PHIL 408Q/PHPE 308D
Fairness

Eric Pacuit, University of Maryland

February 20, 2024

Unfairness

Cailin O'Connor (2022). *Why Natural Social Contracts are Not Fair*. forthcoming in *New Social Contract Theory*.


the
origins of
unfairness
cailin
o'connor
social
categories
and
cultural
evolution

One especially persuasive branch of argumentation shows how fairness, in the sense of equal divisions of resources, is likely to emerge on cultural evolutionary timescales. Models of this sort are used to argue that natural social contracts often emerge that are, for the most part, fair.

One especially persuasive branch of argumentation shows how fairness, in the sense of equal divisions of resources, is likely to emerge on cultural evolutionary timescales. Models of this sort are used to argue that natural social contracts often emerge that are, for the most part, fair.

Of course, when we look at real world conventions and norms regarding the division of resources, fairness is not typically the rule....despite the high ideals and optimism of traditional social contract theorists, the real world is rife with inequity....How do we square these observations with the modeling literature showing that fairness emerges naturally via cultural evolution?

Simplified Nash Bargaining Game

	Low	*Medium*	*High*
Low	3, 3	3, 5	3, 7
Medium	5, 3	5, 5	0, 0
High	7, 3	0, 0	0, 0

Simplified Nash Bargaining Game

	Low	*Medium*	*High*
Low	3, 3	3, 5	3, 7
Medium	5, 3	5, 5	0, 0
High	7, 3	0, 0	0, 0

If the players demand more than the total, then they get a **disagreement point** of 0.

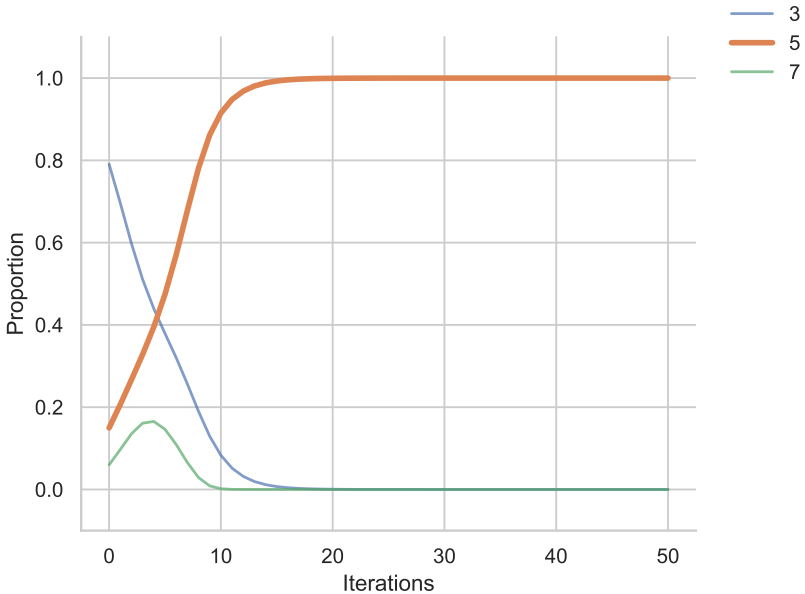
Simplified Nash Bargaining Game

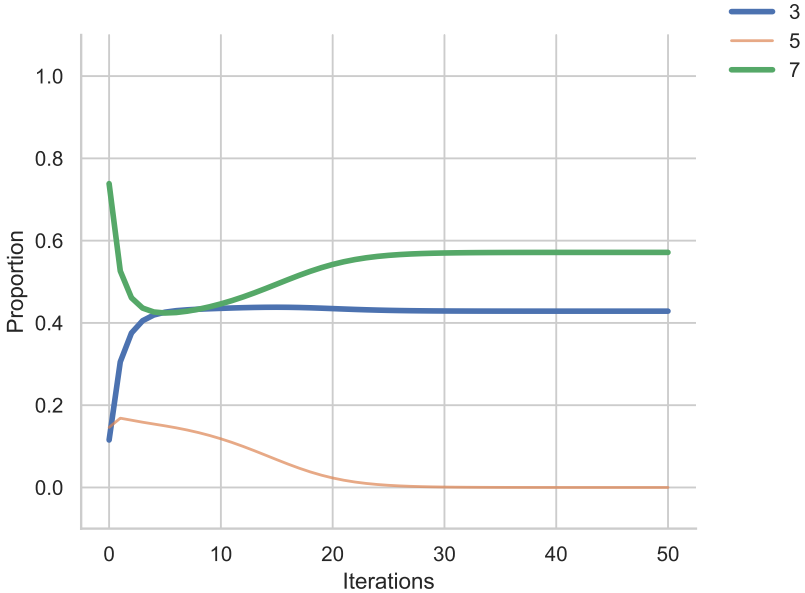
	Low	*Medium*	*High*
Low	3, 3	3, 5	3, 7
Medium	5, 3	5, 5	0, 0
High	7, 3	0, 0	0, 0

