A Report on "Social Goods Dilemmas in Heterogeneous Societies" by McAvoy et al. (2020)

Reviewer 2

February 11, 2026

v2

isitcredible.com

Disclaimer

This report was generated by large language models, overseen by a human editor. It represents the honest opinion of The Catalogue of Errors Ltd, but its accuracy should be verified by a qualified expert. Comments can be made here. Any errors in the report will be corrected in future revisions.

> I am wiser than this person; for it is likely that neither of us knows anything fine and good, but he thinks he knows something when he does not know it, whereas I, just as I do not know, do not think I know, either. I seem, then, to be wiser than him in this small way, at least: that what I do not know, I do not think I know, either.
>
> Plato, *The Apology of Socrates*, 21d

> To err is human. All human knowledge is fallible and therefore uncertain. It follows that we must distinguish sharply between truth and certainty. That to err is human means not only that we must constantly struggle against error, but also that, even when we have taken the greatest care, we cannot be completely certain that we have not made a mistake.
>
> Karl Popper, 'Knowledge and the Shaping of Reality'

Overview

Citation: McAvoy, A., Allen, B., and Nowak, M. A. (2020). Social Goods Dilemmas in Heterogeneous Societies. *Nature Human Behaviour*. Vol. 4, No. 8, pp. 819–831.

URL: https://doi.org/10.1038/s41562-020-0881-2

Abstract Summary: This study explores general evolutionary dynamics for prosocial behaviors in social goods dilemmas on arbitrary spatial structures and with different types of social goods, finding that heterogeneous networks can promote prosociality but also generate large inequality.

Key Methodology: Theoretical evolutionary dynamics modeling using fixation probability under weak selection on graphs (social networks) for three types of social goods (pp, ff, pf) and various update rules (PC, DB, IM).

Research Question: How do arbitrary spatial structures and different types of social goods affect the general evolutionary dynamics of prosocial behaviors in social goods dilemmas?

Editor's Note

Version 2 of this report has been written by an improved model of Reviewer 2.

Summary

Is It Credible?

This study by McAvoy et al. presents a mathematically rigorous examination of how the nature of social goods influences their evolution in structured populations. The authors challenge the monolithic treatment of cooperation in evolutionary game theory by introducing a taxonomy of social goods based on whether benefits and costs are fixed or proportional to the number of recipients. Their central claim is that heterogeneous population structures—such as "rich-club" networks where a few individuals possess many connections—can strongly promote the evolution of prosocial behaviors, particularly for goods with fixed costs and fixed benefits ("ff goods"). Most strikingly, they argue that this structural facilitation can lead to "harmful prosociality." This phenomenon occurs not only when selection favors producers despite total costs exceeding total benefits ($b < c$), but also when the good is wealth-producing overall ($b > c$) yet still leaves the poorest individuals worse off than if no one produced the good at all (p. 820).

The theoretical framework appears robust within the specific constraints of evolutionary graph theory. The derivation of the general condition for selection (Equation 1) and the subsequent analysis of specific update rules (pairwise comparison, death-birth, and imitation) provide a solid mathematical foundation for the authors' conclusions (p. 821). The distinction between rival and non-rival goods is economically grounded and adds necessary nuance to biological models. The finding that highly connected individuals can accumulate wealth at the expense of peripheral individuals is a logical consequence of the network topology and the payoff structures defined. For instance, in a star graph with fixed-cost goods, the hub can evolve to produce because it pays a single fixed cost to benefit many neighbors, while simultaneously reaping the benefits of any peripheral producers. This creates a scenario

where the hub accumulates massive wealth while peripheral individuals may suffer negative payoffs, effectively being exploited by the structural arrangement (p. 823).

However, the credibility of the article's broader sociological implications is constrained by its reliance on static population structures. The model assumes that the physical interaction network is fixed; individuals cannot sever ties with exploitative hubs or reorganize to escape "harmful prosociality." While the authors acknowledge this limitation, noting that "changes to the population size and/or structure... could lead to additional interesting behaviour" (p. 827), this assumption is critical. In many real-world social systems, the extreme inequality described—where peripheral nodes receive negative payoffs—would likely drive individuals to exit the network or alter their connections. Consequently, the results are most applicable to rigid institutional hierarchies or biological systems where interaction partners are determined by physical proximity or unchangeable kinship ties. It is worth noting, however, that the authors do address a form of dynamic interaction through their analysis of reciprocity (Tit-for-Tat), where the "effective" interaction graph changes based on behavior even if the physical graph remains static (pp. 823–824).

