 'adjacent' if one can be obtained from the other by swapping the preference orderings of two adjacent outcomes for one player.

Many of the 2x2 games most frequently the subject of theoretical and experimental analysis – including the PD, Chicken and Stag Hunt – share the characteristic that the choice for each player is between a strategy that is beneficial to the other (Co-operate), and one that is harmful. (Defect). This paper generalises this idea, focusing on the subset of the 2x2 ordinal games, which I shall call the ‘Co-operate-Defect’ (C-D) games, where each player has a consistent preference as to which strategy the *other* player adopts. These games have a natural interpretation in a wide variety of economic and political situations, wherever agents are faced with a choice between actions which carry a clear positive or negative externality for other agents. Examples include arms races, where each player will, generally speaking, prefer their rival to choose a lower rather than a higher level of arms. More generally, they are highly relevant for the analysis of conflict, where they open the way to considerations of threats, promises, rewards, punishments, deterrence, credibility and so forth. (See e.g. Schelling, 1960). Other examples are situations of contribution to a collective endeavour, or solving a common problem, i.e. collective action problems, such as those considered in Todd Sandler’s (2004) recent *Global Collective Action*.

The next section briefly reviews some of the literature relating to the classification of 2x2 games, and specifically relating to the types of situations, such as arms races, that can be described by C-D games. Section 3 introduces the Co-operate-Defect (C-D) subset, and shows that they can be succinctly classified by assigning each player to one of six well-defined preference-ordering types. The nature of these types, and their possible relation to real-world conflict situations is discussed.

One of the considerations developed in this analysis is the possibility that the nature of long-term interactions between two agents (e.g. countries) may change over time, whether as a result of changes in the priorities of the agents, the external constraints, or the underlying strategic situation. This could lead to a change in one or both players’ ‘type’, and thus in the relevant game-form. In some cases, a small change in preference ordering by one player can lead to a radical change in the resulting Nash equilibrium.

This question of ‘nearness’ of game forms to one another corresponds very naturally to a topological structure for the 2x2 games described by Robinson & Goforth (2005), which defines two games as ‘adjacent’ if one is obtained from the other by swapping two adjacent preference orderings for one player. Section 4 describes Robinson & Goforth’s topology, and section 5 locates the C-D games within Robinson & Goforth’s structure, elucidating the topological linkages between them. It is also shown that the specification of 2x2 games by assigning separate player types described in section 3 breaks down when extended from the C-D subset to the full space of 2x2 games. Section 6 completes the analysis of the topology of the C-D games by demonstrating that the graph of this subset can be drawn on the surface of a 4-holed torus, in comparison to the 37-holed torus that Robinson & Goforth demonstrate is required for the full set. Section 7 concludes.

## **2. Literature survey**

The Prisoner’s Dilemma (PD) holds a central position in the study of Game Theory, and continues to exercise a fascination amongst students and practitioners of the social sciences, evolutionary biology, and beyond. It has been applied to a remarkably

diverse range of situations, and has had whole books devoted to it, (E.g. Poundstone 1992, Axelrod 1984). However, the PD is just one of many possible game forms that could plausibly represent such strategic interactions, and considerable analysis has been devoted both to the properties of other simple 2x2 games, and to their classification.

Thomas Schelling's *The Strategy of Conflict* (Schelling, 1963) is one key early work that seeks to apply simple game theoretic models such as the Prisoner's Dilemma, Chicken, Co-ordination games, and many others, to problems of strategic interactions between nations, groups and individuals. This work develops a number of key themes that form the basis of much subsequent analysis, such as threats, credibility, deterrence, and so forth. However it does not seek to give a complete classification of the 2x2 games.

It is not hard to show that there are in fact 78 possible 2x2 games, up to re-labelling of strategies and players – with four outcomes there are 24 possible preference orderings for each player, giving an initial total of 576 possible combinations for the two players. This must be divided by four to allow for re-labelling of either player's strategies, reducing the total to 144. If we then treat the players as interchangeable, this gives the figure of 78. This is shown in Rapoport & Guyer (1966), and these authors provide a more detailed strategic taxonomy in Rapoport & Guyer (1978). Here, they classify games according to whether they have 0, 1 or 2 Nash equilibria, and then further classify those with a single Nash equilibrium according to the 'stability' of the equilibrium, based on whether the equilibrium is vulnerable to either threat or force by one or other player. A threat-vulnerable equilibrium is one where one player could potentially induce the other to change strategy by a *threat* of a (self-punishing) change herself, while a force-vulnerable equilibrium is where a player could induce a change of strategy by actually changing her own strategy away from the equilibrium. This is not entirely satisfactory – for example, the Prisoner's Dilemma is classified as having a 'strongly stable' equilibrium, even though it is well known that the inefficient NE can potentially be escaped from in the iterated game. The classification also seems to group together games with some quite significantly different strategic properties.

Lichbach (1990) and Plous (1993) are amongst those who investigate different possible game forms in relation to military arms races – one of the traditional applications of the Prisoner's Dilemma, especially with regard to the Cold War superpower confrontation. Plous (1993) suggests that the Cold War arms race was not a Prisoner's Dilemma at all, but what he referred to as a 'Perceptual Dilemma' – that is, where the real game was one of 'Stag Hunt' (or Deterrence) – that is, where each player's first choice is (Co-operate, Co-operate), with (Defect, Co-operate) the second choice. That is, each country had a strong desire not to fall behind in the arms race (Co-operate, Defect), but did not have a strong desire to dominate the other. In this game, there are two possible Nash equilibria, the efficient equilibrium of (C,C), and the inefficient equilibrium of (D,D). However, Plous suggests that each country *perceived* the other as having the preference ordering of the Prisoner's Dilemma, that is, of seeking an absolute military advantage. Hence, the inefficient equilibrium results. Lichbach (1990) discusses four different preference orderings that could occur in an arms race situation, based on different utility functions for each country relating to their own and the other country's level of arms. These different utility functions

lead to four different games, of Prisoner's Dilemma, Deadlock, Chicken and Stag Hunt.

Both of these papers, as well as Hamburger (1969)'s analysis of 'separable games', raise the question of analysing games in terms of the *preference ordering of each player separately* – that is, we have the Prisoner's Dilemma ordering of (D,C), (C,C), (D,D), (C,C), the Stag Hunt ordering where the top two are swapped, the Chicken ordering where the bottom two are swapped, and so forth. This idea is the basis of the classification of the 'Co-operate-Defect' games in the next section. Plous's 'perceptual dilemma' also underlines the significance of a small change in one or other player's preference ordering for the strategic outcome – or indeed, of a change in either player's *perception* of the other's preferences. This theme is also taken up in the next section, and also links this analysis to the topological framework developed by Robinson & Goforth (2005).

### **3. The Co-operate-Defect games**

A number of familiar games are characterised by each player having a choice between "Co-operating" and "Defecting", that is a choice between friendly and hostile behaviour. These include the Prisoner's Dilemma, Chicken, and Stag Hunt. We may characterise these games by the fact that each player has a clear (or dominant) preference for a particular strategy to be chosen by the *other* player. Thus, the strategy preferred by the other player can be described as 'Co-operate', and the alternative as 'Defect'. These games are naturally suited to the field of conflict analysis. In particular, in any 'arms race' situation between two more or less hostile powers, Co-operate and Defect have a natural interpretation in terms of the choice between Low and High levels of armaments. It is interesting to consider what are the full possible range of games that can occur in this Co-operate-Defect framework.

Not all games can meaningfully be characterised as having "Co-operate" and "Defect" strategies'; for example in a Co-ordination type game, each player gains or loses depending on the *combination* of strategies employed.