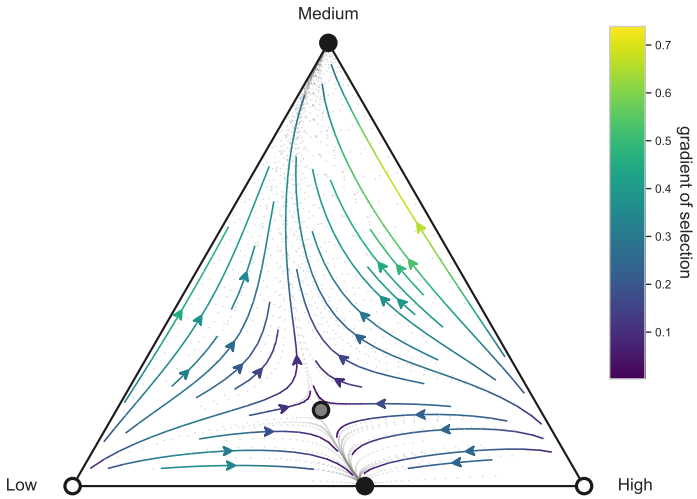
In each of these Nash equilibria the resource is perfectly divided, but only one of them—the Medium vs. Medium equilibrium—is usually considered "fair".

Under most standard evolutionary dynamics (rules for learning and cultural evolution) a population playing this game is most likely to head towards a stable outcome where all agents make the Medium demand. In other words, a fair social contract emerges endogenously....

The other stable evolutionary outcome is sometimes called a "fractious" outcome, because agents sometimes coordinate their demands, but also sometimes miscoordinate by playing High vs High or Low vs Low.


3
5
7
1.0
0.8
0.6
0.4
0.2
0.0
Proportion
0
10
20
30
40
50
Iterations


3
5
7
1.0
0.8
0.6
0.4
0.2
0.0
Proportion
0
10
20
30
40
50
Iterations


Medium
Low
High
gradient of selection
0.7
0.6
0.5
0.4
0.3
0.2
0.1

[T]hese models show that fair conventions of behavior do tend to emerge naturally from an uncoordinated "state of nature". They support the idea that natural social contracts tend to favor equality.

To this point, I have described how evolutionary game theoretic models support claims about the emergence of fair social contracts. But social contracts are not always fair, as a great deal of research in the social sciences has shown....How do we square the models...with these criticisms?

To this point, I have described how evolutionary game theoretic models support claims about the emergence of fair social contracts. But social contracts are not always fair, as a great deal of research in the social sciences has shown....How do we square the models...with these criticisms? The answer is that we need to add *social categories* to these models. A social category is a recognizable group within a society. Most important to us here are primary categories, which Ridgeway (2011) describes as the small number of social categories most generally used for coordinating behavior. Across societies, these always include gender and age, and often also include race, religion, caste, or class.

Tags

Our model will involve a population with two groups (representing social categories) that each have a different arbitrary *tag*. The tags might be “green” and “yellow”, for example, or “star-belly” and “plain belly”.

Tags

Our model will involve a population with two groups (representing social categories) that each have a different arbitrary *tag*. The tags might be "green" and "yellow", for example, or "star-belly" and "plain belly".

