Exploitation and peacekeeping: introducing more sophisticated interactions to the iterated prisoner's dilemma

Toby Ord and Alan Blair

Department of Computer Science and Software Engineering
University of Melbourne,
3010, Australia
tdo@cs.mu.oz.au

School of Computer Science & Engineering
University of New South Wales
2052, Australia
blair@cse.unsw.edu.au

Abstract – We present a new paradigm extending the Iterated Prisoner's Dilemma to multiple players. Our model is unique in granting players information about past interactions between all pairs of players – allowing for much more sophisticated social behaviour. We provide an overview of preliminary results and discuss the implications in terms of the evolutionary dynamics of strategies.

I. INTRODUCTION

Since it was invented by Merrill Flood and Melvin Dresher in 1950, Prisoner's Dilemma (PD) has undergone numerous extensions aimed towards producing better models of social behaviour. Two well-known extensions have been *N*-player Prisoner's Dilemma (NPD) [1]-[2] and Iterated Prisoner's Dilemma (IPD) [3].

NPD is an extension that allows for an arbitrary number of players. It is a collective game between all of the players, in which each player can either cooperate or defect against the group, getting a reward which increases with the number of other players who cooperate. This allows the study of anonymous situations in which agents can choose to help the population as a whole or to help themselves at the expense of the population.

IPD is an extension that allows for multiple turns of the game to be played consecutively. This greatly increases the amount of strategies available and encourages complex and interesting approaches that must balance the potential of exploitation with the dangers of antagonising the other player.

There are also notable attempts to extend Prisoner's Dilemma to multiple players *and* multiple turns. One such attempt is Axelrod's Round-Robin IPD tournament (RR-IPD) [4] which is a collection of games of IPD where each agent plays a game of IPD against each other agent (including itself). This models populations in which agents interact in couples and do not know what transpires in interactions which do not involve them personally. It allows strategies that single out players who should be trusted from those who should not be, but forbids strategies that depend on the manner in which the other players interact among themselves.

N-player Iterated Prisoner's Dilemma (NIPD) [5] is an alternate approach, combining many games of NPD in sequence. This models societies in which agents interact as one group, with no information about which members of the group are cooperating or defecting and no way of singling out members of the group for individual treatment. It allows strategies that take into account the proportion of the population that is defecting but forbids strategies that depend on singling out the individuals who are defecting from those who are not.

In this paper, we introduce *Societal Iterated Prisoner's Dilemma* (SIPD), a new extension of Prisoner's Dilemma to multiple players and multiple rounds which has many characteristics in common with the two above, but has some significant differences which give rise to more possible strategies and, in turn, the ability to model more complex forms of interaction.

II. OUR MODEL

Like IPD, a game of SIPD is comprised of an indeterminate number of rounds. In each of these rounds, each pair of players conducts a single game of Prisoner's Dilemma. When deciding whether to cooperate or defect, the players have complete information on all past interactions between every pair of players. At the end of the game, the score for an agent is calculated as the sum of the scores they achieve in their individual games of PD.

This differs from NIPD in that the players have access to information about which players were defecting and which were cooperating on any given round. They can then use this information to discriminate between the other players, cooperating or defecting on each depending on their history of play.

SIPD differs from RR-IPD in that the players have access to the histories of all their games simultaneously and – more importantly – have access to the histories of other pairs of players. In fact, both NIPD and RR-IPD are very nearly special cases of SIPD with restricted information and (in the case of NIPD) restricted playing choices.

It is the information about interactions that an agent was not involved in that allows the wealth of interesting new strategies. In SIPD, agents learn from the experience of others and get much better understandings of their opponents. The richness of the interactions make this a *society* of players, in which an agent can gain considerable advantage by observing the relationships between the other players, and not just those that involve it personally.

For example, when deciding whether to forgive a defector, agents can use their knowledge of how other victims were treated when they forgave. They could also pre-emptively defect on agents that they have observed being untrustworthy or exploit agents that have been proven to be soft targets.

This opportunity to use information about interactions that do not involve you directly makes SIPD a good model for small populations with public knowledge such as a small town, a circle of friends or global politics (between a small number of countries). In contrast, RR-IPD only models situations where all interactions are private and agent A can never know anything of the relationship between agents B and C. On the other hand, NIPD models communal, anonymous interactions such as littering, in which no-one ever finds out if an individual is cooperating or defecting.