#### **Defining and classifying the C-D games**

I define a game as a Co-operate-Defect (C-D) game if for each player X, there exists a strategy of the other player, which we call 'Co-operate', such that for each strategy for player X, he prefers the other player to choose Co-operate. We call the other strategy for each player 'Defect'.

Hence the Row player for example prefers (C,C) to (C,D), and (D,C) to (D,D), and similarly for column. A game is a C-D game if there exists a labelling of the strategies for each player that satisfy these restrictions. We may say that Row has a dominant preference for Column to Co-operate, analogous to the concept of a dominant strategy.<sup>1</sup>

---

<sup>1</sup> It is interesting to compare this class of games with Hamburger (1969)'s class of 'separable games'. These are games where the outcome of each combination of actions by the two players can be separated into the sums of the outcomes of the individual actions by each player. There are 12 of such games, corresponding to games between players of type 1,4,5 and 6 (Prisoner, Appeaser, Warrior, Pacifist). These are the games where each player has both a dominant strategy, and a dominant

These restrictions leave 6 possible preference orderings, listed below, and since the labelling of the strategies is now significant, each combination of preference orderings for the two players gives a different game, giving a total of 36 possible games. If we treat the players as interchangeable, this number is reduced to 21. The preference orderings (from the point of view of the Row player) are:

- 1) (D,C) (C,C) (D,D) (C,D) (as in the Prisoner's Dilemma)
- 2) (D,C) (C,C) (C,D) (D,D) (as in Chicken)
- 3) (C,C) (D,C) (D,D) (C,D) (as in Stag Hunt)
- 4) (C,C) (D,C) (C,D) (D,D)
- 5) (D,C) (D,D) (C,C) (C,D)
- 6) (C,C) (C,D) (D,C) (D,D)

Where the most-preferred combination is listed first. The mirror image orderings apply from the point of view of the Column player. The last two are perhaps slightly curious, with the second-best outcome involving the other player defecting, but they satisfy the C-D restrictions.

#### Describing the preference orderings

Each of these six orderings thus gives a well-defined player 'type', with its own particular strategic characteristics, and the interaction between these types describes the strategic character of each game. I shall now briefly describe each type, with particular reference to the context of an 'arms race' type situation, or more generally an international conflict situation.

##### 1) The "Prisoner"

Preference ordering (D,C), (C,C), (D,D), (C,D)

This is the classic preference ordering for the Prisoner's Dilemma. They have a dominant strategy of Defecting. In the context of an arms race, this means that they will always consider the benefits of gaining military advantage, or avoiding military disadvantage, to outweigh the costs and risks. It has frequently been argued to be the preference ordering of the two superpowers during the Cold War, though as noted above this has been questioned. While a Prisoner has a dominant strategy of Defecting in a one-shot game, it is well known (e.g. Axelrod, 1984) that it is possible for co-operation to emerge as a long-term strategy in the iterated game between two Prisoners.

##### 2) The "Chicken"

Preference ordering (D,C) (C,C) (C,D) (D,D)

This is obtained from the Prisoner ordering by swapping the two least-preferred outcomes in the preference ordering. In an arms race context, the Chicken is keen to

---

preference for the other player's strategy. Thus each player's action has a fixed effect on both themselves and their opponent.

gain a military advantage if they can get it, but fears (or perhaps cannot sustain) outright confrontation (represented by (D,D)). In other words, they will back down if confronted, but take advantage if they can. It might be conjectured, for example, that the Cold War arms race ended when the Soviet Union's preference ordering switched from that of a Prisoner to that of a Chicken – their economy was no longer able to sustain the enormous costs imposed by the arms race.

### 3) The “Deterrer”

Preference ordering (C,C) (D,C) (D,D) (C,D)

This is the ordering for Stag Hunt, but I have used the term Deterrer to emphasise the arms race/conflict analysis context. This ordering is obtained from the Prisoner ordering by swapping the top two preferences. The Deterrer might be seen as desiring peace, and placing little value on gaining a military advantage, but as fearing being at a disadvantage. They will thus prefer arms control (C,C), but will seek to match the military spending of their rival. Thus, the Deterrer is the mirror image of the Chicken, who places greater value on gaining an advantage than on avoiding a disadvantage. As noted above, Plous (1993) suggested that the US and the Soviet Union actually held the Deterrer ordering for much of the Cold War.

### 4) The “Appeaser”

Preference ordering (C,C) (D,C) (C,D) (D,D)

This ordering is obtained from Chicken by swapping the top two outcomes in the preference ordering, and from Deterrer by swapping the bottom two. The Appeaser desires co-operation and fears confrontation. However, given the choice, they would prefer military advantage to disadvantage. The Appeaser has a dominant strategy of Co-operating.

### 5) The “Warrior”

Preference ordering (D,C) (D,D) (C,C) (C,D)

This is obtained from the Prisoner ordering by swapping the *second and third* preferences. The Warrior has an absolute preference for Defection (higher arms). Thus, unlike the Prisoner, they will necessarily follow a strategy of always Defect, even in the iterated game. An interpretation could be that they are sufficiently confident that they can win any arms race, and desire the power that goes with that more than they desire peace. One might consider, for example, that the USA under President George W. Bush exhibits the Warrior preference ordering; they would rather that potential rivals did not develop more advanced weaponry, but they appear to show little interest in arms control agreements, preferring to proceed with ballistic missile defence technology, new generations of nuclear weapons, etc., as well as massive increases in military spending; they appear to prefer the benefits of global hegemony to those of arms control. One might also apply the “Warrior” analysis to the United States’ strategy in the build-up to the Iraq war; they had essentially decided to attack come what may, hence any degree of “co-operation” by Iraq with the UN weapons inspectors was irrelevant.

## 6) The “Pacifist”

Preference ordering (C,C) (C,D) (D,C) (D,C)

This is obtained from the Appeaser ordering by swapping the 2<sup>nd</sup> and 3<sup>rd</sup> preferences. The opposite of the Warrior, the Pacifist has an unconditional preference for Co-operating. Other things being equal, they prefer Co-operation from the other player, but nothing could possibly convince them to Defect.

While it might be imagined that a Pacifist preference ordering would most likely come from moral conviction, it could equally arise in an ‘arms race’ (or non-arms race) scenario where a country believed itself too small for military defence to have any efficacy in providing security against neighbours. Thus, Costa Rica decided in 1948 to abolish its armed forces. Costa Rica has had disputes with neighbours from time to time, including Nicaragua, but the armaments decisions of any such neighbours are irrelevant, as Costa Rica sees little value in possessing the tiny military force it could afford regardless.

### Other interpretations of the player types

The analysis of conflict situations is not the only application of the C-D game framework. We may also see the player types in terms of their position in a collective action problem. Let us suppose that two or more ‘players’ (countries, groups, individuals) must decide whether to contribute some effort to achieving a mutually desirable goal. Each player will gain a non-excludable *benefit*  $B(n)$ , depending on the number of players  $n$  who contribute to the effort, where  $B(n)$  is increasing in  $n$  and  $B(0)=0$ . (Let us suppose for simplicity that there are only two players), but will incur some private cost  $C$  if they participate. (This analysis bears some relation to an approach taken by Sandler (2004).) The six player types occur in the following circumstances:

Warrior:  $C > B(2)$  – i.e. the private cost is not even worth the benefits of mutual co-operation compared to mutual defection.

Prisoner:  $B(2) > C > B(k)-B(k-1)$  – the private cost is more than the immediate gain from the player herself co-operating, but the benefits of mutual co-operation outweigh the private costs - i.e. the classic collective action problem.

