PHIL 408Q/PHPE 308D
Fairness

Eric Pacuit, University of Maryland

February 6, 2024

Nash Bargaining Game: Two players receive a windfall. Each player makes a demand, and if the two demands do not exceed the total good, both receive their demand. Otherwise, both receive nothing.

Nash Bargaining Game: Two players receive a windfall. Each player makes a demand, and if the two demands do not exceed the total good, both receive their demand. Otherwise, both receive nothing.

Simplified Nash Bargaining Game: Suppose that two players are dividing $1.

Nash Bargaining Game: Two players receive a windfall. Each player makes a demand, and if the two demands do not exceed the total good, both receive their demand. Otherwise, both receive nothing.

Simplified Nash Bargaining Game: Suppose that two players are dividing \$1. The players write their demand on a sheet of paper and hand it to a referee. If the total is less than \$1, then each player gets what they demand. But, if the total is greater than \$1, then each player gets nothing.

Nash Bargaining Game: Two players receive a windfall. Each player makes a demand, and if the two demands do not exceed the total good, both receive their demand. Otherwise, both receive nothing.

Simplified Nash Bargaining Game: Suppose that two players are dividing $1. The players write their demand on a sheet of paper and hand it to a referee. If the total is less than $1, then each player gets what they demand. But, if the total is greater than $1, then each player gets nothing. Assume that each player has 5 available actions:

1. 0: Demand nothing
2. 1/3: Demand 1/3 of $1 (so 33 cents)
3. 1/2: Demand 1/2 of $1 (so 50 cents)
4. 2/3: Demand 2/3 of $1 (so 66 cents)
5. 1: Demand $1

Simplified Nash Bargaining Game

		Player 2				
		0	1/3	1/2	2/3	1
Player 1	0	$0, 0$	$0, \frac{1}{3}$	$0, \frac{1}{2}$	$0, \frac{2}{3}$	$0, 1$
	1/3	$\frac{1}{3}, 0$	$\frac{1}{3}, \frac{1}{3}$	$\frac{1}{3}, \frac{1}{2}$	$\frac{1}{3}, \frac{2}{3}$	$0, 0$
	1/2	$\frac{1}{2}, 0$	$\frac{1}{2}, \frac{1}{3}$	$\frac{1}{2}, \frac{1}{2}$	$0, 0$	$0, 0$
	2/3	$\frac{2}{3}, 0$	$\frac{2}{3}, \frac{1}{3}$	$0, 0$	$0, 0$	$0, 0$
	1	$1, 0$	$0, 0$	$0, 0$	$0, 0$	$0, 0$

Simplified Nash Bargaining Game

		Player 2				
		0	1/3	1/2	2/3	1
Player 1	0	0, 0	0, $\frac{1}{3}$	0, $\frac{1}{2}$	0, $\frac{2}{3}$	**0, 1**
	1/3	$\frac{1}{3}$, 0	$\frac{1}{3}$, $\frac{1}{3}$	$\frac{1}{3}$, $\frac{1}{2}$	**$\frac{1}{3}$, $\frac{2}{3}$**	0, 0
	1/2	$\frac{1}{2}$, 0	$\frac{1}{2}$, $\frac{1}{3}$	**$\frac{1}{2}$, $\frac{1}{2}$**	0, 0	0, 0
	2/3	$\frac{2}{3}$, 0	**$\frac{2}{3}$, $\frac{1}{3}$**	0, 0	0, 0	0, 0
	1	**1, 0**	0, 0	0, 0	0, 0	0, 0

What will people do, when given this problem?

What will people do, when given this problem? I expect that we would all give the same answer — almost everyone will claim half...In fact, the experiment has been done. Nydegger and Owen asked subjects to divide a dollar among themselves. There were no surprises. All agreed to a fifty-fifty split. The experiment is not widely discussed because it is not thought of as an anomaly. Results are just what everyone expected. (p. 4, Skyrms)

R. V. Nydegger and G. Owen. *Two-Person Bargaining: An Experimental Test of the Nash Axioms*. International Journal of Game Theory, 3(4), pp. 239 -249.

We think we know the right answer to the problem, but why is it right? In what sense is it right?

We think we know the right answer to the problem, but why is it right? In what sense is it right? Let us see whether *informed rational self-interest* will give us an answer. If I want to get as much as possible, the best claim for me to write down depends on what you write down. Likewise, your optimum claim depends on what I write down. We have two interactive optimization problems. A solution to our problem will consist of solutions to each optimization problems that are in *equilibrium*.