As the transition from PD to IPD fostered cooperation due to the effects of one's actions being remembered at future times, SIPD offers a chance for an even greater level of cooperation due to the effects of one's actions being known to other agents.

III. THE LANGUAGE

In each of the extensions of PD to include multiple turns, the number of possible strategies is enormous. In fact, the number of strategies available to an agent at a given time is doubly exponential in the number of rounds that have elapsed. It is therefore important to choose a good formalism in which to represent strategies that is relatively easy to understand and gives simple encodings for simple strategies.

We have decided to represent strategies as formulae in first order logic with some special constants and predicates. This allows us to express many interesting strategies in concise and intuitive ways and does not limit the available information to a fixed number of previous rounds.

The formulae used to represent strategies make use of one predicate (D), three constants (r, c, p) and one function symbol (b).

- D is a three place predicate which is true if the agent given in argument #1 defected on the agent given in argument #2 at the time given in argument #3.
- r (standing for 'row') is a constant representing the agent that is considering whether or not to defect.
- c (standing for 'column') represents the agent that r is considering defecting against.
- p represents the previous round. If it is the first round, this represents a special 'round zero' on which it is assumed that all players cooperated.
- b is a one place function symbol that represents the predecessor function. Thus $b(x)$ represents the round before round x. If b is applied to the first round, or to an agent, it represents the special 'round zero' defined earlier.
- Other letters such as x, y, t may be used as variables representing agents in the game or times that have occurred so far.

Strategies are well-formed formulae in first order logic using these symbols with the standard logical connectives and quantifiers. A formula expresses the necessary and sufficient condition for an agent 'row' to defect on another agent 'column'. In other words, row will defect on column if and only if the formula is true.

A simple example of such a formula is $D(c,\square,\square)$ which represents Tit-For-Tat. In other words, 'defect on c if and only if c defected on you last round'.

Here are the representations of several well-known strategies with English translations of when they will defect on c:

Tit-For-Tat (TFT):	(defects on c if and only if)
$D(c, r, p)$	c has defected on me last round
Tit-For-Two-Tats (TF2T):	
$D(c, r, p)$	c has defected on me last round
$\wedge$	and
$D(c, r, b(p))$	the round before
Grim:	
$\exists t\ D(c, r, t)$	c has defected on me at some stage
Cycle-DC:	
$\square D(r, c, p)$	it is an odd numbered round
All-C:	
False	never
All-D:	
True	always

Note how the use of quantifiers allows us to define strategies depending on all previous time steps – in contrast to some earlier IPD formalisms, such as those used in [6], which were limited to a fixed time horizon.

From the above examples, some of the power and simplicity of this representation can be seen[1]. The translation between the natural and formal languages is easy because their structures are very similar. This property can allow us to talk of individual phrases of a strategy in isolation. For example, we can see that TF2T is strictly less inclined to defect than TFT because TFT's condition of defection is just one of two conditions that TF2T must satisfy. When we

[1] Note, however, that certain 'statistical' strategies are cumbersome to write in our notation (such as 'defect if the other player has defected on you in three of the last six rounds'). Others such as 'defect if the player has defected on you more than half of the time' are impossible to express.

discuss some more complex strategies in the next section, we will explicitly use this modularity of strategies to change reasonable strategies into better ones.

An important use for formalisms in IPD is to provide a more objective method of evaluating new strategies through evolution as investigated by Axelrod [6]. Because first order logic has a simple tree structure, there are simple and effective ways to interbreed and mutate successful strategies in the hope of combining their best features [7]. Although the present work focuses on the evolutionary performance of a small number of fixed strategies, in future work, we plan to exploit the full combinatorial power of the framework through crossover and mutation. Such an approach has great potential for understanding the evolution of new strategies within a society.

IV. NEW STRATEGIES

So how can an agent benefit from the additional information available in SIPD? To answer this, we shall first examine the advantages that the additional information provides for strategies which try to exploit other players – that is to defect on those that will not retaliate.

A simple strategy along these lines is:

Bully:

$\forall t\ D(c, r, t)$ — *c* has never defected on me

Bully simply defects on a player until they defect back, at which time it continually cooperates. This very simple strategy works well against populations of All-C and TFT, but poorly against All-D or Grim. Bully can be enhanced, however, by adding a clause to make it defect when its opponent takes advantage of its conciliatory cooperation.

