PHPE 308M/PHIL 209F
Fairness

Eric Pacuit, University of Maryland

September 24, 2025

Unfairness

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

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.

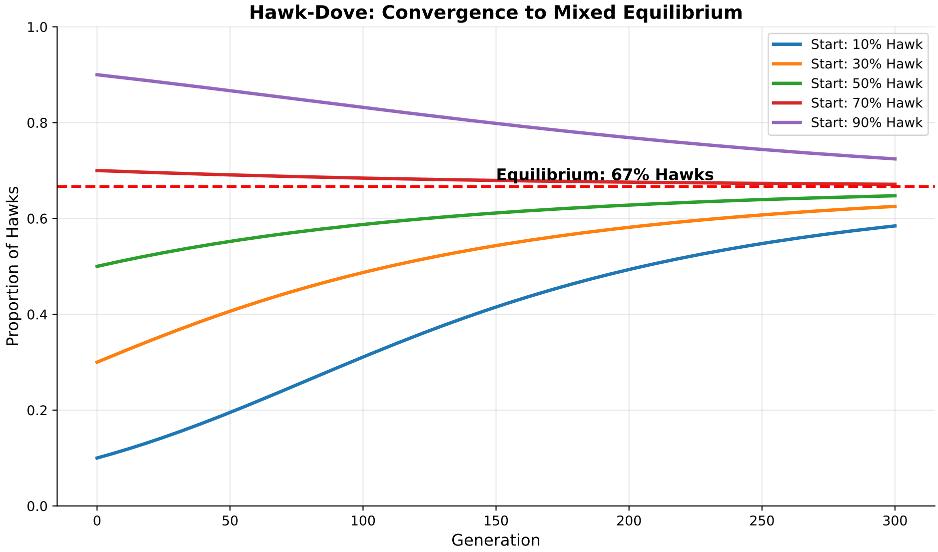
Tags

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.

	Dove	*Hawk*
Dove	$0, 0$	$4, 1$
Hawk	$1, 4$	$2, 2$


Hawk-Dove: Convergence to Mixed Equilibrium
Start: 10% Hawk
Start: 30% Hawk
Start: 50% Hawk
Start: 70% Hawk
Start: 90% Hawk
Equilibrium: 67% Hawks
Proportion of Hawks
Generation
0.0
0.2
0.4
0.6
0.8
1.0
0
50
100
150
200
250
300

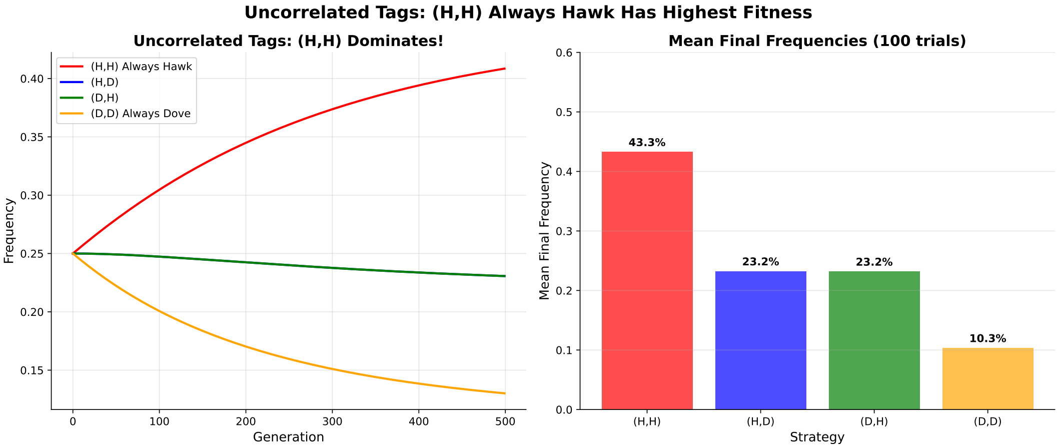
Problem: When 67% of the population is playing Hawk and 33% is playing Dove, there is a stable equilibrium, but this is inefficient, since sometimes Hawk players will play against other Hawk players and the payout will be 0.

Maynard-Smith noted that in such games it might be of interest to the players to find something outside of the game to use as a method for breaking symmetry.

Maynard-Smith noted that in such games it might be of interest to the players to find something outside of the game to use as a method for breaking symmetry.