Furthermore, the mechanism driving these dynamics relies on simplified forms of social learning and replacement. The update rules assume individuals copy successful neighbors or are replaced by their offspring. While the authors argue these rules capture "qualitative features of behaviour imitation" (p. 821), they do not account for strategic reasoning or aspiration levels. In the "harmful prosociality" scenarios, peripheral individuals are not necessarily imitating a wealthy hub to their own detriment; rather, under rules like Death-Birth, they may simply be replaced by the offspring of the successful hub, or they may be "stuck" in a producer state because switching to non-production would yield an even worse outcome given the neighbors' strategies. Finally, while the article proposes institutional interventions like taxes to mitigate inequality (p. 823), these are presented as exogenous fixes. The study successfully identifies a "tyranny of structurelessness" (p. 827) where informal

ties lead to inequality, but the emergence of the redistributive solutions themselves lies outside the model's core evolutionary logic.

The Bottom Line

McAvoy et al. provide a credible and mathematically sound demonstration that the type of social good and the structure of the population fundamentally alter the conditions for cooperation. Their finding that heterogeneous networks can drive the evolution of "harmful prosociality"—where cooperation creates wealth for a few while harming the many—is a significant theoretical contribution. However, these results rely on static network assumptions that may not hold in fluid social groups where exploited individuals can disengage, limiting the direct applicability of the findings to rigid social or biological hierarchies.

Potential Issues

Static network assumption limits the model's applicability to most social systems: The article's analytical framework largely rests on the assumption of a fixed, static population structure. This choice is a significant simplification, as many real-world social networks co-evolve with the behaviors of the individuals within them. For example, individuals might sever ties with non-producers or form new ties with producers, fundamentally altering the network topology. The article's conclusions are therefore most applicable to systems where interaction structures are rigid and not subject to strategic rewiring. The authors acknowledge this as a "main limitation" in their discussion, identifying dynamic population structures as a topic for future research (p. 827). However, the article does explore one form of dynamic interaction in its analysis of reciprocity, where tit-for-tat strategies alter the *effective* interaction network based on past behavior, even while the underlying physical graph remains static (pp. 823–824).

Behavioral models are based on idealized assumptions of social learning: The evolutionary dynamics explored in the article are driven by update rules that, while standard in the literature, rely on simplified models of social learning. These rules generally assume that individuals have access to payoff information from their neighbors and imitate others based on simple comparison rules, without accounting for more complex cognitive assessments or imperfect information gathering. For instance, the pairwise-comparison (PC) rule involves comparing one's own payoff to that of a single, randomly chosen neighbor (p. 821). The authors recognize that these rules are "highly idealized" but argue they "capture important qualitative features of behaviour imitation" (p. 821). It is important to note that these models are not entirely deterministic; the PC rule, for example, explicitly incorporates noise by making the imitation process probabilistic, a feature controlled by the selection intensity parameter (pp. 821, 827). The idealization thus lies not in an absence of

stochasticity, but in the simplicity of the learning heuristics themselves.

The proposed solution for mitigating inequality is not an emergent feature of the model: The article discusses how a "tax" or public pool system could ameliorate the harmful effects of prosociality, particularly for `ff` goods (p. 823; Extended Data Fig. 9). While this is an interesting insight into potential institutional interventions, this mechanism is presented as an exogenous rule imposed on the system rather than an outcome that could emerge from the evolutionary dynamics of the model itself. The authors frame this as a call for the "design of mechanisms to redistribute wealth" (p. 827). The article does analyze the evolutionary stability of cooperation *within* this institutional framework (Extended Data Fig. 9), but it does not model the evolution or emergence of such redistributive institutions, meaning the proposed solution operates outside the core evolutionary framework of the study.

Future Research

Co-evolutionary dynamics: Future work should relax the assumption of static networks to allow the population structure to co-evolve with behavioral strategies. Research could model scenarios where individuals have the agency to sever ties with partners who provide negative payoffs or rewire connections to other producers. This would test whether the "rich-club" structures and the resulting inequality can persist when the "poor" have the option to exit, thereby determining the stability of harmful prosociality in fluid societies.

Endogenous institutional emergence: To address the issue of inequality without relying on exogenous policy interventions, researchers could model the evolution of redistributive mechanisms from within the population. By introducing a second layer of strategy where individuals can vote for or contribute to a "tax" or common pool resource that redistributes wealth, the model could determine if egalitarian institutions can emerge via natural selection to counteract the inequality generated by heterogeneous networks.

Aspiration-based learning: Future studies could replace the standard imitation update rules with aspiration-based learning models, where individuals change their strategy only if their payoff falls below a certain satisfaction threshold, regardless of their neighbors' success. This would determine if "harmful prosociality" is an artifact of imitating wealthy neighbors (or being replaced by them) or if it persists even when individuals act solely based on their own absolute well-being.

isitcredible.com