Spiteful-Bully:

$\forall t\ (D(c, r, t))$ — *c* has never defected on me
or
$\exists s$ — there was a time when:
(
  $D(c, r, s)$ — *c* defected on me
  $\wedge$
  $D(c, r, b(s))$
  $\wedge$
  $D(c, r, b(b(s)))$ — three times in a row
)

Spiteful-Bully is considerably better than Bully, with its results against All-C, TFT and All-D approaching optimality as the number of rounds increases. It still has some trouble against Grim, however, since the optimal behaviour would be to always cooperate. Somewhat surprisingly, there is a simple strategy that can perform optimally with all four of these players:

Vulture:

$D(c, r, p)$ — *c* defected on me last round
or
$\exists j$ — there is someone:
(
  $\exists t\ (D(j, c, b(t))$ — who has defected on *c*
  $\wedge$ — and
  $\forall u\ D(c, j, u))$ — *c* has never retaliated against them
)

If Vulture is playing with All-C, TFT, All-D and Grim it will indeed play optimally in the limit. This cannot be done in RR-IPD because the agent must initiate the conflict with All-C but not with Grim, even though both these strategies act identically until defected on. Vulture uses the information that All-C does not fight back against All-D to break this symmetry and perform optimally. Unfortunately for Vulture, it requires another aggressive strategy in the population in order to know who to exploit. In the absence of such a strategy, it does not perform optimally, but still performs just as well as all the other players because everyone cooperates for the entire game.

Is Vulture guaranteed to perform as well (in the limit) as any other player in the population? No. Consider for instance playing against a TF2T and Cycle-DC. In this case, Cycle-DC will exploit TF2T without any negative consequence, but when Vulture begins its constant defection, it will bring on constant punishment. In the limit, Cycle-DC will outperform the other two players. This shows that while strategies like Vulture are promising, it is not enough for them to see that a player is being exploited without retaliation – they must copy the exact manner in which these strategies are exploited.

As well as new potential for exploitative strategies, SIPD brings some new possibilities for enforcing cooperation. Just as 'grim' strategies hinder a potential exploiter by ignoring how long ago the exploiter defected on them, new *peacekeeping* strategies can cause trouble by ignoring *who* the exploiter defected against. The simplest such strategy is the Vigilante.

Vigilante:

$\exists j\ D(c, j, p)$ — *c* has defected on someone last round

Imagine a game between a Spiteful-Bully, an All-C and a Vigilante. The Spiteful-Bully will defect on both the All-C and the Vigilante, getting no resistance from the All-C and backing off after the initial response from the Vigilante. However, when the Spiteful-Bully continues to defect on the All-C, the Vigilante will keep punishing the Spiteful-Bully in retaliation. Such behaviour prevents exploitative strategies from considering each of their opponents in isolation – the peacekeeper can make them pay for their antisocial behaviour.

Now consider a large group of peacekeepers and All-Cs. If a few exploitative strategies entered this population, they

would perform very poorly, receiving the punishment payoff against all of the many peacekeepers, while each peacekeeper would only suffer the punishment against the few exploiters. They thus share the burden of the defence.

There are two separate ways in which peacekeepers foster cooperation. If the exploiters are quite adaptive, they will realise that the only way to get a good reward is to cooperate with everyone. The peacekeepers (and all others) would thus get very high scores for mutual cooperation. On the other hand, if the exploiters are not this adaptive, there is still an advantage in peacekeeping because it gives the exploiters very low scores and thus an evolutionary disadvantage. These results cannot be accomplished with TFTs because strategies like Bully or Spiteful-Bully would get the highest scores by defecting on the All-Cs and leaving the TFTs alone. Grims would do a better job, and could force the low score on the exploiters, but can never make it rational for them to change their behaviour during the game.

Unfortunately, Vigilantes cannot perform such a perfect peacekeeping role either, because they cannot distinguish between exploitative defection and retaliatory defection. They will see that every Vigilante defected last round and defect on all of them, leading to very low scores all round. This problem can be overcome with a more sophisticated peacekeeping strategy such as Police.

