

Nash equilibria, but that can be handled by adding another parameter to the particle filter. An advantage of using particle filters in this way is that they can learn from observations from different games as long as the opponent stays the same.

Figure 2 shows the performance of two agents which estimate attitude values using a particle filter and reciprocate the attitude of the opponent with a slightly larger attitude value. In each round agents play against their opponent in a new randomly generated game. Agents can rapidly learn to cooperate and achieve good outcomes for both agents.

2. RESTRICTED STACKELBERG RESPONSE WITH SAFETY (RSRS)

Using a particle filter to estimate attitude allows agents to cooperate, but is counterintuitive because the particle filter provides a prediction of opponent behavior which is not then used by the agent. We would like instead for the agent to respond rationally to the prediction (given its chosen attitude).

A best-response to a prediction is easy to calculate, but has no performance guarantees. On the other hand, playing strategies which provide guarantees can reduce performance against the prediction. A number of approaches have been taken to solving this problem [5, 1, 4]; they provide different guarantees and not all are suitable for general-sum games.

We have developed Restricted Stackelberg Response with Safety (RSRS) [2], a parameter driven approach which can smoothly adjust between best-responding to the prediction, dealing with an exploiting opponent, and providing worst-case guarantees. RSRS uses two parameters, w , the prediction weight, which controls the trade-off between exploiting the prediction and dealing with an exploiting opponent, and r , the risk factor, which determines how much to risk in hopes of exploiting the prediction.

RSRS is calculated by creating a modified game, in which the moves are the same as the original game, but the payoffs are adjusted to be a weighted average of the original payoffs and the expected payoff of the calculating players move against the prediction $U'_1(m_1, m_2) = (1 - w) * U_1(m_1, m_2) + w * U_1(m_1, pred)$. Using the modified game, RSRS is the strategy which maximizes performance against a best-responding opponent while guaranteeing a payoff within r of the safety value of the game.

The game we use to show the properties of RSRS is a general-sum modification of Rock/Spock/Paper/Lizard/Scissors – a variant of Rock/Paper/Scissors with 5 moves. Rock/Paper/Scissors/Lizard/Spock was presented in the the TV show The Big Bang Theory; we have modified it to make it general-sum, and changed the name to reflect the precedence relationship between the moves. Each action beats two other actions, and is beaten in turn by the two remaining actions. Players receive a payoff of 1 for a win, -1 for a loss, and 0 for a tie. In addition, both players receive .5 when adjacent moves are played and lose .5 when non-adjacent moves are played. In this game players have conflicting interests but some cooperation is possible, which allows us to distinguish between a best-responding opponent and a worst case outcome.

Figure 3 shows the effect of r and w when using RSRS in RSPLS. The risk factor causes the strategy to gradually shift from the minimax solution to a best-response to the prediction. The prediction weight causes discontinuous changes in the strategy at points where a best-responding opponent changes strategies. It can be proven that when the strategy changes at a point w the ratio of gain against the prediction to the loss against a best-responding opponent is $\frac{1-w}{w}$.

Using the method outlined in [5] to set r values and a combi-

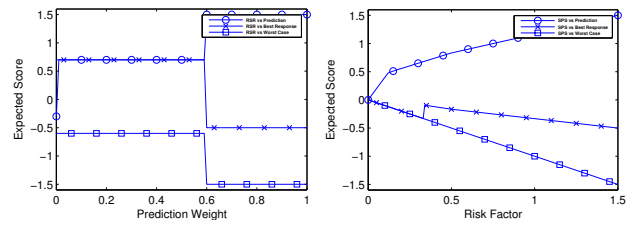


Figure 3: Performance for different values of the w (left) and r (right) parameters in a general-sum game, against the prediction(\circ), against a best-responder (\times) and in the worst-case (\square).

nation of gradient descent with an exponential opponent response model to set w values RSRS can perform well using a flawed predictor against a wide variety of opponents. It is able to coordinate in self-play to arrive at strategies which are beneficial to both agents. However, it is not fully cooperative. RSRS will never select a dominated strategy, so it can't cooperate in games such as Prisoner's Dilemma, where the cooperative move is strictly dominated.

3. FUTURE WORK

We can use attitude to cooperate in self-play at the cost of not responding rationally to our prediction. We can use RSRS to respond effectively to a prediction without cooperating. The next stage is to incorporate RSRS into an attitude-based reciprocating agent to create an agent which can use reciprocation to achieve a cooperative outcome in any general-sum game while avoiding exploitation. We will do this by creating a particle filter which combines attitude and the RSRS model to identify attempts to cooperate.

The final problem is to formalize the process of reciprocation. In symmetric games, it is easy to aim for equal outcomes, but not every game is symmetric, and even when the game appears to be symmetric inaccuracies in the utility function may make an asymmetric outcome desirable. For example, in a game with cash payoffs one player may have an immediate need for a specific amount while the other simply wants as much money as possible - this would affect the achievable cooperative agreements. We don't believe it is possible to fully resolve this problem in the absence of an oracle to provide true utility functions, but formalizing the tradeoffs being made will be an important step.

REFERENCES

- [1] Michael Bowling. Convergence and no-regret in multiagent learning. *Advances in neural information processing systems*, 17:209–216, 2005.
- [2] Steven Damer and Maria Gini. Safely using predictions in general-sum normal form games. In *Proc. Int'l Conf. on Autonomous Agents and Multi-Agent Systems*, 2017.
- [3] N. Frohlich. Self-Interest or Altruism, What Difference? *Journal of Conflict Resolution*, 18(1):55–73, 1974.
- [4] Michael Johanson, Martin Zinkevich, and Michael Bowling. Computing robust counter-strategies. In *Advances in Neural Information Processing Systems (NIPS)*, pages 721–728, 2007.
- [5] Peter McCracken and Michael Bowling. Safe strategies for agent modelling in games. In *AAAI Fall Symposium on Artificial Multi-agent Learning*, October 2004.