The Origins of Unfairness: Social Categories and Cultural Evolution, Cailin O'Connor. Oxford University Press, 2019, 256 pages.
doi:10.1017/S026626712000005X

In *The Origins of Unfairness: Social Categories and Cultural Evolution*, Cailin O'Connor makes a number of excellent contributions to our understanding of social norms, discrimination and inequity. O'Connor blends formal methods from game theory with philosophical discussion and socio-cultural commentary. This combination and the book's accessible style mean it will be of interest to scholars from many disciplines. The methods of evolutionary game theory are used to illuminate social behaviours and attitudes that underpin, reinforce and produce inequality. O'Connor uses an array of models to represent coordination, division of labour and distribution of resources. The central aim is to understand the role that social categories such as race and gender play in these interactions and to understand how significant differences in outcomes between those categories emerge.

Many contributions will be of broad interest, such as a possible explanation for the emergence of minority disadvantages: that these can occur due to subtle differences in how cultural learning operates between groups. Others include defining and measuring conventionality (85–89), fairness (115), robustness (75–79) and power (117–118). O'Connor also uses network modelling to examine homophily (Chapter 7), discusses the evolution of coordination within households (Chapter 8) and suggests methods for reform (Chapter 9). While all of these chapters are valuable additions to the field and warrant discussion, there is too much to cover in a single review. Here we will focus on two key aspects of O'Connor's book: the emergence of types in division-of-labour situations (namely, gender roles) and the significance of the cultural Red King effect, where larger groups gain advantages by being slower to adapt.

1. Evolution of gender: two types or more?

The emergence of types is a key part of O'Connor's explanation for the evolution of inequity. Say a population is trying to solve a complementary coordination game (CCG) (Table 1). In CCGs, the efficient Nash equilibria are pairs of strategies where players are taking complementary, rather than identical, actions. These games represent division of labour problems where players are best served by performing different actions (35). One way to effectively divide labour is to use type-conditioning: change your strategy depending on the other player's 'type'.

Throughout the book O'Connor uses the replicator dynamics (Taylor and Jonker 1978) to model cultural evolution via imitation: more successful strategies spread by imitation at a rate proportional to their success.[1] Members of a group (e.g. women)

[1]Specifically, O'Connor also makes use of the discrete-time replicator dynamics, which we will also use in our analysis below. See Weibull (1995) for details. The form of the dynamics is represented as follows:

new frequency of type i = previous frequency * (i's expected payoff/average payoff)

Table 1. A complementary coordination game ($x > 0$)

	A	B
A	0, 0	1, x
B	x, 1	0, 0

interact both with members of their own group and other groups, but imitate only their own group members. These models show that complementary type-conditioning in CCGs readily emerges: in inter-group interactions, members of one group learn strategy A and members of the other group learn strategy B. This produces a more efficient outcome than populations without types, but also results in inequality since one of the two types ends up in the favoured role (71).

O'Connor argues 'not just that gender facilitates division of labor, but that gender itself *exists* in order to divide labor' (85). This allows her to explain the near-universality of gender inequality among human societies without appealing to innate differences. The explanation, instead, is that division of labour by gender is a stable solution to a ubiquitous type of game (96–97). O'Connor makes the two-part claim that gender types can be utilized to solve coordination games and that gender 'arises for the very purpose of doing so' (97). We will consider this second claim in detail.

Note that ability to distinguish types is partly built into O'Connor's models. Individuals need to be able to distinguish types if they are to imitate based on type. Thus, conditional behaviour emerges, but not the *distinction* between types. O'Connor acknowledges that these models have not provided a 'full account' of the endogenous emergence of types (67). She goes on to provide a sketch of a model that may partially address the issue: 'Consider the following model. There is a homogenous group playing a complementary coordination game. Cultural evolution ... will carry the population to a state where now there are types, and everyone uses these to coordinate. In other words, from a completely homogenous group we can see the endogenous emergence of social categories' (97).