Police:

$D(c, r, p)$	c defected on me last round
	or
$\exists j$	there is an agent
(	
$D(c, j, p)$	who was defected on by c last round
$\wedge$	but
$\neg\exists k\ (D(j, k, b(p))$	had just cooperated with everyone
)	

Police have the property that they will never defect on each other, which makes them much more effective in groups. They are also more forgiving of defections, allowing strategies such as TFT to immediately retaliate when defected on. Vigilantes can be thought of as complete in their peacekeeping (they defect on everyone who should be defected on), but not sound (they defect on players they should not defect on). In contrast, Police are sound but not complete. Their restrictions on who they will defect on give an excuse for defection – if you defect on someone the round after they defect on anyone, the Police agents will let you off, assuming that you could be attempting to keep the peace.

Even if populations are more stable with peacekeepers, is it rational for an individual agent to keep the peace rather than simply playing Tit For Tat? Interestingly, it seems not. For a given agent there are substantial costs to being a peacekeeper but the reward is a communal one and small for each individual. There is a second order NPD at play here because it is communally better to have peacekeepers but personally better to not be one. A population of peacekeepers may thus be expected to drift back into self-interested behaviour such as TFT, allowing exploiters to arise once more.

One solution to this problem would be to have some meta-peacekeepers in the population. For example, *Meta-Police* would defect on someone whenever Police would, but also whenever someone failed to defect when Police would have. This would provide further incentive for peacekeeping strategies, but unsurprisingly it also shifts the dilemma up to the meta-peacekeeping level.

V. NEW PROPERTIES

In *The Evolution of Cooperation* [4] Axelrod advocates the importance of being 'nice' when playing IPD. He calls a strategy nice if and only if it will not defect on someone who has not defected on it. There are many benefits for nice strategies, one of which is that a pair of nice strategies will always cooperate with each other for the whole game.

The concept of niceness has several analogues in SIPD. An obvious distinction is between strategies that are 'individually nice' and those that are 'communally nice'.

Individually nice:
Will not defect on someone who has not defected on it

Communally nice:
Will not defect on someone who has not defected at all

Both of these definitions of niceness have merit. Individual niceness has one of the main properties of niceness as defined by Axelrod because all pairs of individually nice strategies will consistently cooperate. Communal niceness is a weaker property because any individually nice strategy is also communally nice, while the converse is not always true. Communal niceness is still an important property for a strategy to possess, however, because any population containing only communally nice players will never have any defection. This distinction also captures some of the difference between the individually motivated TFT and the communally minded Vigilante or Police.

The distinction between Vigilante and Police can be captured with the following definitions:

Meta-individually nice:
Will not defect on someone who is individually nice

Meta-communally nice:
Will not defect on someone who is communally nice

While Vigilante may end up defecting on the individually nice TFT, we can see clearly that Police never will because it is meta-individually nice. All individually nice strategies such as TFT are also meta-individually nice, but this is not true for communally nice strategies.

These concepts of meta-niceness are particularly useful in describing situations in which mutual cooperation is assured. For example, a communally nice strategy that is also meta-communally nice is assured of never defecting on someone of the same strategy. Never defecting on your own strategy is a very important property and we call such strategies 'loyal'.

Loyal:
Will not defect on someone of the same strategy

TFT and Police are loyal while Vigilante is not. Perhaps surprisingly, Vulture is also loyal even though it lacks all of the other kinds of niceness mentioned above. Loyalty is especially important in evolutionary systems, as being successful creates more copies of your own strategy and you have to be able to work well with yourself.

Each of these definitions we have posed here are equivalent in IPD but diverge in SIPD. They therefore offer a way of examining which aspects of niceness are most important.

As well as being nice, Axelrod characterised TFT as retaliatory and forgiving. These concepts can be generalised in similar ways to niceness.

Individually retaliatory:
Defects on someone who defects on it

Communally retaliatory:
Defects on someone who defects on anyone

Individually forgiving:
Stops defecting on someone if they stop defecting on it

Communally forgiving:
Stops defecting on someone if they stop defecting on everyone

As with niceness, these pairs of definitions contain a weaker and stronger version: all individually forgiving strategies are communally forgiving and all communally retaliatory strategies are individually retaliatory.