We think we know the right answer to the problem, but why is it right? In what sense is it right? Let us see whether *informed rational self-interest* will give us an answer. If I want to get as much as possible, the best claim for me to write down depends on what you write down. Likewise, your optimum claim depends on what I write down. We have two interactive optimization problems. A solution to our problem will consist of solutions to each optimization problems that are in *equilibrium*.

...Such an equilibrium would be even more compelling if it were not only true that one could not gain by unilaterally deviating from it, but also that on such a deviation one would definitely do worse than one would have done at equilibrium. An equilibrium with this additional stability property is a *strict Nash equilibrium*.

(pp. 4-5, Skyrms)

Player 2

Player 1

	0	1/3	1/2	2/3	1
0	$0, 0$	$0, \frac{1}{3}$	$0, \frac{1}{2}$	$0, \frac{2}{3}$	$0, 1$
1/3	$\frac{1}{3}, 0$	$\frac{1}{3}, \frac{1}{3}$	$\frac{1}{3}, \frac{1}{2}$	$\frac{1}{3}, \frac{2}{3}$	$0, 0$
1/2	$\frac{1}{2}, 0$	$\frac{1}{2}, \frac{1}{3}$	$\frac{1}{2}, \frac{1}{2}$	$0, 0$	$0, 0$
2/3	$\frac{2}{3}, 0$	$\frac{2}{3}, \frac{1}{3}$	$0, 0$	$0, 0$	$0, 0$
1	$1, 0$	$0, 0$	$0, 0$	$0, 0$	$0, 0$

In fact, every pair of positive claims that total 100% is a strict Nash equilibrium. There is a profusion of strict equilibrium solutions to our problem of dividing the cake, but we want to say that only one of them is just. Equilibrium in informed rational self-interest, even when strictly construed, does not explain our conception of justice. (Skyrms, p. 5)

Veil of Ignorance

Suppose that you and I are deciding how to divide M between individuals A and B, under the condition that a referee will later decide whether you are A and I am B or conversely.

Veil of Ignorance

Suppose that you and I are deciding how to divide M between individuals A and B, under the condition that a referee will later decide whether you are A and I am B or conversely.

What is the rational choice under this veil of ignorance?

Veil of Ignorance: Harsanyi

Suppose that the referee will flip a fair coin to decide whether I am A or B.

Veil of Ignorance: Harsanyi

Suppose that the referee will flip a fair coin to decide whether I am A or B.

If *all you care about is the expected amount of M*, then you should evaluate a division p for A and $M - p$ for B by calculating the expected utility:

$$p\frac{1}{2} + (M - p)\frac{1}{2} = \frac{M}{2}$$

Veil of Ignorance: Harsanyi

Suppose that the referee will flip a fair coin to decide whether I am A or B.

If *all you care about is the expected amount of M*, then you should evaluate a division p for A and $M - p$ for B by calculating the expected utility:

$$p\frac{1}{2} + (M - p)\frac{1}{2} = \frac{M}{2}$$

But this means that every strict Nash equilibrium has the same expected utility: “The Harsanyi-Rawls veil of ignorance has not helped at all with this problem (though it would with others.)”

Veil of Ignorance: Rawls

Rawls doesn't have the referee flip the coin. We don't know anything at all about Ms. Fortuna. In my ignorance, he argues, I should guard myself by acting as if she doesn't like me. So should you. We should follow the decision rule of maximizing minimum gain. Then we will both agree on the 50%-50% split.

Veil of Ignorance: Rawls

Rawls doesn't have the referee flip the coin. We don't know anything at all about Ms. Fortuna. In my ignorance, he argues, I should guard myself by acting as if she doesn't like me. So should you. We should follow the decision rule of maximizing minimum gain. Then we will both agree on the 50%-50% split. This gets us the desired conclusion, but on what basis? Why should we both be paranoid? After all, if there is an unequal division between A and B, Fortuna can't very well decide against both of us. (Skyrms, p. 7)

An Evolutionary Model

Individuals, paired at random from a large population, play our bargaining game.

An Evolutionary Model

Individuals, paired at random from a large population, play our bargaining game.

The total M represents a quantity of *Darwinian fitness* that can be divided and transferred.

An Evolutionary Model

Individuals, paired at random from a large population, play our bargaining game.

The total M represents a quantity of *Darwinian fitness* that can be divided and transferred.

