

SHORT REPORTS

Normative assembly rule reveals fairness in microbial communities

Teemu Kuosmanen^{1,2*}, Juhani Rantanen³, Dovydas Kičiatovas^{1,2}, Sanna Pausio³, Ville-Petri Friman⁴, Teppo Hiltunen³, Ville Mustonen^{1,2*}

1 Department of Computer Science, University of Helsinki, Helsinki, Finland, **2** Department of Organismal and Evolutionary Biology, University of Helsinki, Helsinki, Finland, **3** Department of Biology, University of Turku, Turku, Finland, **4** Department of Microbiology, University of Helsinki, Helsinki, Finland

* teemu.kuosmanen@helsinki.fi (TK); v.mustonen@helsinki.fi (VM)



OPEN ACCESS

Citation: Kuosmanen T, Rantanen J, Kičiatovas D, Pausio S, Friman V-P, Hiltunen T, et al. (2026) Normative assembly rule reveals fairness in microbial communities. PLoS Biol 24(6): e3003872. <https://doi.org/10.1371/journal.pbio.3003872>

Academic Editor: Britt Koskella, University of California Berkeley, UNITED STATES OF AMERICA

Received: January 28, 2026

Accepted: June 8, 2026

Published: June 24, 2026

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pbio.3003872>

Copyright: © 2026 Kuosmanen et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution,

Abstract

Understanding and predicting how communities assemble is a paramount challenge in ecology. Here we address these questions normatively by comparing the observed species abundance distribution to a game-theoretically fair distribution based on each species' Shapley value. By analyzing in total 56 distinct community outcomes, we assess how fairly biomass is distributed in microbial communities displaying both competitive and cooperative interactions in different growth conditions. We find examples of fair communities that closely follow their Shapley value across all environments as well as counterexamples where the true abundances deviate from the species' objective contribution to community biomass. Next, we develop a fair assembly rule based on the recursive definition of Shapley value and show that also unfair community compositions are consistent with the principles of fair assembly after the lower-level competitive outcomes are known. Our results give unique empirical insights into the distributive function of ecological dynamics and lay down the theoretical foundations of what might become a normative community assembly theory.

Introduction

Organisms interact with each other and their surrounding environment to grow, adapt, and assemble to biological populations, ecological communities, and economic coalitions which in turn self-organize to functioning ecosystems and technological societies with advanced economies [1]. Such complex systems display various emergent phenomena ranging from cyclicity and chaos [2,3] to robustness and resilience [4,5], and evolution and diversification [6,7]. By exploiting available free energy reservoirs, all living organisms produce surplus output that is distributed in the community context by some distinct mechanisms. In nature, the ecological game of growth is largely based on competitive success in utilizing limited shared resources whereas cooperative interactions among social groups, such as seen in human societies, might lead to a different payoff distribution.

and reproduction in any medium, provided the original author and source are credited.

Data availability statement: All data and analysis notebooks are available at: <https://osf.io/a7yzp/overview>.

Funding: This work was funded in part by the Research Council of Finland (Multidisciplinary Center of Excellence in Antimicrobial Resistance Research, grants 346126 and 364232 to TH; grants 346128 and 364234 to VM). <https://www.aka.fi/en/> The funders did not play any role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: AMP, ampicillin; CON, control condition; MOX, moxifloxacin; PPY, protease-peptone-yeast; STR, streptomycin.

By studying competition in its purest form, ecological and evolutionary theory have since their adoption broadly influenced economic thought by shedding light to properties of the ‘state of nature’. While the distribution of income, wealth, and consumption in human societies has been a prominent theme in economic literature [8,9], species abundance distributions [10,11] have similarly been some of the most intensively studied macroecological patterns ever since Preston’s observation that different physical, economic, and ecological ‘wealth distributions’ all seem to take similar form [12,13]. But whereas the distributive properties of political and economic systems have also given rise to normative theories addressing questions of distributive justice in terms of fair resource allocation, corresponding ecological theories have remained mainly descriptive, characterizing biodiversity with statistical scaling laws [14,15]. However, redistribution of resources has also been observed in non-human species, where uncertainty about kinship can reduce inequality among offspring, a finding consistent with Rawlsian notion of fairness [16].

The field of game theory can be roughly divided to non-cooperative game theory, which focuses on modeling individual strategic behavior in a competitive setting, and cooperative game theory, which in contrast focuses solely on the outputs of different coalitions of players and how the total output is distributed among them (either by bargaining or fair redistribution). While game theory has been extensively applied, and indeed also developed, within the broader fields of evolution and ecology, existing eco-evolutionary theory is predominantly making use of the concepts of non-cooperative game theory by attempting to model microscopic interactions and strategy dynamics either at the level of individuals, species, or genes as players [17]. However, measuring these underlying mechanistic interactions can be generally very difficult in case of complex biological systems even for simple minimal models, such as the generalized Lotka-Volterra model commonly employed for studying ecological communities [18,19]. Here, we demonstrate, perhaps counterintuitively, that concepts of cooperative game theory can be meaningfully applied also in an inherently non-cooperative biological setting, where competition and exploitation are the norm [20,21]. As we will show, this opens new, top-down methods to model and predict assembly of complex ecological communities.

We introduce the ecological *Game of Growth* where we first measure the biomass growth—the universal currency of living systems—in different species assemblages and then analyze the game’s distributional properties using the concept of Shapley value (see Fig 1). The Shapley value [22] is an established solution concept in cooperative game theory, which provides an objective way to distribute the total payoff between the players fairly in the sense that, among all possible payoff distributions, the Shapley distribution is the only one satisfying the mathematical properties of efficiency, symmetry, and linearity [22] as well as admitting a unique potential function [23]. The Shapley value has been extensively applied and extended in the economics game theory literature [24] as well as more recently in machine learning as a principled way to quantify the feature importances of different predictors [25]. In the context of biology, analysis inspired by the Shapley value has found applications, for example, in the fair proportion index used in constructing phylogenetic trees, to assess the