Deterrer:  $B(2) - B(1) > C > B(1)$  – this implies some non-linearity in the benefits function, so that the gains from collective action are greater than those from individual action – perhaps there is some ‘threshold’ level below which action will have limited efficacy. Another possibility is if the Deterrer is motivated by a sense of fairness, so that they would incur a subjective cost in being the only one to contribute.

Chicken:  $B(1) > C > B(2)-B(1)$  – the opposite of the deterrer; perhaps there is some major gain from *something* being done, but diminishing marginal returns to effort thereafter.

Appeaser:  $B(k) - B(k-1) > C > 0$  – always worth taking action, but if there is only going to be one person doing it, you'd still rather the other person bears the cost.

Pacifist:  $C < 0$  – there are actually private *benefits* to taking action.

Differences between player types in a particular situation could most obviously occur through differences in the private costs to each player of taking action. However they could also occur through different perceptions of the common benefits.

### Interactions between the types

The symmetric games, Prisoner's Dilemma, Chicken and Stag Hunt (Deterrence) have been fairly thoroughly analysed by other writers (e.g. Schelling, 1963, Lichbach, 1990, Axelrod, 1984, etc.) In the PD, most attention focuses on the iterated game, on which there is a vast literature, perhaps most famously Axelrod (1984), which argues for the effectiveness of the 'Tit-for-Tat' strategy. The game of Chicken has received less attention, but there have been some interesting experimental studies, e.g. Possajennikov (2002), Rapoport & Chammah (1966). Less attention has been given to asymmetric games, although Robinson & Goforth (2005) for example discuss the 'Alibi' games. But since there is no particular reason to suppose that both actors in a given situation face identical priorities and constraints, the asymmetric games deserve as much attention.

Games involving Warriors and/or Pacifists are mostly uninteresting from a game theoretic point of view. Warriors will always Defect, and Pacifists will always Cooperate, in both the one-shot and the repeated game, and whether or not enforceable agreements are possible.<sup>2</sup> The other player, whatever type they are, must make the best choice available to them knowing this. The significance of these types, perhaps especially the Warrior, is in realising when you are dealing with them, and that therefore approaches suited to, say, a PD situation, are no longer relevant.

### Prisoner vs Chicken

In the one-shot game, the Prisoner has a dominant strategy of defecting. As the Chicken's lowest preference is (D,D), this player will therefore be forced to cooperate. The Nash Equilibrium is therefore (D,C) with the Prisoner listed first, giving a payoff of 4 to the Prisoner, but 2 to the Chicken. The Prisoner 'wins' this game.

It is not *a priori* obvious what would be the outcome of the iterated game. On the one hand (assuming each player knows the others' preference orderings), the Prisoner can easily adopt a strategy of always Defect, and the Chicken would appear to have little

---

<sup>2</sup> This may not necessarily always be the case. For example a Pacifist playing a Chicken (or with more difficulty a Chicken) might still choose to play 'against type' in an iterated game, carrying out punishment defections in response to defections by the other player, in the hope of inducing cooperation, to obtain the best possible payoff. However, this strategy is highly costly, as the punishment involves dropping down *two* places in the Pacifist's own preference ordering, compared to what they would get if they accept the other player's defection. It is also worth noticing in passing that Warrior vs Prisoner is slightly interesting in that it is the only C-D game that is also zero-sum.



choice but to Co-operate. The 'Tit-for-Tat' strategy for the iterated PD would appear not to be available to the Chicken, as they lack a 'credible threat'. On the other hand, it might be possible that if a Chicken were to resolutely stick to a tit-for-tat strategy, in spite of the short-term cost, they could induce the Prisoner to compromise on long-term mutual co-operation. Such a question can perhaps best be answered with experimentation, either with computer simulations or live subjects. It seems likely though that, in comparison with the PD, the Prisoner has an advantage over the Chicken.

The Prisoner-Chicken game could be seen as a model of how an arms race or conflict situation *ends*. In an ongoing Prisoner's Dilemma, if one side finds themselves unable to sustain a continuing arms race or conflict, so that (D,D) becomes their lowest preference (swapping the bottom two preferences), then this player turns into a Chicken, and thus backs down. This illustrates how a small change in one player's priorities or constraints can fundamentally change the nature of the game.

#### Prisoner vs Deterrer

This is sometimes called the 'Alibi' game, the idea being that one of the 'prisoners' (the Deterrer) has a cast-iron alibi that will get them off the lesser charge that they'd face if both prisoners remain silent.

In this case, the change in preference ordering of one player does not change the Nash Equilibrium of (D,D), since the Prisoner still has a dominant strategy of Defecting. The interesting differences might come about in the repeated game. On the one hand, the Prisoner need have less fear of cheating by the Deterrer, since (C,C) is the latter's preferred option, so one might expect it to be easier to establish long-term patterns of co-operation. On the other hand, the Deterrer has more to lose from a break-down of co-operation, so the Prisoner might do well to attempt occasional Defections (as opposed to a simple Tit for Tat strategy), as the Deterrer has more incentive to be forgiving and attempt to restore a Co-operative pattern. Tit for Tat is clearly available to the Deterrer as a meta-strategy however, since their threat of Defection is credible.

#### Chicken vs Deterrer

This game is unique amongst the C-D games in that it doesn't have a Nash Equilibrium. It is therefore extremely difficult to predict what outcome will result in either the one-shot or the iterated game. In the latter it is possible that, somewhat paradoxically, a strategy of "always Defect" could be quite successful for the Deterrer, as this would compel the Chicken to Co-operate, while a strategy of "always Co-operate" could work well for the Chicken. The problem is that neither strategy is credible, in that each player has an incentive to "cheat". It is also possible that a tit-for-tat strategy could work for either player, although if this did achieve mutual co-operation, the Deterrer would have some reason to distrust the Chicken, who would have a continuing temptation to cheat.

#### Games involving the Appeaser

If an Appeaser plays either another Appeaser or a Deterrer, the unique Nash Equilibrium is (C,C), the most preferred outcome for each player. It is hard to see why

either player would ever deviate from this. In Prisoner vs. Appeaser, each player will surely follow their dominant strategy of Defecting and Co-operating respectively (Prisoner “wins”), and it is hard to see what the Appeaser might have to offer or threaten the Prisoner in a repeated game. (Though as with Prisoner vs. Chicken, it is just possible that the Appeaser could learn to ‘play against type’ and play tit-for-tat to try to induce co-operation in the Prisoner.)

Chicken vs. Appeaser is slightly more interesting. The Appeaser’s dominant strategy is to Co-operate, and therefore the Chicken will defect, giving them their most preferred option as the Nash Equilibrium. The question is whether the Appeaser has any way out of this in the iterated game; the problem is they do not have a credible threat of Defection as a punishment strategy (at least myopically), but on the other hand if such a player does go against type and adopt, say, a ‘tit for tat’ strategy (or even an always defect strategy), then although they would be harming themselves, they would also be harming the Chicken, and so might be able to induce a change in behaviour. Thus the Appeaser has some incentive to attempt to, as it were, ‘cut off their nose to spite their face’.

#### Discussion, and ideas for experimentation

The question of what would happen in repeated versions of each of these games is interesting, and could be a fruitful subject for both theoretical modelling and experimentation. What is interesting is that in most of these games, there is at least the potential for ‘tit for tat’ strategies to be effective, *even by a weaker player* – that is, one whose best (myopic) response to Defect is to Co-operate. The question is whether such a long-term strategy, which involves potentially accepting considerable short-term pain, can be sustained in the face of an equally determined Prisoner or Chicken, for example. (Punishment strategies are of course totally ineffective against a Warrior). For example, can an Appeaser go against their type and, as it were, acquire a backbone, in the hope of inducing long-term co-operation?

