Altruism and Evolutionarily Stable Strategies

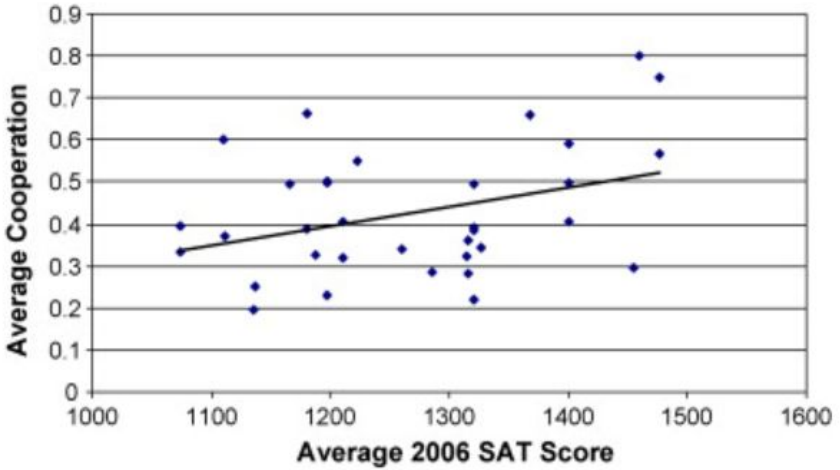
Prisoner's Dilemma

- cooperating is a strictly dominated strategy
- Nash equilibrium predicts both should defect

		P_2	
		C	D
P_1	C	$(-1,-1)$	$(-4,0)$
	D	$(0,-4)$	$(-3,-3)$

Prisoner's Dilemma - Experimental Results

- experimental results repeatedly show that people cooperate much more than they should while actually playing the game
- are they just not thinking about how the game works?



Jones 2008

Prisoner's Dilemma - Why Cooperate?

- the outcome where both people cooperate Pareto dominates the outcome where both people defect
- however, by choosing to cooperate you sacrifice some of your own gain for the gain of others
- this is called **altruism**

Would you cooperate in a prisoners' dilemma?

Not Just a Hypothetical...

- people go into burning buildings to rescue others
- people take care of others when they are sick
- people give to charity
- and so on…

The Big Questions - and Answers

- Is altruism rational?
  - Yes! We can create utility functions to model it
- What would those utility functions look like?
  - They incorporate both your own gain and the gain of your opponent
- Can people who adopt altruistic strategies survive, or is it better to maximize selfish gain?
  - Depends on the game! Sometimes being altruistic aids survival - and sometimes it doesn't

What Is Altruism?

- taking an action that **reduces your gain** and **increases the gain of others**
- contrasts with **egoism** - playing purely to maximize your own gain
- but rationality involves maximizing your own payoff
- is altruism irrational?

Why Are People Altruistic?

- a sense of fairness and equality
- wanting to preserve a social reputation
- believing that their altruistic actions will be reciprocated

Utility Functions Are What We Want Them to Be

- altruism can be perfectly rational when we **modify utility functions** to explain it
- if people have reasons for being altruistic, we can bake those in
- this models a world where not all is determined by entirely selfish gain

Utility Function Example: Fairness

$$U_M(x_M, x_J) = x_M - \alpha_M \max[x_J - x_M, 0] - \beta_M \max[x_M - x_J, 0]$$

- Mary and John are playing a game where x_M and x_J represent their material game
- Mary wants to be fair
- But maybe she wants to be a little more fair to herself than to John

Utility Function Example: Bester and Guth

$$V_1(x,y) = \alpha U_1(x,y) + (1-\alpha)U_2(x,y), \quad V_2(x,y) = \beta U_2(x,y) + (1-\beta)U_1(x,y)$$

$$1/2 \leq \alpha \leq 1, \quad 1/2 \leq \beta \leq 1$$

- for egoists, alpha and beta are 1
- altruists still care at least as much about themselves as they care about the other person

Is Altruism a Good Idea?

- we know how to model altruism…
- but just because people feel good about doing something, it doesn’t mean it’s good for them
- is a population of altruists better off than a population of egoists?
- can a group of altruists survive if dastardly egoists show up?