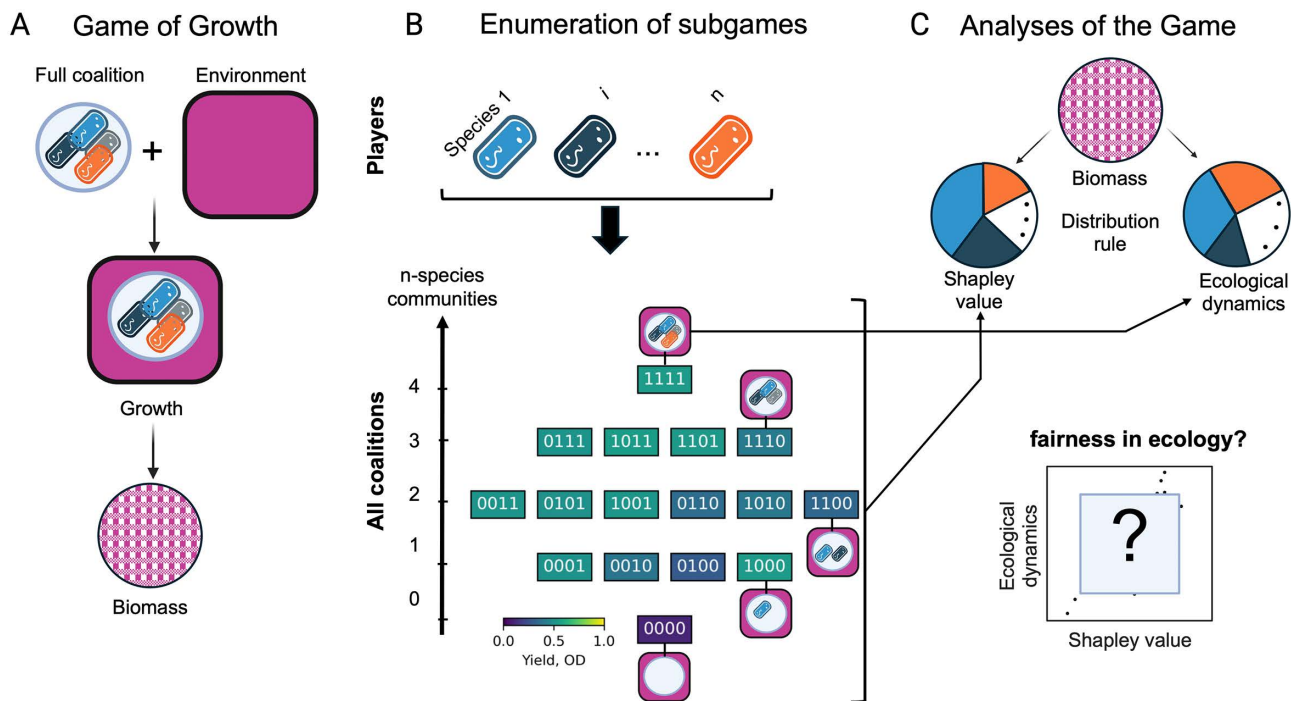


Fig 1. Experimental design. (A) The observed growth biomass (yield) depends on both the community members, their composition, and the abiotic environment, which together define an instance of *Game of Growth*. (B) Here, we first formed 14 distinct 4-species communities and then analyzed their growth in four different growth conditions, which leads to altogether 56 games. Each game is analyzed by systematically measuring the growth in all possible subcommunities, i.e., coalitions. For a 4-species communities this results in 16 subgames (binary values 1/0 indicate the presence or the absence of each member species) including the full community, i.e., the grand coalition, and the null condition, i.e., the empty coalition. These measurements allow us to quantify the objective contribution of each species by computing its Shapley value related to growth. (C) By determining the relative proportions of the species in the end of each game by genomic sequencing, we can compare the observed ecological distribution to the game-theoretically fair distribution based on the Shapley value and thus assess the fairness of ecological dynamics as a distribution rule across the communities and environments. Created in BioRender. Mustonen, V. (2026) <https://BioRender.com/55mv1na>.

<https://doi.org/10.1371/journal.pbio.3003872.g001>

importance of genes within a metabolic pathway, or to determine the individual contribution of each taxon within a multi-taxa setting to microbiome functional profiles [26–32].

In this paper, we investigate how biomass as a proxy for cumulative consumption of shared resources is distributed in microbial communities and further pose the question of how fair the resulting ecological outcome is from a normative point of view. By systematically measuring the growth of 14 distinct 4-species communities across all possible subcommunities, we can compute the Shapley value of each species and assess the fairness of the community assembly by comparing these to the actual end-population shares of the species. We also applied varying antibiotic stress to each community to see whether the observed species abundance distribution and fairness change in more challenging growth conditions, where cooperation can readily emerge via production of public goods, detoxification, or nutrient availability promotion, and thus potentially play a pivotal role in the community outcomes [33]. By using the established concept of Shapley value, we first empirically quantify the fairness of altogether 56 community outcomes and then develop a normative assembly rule that can also predict unfairly assembled communities if the realized relative abundances in lower-level subcommunities are known.

Results

The Shapley value of player i in a game with n players forming a grand coalition $N = \{1, \dots, n\}$ is given by [22]

$$\phi_i^N(v) = \sum_{S \subset N} \frac{(|S|-1)!(n-|S|)!}{n!} (v(S) - v(S \setminus \{i\})) \quad (1)$$

where v is a function that gives the output of every 2^n possible (sub)coalitions S that can be formed from the members of the grand coalition. In the context of ecology, we can interpret $v(S)$ as some community function of community S , which we assume can be directly measured. Here, we focus on the community yield, denoted by Y_S , which is the total biomass at the end of the experiment. Now, to compute the Shapley values, we need not make any assumptions about the ecological dynamics that give rise to the species abundances, but instead simply experimentally measure the observable community yield in all subcommunities (see Fig 1).

In practice, the Shapley value Eq. (1) can be computed more efficiently using the following recursive formula

$$\phi_i^S(v) = \frac{1}{|S|} \left(v(S) - v(S \setminus \{i\}) + \sum_{j \neq i} \phi_j^{S \setminus \{i\}}(v) \right) \quad (2)$$

with initial condition $\phi_i^{\emptyset}(v) = 0$ for empty coalitions (for a proof, see Ref. [34]). This more interpretable formula shows how the Shapley value $\phi_i^N(v)$ is built bottom-up from the subcoalition Shapley values based on the species' marginal value on the community function.

We formed 14 distinct communities by randomly selecting four species to each community from a pool of 16 microbial species that are known to be able to coexist as a community in laboratory microcosm experiments (see Materials and methods for more details). We then measured the growth curves and determined the yields from the combinatorially complete set of experiments in four different growth conditions, which allowed us to compute the Shapley values of the community members across the different communities and environments. These results are summarized in Fig 2.

First, we investigated the yield distribution within each community without focusing on the species identities: by ranking the Shapley values from largest to smallest independently for each community and environment, we gain a first look on how the fair yield shares based on the Shapley value (Fig 2A) compare to the actual yield distribution (Fig 2B). Note that by the property of efficiency, the Shapley value distributes the total yield fully among the community members, so we are comparing two different ways to distribute the same surplus output. By averaging these proportions across all communities (Fig 2C), we noticed that the most prevalent species in each community typically attains higher abundance compared to ranking based on Shapley value, the second-ranked species obtains roughly the same share under both distribution rules, and the third and fourth-ranked species are left with lower shares than what their Shapley values would warrant. Therefore, the observed yield shares are more concentrated to the most abundant species than the corresponding yield distribution obtained from the Shapley values which would distribute the biomass more equally. However, in the more challenging growth condition of moxifloxacin stress, we noticed that both the Shapley values and the observed yield shares become increasingly concentrated on a single species. In this case, two of the communities produce no observable biomass, and some species even have negative values, meaning that their presence lowered the community output.