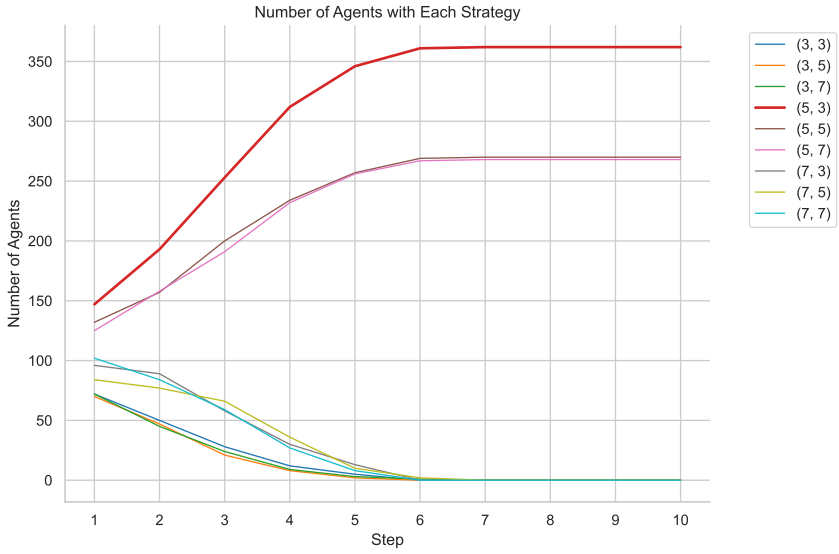
Agents in this model play the bargaining game...but in doing so may condition their strategy on the tag of their partner.

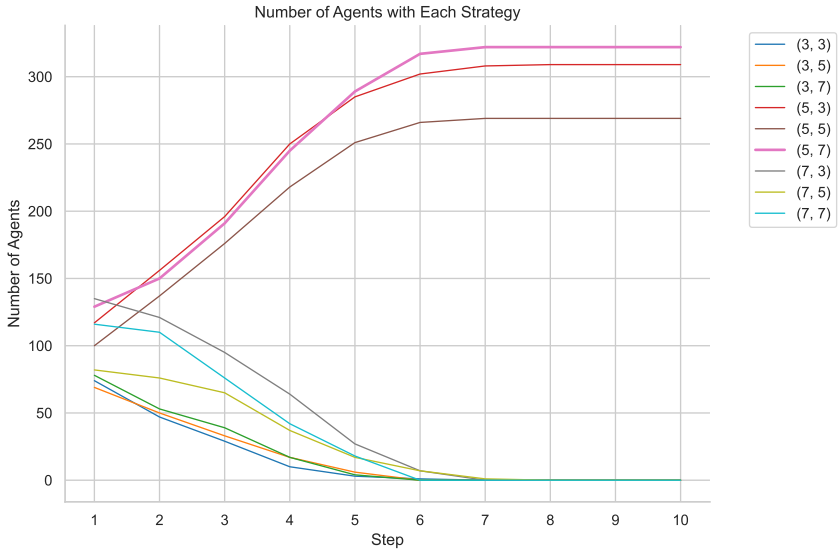
Tags

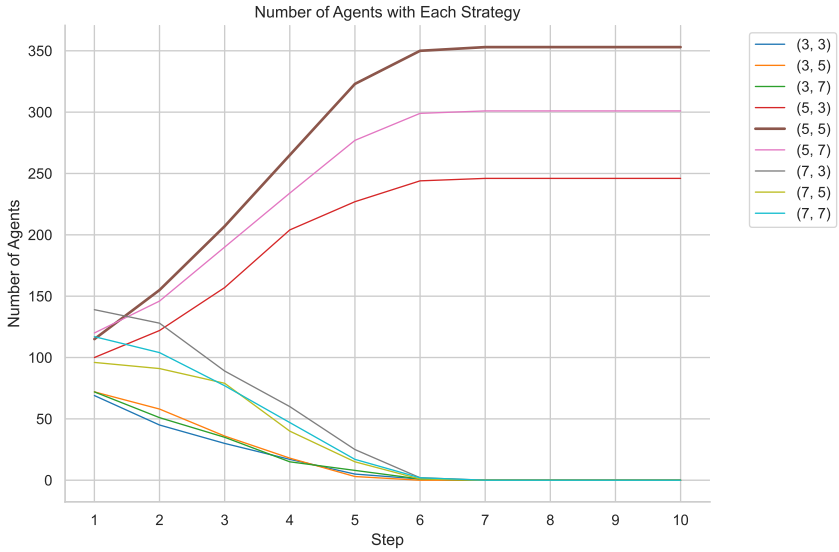
Our model will involve a population with two groups (representing social categories) that each have a different arbitrary *tag*. The tags might be "green" and "yellow", for example, or "star-belly" and "plain belly".