In this simple model, individuals have strategies programmed in, and the strategies replicate themselves in accord with the *evolutionary fitness* that they receive in the bargaining interactions.

An Evolutionary Model

Individuals, paired at random from a large population, play our bargaining game.

The total M represents a quantity of *Darwinian fitness* that can be divided and transferred.

In this simple model, individuals have strategies programmed in, and the strategies replicate themselves in accord with the *evolutionary fitness* that they receive in the bargaining interactions.

Notice that in this setting it is the strategies that come to the fore; the individuals that implement them on various occasions recede from view.... The identity of the individuals playing is unimportant and is continually shifting. This is the *Darwinian Veil of Ignorance*.

Suppose that there is a population of individuals all demanding 60% of M.

Suppose that there is a population of individuals all demanding 60% of M.

- Meeting each other they get nothing.
- If anyone were to demand a positive amount less than 40%, she would get that amount and thus do better than the population average.
- Thus, any mutant demanding less than 40% will eventually take over the population

Likewise, for any population of individuals that demand more than 50% (and less than 100%).

Suppose we have a population demanding 30%.

Suppose we have a population demanding 30%.

- Meeting each other they get 30%.
- Anyone demanding a bit more will do better than the population average.
- Thus, any mutant demanding a bit more than 30% will eventually take over the population

Likewise, for any population of individuals that demand less than 50%.

The only strategies that can be equilibrium strategies under the Darwinian veil of ignorance are Demand 50% and Demand 100%.

The strategy Demand 100% is an equilibrium, but an *unstable one*.

The strategy Demand 100% is an equilibrium, but an *unstable one*.

- In a population in which everyone demands 100%, everyone gets nothing, and if a mutant popped up who made a different demand against 100 percenters, she would also get nothing.

The strategy Demand 100% is an equilibrium, but an *unstable one*.

- In a population in which everyone demands 100%, everyone gets nothing, and if a mutant popped up who made a different demand against 100 percenters, she would also get nothing.
- But suppose that a small proportion of modest mutants arose who demanded, for example, 45%.

The strategy Demand 100% is an equilibrium, but an *unstable one*.

- In a population in which everyone demands 100%, everyone gets nothing, and if a mutant popped up who made a different demand against 100 percenters, she would also get nothing.
- But suppose that a small proportion of modest mutants arose who demanded, for example, 45%.
- Most of the time they would be paired with 100 percenters and get nothing, but some of the time they would be paired with each other and get 45%.
- On average their payoff would be higher than that of the population, and they would increase.

Demand 50% is a *stable equilibrium*: In a population in which everyone demands half of the cake, any mutant (or group of mutants) who demanded anything different would get less than the population average.

The strategy Demand 50% is the unique *evolutionarily stable strategy*

J. McKenzie Alexander. *Evolutionary Game Theory*. The Stanford Encyclopedia of Philosophy (Summer 2021 Edition), Edward N. Zalta (ed.).

Fair division will be stable in any dynamics with a tendency to increase the proportion (or probability) of strategies with greater payoffs, because any unilateral deviation from fair division results in a strictly worse payoff. For this reason, the Darwinian story can be transposed into the context of *cultural evolution*, in which imitation and learning may play an important role in the dynamics. (Skyrms, p. 11)

Polymorphism

If we look more deeply into the matter, however, complications arise. In both the case of sex ratio and dividing the cake, we considered the evolutionary stability of pure strategies. We did not examine the possibility that evolution might not lead to the fixation of a pure strategy, but rather to a *polymorphic* state of the population in which some proportion of the population plays one pure strategy and some proportion of the population plays another. (Skyrms, p. 11)

Polymorphism Equilibrium

Suppose that half of the population claims 2/3 of the cake and half the population claims 1/3. Let us call the first strategy Greedy and the second Modest.

- A greedy individual stands an equal chance of meeting another greedy individual or a modest individual.
- If she meets another greedy individual she gets nothing because their claims exceed the whole cake, but if she meets a modest individual, she gets 2/3. Her average payoff is 1/3.
- A modest individual, on the other hand, gets a payoff of 1/3 no matter whom she meets.

A Polymorphism Equilibrium is Stable

- If the proportion of greedys should rise, then greedys would meet each other more often and the average payoff to greedy would fall below the $1/3$ guaranteed to modest.