We then turned to analyze the roles of the individual species by collecting their Shapley values from all the communities they were present. Note that by the property of linearity, averaging the Shapley values of any given coalition across different environments constitutes a Shapley value of the entire game across environments. However, in our experiment, some species can belong to multiple different coalitions, and averaging over them is not anymore formally a Shapley value. Nevertheless, by averaging over all the individual Shapley values per species, the obtained mean Shapley value

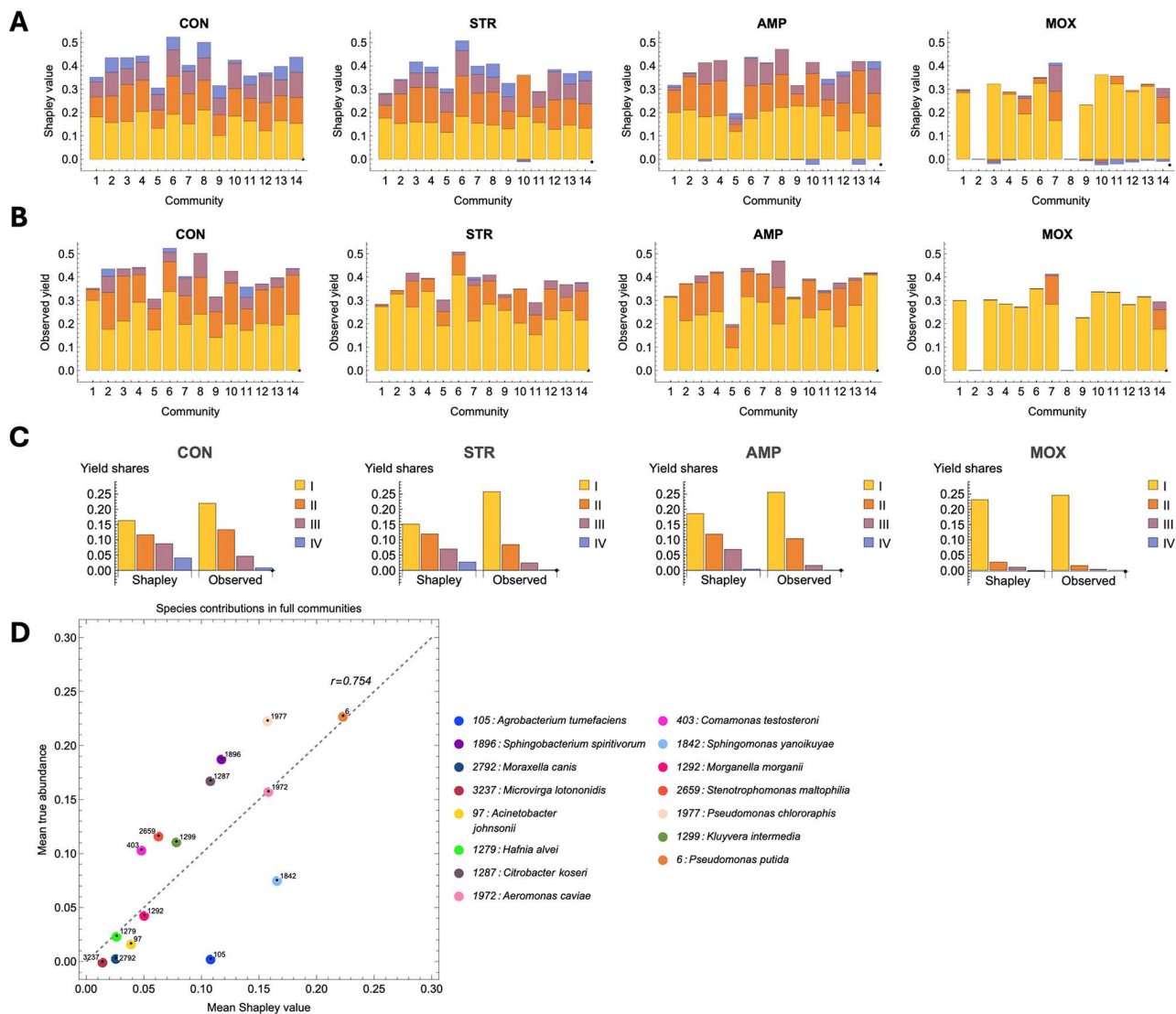


Fig 2. Shapley value analysis of growth across communities and environments. (A) The Shapley values of each species in different communities and environments. In each bar, the Shapley values have been independently ranked from largest (I, yellow) to smallest (IV, blue) and the colors refer to the ranking, not species identity. Columns refer to different growth conditions (CON, control condition; AMP, ampicillin; STR, streptomycin; MOX, moxifloxacin). (B) The corresponding true abundances of each species ranked independently. (C) The distribution of yield shares of the ranked four species averaged over all the communities. The observed ecological distribution gives a higher share to the most abundant species compared to their Shapley value. (D) The mean Shapley value and true abundance of each species plotted against each other as averaged over all the communities and environments in which the species were present. Species falling below the diagonal receive overall a lower share than what their objective contribution would warrant while the species above the diagonal gain unfairly high abundances. The data and code needed to create this Figure can be found in <https://osf.io/a7yzp/overview>.

<https://doi.org/10.1371/journal.pbio.3003872.g002>

gives a measure of the species overall deserved share of the games' output (performance). We show that the mean Shapley value of each species is correlated quite strongly with the true abundances ($r=0.754$), but note that some species receive on average less than what their objective contribution to the community output is, while others gain a larger share than what could be considered fair (see Figs 2D and S1 where the same is shown separately for each environment).

Next, we investigated how closely the observed frequency distribution follows the game-theoretically fair end-distribution φ , which can be obtained from the optical density data by normalizing with the sum of the Shapley values, which equals the yield of the grand coalition (see [Materials and methods](#) for more details). For example, in the case of only two species, the fair end-distribution is given by

$$\varphi_1^{\{1,2\}} = \frac{1}{2} \left(1 + \frac{Y_1 - Y_2}{Y_{12}} \right) \quad (3.1)$$

$$\varphi_2^{\{1,2\}} = \frac{1}{2} \left(1 + \frac{Y_2 - Y_1}{Y_{12}} \right) \quad (3.2)$$

where Y_1 and Y_2 are the monoculture yields of species 1 and 2, respectively, and Y_{12} is the biculture yield. This uniquely defines the fair distribution in case of two species for any ecological dynamics, assigning a higher fair proportion for the species with more efficient resource conversion efficiency while the biculture value scales the monoculture difference more strictly in case of antagonism (lower biculture yield) and more loosely in case of synergy (higher biculture yield) allowing for more equal proportions. The fair distribution therefore has the interesting feature that as the total output increases, the fair proportion of the less abundant species increases not only in absolute, but also in relative terms.

In reality, the true end-frequencies after the community growth can deviate from the fair distribution because, in a competitive setting, resource efficiency (monoculture yield) is not the only important factor: for example, shorter lag time, higher resource uptake rate, consumption of secondary metabolites, exploitation of other public goods without reciprocal contribution to their production (cheating), and direct negative interactions, such as bacterial warfare [20], can all significantly change the competitive dynamics such that the monoculture yield alone is not predictive enough for success in a community setting. However, species that reach high abundances mainly by exploiting others without contributing themselves cannot grow successfully in harsher environments without the help of the contributing species. Thus, by observing growth systematically across all possible coalitions, we can effectively dissect the causal impact and role of each species on the community output, which defines the fair proportions of the species. By sequencing the community after growth, we can estimate the true end-frequencies, which allows us to compare how ecological reality relates to what would be fair community outcomes.