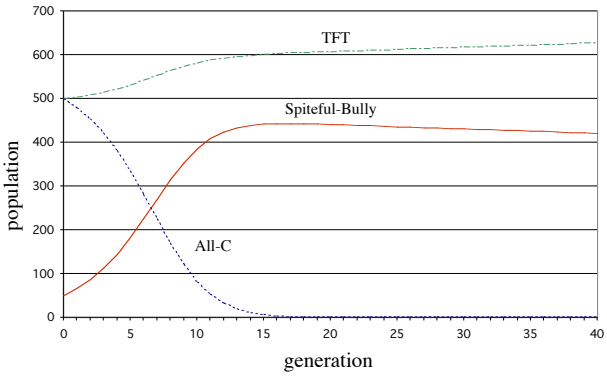
Where TFT can be characterised as individually nice, retaliatory and forgiving, Vigilante is communally nice, retaliatory and forgiving. Each of these three properties makes Vigilante strictly more inclined to defect than TFT.

As with the IPD versions of the above properties, there is considerable scope for using such properties to analyse existing strategies and to create new ones

VI. EXPERIMENTAL RESULTS

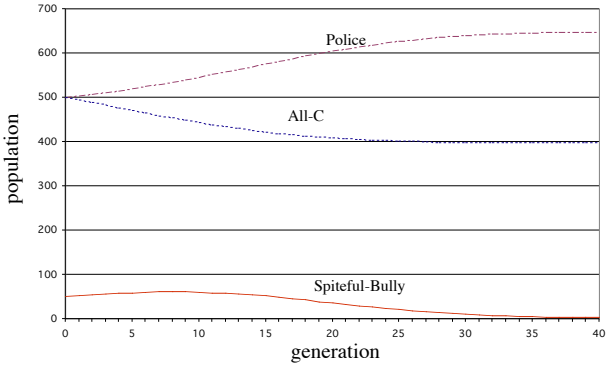
To make some of the qualitative claims above somewhat more concrete, we performed several evolutionary experiments of the type performed by Axelrod in [4]. We took several different initial populations of agents and for each of these, played a series of 100-round games. Each game was treated as one generation of an evolving system. At the end of each generation, the population of agents using each strategy was changed in proportion to its relative success. That is, the new population of agents using a given strategy equals the old population multiplied by a factor representing the strategy's relative success. This factor is simply the score for agents using that strategy, divided by the average score for the whole population.

Figure 1 shows the results with an initial population of 500 TFTs, 500 All-Cs and 50 Spiteful-Bullies. The Spiteful-Bullies initially get very high scores, increasing their population dramatically at the expense of the All-Cs. All of these are classical IPD strategies and this simulation shows how exploiters can mostly leave TFTs alone, but gain a large population by defection on a non-retaliatory strategy. (Note however that in this case, when no All-Cs are left, the population of Spiteful-Bullies very slowly falls back to zero due to a very minor difference in performance with TFT).



(1)

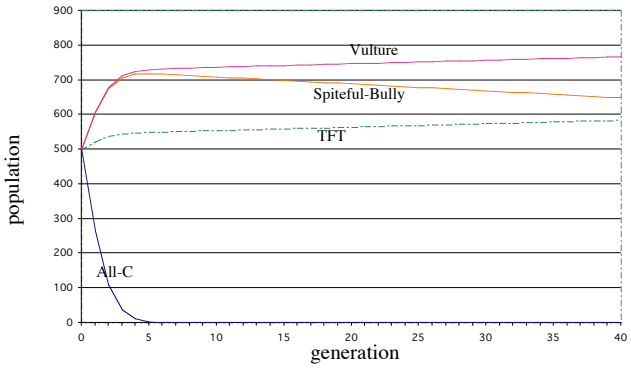
Figure 2 shows the results with Police replacing the TFTs. In this case, the Spiteful-Bullies cannot get an evolutionary foothold and are wiped out in 40 generations, leaving the Police and All-Cs in equilibrium. This shows how Police can make a large difference to a non-retaliatory population even when the exploitative strategy is not sophisticated enough to begin cooperation with all the agents to avoid the punishment from the Police.



(2)

Figure 3 shows the results with an initial population of 500 TFTs, 500 All-Cs, 500 Spiteful-Bullies and 500 Vultures. The exploitative strategies quickly drive the All-Cs to extinction, but while doing so, they get a large population increase over the individually nice TFT. As the generations go by, Spiteful-

Bully slowly loses its advantage due to its bad performance in the first couple of rounds of each game, eventually reaching a population of zero.