O'Connor provides two explanations of how gender could emerge in this fashion. First, she says that gender involves 'the best possible conditions for the use of types to solve complementary coordination problems' (98). These conditions include (a) two types in equal proportions in the population, which maximizes the chances that types will randomly interact with one another, and (b) frequent grouping of population members into partnerships with one member of each type, which leads to (nearly) perfect coordination (98). The second explanation is that sex, with which gender is correlated, is already salient for reproductive purposes (98). While we agree that salience of sex may be a key reason for the emergence of gender in solving coordination problems, we have two points of contention.

Regarding the first explanation, O'Connor has provided a reason why the types should be equal in proportion. But why *two* types? The biological realm does not limit itself to two sexes and shows a surprising variety of sexual behaviours (Roughgarden 2009). Some fungi species appear to have thousands of sexes (Kothe 1996). These considerations raise a question that O'Connor does not address: why

https://doi.org/10.1017/S026626712000005X Published online by Cambridge University Press

expect individuals to fall into *two* types rather than three or more? Moreover, some might argue that the salience of sex for purposes of reproduction and condition (b) above are not independent, because people often group into households with one member of each type for reproductive reasons. We don't think this is plausible: many sexually reproducing species have no household-type social structures at no cost to reproductive capabilities (for primate examples, see Hrdy 2009).

Regarding the second explanation, if gender emerged to solve coordination problems because of the salience of sex, then it is not the case that gender emerged endogenously from a homogenous population. Instead, the pre-existence of types (sex types) provides the basis for the emergence of other types (gender types) for use in coordination. If O'Connor's claim about the emergence of gender hinges on this second reason – the salience of sex – then she has not substantiated her claim that gender emerged endogenously to solve coordination problems.

Here, we develop a model where there are no salient or pre-established types and there are many ways to divide the population (by more than two types). Suppose individuals are in an infinite, randomly mixing population, playing the game in Table 1 (let $x = 3$), and that 50% of its members are women and 50% are men. However, individuals initially cannot distinguish these types (either for imitation or behaviour). The replicator dynamics will lead the population to an equilibrium of 1/4 playing A, 3/4 playing B. Now, suppose a new strategy (G1) is introduced that distinguishes gender-roles: Play A if you are a man matched with a woman, play B if you are woman matched with a man, and play A otherwise. This strategy will quickly proliferate – it outperforms either other strategy. If a variant, G2, is introduced that plays B when matched with a same-sex partner, these two type-conditional strategies will converge on an equilibrium where 1/4 uses G1 and 3/4 uses G2. Note that the dynamics does not presuppose types here as everyone imitates everyone else; a man's strategy includes how to act as a woman (and vice versa), even though they never employ that part of their strategy. Randomizing initial populations has no effect: all populations invariably converge to the equilibrium mix of G1 and G2. This result mirrors O'Connor's argument in Chapter 4.

However, complications arise when we expand the model to include the possibility of more types. Suppose that, in addition to gender, there is another, uncorrelated trait, say eye-colour, where 1/3 are Red, 1/3 are Green and 1/3 are Blue. We can introduce another discriminating strategy, E1, that distinguishes eye-colour rather than gender (Table 2).

Each individual is given an eye-colour that does not change, but their strategy provides a plan for all combinations: the strategy in Table 2 is not making simple, two-category same/different determinations, but rather distinguishes each of the possible six pairings of different eye-coloured individuals. As such, there are many variants that could be included; for simplicity, we will introduce only E1 and a variant E2 that plays B whenever paired with an individual of the same eye-colour. If we initialize a population with non-discriminatory types that only play A or B, and then introduce E1 and E2 into the population, the discriminatory types quickly take over the population and we reach an equilibrium of 1/4 E1 and 3/4 E2.

If we include both the types that distinguish genders (G1 and G2) and the types that distinguish eye-colours (E1 and E2), with random initial populations in 1000

https://doi.org/10.1017/S026626712000005X Published online by Cambridge University Press

Table 2. Example eye-colour discriminator strategy (E1)

Eye Colour / Opponent Eye Colour	Red	Green	Blue
Red	Play A	Play B	Play B
Green	Play A	Play A	Play B
Blue	Play A	Play A	Play A

Table 3. A game including multiple kinds of conditional strategies. G1 and G2 distinguish two genders. E1 and E2 distinguish three eye colours. H distinguishes based on a continuous height trait. Numbers are rounded to the nearest hundredth