To quantify the fairness of each community across different environments, we computed the cosine similarity of the fair distribution and the true frequency vector ([Fig 3A](#)). We found communities that systematically followed the fair distribution across all environments; communities that were fair in some conditions, but not others; and communities that deviated significantly from the fair distribution across all conditions. We then quantified both the actual and implied community diversity of the true and fair distributions, respectively, using the Gini–Simpson index, which can be interpreted as the probability of observing two different species when randomly sampling a pair of individuals from the population, i.e., the probability of interspecific encounter (note that the Gini–Simpson index is inversely related to the Gini index with higher values meaning higher diversity). First, we found that the implied community diversity of the fair distribution could be either high or low, with the latter being the default mode in the presence of moxifloxacin ([Fig 3D](#)). However, in other conditions, the fair distribution displayed non-trivial coexistence as showcased in [Fig 3C](#), which makes the observed fairness of these communities intriguing. We then compared the properties of the fair distribution and the true ecological distribution by contrasting their community diversity against each other ([Fig 3E](#)) and found that the Gini–Simpson index of actual ecological communities was systematically lower or at most as high as that of the fair distribution. This suggests that the ecological reality seems to be more unequal than the fair distribution.

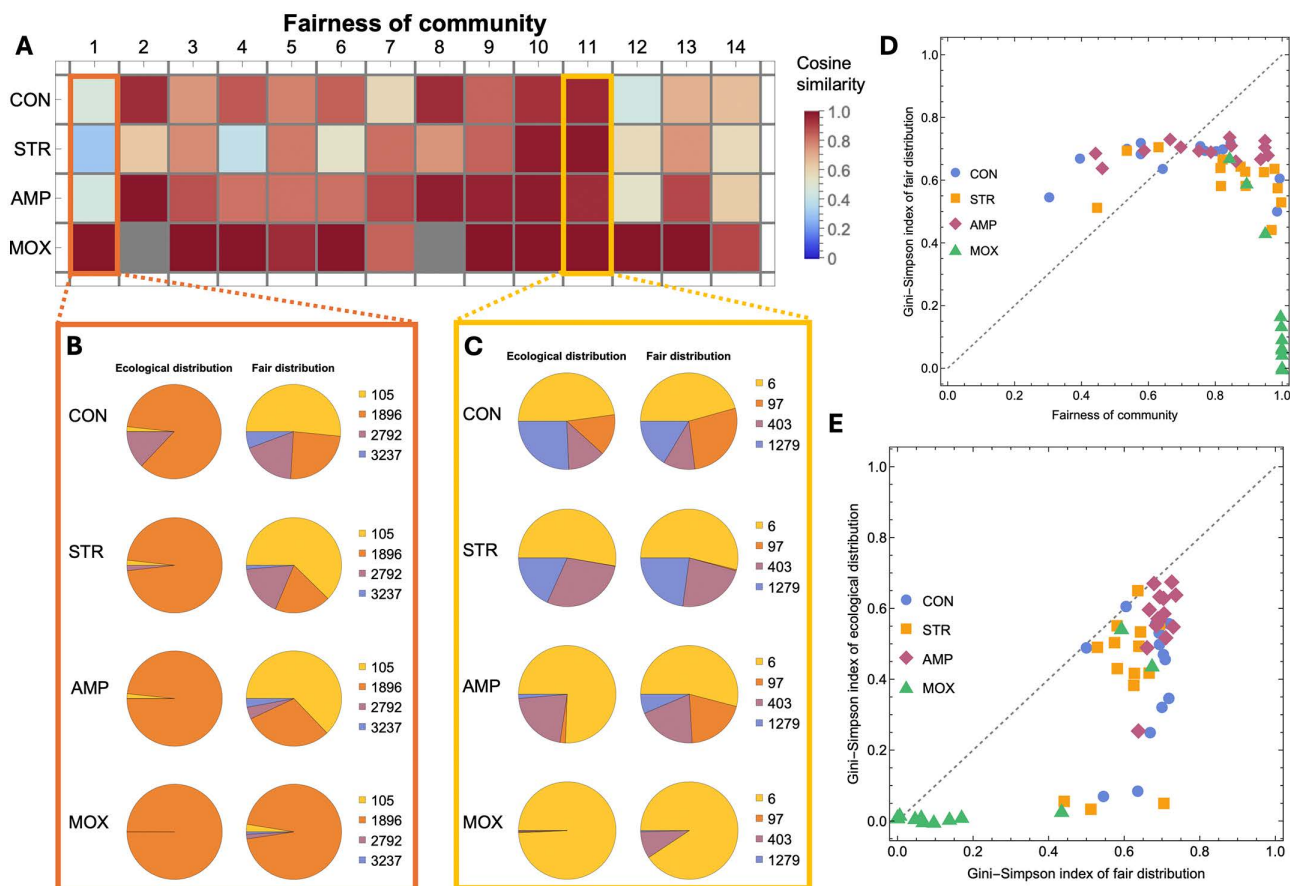


Fig 3. Distributional properties of the ecological vs. fair end-distributions. (A) We quantified the fairness of each community across all environments by taking the cosine similarity of the true end-frequency vector and the corresponding fair end-frequency vector as given by the normalized Shapley values. Higher cosine similarity means that the community is closer to the fair distribution. Gray color indicates that there was no observed growth in communities 2 and 8 in the MOX-environment. (B) Example of an unfair community, which clearly deviates from the fair distribution in all but the MOX-environment. (C) Example of a fair community, which closely follows the fair distribution in all growth conditions. (D) The fairness of the community is not strongly correlated with community diversity of the fair distribution, and in the stronger selective environment (MOX), the fair end-distribution leads to low diversity. (E) The community diversity of the fair distribution plotted against the diversity of the true ecological distribution. All communities fall below the diagonal meaning that the diversity of the fair distribution seems to set an upper bound for the true diversity observed in this experimental system. The data and code needed to create this Figure can be found in <https://osf.io/a7yzp/overview>.

<https://doi.org/10.1371/journal.pbio.3003872.g003>

A property of the fair distribution is that any discrepancies present with respect to true ecological outcomes already in the biculture competitions get propagated forward via the recursive logic of the Shapley value as shown by Eq. (2). However, if we have access to such additional biologically relevant information besides the yields, can we use this information to predict the full community assembly?

To test this, we also sequenced all the subcommunities and modified the recursive formula Eq. (2) to take the estimated true abundances x_i^S at some subcommunity size L as the initial condition for the recursion, and then computed the higher-order Shapley values as before. This leads to the following recursive algorithm

$$\phi_i^S = \begin{cases} \frac{1}{|S|} \left(Y_S - Y_{S \setminus \{i\}} + \sum_{j \neq i} \phi_j^{S \setminus \{i\}} \right), & \text{if } |S| > L \\ x_i^S, & \text{if } |S| = L \end{cases} \quad (4)$$

which can be used to predict the species abundances in the full community as $\hat{x}_i^N = \phi_i^N$. Note that this fair assembly rule contains no free parameters but instead assumes that after the initial unfairness present in the lower subcommunities has been accounted for, the community assembly thereafter follows the rules of the fair assembly.