- $\langle H, H \rangle$: Play Hawk if playing against the same tag, else play Hawk
- $\langle H, D \rangle$: Play Hawk if playing against the same tag, else play Dove
- $\langle D, H \rangle$: Play Dove if playing against the same tag, else play Hawk
- $\langle D, D \rangle$: Play Dove if playing against the same tag, else play Dove

Uncorrelated Tags



Correlation

If the players could use some feature observable to both players to which they can correlate their strategy, evolution might select for strategies which use this cue.

Individuals can evolve role-based strategies of the form

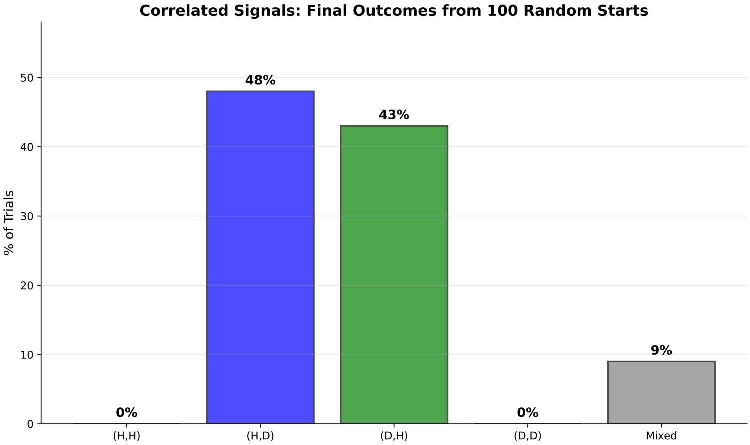
$$\langle \text{do this if owner, do that if intruder} \rangle$$

Correlation

Nature sends a signal to individuals, the signals are (anti)correlated, and individuals have strategies that are conditional on the signal. In the population states All $\langle H, D \rangle$ or in All $\langle D, H \rangle$ we have realizations of a special case of what Aumann (1974) calls a **correlated equilibrium**.... For situations where the role is unclear—an ambiguous signal or no signal at all—we should expect a polymorphism of Hawks and Doves. For situations with a clear signal, we should expect a correlated equilibrium.

Brian Skyrms and Kevin Zollman (2010). *Evolutionary Considerations in the Framing of Social Norms*. Politics, Philosophy, and Economics 9(3), pp. 265-273.

Correlated Tags



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).

