Marissa Graham

CS 670 Midterm 1

I chose to read "Satisficing and Learning Cooperation in the Prisoner's Dilemma", which models an (arbitrarily) iterated version of the prisoner's dilemma. Since the aim of the paper is to investigate mutual cooperation over time under specific conditions for modeling player's actions, it would not make sense to use a stage game model.

The payoff matrix is as follows. The game is initially presented and the algorithm for modeling player's behavior is illustrated using a standard payoff matrix for the prisoner's dilemma, and a modified version is used to investigate the effect of the payoff matrix on the prevalence of mutual cooperation.

		Player 2's choice	
		Cooperate	Defect
Player 1's choice	Cooperate	(σ, σ)	$(0, 1)$
	Defect	$(1, 0)$	(δ, δ)

Generalized payoff matrix used to test the model, with $0 < \delta < \sigma < 1$ and $\sigma < 0.5$.

The interesting idea in this paper is to modify the idea of "rationality", where agents' behavior is assumed to have the goal of maximizing a utility function. In the real world, however, the structure of a game and its potential payoffs are not always known. It takes time and effort for an agent to even determine what their own maximum potential payoff might be, especially in scenarios where players may not even be aware that the actions of other agents are affecting their outcomes. Creating a full model of the relationships between their own actions, other player's actions, and the resulting payoffs is even more difficult and potentially unrealistic.

Modeling the prisoner's dilemma in a context without strict game-theoretic assumptions requires a reexamination of our standard solution concepts for it. A player cannot determine their maximin strategy, or whether a strategy is strategically dominant, without knowing how the actions of other players affect their own outcomes. Finding a best response strategy requires a player to not only know how other player's actions affect their own payoffs, but to have some model of the other players. A Nash equilibrium can only be a practical equilibrium if both players are aware of the actions of the other player and are disincentivized to change by being aware of how the other player's actions affect their own payoffs. Finally, agents only have the ability to move from their current strategy to one which is a Pareto improvement if they are both aware of all possible outcomes.

Overall, none of our solution concepts can be directly used by game players in their decision-making process. Players may still reach a Pareto optimal solution or a Nash equilibrium, but it will not be on purpose.

As an alternative to our previous solution concepts, agents can choose to make decisions with a satisficing algorithm. Instead of attempting to choose the *best* option, an agent attempts to choose an action that is *good enough*, so as to avoid the potential search costs involved in learning what the optimal outcome is and how to attain it. An agent will continue with a certain action for as long as his or her "good enough" aspiration levels are being met, and will otherwise search for alternatives until they are. If finding satisfactory alternatives is easy, aspiration levels will be raised, and if it is difficult, aspiration levels are lowered until a satisfactory alternative can be found (e.g. eating dry cereal for dinner after trying to think of a restaurant to eat at and nothing sounds good). The relationship of this concept to our previous solution concepts is summarized in the following table.

Solution Concept	Idea	Limitations	Relationship to Satisficing
Best response	What's the best thing for me to do, given my guess about what you're going to do?	Requires a guess about what you're going to do AND knowledge of how that will affect me.	I trade "the best thing" for a "good enough thing" to avoid modeling your behavior.
Maximin	What's the best I can guarantee myself, given that I don't know what you're going to do?	Requires knowing how your actions will affect me.	I settle for a (potentially) suboptimal payoff to avoid the search process instead of to avoid the risk of a worse payoff.
Strategic dominance	I'm better off doing this than something else, regardless of what you do.	Requires knowing how your actions will affect me.	"I like what I get from what I'm doing, so I'm going to keep doing that."
Pareto optimality	Neither of us can do any better without at least one of us being worse off.	Requires knowing how our actions affect each other.	Still can exist, but isn't part of agents' decision making process.
Nash equilibrium	If we find out each other's current strategy, neither of us will be tempted to switch to another strategy.	Requires knowing how our actions affect each other	Still can exist, but isn't part of agents' decision making process.

The satisficing algorithm itself is simple to implement, but it produces interesting dynamics between two agents that resist a simple mathematical characterization. However, the results of a game between two satisficing agents are deterministic based on the payoff matrix, the weight $\lambda \in [0, 1]$ we use to update aspiration levels to an average between player's current aspiration level and received payoff, and the player's initial actions and aspiration levels, meaning that for a two-player two-action game the results can be thought of as a function from a 7-dimensional space to the space of all possible infinite action sequences.

Generalizing the payoff matrix further can easily extend this model to other domains, and would likely result in interesting modifications to the topology of this function. We could also extend the problem to higher numbers of actions and players, and run further simulations there. Another potential next step in research could be to seek a closed form for our results function, using the techniques of dynamical systems, and to investigate its range. The results in Figure 4 suggest that the relationship between initial aspirations and final outcomes might have a fractal structure. What is this structure? Can we find initial conditions which do not converge to some stable sequence of actions? It seems possible that this is related to the rationality of the initial parameters, analagously to how rotation by a rational angle will always be periodic, but rotation by an irrational number never will. If this is the case, such behavior would never be observed in simulations, as floating point rational numbers must be rational. This would therefore need to be proved theoretically.