By incorporating the true biculture and triplet frequencies, we found that the fair assembly rule Eq. (4) was able to predict the observed full community outcomes up to very high accuracy (see Fig 4). We quantified the prediction accuracy using the Euclidean distance between the predicted and observed community composition and found that the mean squared error of the fair distribution with only yields was 0.367 ± 0.034 (SE, $N=54$). By using the fair assembly rule with the correct pairwise frequencies, this improved to 0.233 ± 0.025 (SE, $N=54$), and to 0.132 ± 0.015 (SE, $N=54$) with triplet frequencies. This suggests that the observed assembly process is consistent with the recursive logic of the fair assembly rule: the improved predictive performance is not driven by simply feeding in more data, but by adjusting the initial condition of the recursion based on observed, as opposed to fair, subcommunity outcomes.

Discussion

Although much empirical research has been devoted to characterizing biodiversity in terms of species abundance distributions, there is surprisingly little knowledge about the normative status of the “state of nature”. Here, we asked how fair ecological outcomes are in microbial communities by comparing the observed species abundance distribution to a fair distribution where each species claims only their fair share of the total surplus given by their Shapley value. We found examples of communities where the true ecological distribution followed the fair distribution very closely both in the control conditions and under antibiotic stress, while in other cases some significant species reached unfairly high abundances at the expense of “exploited species” which had a higher objective contribution to the community function than their observed share. One potential mechanism leading to unfairness in our system could be the varying lag time before the onset of growth: we found that species with shorter lag time were generally able to claim their fair share according to their Shapley value while species with a longer lag grew either better or worse after the initial growth of other species (see S2 Fig in the

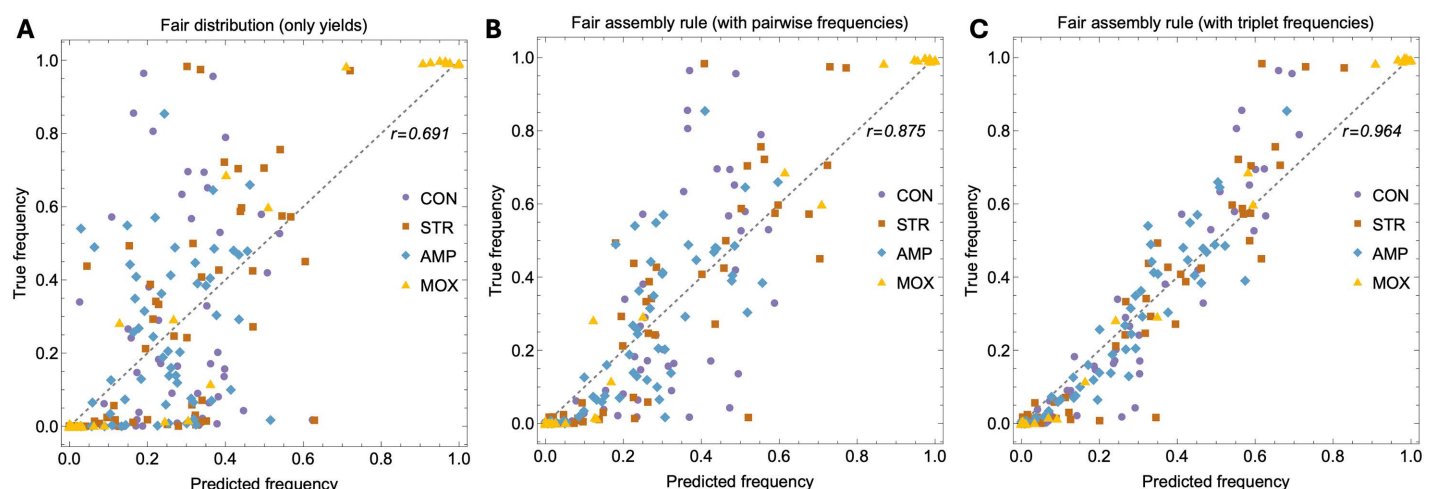


Fig 4. Fair assembly rule predicts community outcomes. The fair assembly rule Eq. (4) provides a simple method to predict the full community assembly using the recursive logic of the Shapley value together with information about observed proportions in smaller subcommunities. **(A)** The fair distribution given by the Shapley values captures some aspects of the true assembly process (Pearson's $r=0.691$). **(B)** When the fair assembly rule with true rather than fair biculture proportions is used, the full community assembly can be predicted much better (Pearson's $r=0.875$). **(C)** When the fair assembly rule with the correct triplet frequencies is used, the predictive performance further improves (Pearson's $r=0.964$). The data and code needed to create this Figure can be found in <https://osf.io/a7yzp/overview>.

<https://doi.org/10.1371/journal.pbio.3003872.g004>

Supplementary material). This observation may be explained by the varying degree of niche overlap, where the growth of the species with longer lag can be hindered in case of high overlap, but improved in case of low overlap, for example, due to cross-feeding.

While very unequal payoff distributions can also be fair, as seen in the stronger selective environment (Fig 3D), we found that the community diversity of the true ecological distribution was systematically lower or at most the same as the corresponding diversity of the fair distribution as measured by the Gini–Simpson index. This tentatively suggests that at least in our microbial model system, the ecological reality seems to be even harsher than that of a purely meritocratic system. However, an important open question remains, whether this applies also in real ecosystems where species have co-evolved together as opposed to the random assemblages used here. Therefore, it may well be plausible that in the absence of “unfair” asymmetries, such as historical priority effects, nature is indeed organized increasingly ‘fairly’ in the Shapley sense. To what extent species sorting and evolutionary mechanisms invoking community state shifts [35] can act on fairness remains an intriguing topic for future studies, especially if fairness indeed affects community function. For example, it has been shown that community evenness can facilitate higher robustness in metrics such as invasion resistance and functional redundancy [36,37].

We believe that the introduced framework based on ideas of cooperative game theory and the Shapley value can become broadly useful in analyzing ecological data and developing ecological theory based on the several key strengths highlighted by our work. First, it provides a fully general way to make a non-trivial, game-theoretical baseline prediction for community outcomes and as such potentially serve as a useful null model that is driven by experimental data rather than any assumptions about the underlying ecological dynamics. Indeed, the fair assembly rule Eq. (4) based on the recursive definition of the Shapley value Eq. (2) could accurately predict complex community outcomes with zero free parameters. To emphasize, we did not *fit* any mathematical or statistical model to data but rather derived the predictions from the basic principles of fair distribution, further showing that the recursive logic of the Shapley value provided a very parsimonious description for the observed data and a new lens for understanding community composition. This success is particularly interesting as it remains unclear when microbial assembly follows general rules in contrast to cases where the community member species succeed in “surprising” the experimenters by exhibiting the so-called emergent properties [38,39]. Thus, our assembly rule based on fairness opens a new way to identify at what level of hierarchy the possible emergent properties arise. The methodology can also be directly applied to other community functions or properties of interest besides the yield, which was our focus here. By construction, the Shapley value compares community outcomes with and without the presence of each species, thus allowing to make causal inferences of the species contributions as well as their interactions in a model-agnostic way.