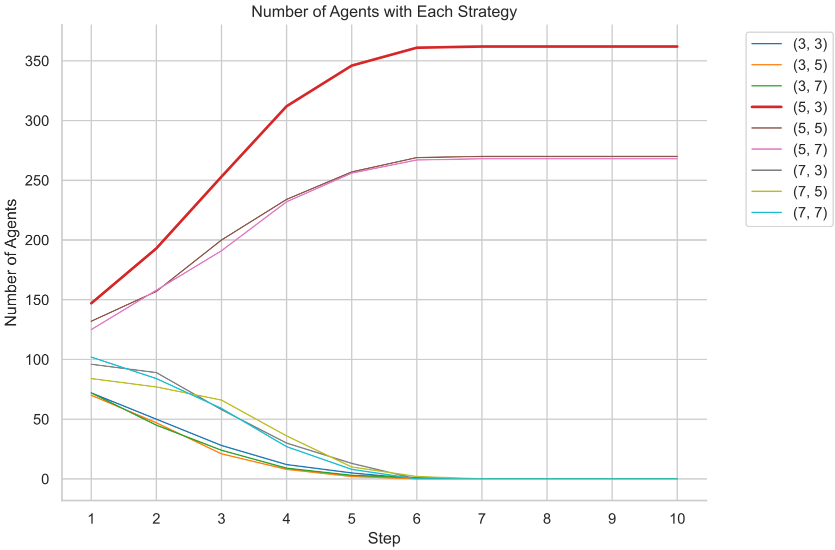
Complimentary Coordination

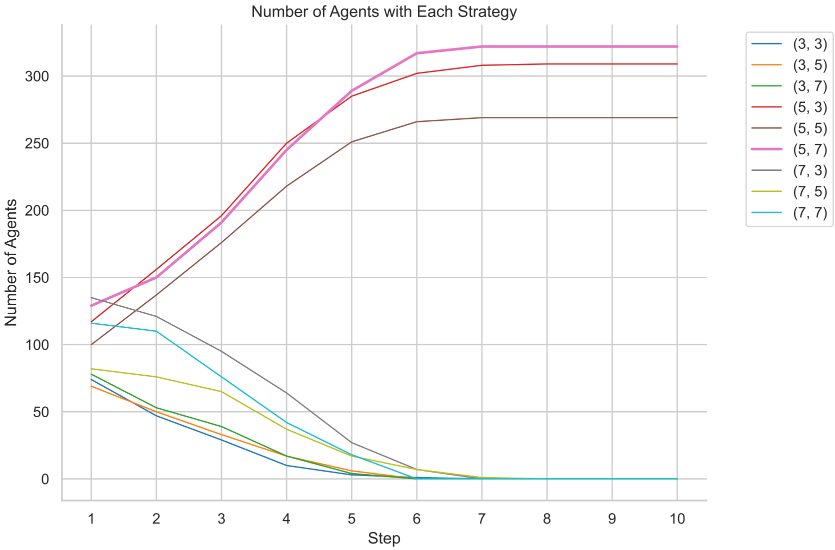
	A	B
A	$0, 0$	α, β
B	β, α	$0, 0$

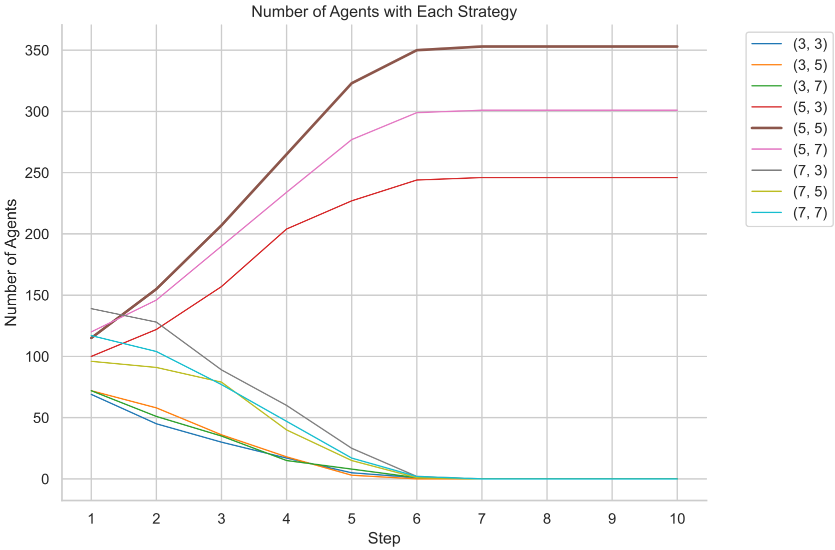
In this scenario, two actors must take complementary strategies, A and B, to succeed. This might represent division of labor, where A involves one set of jobs and B a complementary set. A population with two groups, say men and women, might evolve to solve this problem when one group always plays A (engages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

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


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
0
50
100
150
200
250
300
Step
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
350
300
250
200
150
100
50
0
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.

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.

Minority Disadvantage

Justin Bruner (2019). *Minority (dis)advantage in population games*. Synthese, 196(1), pp. 413-427.

If agents are tasked to just interact with out-group members (i.e., Blues never interact with fellow Blues), then inefficient arrangements can be completely avoided. In this case, play evolves to either the equal split, in which both those from the Blue and Green group demand five, or an asymmetric split, where Blues (Greens) always demand the high amount of six and Greens (Blues) acquiesce and demand four.

Consider the case in which both Blue and Green populations consist of a mixture of those who demand four and demand six.

Consider the case in which both Blue and Green populations consist of a mixture of those who demand four and demand six. If 60% of Blues demand the high amount of six while only 30% of Greens demand six, the vast majority of Blues who demand the high amount will interact with a Green agent demanding four.

Consider the case in which both Blue and Green populations consist of a mixture of those who demand four and demand six. If 60% of Blues demand the high amount of six while only 30% of Greens demand six, the vast majority of Blues who demand the high amount will interact with a Green agent demanding four. Thus the payoff associated with being a Blue who demands six will be particularly high. Likewise, Greens who demand six will interact for the most part with agents who similarly demand six, resulting in a rather paltry payoff.

Consider the case in which both Blue and Green populations consist of a mixture of those who demand four and demand six. If 60% of Blues demand the high amount of six while only 30% of Greens demand six, the vast majority of Blues who demand the high amount will interact with a Green agent demanding four. Thus the payoff associated with being a Blue who demands six will be particularly high. Likewise, Greens who demand six will interact for the most part with agents who similarly demand six, resulting in a rather paltry payoff.

This gives rise to a positive feedback loop, whereby the proportion of those who demand six in one group rises, in turn increasing the average payoff these agents receive, which leads to them further proliferating.