(3)

On the other hand, Vulture now fares just as well as TFT and will never lose the population advantage gained by its exploitation in the early generations. This is an example of a strategy that is not individually or communally nice rising to dominance. This highlights the distinctions made earlier about the different kinds of niceness. By combining willingness to defect first on some occasions with an unwillingness on others, Vulture can sometimes get the best of both worlds.

VII. CONCLUSIONS

SIPD promises many interesting things. Most importantly, it is a much more sophisticated model of interactions between multiple members of a society than any of the other variants of PD. The extra information available makes a tremendous difference in the interconnectedness of the society. Agents must bear more responsibility for their actions because they will be seen by many players – a single betrayal could cause all other acquaintances to cease cooperation immediately.

Will this information encourage peacekeepers to arise? If so, exploitation will be much more difficult as exploiters are punished for all defections and can no longer afford to consider their targets individually. Even if peacekeepers cannot be encouraged, there will probably be more sophisticated versions of TFT, which base their forgivingness and retaliation at least partly on how players have been treating each other.

There is also considerable room for mixed exploitative / nice strategies such as Vulture who can pick on the weak while avoiding those who will fight back. Such strategies that are not individually nice, yet are loyal to others of the same kind, may be able to get the evolutionary success that was lacking for exploiters in classical IPD.

Through strategies like these, SIPD allows us to better understand the important strategic properties discussed in IPD. It allows intuitively different interpretations of niceness, retaliation and forgiveness to represent substantively different properties which can be given independent analysis.

SIPD provides a framework in which these interesting questions can be addressed using the tools of IPD and a clean, simple model.

VIII. FURTHER WORK

As this paper is an introduction to a new model, there is much scope for further study. A more thorough investigation of possible strategies would be a very enlightening task. We have discovered many interesting ones that have not been discussed here and continue to find new ways of using the available information.

A related project is to look at a more realistic evolutionary model as hinted earlier. With this first order logic formulation of strategies it is fairly simple to construct crossover and mutation operations and explore the evolutionary stability of various strategies when new strategies can come into the environment. This would greatly help to avoid the arbitrariness of the initial population and give a more objective description of the quality of different strategies.

There are also many ways in which the model could be adjusted to give more sophisticated models of society. Of particular note is the limiting of information to a more realistic level. The interconnections between agents could also be altered. Currently there is a game of PD for each pair of agents, but this could easily be limited to prevent some pairs of agents from playing and thus imposing a different topology of interaction on the society. This would be essential to give a good model of the clustering of social groups.

Finally, the many modifications that have been suggested for IPD can also be applied to SIPD such as making the results of actions uncertain [8]-[9] or allowing a continuum of choices between cooperation and defection [10].

IX. REFERENCES

[1] Thomas Schelling, *Micromotives and Macrobehaviour*, New York, Norton, 1978.

[2] Garret Hardin, "The Tragedy of the Commons", *Science*, Vol. 162, pp.1243-1248, 1968.

[3] R. Axelrod & W. D. Hamilton, "The evolution of Cooperation", *Science*, Vol. 211, pp.1930-1936, 1981.

[4] R. Axelrod, *The evolution of Cooperation*, New York, Basic Books, 1984.

[5] X. Yao, "Evolutionary stability in the *N*-person prisoner's dilemma", *Biosystems*, Vol. 37, pp. 189-197, 1996.

[6] R. Axelrod, "The evolution of strategies in the iterated prisoner's dilemma", in *Genetic Algorithms and Simulated Annealing* (L. Davis, ed.), pp. 32-41, San Mateo, CA, Morgan Kaufmann, 1987.

[7] J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, MIT Press, MA, 1992.

[8] R. Axelrod and D. Dion, "The further evolution of cooperation", *Science*, Vol. 242, pp.1385-1390, 1988.

[9] J. Bendor, R. Kramer, P. Swistak, "Cooperation under uncertainty: what is new, what is true and what is important?", *American Socialogical Review*, Vol. 61, pp.333-338, 1996.

[10] Marcus Frean, "The evolution of degrees of cooperation", *Biosystems*, Vol. 44, pp.135-152, 1996.