	A	B	G1	G2	E1	E2	H
A	0	1	0.25	0.75	0.33	0.67	0.5
B	3	0	2.25	0.75	2	1	1.5
G1	0.75	0.75	1	1.5	0.75	0.75	0.75
G2	2.25	0.25	2.5	1	1.58	0.92	1.25
E1	1	0.67	0.92	0.75	1.33	1.67	0.83
E2	2	0.33	1.58	0.75	2.33	1.33	1.17
H	1.5	0.5	1.25	0.75	1.17	0.83	2

simulations, 38.9% resulted in populations that distinguished gender and 61.1% in populations that distinguished eye-colours. Distinguishing three types, rather than two, allows for effective coordination within a larger proportion of the population (2/3 compared to 1/2). All else equal, three-type distinctions will evolve more readily than two-type distinctions.

Going further, suppose all individuals have a continuous trait, height, and no two individuals have the same height. There are now very effective strategies available (see e.g. Skyrms 1996 on correlated strategies): for example, a strategy that adopts A when they are the taller individual, and B when they are the shorter individual (strategy H). Introducing H into the population, along with G1, G2, E1 and E2, further reduces gender distinctions. Of 1000 random initial populations, 29.6% of populations evolved gender distinctions, 45% evolved eye-colour distinctions, and 25.3% evolved height distinctions. The H strategy is maximally efficient: as O'Connor notes, using 'tags where the tags do not form categories or types, but instead have gradient values … can allow for perfect coordination in every interaction' (52). Table 3 shows the game ($x = 3$).

The two-type distinction is stable and sometimes does emerge despite the advantages of three-type or continuous distinctions. When the two-type distinction does evolve it is because the evolutionary process randomly started closer to that equilibrium, or it was introduced before the other strategies. In either case, two-type distinctions result when they are more salient to the evolutionary process at the outset, not because they are more effective or efficient.

https://doi.org/10.1017/S026626712000005X Published online by Cambridge University Press

Despite O'Connor's arguments to the contrary, we have shown that the emergence of two types is less efficient and less likely (in a neutral setting) than the emergence of three or more types. There are better ways to solve coordination games, but we happened to start with preconfigured, suboptimal categories. This casts doubt on O'Connor's claim that gender arose for 'the very purpose' of solving these games, and we are left only with the salience of sex as the explanation. However, our results also suggest these cultural categories are conventional in nature, which aligns with many of O'Connor's claims about the social significance of conventions.

Throughout her book, O'Connor sometimes claims to be giving how-possibly explanations, sometimes how-plausibly explanations (O'Connor uses 'how-potentially'), and other times how-minimally explanations, where a model illustrates the minimal conditions needed to produce the phenomena of interest (8–9). Our analysis shows that O'Connor *has* given a how-possibly story about the emergence of conditioning on two types. Whether O'Connor has given a how-plausibly explanation depends on whether the conditions needed to result in these two-type distinctions are plausible. However, O'Connor has not provided a how-minimally story of the emergence of types as she has not included the possibility of three or more types; the salience of the types seems crucially important to which strategies evolve.

2. Significance of the cultural Red King

Another important aspect of O'Connor's book is the use of the cultural Red King effect, which she proposes as one explanation for minority groups being disadvantaged in bargaining contexts (Chapter 6). The Red King effect occurs when groups that evolve slowly reap a benefit. This is the inverse of the better-known Red Queen effect, in which it is advantageous to evolve quickly.

Here is how the Red King effect works. Suppose individuals of different groups are matched to play a bargaining game where players can make higher or lower demands on some shared resource. Higher demands risk conflicting with other higher demands and can yield nothing. Lower demands are less risky but less rewarding. If there are two groups and one is slower to learn, the slower group will maintain initial variation in their behaviour longer. This typically leads the faster population to learn the less risky low-demand behaviour. The slower population then reacts by converging on the advantageous high-demand behaviour when interacting with the low-demanding group. In O'Connor's models, the size of a group affects how quickly its members learn. Minority groups learn inter-group behaviour more quickly than majorities because they are interacting more frequently with the majority group than vice versa. This means, in these cases, the minorities end up disadvantaged simply because they are the minority.