A Polymorphism Equilibrium is Stable

- If the proportion of greedys should rise, then greedys would meet each other more often and the average payoff to greedy would fall below the $1/3$ guaranteed to modest.

- If the proportion of greedys should fall, the greedys would meet modests more often, and the average payoff to greedys would rise above $1/3$.

A Polymorphism Equilibrium is Stable

- Suppose that a Supergreedy mutant who demands more than 2/3 arises in this population. This mutant gets payoff of 0 and goes extinct.

A Polymorphism Equilibrium is Stable

- Suppose that a Supergreedy mutant who demands more than 2/3 arises in this population. This mutant gets payoff of 0 and goes extinct.
- Suppose that a Supermodest mutant who demands less than 1/3 arises in the population. This mutant will get what she asks for, which is less than greedy and modest get, so she will also go extinct - although more slowly than supergreedy will.

A Polymorphism Equilibrium is Stable

- Suppose that a Supergreedy mutant who demands more than 2/3 arises in this population. This mutant gets payoff of 0 and goes extinct.
- Suppose that a Supermodest mutant who demands less than 1/3 arises in the population. This mutant will get what she asks for, which is less than greedy and modest get, so she will also go extinct - although more slowly than supergreedy will.
- The remaining possibility is that a middle-of-the-road mutant arises who asks for more than modest but less than greedy. A case of special interest is that of the Fair-minded mutant who asks for exactly 1/2. All of these mutants would get nothing when they meet greedy and get less than greedy does when they meet modest. Thus they will all have an average payoff less than 1/3 and all - including our fair-minded mutant - will be driven to extinction.

This is unhappy news, for the population as well as for the evolution of justice, because our polymorphism is inefficient. Here everyone gets, on average, 1/3 of the cake — while 1/3 of the cake is squandered in modest encounters.

This is unhappy news, for the population as well as for the evolution of justice, because our polymorphism is inefficient. Here everyone gets, on average, 1/3 of the cake — while 1/3 of the cake is squandered in modest encounters.

Compare this equilibrium with the pure equilibrium where everyone demands and gets 1/2 of the cake.

This is unhappy news, for the population as well as for the evolution of justice, because our polymorphism is inefficient. Here everyone gets, on average, 1/3 of the cake — while 1/3 of the cake is squandered in modest encounters.

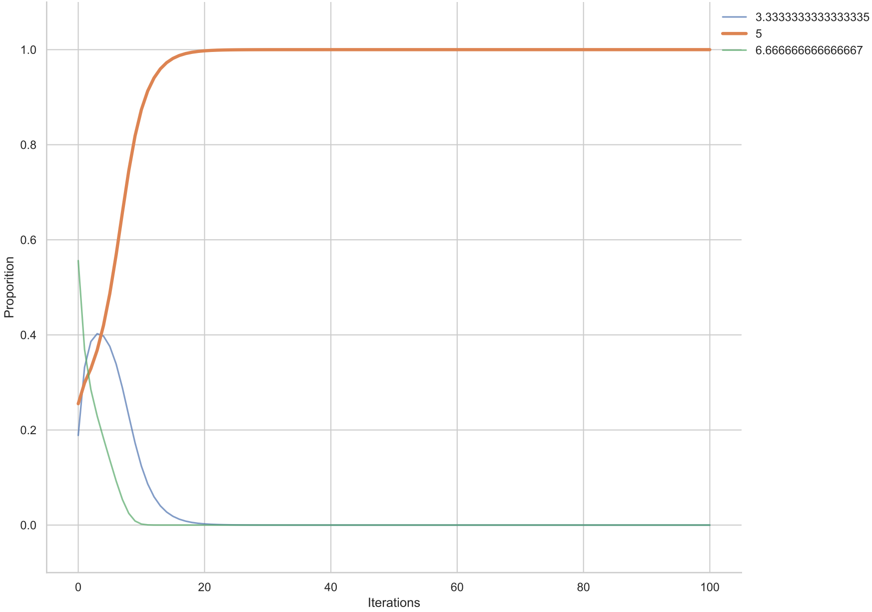
Compare this equilibrium with the pure equilibrium where everyone demands and gets 1/2 of the cake.