There are, of course, also some limitations to the approach taken here, the most notable being the fact that to compute the Shapley value in practice, one must experimentally measure the yield (or other value of interest) from all the 2^n possible subcommunities, which grows quickly with the number of present species. However, the technology to make combinatorially complete experiments even for moderately large communities already exists [40] and is likely to become easier in the future. Similarly, computer scientists actively develop approximative analysis ideas trying to overcome enumeration challenges related to Shapley values [30] which may be transferable to experimental design also in biology. Additionally, if one has access to the realized species distribution at some level, the introduced fair assembly rule can then be used to make predictions for the full community of any size without having to measure all the smaller subcommunities. This makes leave-one-out experiments (see, e.g., Ref. [41]) particularly interesting for testing how these ideas might translate to larger communities.

In conclusion, we have demonstrated how a purely normative concept can be successfully applied to studying a complex system, where a central problem in the typical bottom-up modeling approaches has been the number of unknown parameters that are very difficult to infer even for a minimal model. The suggested top-down approach of using the Shapley value and fairness to understand and predict community assembly is analogous to methods based on optimization

principles [42] or principle of maximum entropy [43], which avoid this problem by deriving the model parameters and predictions from first principles. We anticipate that community assembly theory could be advanced with new types of normative models and identify the evolution of fairness as a particularly interesting area of future research both theoretically related to the evolutionary stability and reachability of fair community compositions as well as empirically in experimental evolution setups.

Materials and methods

Sampling logic and experimental setup

In this experiment, we had a pool of 16 bacterial species from which we randomly selected 4 species to each set. Randomized sets with duplicates of any given strain were discarded, and thus all sets have 4 different species in them. We then formed all possible combinations from these 4 species ($n = 15$), and this grand coalition is referred to as a block. We formed 25 blocks for this experiment, but sequenced 14 of them, and thus only data from 14 blocks is used in this article.

Strain availability and pre-culturing

All bacterial strains used in this experiment are from the University of Helsinki Culture Collection (Microbial Domain Biological Resource Centre, HAMBI). The strains chosen into the species pool were a subset of species based on prior research (see Table 1 and Ref. [44]) and they were known to be able to coexist even for extended periods of time, with their antibiotic resistance profiles well-characterized.

Prior to the experiment, we revived the frozen glycerol stocks by axenically cultivating them in 6 ml proteose-peptone-yeast (PPY) medium for 22 hours under experimental conditions. Due to its extended lag phase, we revived HAMBI 3,237 in two stages: following an initial 48 h growth period, 500 μ l of culture was used to inoculate 6 mL fresh PPY medium for the main pre-cultivation step.

Table 1. Strains in the bacterial pool. HAMBI strains in the pool of 16 bacterial species used in this experiment. Adapted from supplementary material (S2 Table) of Ref. [44].

Strain ID	Species Name	Origin
6	<i>Pseudomonas putida</i>	N/A
97	<i>Acinetobacter johnsonii</i>	N/A
105	<i>Agrobacterium tumefaciens</i>	Soil, USA
262*	<i>Brevundimonas bullata</i>	Soil
403	<i>Comamonas testosteroni</i>	Soil, Berkeley, USA
1279	<i>Hafnia alvei</i>	N/A
1287	<i>Citrobacter koseri</i>	N/A
1292	<i>Morganella morganii</i>	Human with summer diarrhea, London, UK
1299	<i>Kluyvera intermedia</i>	Surface water, France
1842	<i>Sphingobium yanoikuyae</i>	Clinical specimen
1896	<i>Sphingobacterium spiritivorum</i>	Human, intra-uterine, KS, USA
1972	<i>Aeromonas caviae</i>	Epizootic of young guinea pigs, KY, USA
1977	<i>Pseudomonas chlororaphis</i>	Farm soil, Peoria, IL, USA
2659	<i>Stenotrophomonas maltophilia</i>	Oropharyngeal region of patient with mouth cancer
2792	<i>Moraxella canis</i>	Dog's beard, Finland
3237	<i>Microvirga lotononidis</i> sp. Nov.	Nitrogen-fixing nodule of <i>Lotononis angolensis</i> , Chibala, Zambia

*Strain 262 was included in the pool, but it was not present in any of the 14 randomized four-species blocks used in the article.

<https://doi.org/10.1371/journal.pbio.3003872.t001>

Growth assay

Prior to setting up the growth assay, we adjusted the cell density of each culture to 2.4×10^7 cells/ mL by measuring each culture's OD₆₀₀ and diluting the cultures with M9 minimal medium. The dilution factors were determined based on the relationship between OD₆₀₀ and cell density, established from previously done flow cytometer measurements and CFU (colony-forming unit) counts.

After density adjustment, we assembled the combinations by dispensing 500 μ l density-leveled axenic cultures to the corresponding combination tubes with species number-specific volumes of M9. This resulted in the cell density of 6×10^6 cells/ mL per community.

For the growth assay, we used liquid 100% R2A (Reasoner's 2A) medium. In this experiment, we have 4 treatment groups: control, ampicillin, streptomycin, and moxifloxacin. The final media antibiotic concentrations were 4.24 μ g/ mL for ampicillin, 6.3 μ g/ mL for streptomycin, and 0.06 μ g/ mL for moxifloxacin. These concentrations are based on median IC50 values across the 16 strains in the bacterial pool the combinations were randomized from. For the control treatment, we added sterile MQ water to dilute the media to have the same amount of nutrients and salts as the antibiotic treatments. The media volume was 140 μ l/ well on 96-well cell culture microplates (Corning). We inoculated 10 μ l of the assembled bacterial consortia into each well. The growth assays were conducted on Agilent BioTek LogPhase 600 Microbiology Readers at 600 nm, 30 °C for 48 h with shaking on at 800 RPM. The OD₆₀₀ was measured every 10 min. After the growth assay, the culture plates were frozen at -20 °C prior to 16S amplicon library preparation.

Library preparation

We constructed the 16S amplicon libraries for community composition analysis using an in-house library preparation method derived from the protocol described in [45]. Briefly, the V3 region of the 16S gene was amplified with fusion primers which included iTru fusion adapters and an internal sample-specific and combinatorial index. Amplicons were purified with NGS Normalization 96-Well Kit (Norgen Biotek), pooled in sets of 90 reactions, and subsequently amplified with standard TrueSeq primers and indexes. The reaction products were purified with Sera-Mag particles (Cytiva), and equimolarly pooled. The quality of the library was evaluated using Bioanalyzer High sensitivity DNA kit (Agilent) and sequenced on Illumina Miseq at the Finnish Functional Genomics Centre (Turku, Finland), using the MiSeq Reagent Micro Kit v2 (2 $\times$ 150 bp reads). DNA template for PCR was obtained by thermal lysis of diluted samples in dH2O (10 min at 99.9 °C) and subsequent removal of cellular debris by centrifugation.

