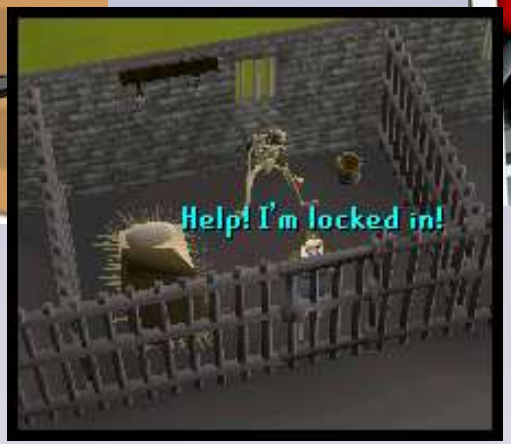

PRISONER'S DILEMMA

THE MATHEMATICS OF BETRAYAL

(AND DATING AND VARIOUS OTHER THINGS)

Math Club 9/26/2011

Prisoner's Dilemma?

- Two criminals are arrested, but the police do not have enough information to convict them.
- The two men are put into separate rooms and are given a deal by the police: to give information (in other words, betray the other) in exchange for freedom.
- If neither prisoner betrays the other, then the police cannot convict them and has to let them both off with a light, 1 month sentence.
- If one prisoner betrays the other, then the 'betrayer' gets away free, and the 'betrayee' gets a heavy, 1 year sentence.
- If both prisoners betray each other, then both get a medium, 3 month sentence.

Payoff Matrix

	Prisoner B stays silent (*cooperates*)	**Prisoner B confesses (*defects*)**
Prisoner A stays silent (*cooperates*)	Each serves 1 month	Prisoner A: 1 year Prisoner B: goes free
Prisoner A confesses (*defects*)	Prisoner A: goes free Prisoner B: 1 year	Each serves 3 months

	Cooperate	**Defect**
Cooperate	(3,3)	(0,5)
Defect	(5,0)	(1,1)

Let's try this!

- We will play five rounds.
- The goal is to get as many points as possible, **not** to beat your opponent.

(Naïve) Mathematical Analysis

	Cooperate	Defect
Cooperate	(3,3)	(0,5)
Defect	(5,0)	(1,1)

- If they cooperate, you're better off defecting.
- If they defect, you're still better off defecting.
- Result: you should always defect.
- Yet cooperating yields the most overall *net* gain…

Dating…

- … can be extremely problematic.

Dating

- You arrange a date with a friend at the Stampede.
- But when you get there, you discover that the Stampede is huge and you have no idea where to look for her.
- You prefer to go to the *horseracing show*, while your date prefers to go to the *cooking show*.
- However, both of you prefer to be with each other and go to the same place, rather than be alone.
- You know that your date has to make the same decision.
- Should you go to the horseracing show or the cooking show?

Payoff Matrix

	Cooperate (Horseracing)	Defect (Cooking)
Cooperate (Cooking)	(1,1)	(4,5)
Defect (Horseracing)	(5,4)	(2,2)

Let's try this!

- We will play five rounds.
- The goal is to get as many points as possible, **not** to beat your opponent.

(Naïve) Mathematical Analysis

	Cooperate (Horseracing)	Defect (Cooking)
Cooperate (Cooking)	(1,1)	(4,5)
Defect (Horseracing)	(5,4)	(2,2)

- If they cooperate, you're better off defecting.
- If they defect, you're better off cooperating.
- But if *you* cooperate, *they're* better off defecting and if *you* defect, *they're* better off defecting.
- Result: ?????????

Did math fail us?!

- Well, sort of.
- But if we are God…

Kant's Categorical Imperative

"Act only according to that maxim whereby you can, at the same time, will that it should become a universal law." – Immanuel Kant

- In other words, assume that everyone in the world acts and makes decisions in the same way that you do.
- Then how should you act?

Prisoner's Dilemma

	Cooperate	Defect
Cooperate	(3,3)	(0,5)
Defect	(5,0)	(1,1)

- Clearly it's better if everyone cooperates all the time than if everyone defects all the time.

Dating

	Cooperate (Horseracing)	Defect (Cooking)
Cooperate (Cooking)	(1,1)	(4,5)
Defect (Horseracing)	(5,4)	(2,2)

- Suppose that everyone cooperates with x probability, and defects with $1 - x$ probability. What should x be?
- The probability that both of you cooperate is x^2, so that payoff is x^2.
- The probability that you cooperate and they defect is $x(1 - x)$ with payoff $4x(1 - x)$.
- The probability that you defect and they cooperate is $x(1 - x)$ with payoff $5x(1 - x)$.
- The probability that both of you defect is $(1 - x)^2$ with payoff $2(1 - x)^2$.

Some algebra

- We want to choose x between 0 and 1 as to maximize
$$x^2 + 9x(1-x) + 2(1-x)^2$$
- $-7x^2 + 7x + 1$
- $-\frac{7}{2}\left(x - \frac{1}{2}\right)^2 + \frac{11}{4}$
- So x is maximized at $\frac{1}{2}$.
- You should cooperate exactly half the time!

Optional Exercise

- What if you're '*meh*' about your date and you would rather watch horseracing alone rather than cooking with your date?

	Cooperate (Horseracing)	Defect (Cooking)
Cooperate (Cooking)	(0,0)	(2,5)
Defect (Horseracing)	(5,2)	(3,3)

- How often should you cooperate, assuming Kant's categorical imperative?