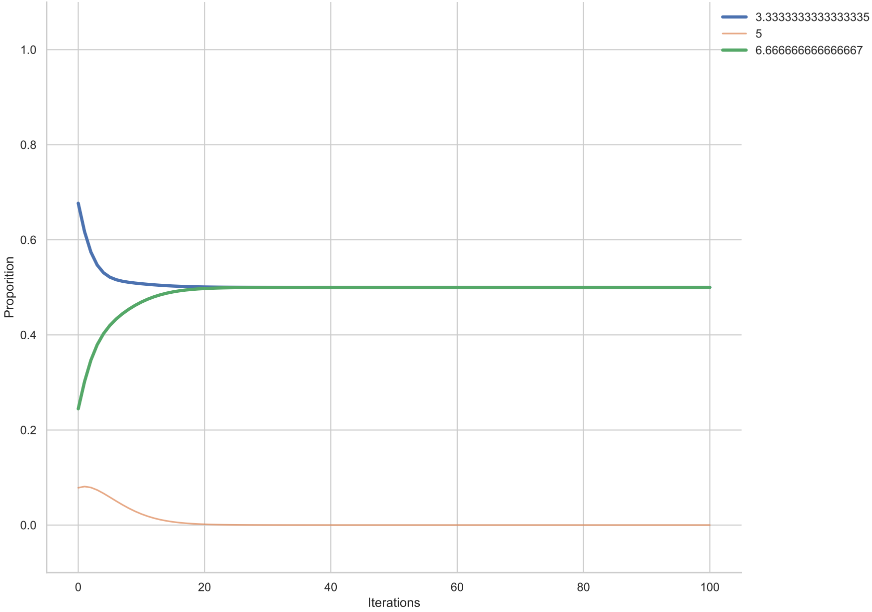
In view of both the inefficiency and the strong stability properties of the 1/3-2/3 polymorphism, it appears to be a kind of trap that the population could fall into, and from which it could be difficult to escape. (Skyrms, p. 13)

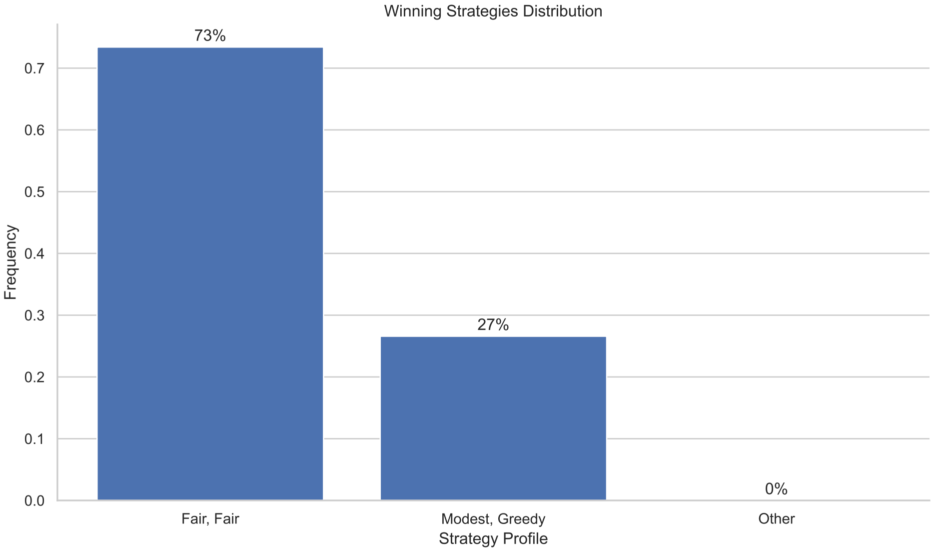
There are many polymorphic traps: For any number, x between 0 and 1, there is a polymorphism of the two strategies Demand x, Demand $1 - x$, which is a stable equilibrium by essentially the same reasoning.

There are many polymorphic traps: For any number, x between 0 and 1, there is a polymorphism of the two strategies Demand x, Demand $1 - x$, which is a stable equilibrium by essentially the same reasoning.

The existence of polymorphic traps does not make the situation hopeless....We would like to know how probable it is that a population would evolve to the rule of share and share alike, and how probable it is that it will slip into a polymorphic trap.


3.3333333333333335
5
6.666666666666667
1.0
0.8
0.6
0.4
0.2
0.0
Proporition
0
20
40
60
80
100
Iterations


3.3333333333333335
5
6.666666666666667
1.0
0.8
0.6
0.4
0.2
0.0
Proportion
0
20
40
60
80
100
Iterations


Winning Strategies Distribution
Frequency
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
73%
27%
0%
Fair, Fair
Modest, Greedy
Other
Strategy Profile