Bioinformatics/downstream sequence analysis

We demultiplexed the pooled 16S amplicon paired-end sequencing data by: (1) identifying each experimental community sample by the forward and reverse primer sequence combination; (2) trimming paired-end adapter sequences; (3) merging paired-end reads; and (4) filtering merged reads. Scripts to run this pipeline are based on scripts available at <https://gitlab.utu.fi/slhogl/hambiDemuxQC>.

Identification of samples by primer combinations was done using *cutadapt* [46] (version 4.9) software to search sequencing data for primer combinations corresponding to specific experimental communities, as defined during library preparation stage (maximum error rate 0.15). Likewise, *cutadapt* [46] (version 4.9) was used to trim paired-end adapter sequences (forward: CCTACGGGAGGCAGCAG; reverse: ATTACCGCGGCTGCTGG; maximum error rate 0.2). Merging of the trimmed paired-end reads was done using *NGmerge* [47] (version 0.3) software in default ("stitch") mode. Filtering of the merged reads was done using *vsearch* [48] (version 2.22.1) software (maximum expected error 1; minimum sequence length 115; maximum sequence length 165; "N" characters in sequences are not allowed).

We derived the absolute abundances of the bacterial species within each experimental community using the demultiplexed merged reads with *Rbec* [49] (version 1.8.0; as part of R Bioconductor version 3.17) software. Relative

abundances obtained by first normalizing species-specific counts by the expected number of 16S gene copies in their genome, then dividing by the total number of counts in the sample. Scripts to run this software are based on the script available at <https://gitlab.utu.fi/slhogl/hambiAmplicon>.

Data analysis

When computing the yields, we subtracted the measured background optical density obtained from the associated control condition. As the Shapley value can become negative for some species, unlike true abundances, we compute the fair end-distribution as $\varphi_i = \max(0, \phi_i) / \sum_k \max(0, \phi_k)$, where the max-operator renormalizes the potentially negative values to give well-defined fair frequencies that sum up to unity. We quantified community diversity using the Gini–Simpson index defined as $1 - \sum_i \varphi_i^2$.

Supporting information

S1 Fig. Species' average Shapley value versus true abundance across environments. The Shapley values of each species (given by the color label) was averaged over all the communities it was present and plotted separately for each environment (given by the plot marker) against the corresponding mean true abundance. [Fig 2D](#) is obtained from this plot by further averaging these points across the environments per species.

(TIFF)

S2 Fig. Lag time is correlated with unfairness. Using the optical density measurements, we first computed the lag time associated with each growth curve as the time until maximum growth rate. We then plotted the median lag time of each species against the observed unfairness (absolute distance to the diagonal in [Fig 2D](#)). This shows that species with shortest lag time (that is, those that start to grow earlier) are also those which on average obtain their fair share according to their Shapley value, while species with longer lag time reach either unfairly high or low abundances.

(TIFF)

Acknowledgments

We would like to acknowledge members of the Bioinformatics and Evolution group for helpful discussions and comments as well as Shane Hogle and Johannes Cairns for their help related to bioinformatics. The authors wish to acknowledge CSC—IT Center for Science, Finland, for computational resources.

Author contributions

Conceptualization: Teemu Kuosmanen, Juhani Rantanen, Teppo Hiltunen, Ville Mustonen.

Formal analysis: Teemu Kuosmanen, Ville Mustonen.

Funding acquisition: Teppo Hiltunen, Ville Mustonen.

Investigation: Teemu Kuosmanen, Juhani Rantanen, Dovydas Kičiatovas, Sanna Pausio, Ville-Petri Friman, Teppo Hiltunen, Ville Mustonen.

Methodology: Teemu Kuosmanen, Juhani Rantanen, Dovydas Kičiatovas, Ville Mustonen.

Project administration: Sanna Pausio, Teppo Hiltunen.

Resources: Teppo Hiltunen, Ville Mustonen.

Software: Teemu Kuosmanen.

Supervision: Teppo Hiltunen, Ville Mustonen.

Validation: Teemu Kuosmanen.

Visualization: Teemu Kuosmanen, Ville Mustonen.

Writing – original draft: Teemu Kuosmanen, Juhani Rantanen, Ville Mustonen.

Writing – review & editing: Teemu Kuosmanen, Juhani Rantanen, Dovydas Kičiatovas, Sanna Pausio, Ville-Petri Friman, Teppo Hiltunen, Ville Mustonen.