This suggests an experimental setting where the participants are assigned different types (say, amongst the four middle types of Prisoner, Chicken, Deterrer and Appeaser), and matched against each other in iterated games. A variation on this treatment would be to make each player’s type a matter of *private knowledge*, so that a player might have an incentive to try to ‘mask’ their type to induce desirable behaviour in the other player; each player, meanwhile, would have to try to deduce their opponent’s type from their behaviour. I advance the tentative hypothesis that (in infinite duration games with sufficiently high discount factors) *tit-for-tat may be a generally effective and hard-to-beat strategy for this scenario, for all player types*. The attractiveness of this strategy comes from the nature of the C-D games: for all player types, Co-operate represents a ‘reward’ and Defect a ‘punishment’, so that tit-for-tat neatly encapsulates the notion of reciprocity, rewarding good behaviour and punishing bad.

#### **4. Robinson & Goforth’s Classification and Topology**

The above classification of the C-D games deals with 21 of the 78 2x2 games. It provides a compact classification, and also illustrates the interconnections between the games; each player type can be obtained from other player types by swapping two

adjacent preferences. Such a swap in a player's preference ordering could easily be seen as resulting from a marginal change in their priorities, constraints, or perceptions of the strategic situation. Thus, the analysis of the relationship between games formed from 'nearby' preference orderings is potentially of considerable strategic significance.

This concept of 'nearness' or adjacency of games forms the basis of Robinson & Goforth's (2005) classification, indexing and topology of the 2x2 games. In *The Topology of the 2x2 Games*, the authors produce a 'periodic table' of the 2x2 games that underlines the linkages between the games, consistent with a natural topological structure. This section briefly describes this classification and topological structure, and the next section will locate the C-D games within that structure.

### Indexing the games

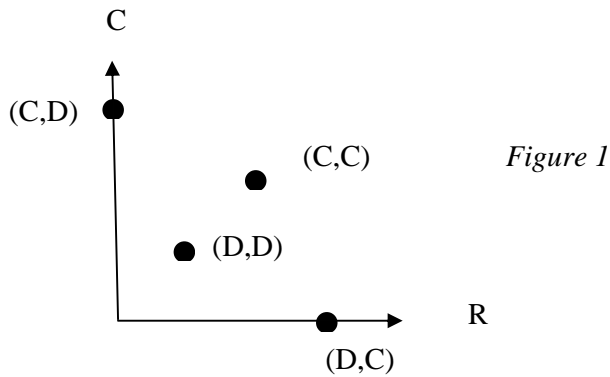
As discussed above, various classifications of the 2x2 games have been produced, including Rapoport & Guyer (1978), who classify the 78 distinct games (up to relabelling of strategies and players) according to their strategic properties. Robinson & Goforth take a rather different approach, indexing the games according to their topological relationship, based on nearness of preference orderings. In contrast to Rapoport & Guyer, they base their classification on the 144 distinct games that result if we do not treat the players as interchangeable; this is to fit in with the topological structure they impose on the games.

Robinson & Goforth index each 2x2 ordinal game using three index numbers, as follows. Consider first the pay-off matrix representation. Suppose first that we fix, without loss of generality, Row's first preference (payoff 4) to a particular position in the matrix. There are then first of all four possibilities for where Column's first preference occurs: (1) in the diagonally opposite corner, (2) in the same row (3) in the same position or (4) in the same column. These four possibilities give the first of the three index numbers, from 1 to 4. This divides the space of games into four 'layers' of 36 games each.

The second and third index numbers reflect the orientation of the other three preferences of the Row and Column players respectively. The Prisoner's Dilemma is given index number (1,1,1) (the first number being 1 as in this game the players' top preferences are diagonally opposite). The other row indices are generated by alternately swapping the bottom two preference orderings, and the second and third preference orderings, for the Row player, leaving the first preference fixed, while the column indices are generated by doing the same thing for the Column player. This cycles through all 6 possible preference orderings for each player, giving all the games with index (1,Y,Z). The other (X,1,1) games are obtained by swapping the rows and/or columns of the Prisoner's Dilemma payoff matrix as appropriate, and the other games on each layer generated in the same manner as on layer 1. The full set of 144 games, arranged in their four layers, is shown in Appendix 1, taken from Robinson & Goforth (2005).

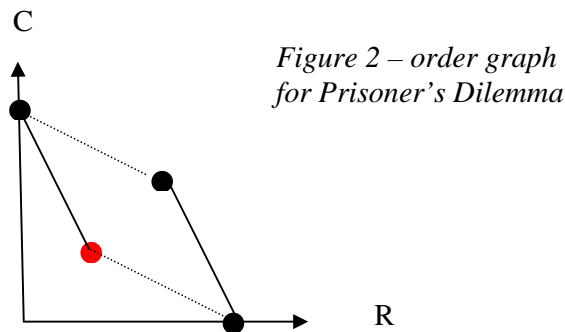
### Representing the games

Robinson & Goforth employ a convenient graphical representation of the 2x2 games that gives each distinct game a unique representation. First of all the four outcomes



are represented as points on a 4x4 discrete grid of dots, with Row's payoff measured along the horizontal axis and Column's along the vertical, with (1,1) rather than (0,0) at the origin. For example, figure 1 above shows the outcomes for the Prisoner's Dilemma. We have the points (3,3) and (2,2) if both players Co-operate or both Defect respectively, (4,1) for (C,D), and (1,4) for (D,C):

The players' strategies are shown as follows: we join pairs of points that represent the same strategy choice by Row with dotted lines, and pairs of points that represent the same strategy choice by column with solid lines (See figure 2). Thus, (C,C) is connected to (C,D) with a dotted line, and (D,D) to (D,C), while the pairs ((C,C),(D,C)) and ((C,D),(D,D)) are joined with solid lines. Finally, the Nash equilibria, if any, are marked. (I have marked it with a red dot below).



As well as providing a handy visual form for a game, this representation gives a unique 1-1 correspondence between distinct 2x2 ordinal games and their graphs; whereas, with the traditional pay-off matrix representation, each game has four possible representations, depending on the order of the rows and columns.

The topological structure

The games are given a topological structure by defining two games as being adjacent if one is obtained from the other by swapping the preference ordering of two adjacent outcomes for one of the two players. Thus the games are made the vertices of a graph, with an edge drawn between each pair of adjacent games. Each game therefore has six

neighbours. We call the operation of swapping the bottom two preferences for player  $X$  by  $X_{12}$ , the middle two by  $X_{23}$ , and the top two,  $X_{34}$ . (Where  $X = R$  for the Row player,  $C$  for the Column player).

Robinson & Goforth use this system to explore the properties of the graph. Each layer is generated from the game  $(X,1,1)$  by the operations  $R_{12}$ ,  $R_{23}$ ,  $C_{12}$  and  $C_{23}$ , that is, those operations that leave the top preference for each player unchanged. Within each layer, each row or column of 6 is connected cyclically, since the sequence of operations  $R_{12}$ ,  $R_{23}$ ,  $R_{12}$ ,  $R_{23}$ ,  $R_{12}$ ,  $R_{23}$  lead back to the identity operation, and the same of course for column. The resulting graph structure is therefore topologically a torus.