Agents in this model play the bargaining game...but in doing so may condition their strategy on the tag of their partner.

For example, an agent in the green group might play Medium against other greens, and Low against yellows. We can label this two part strategy, listing the in-group strategy first, as follows: $\langle Medium, Low \rangle$. For now, we can also assume that agents learn from in-group members only. I.e., a yellow will only copy the strategies of other yellows.


Number of Agents with Each Strategy
Number of Agents
Step
0
50
100
150
200
250
300
350
1
2
3
4
5
6
7
8
9
10
(3, 3)
(3, 5)
(3, 7)
(5, 3)
(5, 5)
(5, 7)
(7, 3)
(7, 5)
(7, 7)


Number of Agents with Each Strategy
Number of Agents
Step
0
50
100
150
200
250
300
1
2
3
4
5
6
7
8
9
10
(3, 3)
(3, 5)
(3, 7)
(5, 3)
(5, 5)
(5, 7)
(7, 3)
(7, 5)
(7, 7)


Number of Agents with Each Strategy
Number of Agents
Step
0
50
100
150
200
250
300
350
1
2
3
4
5
6
7
8
9
10
(3, 3)
(3, 5)
(3, 7)
(5, 3)
(5, 5)
(5, 7)
(7, 3)
(7, 5)
(7, 7)

What happens when we evolve this game?

What happens when we evolve this game?

The stable end points, or evolutionary equilibria, are different from those described in the single population model: Within each group, the stable equilibria mimic those for a single population. The greens, for example, might all make fair demands of each other, or settle on the fractious equilibrium. And ditto the yellows. This is because within-group evolution just recreates the conditions of a single population.

Between groups, there are three stable equilibria, one where both groups make fair demands of the other, one where the yellows demand High and greens Low, and one where the yellows demand Low and the greens High.

These latter two equilibria can be thought of as bare bones representations of a discriminatory convention or norm.

Another set of relevant results focuses not on the Nash demand game, but on breaking symmetry in social coordination problems. In some situations, actors need to use complementary strategies in a game to do well, but one strategy is preferable.

Hawk-Dove

	Dove	*Hawk*
Dove	$1, 1$	$1, 3$
Hawk	$3, 1$	$0, 0$

Hawk-Dove

	Dove	*Hawk*
Dove	1, 1	1, 3
Hawk	3, 1	0, 0

The two equilibria, bolded in this figure, are Hawk vs Dove and Dove vs Hawk. At either of these equilibria it is better to be a hawk. Hoffmann (2006) and Amadae and Watts (2022) both illustrate how groups with categories playing this game can evolve to situations where one side always plays dove, to their disadvantage.

Complimentary Coordination

	A	B
A	$0, 0$	α, β
B	β, α	$0, 0$

Complimentary Coordination

	A	B
A	$0, 0$	α, β
B	β, α	$0, 0$

In this scenario, two actors must take complementary strategies, A and B, to succeed. This might represent division of labor, where A involves one set of jobs and B a complementary set. A population with two groups, say men and women, might evolve to solve this problem when one group always plays A (engages in market labor) and the other B (focuses on household labor).

Complimentary Coordination

	A	B
A	$0, 0$	α, β
B	β, α	$0, 0$

In this scenario, two actors must take complementary strategies, A and B, to succeed. This might represent division of labor, where A involves one set of jobs and B a complementary set. A population with two groups, say men and women, might evolve to solve this problem when one group always plays A (engages in market labor) and the other B (focuses on household labor). But when one outcome is preferable, say $\beta > \alpha$, this leads to persistent advantage for one group (O'Connor, 2019).

Thus categories allow for coordination on a new set of efficient equilibria. But they also allow for categorical inequity that would not otherwise be possible.

Taken together, the results described here show that critiques from theorists like Mills and Pateman are supported by models of cultural evolution.