O'Connor takes these Red King cases to be significant for two reasons. First, she argues that this shows inequity can emerge with very 'minimal' conditions (161). Second, she believes this provides a plausible explanation for many real cases of inequity (135).

The cultural Red King effect is a fascinating *how-possible* explanation; however, we believe O'Connor has oversold its significance as a how-minimally and

https://doi.org/10.1017/S026626712000005X Published online by Cambridge University Press

how-plausibly explanation. Regarding the former, there is reason to think these explanations are not as minimal as they seem. O'Connor admits that whether the Red King effect occurs depends not only on differences in group-size, but also on the specific payoffs of the game being played and on the specific learning dynamics used (143; see also O'Connor 2017). In some bargaining games, with slightly different payoffs we see systematic minority advantages rather than disadvantages (Zucker *et al.* 2019). Therefore, cultural dynamics between groups of different sizes does not alone produce inequity; rather, it is certain dynamics in certain games with certain payoffs that does so. This casts doubt on O'Connor's model as a how-minimally explanation of minority disadvantage.

Second, for the Red King effect to explain real-world cases of inequity, O'Connor must show that the specific games and payoffs are plausibly similar to actual populations' interactions. On this point, more detail is needed to assess whether this is a viable how-plausibly explanation. Moreover, many instances of inequity occur where the majority group is oppressed by a powerful minority, e.g. Apartheid, along with countless cases of economic stratification. These cases are much more naturally explained by power or historical precedent than by differences in learning dynamics. O'Connor does consider these other aspects as well (power in Chapter 5 and historical precedent implicitly by examining initial conditions of models). Incorporating such differences into the models is straightforward and has an obvious effect: the more powerful or initially advantaged groups tend to end up advantaged, usually regardless of any differences in size of populations or learning speed. We think power and precedent seem the more plausible explanations in most real cases of inequality, which undermines the purported significance of the Red King effect.

Aja Watkins
Boston University, 745 Commonwealth Avenue,
Boston, MA 0221, USA
Email: ajawatki@bu.edu

Rory Smead
Northeastern University, 360 Huntington Ave,
Boston, MA 02115, USA
Email: r.smead@northeastern.edu.

Acknowledgements. We would like to thank Cailin O'Connor for several stimulating discussions prior to writing this review. We would also like to thank Patrick Forber for some helpful comments and suggestions on an earlier draft.

References

Hrdy S.B. 2009. *Mothers and Others: The Evolutionary Origins of Mutual Understanding*. Cambridge, MA: Harvard University Press.

https://doi.org/10.1017/S026626712000005X Published online by Cambridge University Press

Kothe E. 1996. Tetrapolar fungal mating types: sexes by the thousands. *FEMS Microbiology Reviews* **18**, 65–87.

O'Connor C. 2017. The cultural red king effect. *Journal of Mathematical Sociology* **41**, 1–23.

Roughgarden J. 2009. *Evolution's Rainbow: Diversity, Gender, and Sexuality in Nature and People*. Berkeley, CA: University of California Press.

Skyrms B. 1996. *Evolution of the Social Contract*. Cambridge: Cambridge University Press.

Taylor P.D. and L.B. Jonker 1978. Evolutionarily stable strategies and game dynamics. *Mathematical Biosciences* **40**, 145–156.

Weibull J.W. 1995. *Evolutionary Game Theory*. Cambridge, MA: MIT Press.

Zucker J., D. Rassaby, A. Watkins and R. Smead 2019. Bargaining and intersectional disadvantage: Reply to O'Connor, Bright, and Bruner. *Social Epistemology Review and Reply Collective* **8**(7), 1–8.

Aja Watkins is a PhD student in Philosophy at Boston University. She works in the philosophy of biology and philosophy of science. Her current work focuses on developmental biology and historical sciences such as palaeontology.

Rory Smead is an Associate Professor of Philosophy and the Rossetti Professor for the Humanities at Northeastern University. He works primarily in philosophy of biology and philosophy of social science, where he uses game theory to understand social evolution. His current work is on spite and related harmful social behaviour. URL: https://cssh.northeastern.edu/people/faculty/rory-smead/

https://doi.org/10.1017/S026626712000005X Published online by Cambridge University Press