Things are more complicated, however, if we allow for interactions between both in-group and out-group members. In this case we assume individuals can accurately determine the group membership of their counterpart (i.e., whether they are Green or Blue) and group membership is fixed (that is, one cannot switch from Green to Blue).

Things are more complicated, however, if we allow for interactions between both in-group and out-group members. In this case we assume individuals can accurately determine the group membership of their counterpart (i.e., whether they are Green or Blue) and group membership is fixed (that is, one cannot switch from Green to Blue). Additionally, individuals can condition their behavior on the group membership of their strategic partner. In other words, agents can select to offer an equal split to Blue agents, and demand the bulk of the resource when interacting with Green agents.

In what follows, we consider the situation in which Green and Blue groups are not of the same size. We find that this has rather dramatic effects that can, in many circumstances, result in the minority systematically demanding the low amount when interacting with members of the majority.

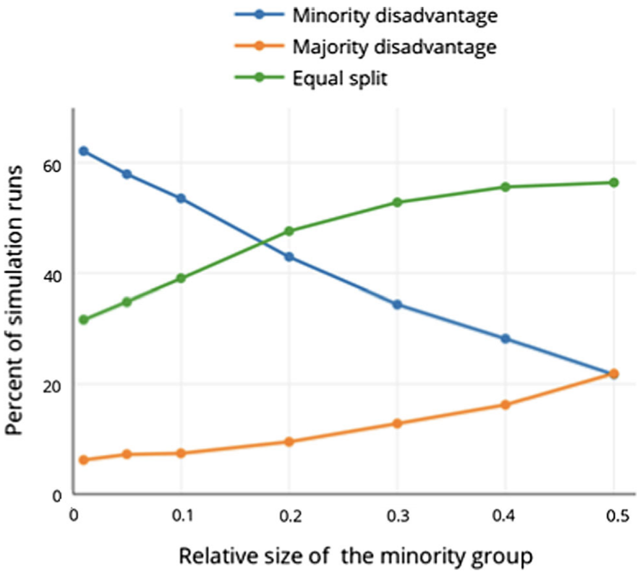


Fig. 2 Simulation results of the modified two-population replicator dynamic

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*
	Low	3, 3	3, 5	3, 7
Player 1	*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.

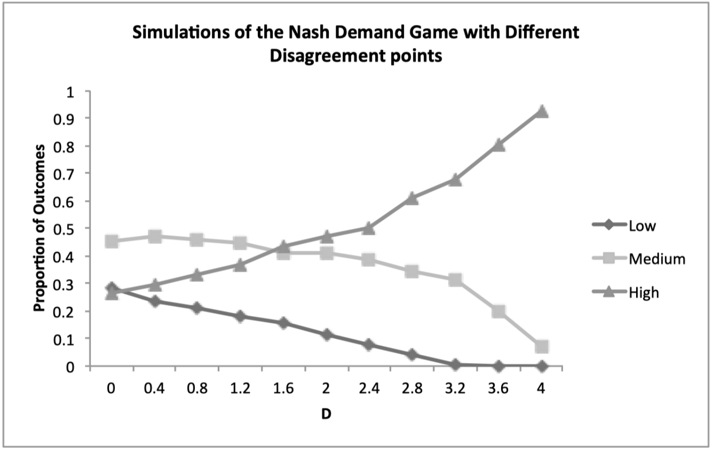


Figure 5: Simulation results for the game presented in figure 4. The y-axis shows the proportion of simulations that result in senior academics demanding Low, Medium, and High as D increases (x-axis) and $d = 0$.

An important take-away, emphasized by O'Connor (2019), is that inequity emerges robustly across a wide range of models under very minimal preconditions.

An important take-away, emphasized by O'Connor (2019), is that inequity emerges robustly across a wide range of models under very minimal preconditions. These preconditions are that 1) individuals recognize social categories, 2) they condition their behavior on social categories, and 3) they learn to take actions that benefit themselves.

An important take-away, emphasized by O'Connor (2019), is that inequity emerges robustly across a wide range of models under very minimal preconditions. These preconditions are that 1) individuals recognize social categories, 2) they condition their behavior on social categories, and 3) they learn to take actions that benefit themselves.

Since all these features are likely to be present in most social groups, we should expect that underlying social dynamics will tend to persistently push towards inequity. Thus, attempts to eradicate inequity are unlikely to be permanently successful. We should thus adopt a model where unfairness is something to be continually watching for, and continually combating, rather than something that will someday be "fixed". Inequity is a hydra whose heads grow back.