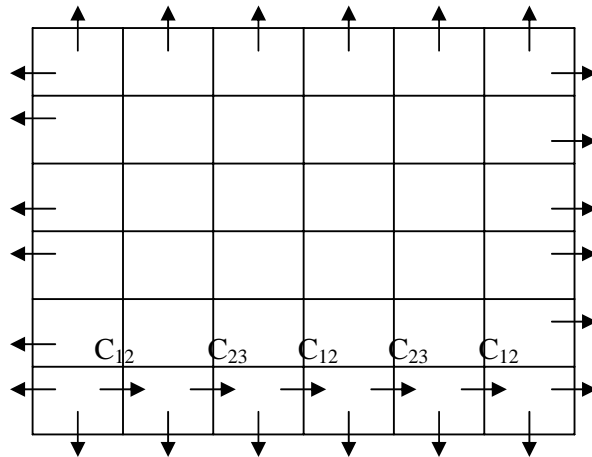


Figure 3

(See figure 3).

The arrows at the left-hand edge of each row indicate that the leftmost square of the row is connected to the rightmost (via a  $C_{23}$  operation), while the top and the bottom of each column are connected via a  $R_{23}$  operation.

Each Layer is divided into two-by-two “tiles” of games, connected using only the operations  $R_{12}$  and  $C_{12}$ . The operations  $R_{34}$  and  $C_{34}$  connect the layers together in a series of “Pipes” and “Hotspots”; a Pipe is where four tiles, one on each layer, are circularly connected to each other using the  $R_{34}$  and  $C_{34}$  operations. Each Pipe is also topologically a torus: four circles of four, connected in a circle. A Hotspot is where two tiles on opposite Layers are doubly connected to each other. These occur in matching pairs. When all the Pipes and Hotspots are added to the four Layers, the result is a 37-holed torus that is required to draw the full graph of 144 games.

For more details of Robinson & Goforth’s topology and analysis, consult the aforementioned book.

### **5. The C-D games within the Robinson & Goforth topology**

The C-D games find a natural place within the graph structure and the indexing system used by Robinson & Goforth. First of all, it is worth considering the graph structure of the six individual player types. Figure 4 shows the connections between the types, with the operations involved in moving between them. Here  $X_{ij}$  represents

swapping the  $i$ 'th and  $j$ 'th payoff for player X. Note that the graph edges are undirected, as a simple swap operation is its own inverse.

The full graph of the 36 C-D games (with players treated as non-interchangeable) involves meshing together two copies of this structure. This will be looked at more

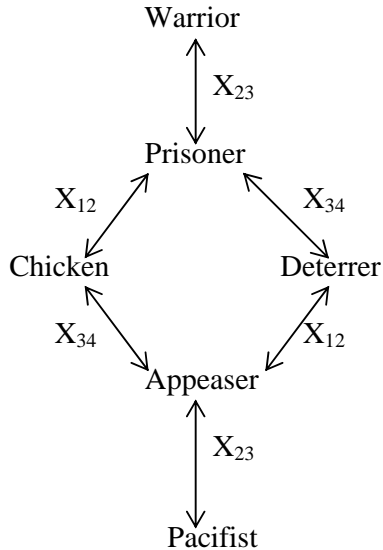


Figure 4

closely in the following section.

What happens outside of these types, for example if you apply the operation  $X_{23}$  to a Chicken or a Deterrer, or an  $X_{12}$  or  $X_{34}$  to a Warrior or Pacifist? Does this lead to a well-defined player 'type'? These moves take us outside the realm of C-D games, as the resulting preference ordering will have no dominant preference for the other player to choose a particular strategy. We shall see what this means for analysis of player types later.

#### The position of the C-D games on the Robinson-Goforth Periodic Table

Figure 5 shows the place of the C-D games on the Robinson-Goforth grid. The C-D games are shaded grey. As can be seen, the six player types occupy specific rows and columns of the grid. At first sight they may appear to occupy rather disconnected regions of the graph; however the toroidal nature of the layers means the C-D games are in fact connected, forming a 3x3 block on each layer. The 'aggressive' player types, whose first preference is (D,C) occur on the bottom and left-hand layers for Row and Column respectively, while the 'peaceful' types (whose first preference is (C,C)) occur on the upper and right-hand layers.

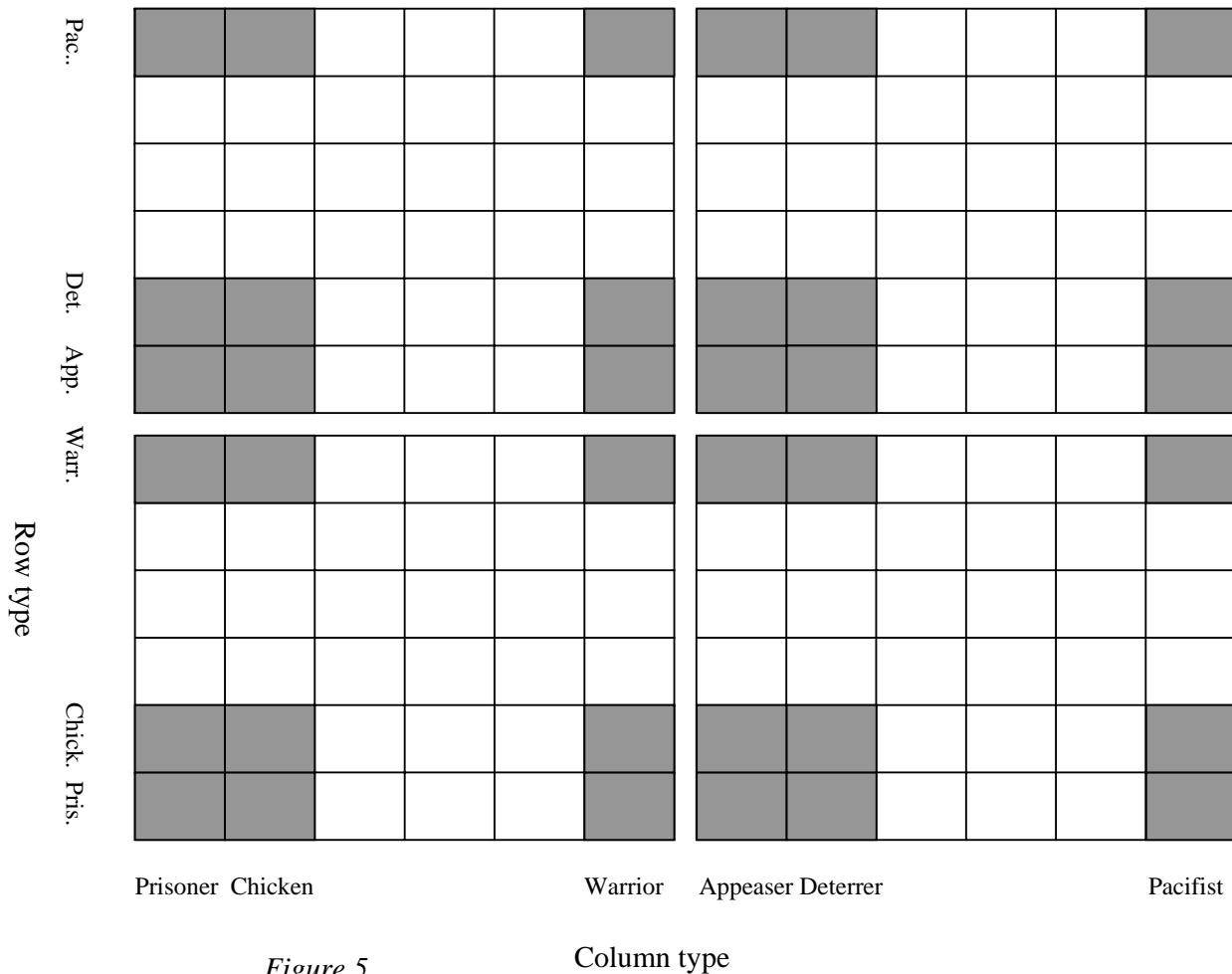


Figure 5

While the full topological structure of the C-D games is quite complicated, a central subset of them actually form a very simple subset, that plays an important role in the Robinson-Goforth topology. That is, the 16 games involving only the Prisoner, Chicken, Appeaser and Deterrer types, that form the 2x2 blocks at the bottom left of each layer. These four types are the ones that share an absolute preference for the other player to Co-operate, i.e. the other player Co-operating gives the top two outcomes for this type.