Taken together, the results described here show that critiques from theorists like Mills and Pateman are supported by models of cultural evolution. Whenever social categories like gender and race are in place, fairness is not what we typically expect from natural social contracts. Instead, we often expect unfairness.

Taken together, the results described here show that critiques from theorists like Mills and Pateman are supported by models of cultural evolution. Whenever social categories like gender and race are in place, fairness is not what we typically expect from natural social contracts. Instead, we often expect unfairness. Given that social categories are culturally ubiquitous we thus should not expect emergent social contracts to be fair.

Taken together, the results described here show that critiques from theorists like Mills and Pateman are supported by models of cultural evolution. Whenever social categories like gender and race are in place, fairness is not what we typically expect from natural social contracts. Instead, we often expect unfairness. Given that social categories are culturally ubiquitous we thus should not expect emergent social contracts to be fair.

To clarify, the claim here is not that these models exactly match the pictures presented by either author. Rather, they confirm a picture where cultural divisions that ought not impact resource distribution in an ideal, just social contract tend to nonetheless become deeply relevant to natural, emergent contracts.

Power

[O]ne thing missing from the models discussed to this point is the coercive nature of the way inequitable contracts are often formed in reality. There is no coercion in these models, and there is no sense of power inequity between groups.

Power

[O]ne thing missing from the models discussed to this point is the coercive nature of the way inequitable contracts are often formed in reality. There is no coercion in these models, and there is no sense of power inequity between groups. Part of what makes them such effective epistemic tools, in fact, is the way that otherwise entirely identical groups starting from neutral states ("of nature") can evolve to stable, discriminatory norms.

Power

[O]ne thing missing from the models discussed to this point is the coercive nature of the way inequitable contracts are often formed in reality. There is no coercion in these models, and there is no sense of power inequity between groups. Part of what makes them such effective epistemic tools, in fact, is the way that otherwise entirely identical groups starting from neutral states ("of nature") can evolve to stable, discriminatory norms. But we still might wish to know: what happens if we add power to these models?

Power is most often included in bargaining models via disagreement points.

Agents with different disagreement points will end up in different positions should bargaining fail, which impacts how much the bargain matters to them, and which, in turn, impacts their power over the bargain.

Power is most often included in bargaining models via disagreement points.

Agents with different disagreement points will end up in different positions should bargaining fail, which impacts how much the bargain matters to them, and which, in turn, impacts their power over the bargain.

On his interpretation, before bargaining, each agent makes some threat about what they will do if bargaining fails. More powerful threats translate into lower disagreement points (or *threat-points*) for the other agent. Thus agents can use coercion (or power-over) to reshape their opponent's bargaining position.

Power is most often included in bargaining models via disagreement points.

Agents with different disagreement points will end up in different positions should bargaining fail, which impacts how much the bargain matters to them, and which, in turn, impacts their power over the bargain.

On his interpretation, before bargaining, each agent makes some threat about what they will do if bargaining fails. More powerful threats translate into lower disagreement points (or *threat-points*) for the other agent. Thus agents can use coercion (or power-over) to reshape their opponent's bargaining position.

Alternatively, agents might have different disagreement points because of material or political differences in their lives that make bargains more or less important to them.

		Player 2		
		Low	*Medium*	*High*
Player 1	*Low*	3, 3	3, 5	3, 7
	Medium	5, 3	5, 5	0, 0
	High	7, 3	0, 0	0, 0

		Player 2		
		Low	*Medium*	*High*
Player 1	*Low*	3, 3	3, 5	3, 7
	Medium	5, 3	5, 5	D, d
	High	7, 3	D, d	D, d

When $D > d$, Player 1 has power over Player 2 in the bargaining game.

The power imbalance systematically advantages the more powerful group, who tend to end up at the outcome where they demand High more often. The greater the power, the greater the discrepancy. This happens because powerful individuals have relatively little incentive to adopt low demands—their disagreement point is not much worse. As a result they move towards such demands more slowly, and tend to end up adopting higher demands instead.

Justin Bruner and Cailin O'Connor (2018). *Power, Bargaining, and Collaboration*. in *Scientific Collaboration and Collective Knowledge* Ed. Conor Mayo-Wilson Thomas Boyer and Michael Weisberg. Oxford University Press.