Evolutionarily Stable Strategies

- ESS are a stricter form of Nash equilibrium
- drastically different in terms of motivation
- Nash equilibria ask for all players to be aware of the game structure and rationalize their way into maximizing their payoffs
- but no one is going around the world thinking of their life as one big game theory problem…

Evolutionarily Stable Strategies - Motivation

- when we are born into the world, we do not think of ourselves as game players
- yet we have strategies of how deal with different situations
- to reason about whether a strategy is “good”, we can consider whether people playing the strategy will continue to be better off **even if** people playing a new, different strategy show up
- these new people can be considered to have a **mutation**
- if the mutants aren’t able to take over the world, the original strategy was evolutionarily stable

Evolutionarily Stable Strategies - Formalism

- consider two strategies S and T
- if (S, S) is a Nash equilibrium in a two player game, then $U(S, S) \geq U(T, S)$ for all possible strategies T
- if (S, S) is an ESS, then by Thomas' definition, $U(S, S) \geq U(T, S)$ for all $T \neq S$ **and** $U(S, T) > U(T, T)$
- essentially - more players changing to strategy T reduces their payoff, so strategy S is stable
- if everyone plays S, no mutant strategy can invade

Is Altruism an Evolutionarily Stable Strategy?

- sometimes!
- depends on *how* altruistic we're being, the nature of the initial payoff functions, and so on
- we will continue with the prisoner's dilemma example

Prisoner's Dilemma Revisited

- what if we play the same game **more than once**?
- consider three strategies:
  - Always Cooperate
  - Always Defect
  - Tit for Tat (cooperate the first round, then do whatever your opponent did the last round in your next round)

		P_2	
		C	D
P_1	C	$(-1,-1)$	$(-4,0)$
	D	$(0,-4)$	$(-3,-3)$

Prisoner's Dilemma Revisited

- consider a population of people that **always cooperate**
- if mutants shows up that **always defect**, they will exploit everyone and take over!
- but what about an initial population that plays **tit for tat**?
- mutants that always defect **cannot take over**, because their initial minor gain in the first round is far offset by the big losses of mutual defection in subsequent rounds

		P_2	
		C	D
P_1	C	$(-1,-1)$	$(-4,0)$
	D	$(0,-4)$	$(-3,-3)$

Prisoner's Dilemma Revisited

- when considering the two strategies **tit for tat** and **always defect**, tit for tat is **evolutionarily stable**
- always cooperate was a little *too* altruistic
- but tit for tat is altruistic as well
- it follows the model of **reciprocal altruism** - being altruistic with the expectation that the person you are being altruistic towards will return the favor

		P_2	
		C	D
P_1	C	$(-1,-1)$	$(-4,0)$
	D	$(0,-4)$	$(-3,-3)$

Altruists vs Egoists

- in general, a single altruist always does worse against a single egoist
- nut a *population* of altruists can do better than a population of egoists
- Intuition
  - two altruistic friends - each helps the other when they are sick. helping comes at a small cost, but is repaid, to both their benefits
  - two egoist friends - neither helps the other when they are sick. they avoid paying a small price to help, and then they are sad when they're sick themselves. are they really friends?
- Formalism - Bester and Guth - some fun homework

Conclusion

- altruism involves a payoff that increases with the material gains of other players
- defined this way, altruism is rational
- an evolutionarily stable strategy is one that can withstand potential invasion by mutants
- if the game structure allows for increasing mutual benefit over time through altruistic actions, a certain degree of altruism is evolutionarily stable

References

- Jones, Garett, (2008), Are smarter groups more cooperative? Evidence from prisoner's dilemma experiments, 1959-2003, *Journal of Economic Behavior & Organization*, 68, issue 3-4, p. 489-497
- Derek J. Koehler, Nigel Harvey, (2004), *Blackwell Handbook of Judgment and Decision Making*, Chapter 24
- Bester, Helmut and Guth, Werner, (1998), Is altruism evolutionarily stable?, *Journal of Economic Behavior & Organization*, 34, issue 2, p. 193-209