Each of these 2x2 blocks form a ‘tile’ in Robinson & Goforth’s terminology, that is a block of 4 games connected using the  $R_{12}$  and  $C_{12}$  operations. For example, the bottom-left tile in layer 1 is connected as shown in figure 6.

The four tiles are connected cyclically to each other via the  $R_{34}$  and  $C_{34}$  operations to form a Pipe, one of six such toroidal subspaces that connect the four layers. For example, the operation  $R_{34}$  takes each game in the above figure into the corresponding position in the following figure, which forms the bottom-left tile in layer 2, as shown in figure 7.

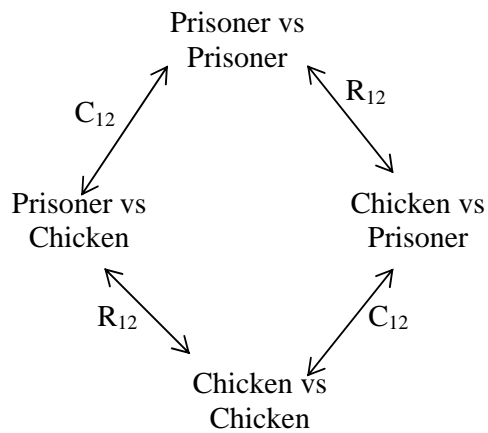
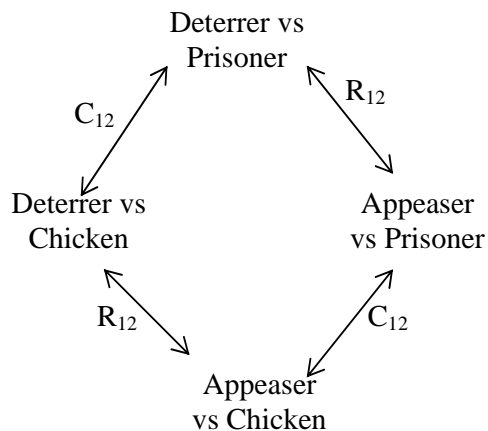


Figure 6

Figure 7



The  $C_{34}$  operation then turns the Column player from Prisoner/Chicken to Deterrer/Appeaser,  $R_{34}$  takes the Row player back to Prisoner/Chicken, and  $C_{34}$  cycles us back to where we started in layer 1. We thus have a circle of four circles of four, creating a torus.

### 5a) Extending the player type classification

The above grid shows each of the six player types occurring on a particular row or column of the grid, for the Row and Column players respectively. One might wonder if it is possible to extend this to a consistent definition of player types for the rest of the  $2 \times 2$  games, by treating each row or column as a distinct type. In fact this is not the case, since once the C-D conditions are violated there is no longer a meaningful labelling of the strategies, and so the unique definition of a player type by preference ordering is lost. For example, suppose we start with the Pacifist preference ordering for Column:

(C,C) (D,C) (C,D) (D,D)

Where Row's strategy is listed first. Suppose we apply the operation  $C_{12}$  to this ordering, giving



(C,C) (D,C) (D,D) (C,D)

This could be seen as giving some sort of ‘inconsistent Pacifist ordering’, that is Column still has an absolute preference for “Co-operating”, but they no longer have a dominant preference for Row to Co-operate, as (C,D) is now the worst outcome for Column. We are outside of the space of C-D games.

Now let us suppose that Row is a Warrior, that is, she has a preference ordering of (D,C) (D,D) (C,C) (C,D). In the original Warrior vs. Pacifist game, the outcome is (D,C), giving Row a payoff of 4, and Column a payoff of 3. This game occurs in the top right corner of layer 4 (the bottom right layer).

If the  $C_{12}$  operation is performed, swapping Columns bottom two preferences, Column now becomes an ‘inconsistent Pacifist’, and sure enough we have shifted one to the left on the grid. (see figure 8.)

Suppose however that, starting from our initial Warrior vs. Pacifist game, we instead apply the operation  $C_{34}$ . The result, surprisingly enough, is that we now have the game Pacifist vs. inconsistent Pacifist! This remarkable piece of preference-ordering

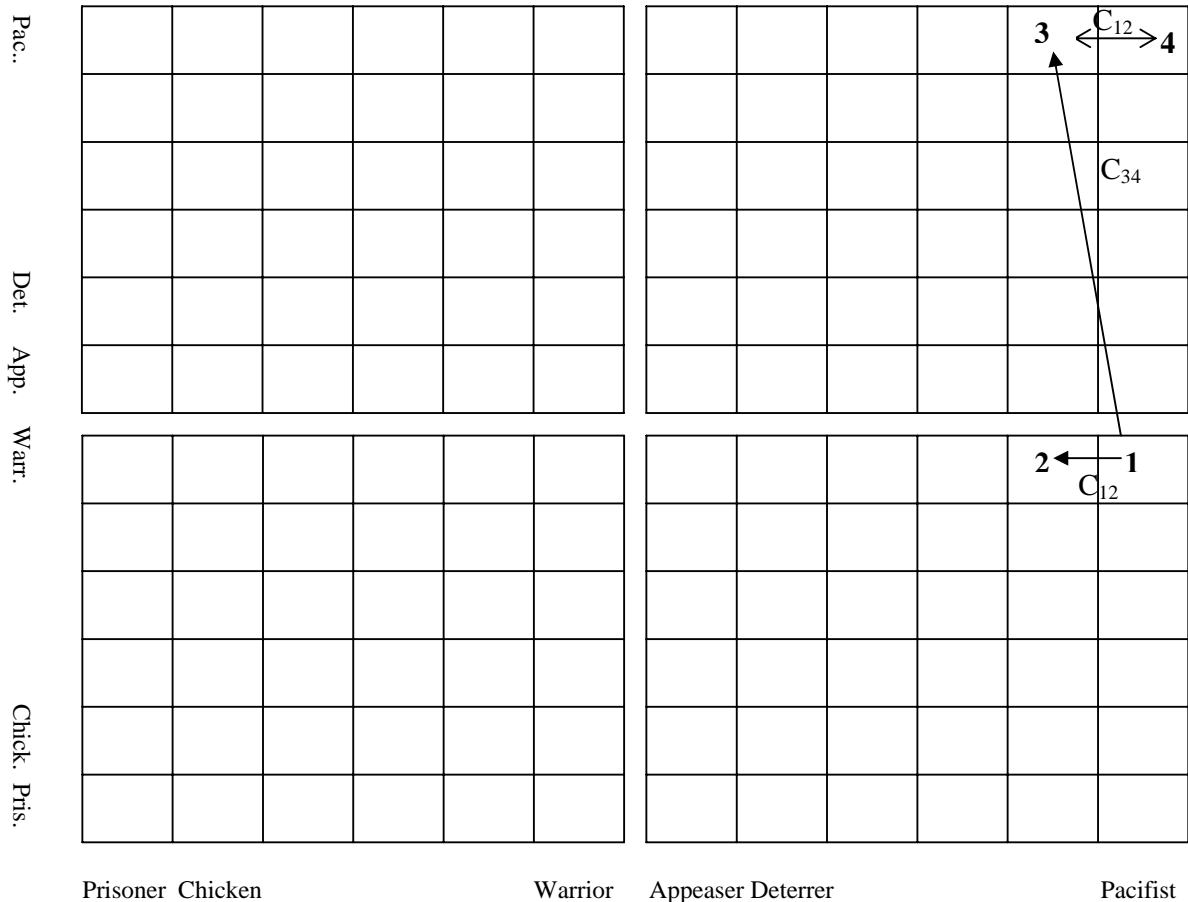


Figure 8

1) Warrior vs. Pacifist 2) Warrior vs “inconsistent Pacifist” 3) “Pacifist vs inconsistent Pacifist” 4) Pacifist vs. Pacifist

judo by Column has calmed the savage beast, and turned Row's Warrior into a Pacifist!

To demonstrate this rather surprising fact, consider figure 9, the order graph for Warrior vs. Pacifist:

Recall, Row chooses between the dotted lines, with the left-most line representing Co-operate in this case, and the right-most line Defect. Row clearly chooses Defect, as Row prefers the outcomes further to the right. Column chooses between the solid lines, the upper one representing Co-operate, the lower one Defect. Column, the Pacifist, Co-operates.

If we apply the operation  $C_{12}$ , swapping Column's bottom two preferences – that is

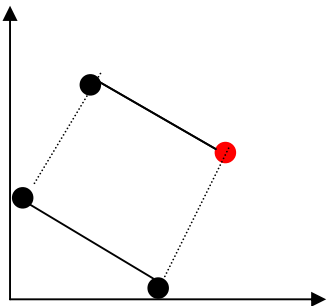


Figure 9

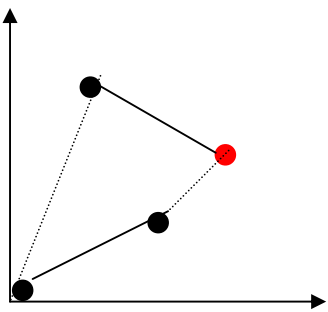


Figure 10

swapping the vertical position of the two lowest dots on the order graph, we get the Warrior vs. “inconsistent Pacifist” order graph as shown in figure 10.

The outcome is the same; though now Row's choice of the right-hand dotted line can no longer be so meaningfully described as ‘Defecting’, since it includes Column's middle two preferences.

If instead we swap Column's top two payoffs, that is we perform the operation  $C_{34}$  on figure 9, however, we get the order graph shown in figure 11.

It is not immediately clear where this game sits in the grid, though the fact that the Nash Equilibrium is now the preferred outcome for each player suggests that we are in the upper-right layer, and this is indeed the case. For if we now perform  $C_{12}$ , which always moves us one to the left or to the right as appropriate (remembering that the rows and columns circle round within each layer), we obtain figure 12 below.

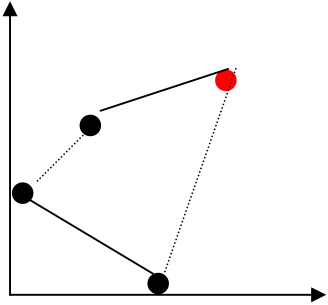


Figure 11

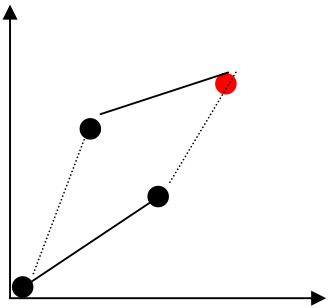


Figure 12

This is the graph for the Pacifist-Pacifist game. For Row, the rightmost dotted line is Co-operate (as it gives the higher payoff for Column for each of his choices), and the leftmost dotted line is Defect. Likewise, the upper solid line is Co-operate for Column, and the lower one is Defect. It is not hard to show that each player now has the Pacifist preference ordering.

Hence the previous game is the one which is a  $C_{12}$  operation away from Pacifist-Pacifist, namely the game marked 3) in the grid in figure 8 – which is the position for Pacifist vs. “Inconsistent Pacifist”.

Hence, a change in preference ordering for Column, using the  $C_{34}$  operation, has also changed the Row player’s type. This rather negates the sense of viewing each game as being one between two well-defined player types; for this to be the case, a change in one player’s preference ordering should only change their own player type. The conclusion is that player types cease to be well-defined when we move outside the realm of the C-D games.

## **6. The topological structure of the C-D games**

We have seen that the 16 games involving only the Prisoner, Deterrer, Chicken and Appeaser types form a torus (that is to say, their graph can be drawn on the surface of a torus.) What about the full graph of all 36 C-D games?

A crucial property of a graph is the number of ‘holes’ that is necessary to have in a surface to be able to draw the graph on it. A planar graph is one that can be drawn on a plane sheet of paper, or equivalently, the surface of a sphere. The more complicated a graph, the more holes it is necessary to put in the sphere to be able to draw the graph on its surface. This is determined by the graph’s Euler number, defined as:

$V + F - E$ , where  $V$  is the number of vertices,  $F$  the number of faces, and  $E$  the number of edges. A planar graph has an Euler number of 2. A graph that can be drawn on a 1-holed torus (but not a plane or a sphere) has an Euler number of 0. Each additional 'hole' that is required subtracts 2 from the Euler number. Robinson and Goforth show that the full graph of 144  $2 \times 2$  ordinal games can be drawn on the surface of a 37-holed torus, that is, the graph has an Euler number of -72.

I shall show that this graph can be drawn upon the surface of a 4-holed torus; or, equivalently, a sphere with four handles.

First of all, we may draw the graph of the 20 remaining games involving either a Warrior and/or a Pacifist on a plane (or equivalently, the surface of a sphere). This is shown in figure 13.

It remains to show how this graph is to be connected to the 16-game torus involving the four central types. Each of the four squares of four games in the corners needs to be connected to a corresponding square on the surface of the torus. For example, the top-left square needs to be connected to the square shown in figure 14 with each  $(W,X)$  connected to  $(Pr,X)$  via a  $R_{23}$  operation. It is not hard to show that the  $4 \times 4$  toroidal grid of games can be drawn so that each of the four squares that need to be connected to the Warrior-Pacifist structure appears as a square on a plane piece of surface, that is, not circling round the torus.

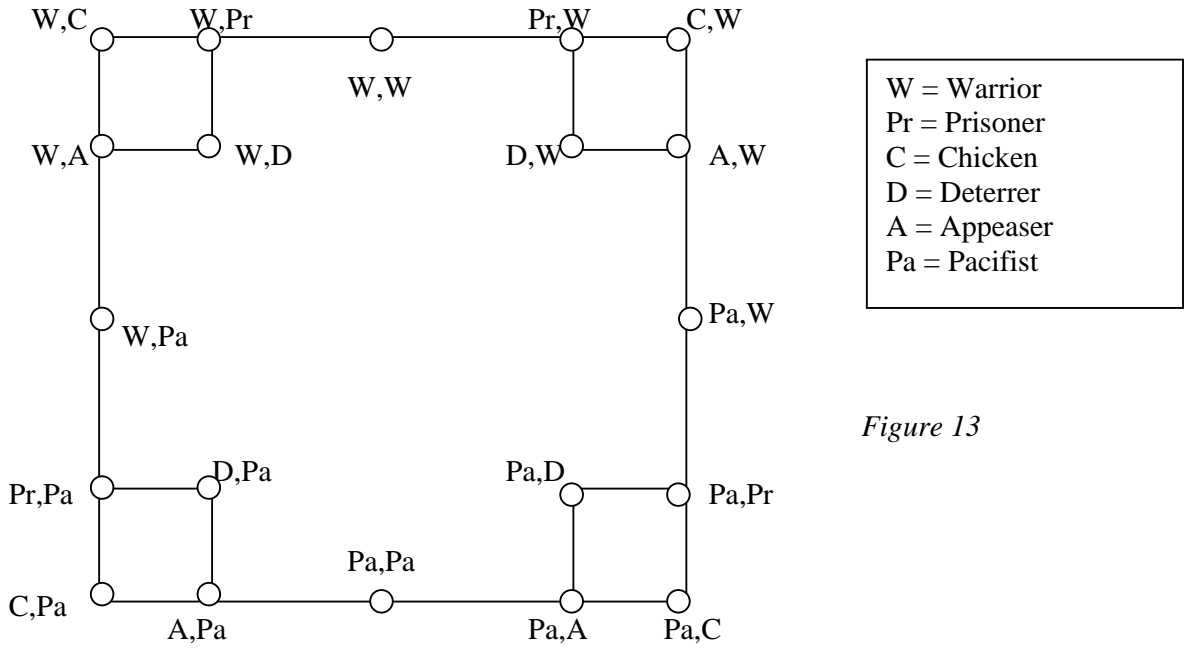


Figure 13

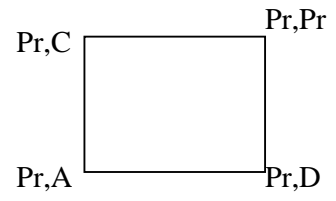


Figure 14

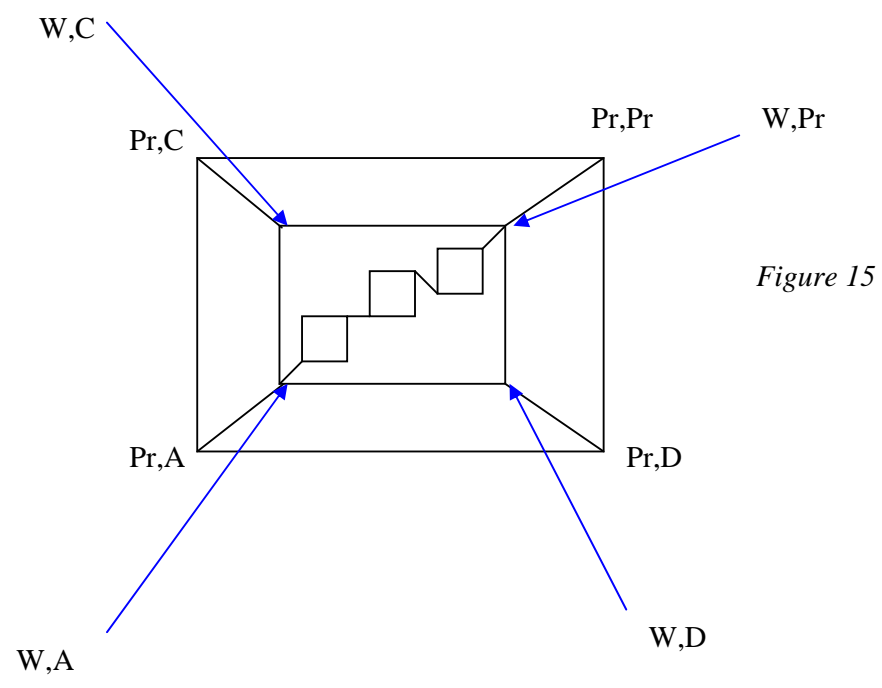


Figure 15

This would at first sight appear to add four connecting handles, and therefore four holes to the torus, for a total of five. However, the first connecting handle does not

actually add any additional holes. Alternatively, we may redraw the 20-game graph above so that the other three squares are inside the top-left square. We may then draw this inside the relevant square on the original torus, as in figure 15.

The inner rectangle with the three squares inside it is topologically equivalent to the original 20-game graph, except that in each of the short lines connecting the inner squares there is an additional vertex in the middle.

Now each of the three inner squares can be connected to its corresponding square on the torus via a handle (along which four edges can be drawn, connecting each corner of the inner square to its target on the torus), so that three additional handles must be added to the torus to draw the complete graph. The result is therefore a 4-holed torus. Counting the edges, faces and vertices of this graph confirms that the Euler number is -6, which also implies a 4-holed torus.

## **7. Conclusions**

The Co-operate-Defect games, where each player has a dominant preference for a particular strategy by the other, form a theoretically and practically interesting subset of the 2x2 games. They have a natural interpretation in the field of international conflict and arms rivalries (amongst others), and can be linked together in a natural and concise way using the topological approach of Robinson & Goforth. There are 36 possible C-D games, generated by the combination of 6 possible preference-ordering types by the 2 players. These types can each be seen to be reflected in possible real-world scenarios, and many of the possible games formed by the interaction of the types pose interesting problems worthy of theoretical and empirical study. As well as the familiar Prisoner's Dilemma, Stag Hunt and Chicken games, perhaps particularly interesting are the Prisoner vs Chicken, Prisoner vs Deterrer, Chicken vs Deterrer and Chicken vs. Appeaser games.

When located within the Robinson & Goforth indexing system and topology for all the 2x2 games, the C-D games form a compact block, appearing in specific rows and columns of the indexing grid used by those authors. However, the 'player type' classification used for the C-D games cannot consistently be extended to the other 2x2 games, as it breaks down once there is no longer a well-defined meaning to the labels 'Co-operate' and 'Defect' on the two strategies.

The C-D games, comprising one quarter of the total number of 2x2 games, require a topologically much simpler graph to display their inter-linkages, one that can be drawn on the surface of a 4-holed torus, as opposed to the 37-holed torus required for the full graph of 144 games. A central subset of the C-D games, those involving only the Prisoner, Chicken, Deterrer and Appeaser types, form a particularly simple structure, a 1-holed torus, forming one of the "Pipes" in Robinson & Goforth's terminology that connect the four toroidal "layers" defined for their structure.

A fruitful possibility for future research could be to investigate experimentally the interactions between the different player types (at least the Prisoner, Deterrer, Chicken and Appeaser types) in iterated games; at least two significantly different treatments suggest themselves, one with full information about other player's types, the other with player type known only to each player themselves.

## References

- Axelrod, Robert (1984), *The Evolution of Co-operation*, Basic Books.
- Hamburger, H. (1969), "Separable Games", *Behavioural Science*, 14: 121-132.
- Lichbach, Mark I. (1990), "When is an Arms Rivalry a Prisoner's Dilemma? Richardson Models and 2x2 Games", *Journal of Conflict Resolution*, Vol. 34, no. 1, pp29-56.
- Plous, S. (1993), "The Nuclear Arms Race: Prisoner's Dilemma or Perceptual Dilemma?", *Journal of Peace Research*, vol. 30, no. 2, pp163-179.
- Possajennikov, Alex (2002), "Co-operative Prisoners and Aggressive Chickens: Evolution of Strategies and Preferences in 2x2 Games", <http://www.sfb504.uni-mannheim.de/~possajen>
- Poundstone, W. (1992), *Prisoner's Dilemma*, Doubleday.
- Rapoport, Anatol, and Albert M. Chammah (1966), "The Game of Chicken," *American Behavioral Scientist*, 1010-28.
- Rapoport, A., Guyer, M. (1966). A taxonomy of 2 x 2 games. *Gen. Syst.* 11, 203-214
- Rapoport, Anatol & Guyer, Melvyn J., (1978), "A Taxonomy of 2x2 Games", *General Systems*, Vol. XXIII, pp125-136.
- Robinson, David & Goforth, David, (2005), *The Topology of the 2x2 Games: a New Periodic Table*, Routledge.
- Sandler (2004), *Global Collective Action*, Cambridge University Press.
- Schelling, Thomas C. (1963), *The Strategy of Conflict*, Oxford University Press.