References

- Smil V. Growth: from microorganisms to megacities. MIT Press; 2019.
- Hastings 2 A, Hom CL, Ellner S, Turchin P, Godfray HCJ. Chaos in ecology: is mother nature a strange attractor?. *Annu Rev Ecol Syst*. 1993;24:1–33.
- Grinin L, Korotayev A, Tausch A. Economic cycles, crises, and the global periphery. 1st ed. Cham, Switzerland: Springer International Publishing; 2016.
- Thorogood R, Mustonen V, Aleixo A, Aphalo PJ, Asiegbu FO, Cabeza M, et al. Understanding and applying biological resilience, from genes to ecosystems. *NPJ Biodivers*. 2023;2(1):16. <https://doi.org/10.1038/s44185-023-00022-6> PMID: 39242840
- Arani BMS, Carpenter SR, Lahti L, van Nes EH, Scheffer M. Exit time as a measure of ecological resilience. *Science*. 2021;372(6547):eaay4895. <https://doi.org/10.1126/science.aay4895> PMID: 34112667
- Anderson PW, Arrow K, Pines D, editors. The economy as an evolving complex system. Addison-Wesley; 1988.
- Geritz SAH, Kisdi E, Mesze NA G, Metz JAJ. Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evol Ecol*. 1998;12(1):35–57. <https://doi.org/10.1023/a:1006554906681>
- Kuznets S. Economic growth and income inequality. *Am Econ Rev*. 1955;45:1–28.
- Piketty T, Saez E. Inequality in the long run. *Science*. 2014;344(6186):838–43. <https://doi.org/10.1126/science.1251936> PMID: 24855258
- MacArthur R. On the relative abundance of species. *Am Nat*. 1960;94(874):25–36. <https://doi.org/10.1086/282106>
- Locey KJ, White EP. How species richness and total abundance constrain the distribution of abundance. *Ecol Lett*. 2013;16(9):1177–85. <https://doi.org/10.1111/ele.12154> PMID: 23848604
- Nekola JC, Brown JH. The wealth of species: ecological communities, complex systems and the legacy of Frank Preston. *Ecol Lett*. 2007;10(3):188–96. <https://doi.org/10.1111/j.1461-0248.2006.01003.x> PMID: 17305802
- Preston FW. The commonness, and rarity, of species. *Ecology*. 1948;29:254–83.
- Grilli J. Macroecological laws describe variation and diversity in microbial communities. *Nat Commun*. 2020;11(1):4743. <https://doi.org/10.1038/s41467-020-18529-y> PMID: 32958773
- Gao Y, Abdullah A, Wu M. The powerbend distribution provides a unified model for the species abundance distribution across animals, plants and microbes. *Nat Commun*. 2025;16(1):4035. <https://doi.org/10.1038/s41467-025-59253-9> PMID: 40301372
- Marshall HH, Johnstone RA, Thompson FJ, Nichols HJ, Wells D, Hoffman JI, et al. A veil of ignorance can promote fairness in a mammal society. *Nat Commun*. 2021;12(1):3717. <https://doi.org/10.1038/s41467-021-23910-6> PMID: 34162841
- Broom M, Rychtář J. Game-theoretical models in biology. Boca Raton: Chapman and Hall/CRC; 2022.
- Meacock OJ, Mitri S. Environment-organism feedbacks drive changes in ecological interactions. *Ecol Lett*. 2025;28(1):e70027. <https://doi.org/10.1111/ele.70027> PMID: 39737705
- Cui W, Marsland R, Mehta P. Les Houches lectures on community ecology: From niche theory to statistical mechanics. *arXiv Preprint*. 2024. <https://doi.org/10.48550/arXiv.2403.05497>
- Granato ET, Meiller-Legrand TA, Foster KR. The evolution and ecology of bacterial warfare. *Curr Biol*. 2019;29(11):R521–37. <https://doi.org/10.1016/j.cub.2019.04.024> PMID: 31163166
- Palmer JD, Foster KR. Bacterial species rarely work together. *Science*. 2022;376(6593):581–2. <https://doi.org/10.1126/science.abn5093> PMID: 35511986
- Shapley LS. A value for n-person games. In: Kuhn HW, Tucker AW, editors. Contributions to the theory of games. Princeton: Princeton University Press; 1953. p. 307–17.
- Hart S, Mas-Colell A. Potential, value, and consistency. *Econometrica*. 1989;57:589.
- Roth AE. The shapley value: essays in honor of Lloyd S. Shapley. Cambridge, England: Cambridge University Press; 1988.
- Scott M, Su-In L. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst*. 2017;30:4765–74.
- Bohl K, Hummert S, Werner S, Basanta D, Deutsch A, Schuster S, et al. Evolutionary game theory: molecules as players. *Mol Biosyst*. 2014;10(12):3066–74. <https://doi.org/10.1039/c3mb70601j> PMID: 25248454
- Deutscher D, Meilijson I, Schuster S, Ruppin E. Can single knockouts accurately single out gene functions?. *BMC Syst Biol*. 2008;2:50. <https://doi.org/10.1186/1752-0509-2-50> PMID: 18564419

28. Keinan A, Sandbank B, Hilgetag CC, Meilijson I, Ruppin E. Fair attribution of functional contribution in artificial and biological networks. *Neural Comput.* 2004;16(9):1887–915. <https://doi.org/10.1162/0899766041336387> PMID: [15265327](#)
29. Manor O, Borenstein E. Systematic characterization and analysis of the taxonomic drivers of functional shifts in the human microbiome. *Cell Host Microbe.* 2017;21(2):254–67. <https://doi.org/10.1016/j.chom.2016.12.014> PMID: [28111203](#)
30. Pham G, Milazzo P. A comprehensive review of the use of Shapley value to assess node importance in the analysis of biological networks. *Comput Methods Programs Biomed Update.* 2025;7:100185. <https://doi.org/10.1016/j.cmpbup.2025.100185>
31. Waldo C, Wegscheider B, Josi D, Calegari BB, Brodersen J, Jardim de Queiroz L, et al. Deconstructing the geography of human impacts on species' natural distribution. *Nat Commun.* 2024;15(1):8852. <https://doi.org/10.1038/s41467-024-52993-0> PMID: [39402017](#)
32. Hartmann K. The equivalence of two phylogenetic biodiversity measures: the Shapley value and Fair Proportion index. *J Math Biol.* 2013;67(5):1163–70. <https://doi.org/10.1007/s00285-012-0585-y> PMID: [22986892](#)
33. Piccardi P, Vessman B, Mitri S. Toxicity drives facilitation between 4 bacterial species. *Proc Natl Acad Sci U S A.* 2019;116(32):15979–84. <https://doi.org/10.1073/pnas.1906172116> PMID: [31270235](#)
34. Dragan I. Some recursive definitions of the Shapley value and other linear values of cooperative TU games. *Working Paper 328 of the University of Texas at Arlington*; 1997.
35. Kivikoski M, Cairns J, Hogle SL, Pausio S, Becks L, Mustonen V, et al. Evolution induced state shifts in a long-term microbial community experiment. *Proc Natl Acad Sci U S A.* 2026;123(22):e2533269123. <https://doi.org/10.1073/pnas.2533269123> PMID: [42201971](#)
36. Wittebolle L, Marzorati M, Clement L, Balloi A, Daffonchio D, Heylen K, et al. Initial community evenness favours functionality under selective stress. *Nature.* 2009;458(7238):623–6. <https://doi.org/10.1038/nature07840> PMID: [19270679](#)
37. De Roy K, Marzorati M, Negroni A, Thas O, Balloi A, Fava F, et al. Environmental conditions and community evenness determine the outcome of biological invasion. *Nat Commun.* 2013;4:1383. <https://doi.org/10.1038/ncomms2392> PMID: [23340423](#)
38. Chang C-Y, Bajić D, Vila JCC, Estrela S, Sanchez A. Emergent coexistence in multispecies microbial communities. *Science.* 2023;381(6655):343–8. <https://doi.org/10.1126/science.adg0727> PMID: [37471535](#)
39. Friedman J, Higgins LM, Gore J. Community structure follows simple assembly rules in microbial microcosms. *Nat Ecol Evol.* 2017;1(5):109. <https://doi.org/10.1038/s41559-017-0109> PMID: [28812687](#)
40. Diaz-Colunga P, Catalan M, San Roman A, Arrabal A, Sanchez. Full factorial construction of synthetic microbial communities. *eLife.* 2024;13:RP101906.
41. Ansari AF, Reddy YBS, Raut J, Dixit NM. Predicting structures of microbial communities. *Nat Comput Sci* 2021;1:619–28.
42. Sokolowski TR, Gregor T, Bialek W, Tkačik G. Deriving a genetic regulatory network from an optimization principle. *Proc Natl Acad Sci U S A.* 2025;122(1):e2402925121. <https://doi.org/10.1073/pnas.2402925121> PMID: [39752518](#)
43. Harte J, Newman EA. Maximum information entropy: a foundation for ecological theory. *Trends Ecol Evol.* 2014;29(7):384–9. <https://doi.org/10.1016/j.tree.2014.04.009> PMID: [24863182](#)
44. Cairns J, Jokela R, Hultman J, Tamminen M, Virta M, Hiltunen T. Construction and characterization of synthetic bacterial community for experimental ecology and evolution. *Front Genet.* 2018;9:312. <https://doi.org/10.3389/fgene.2018.00312> PMID: [30154827](#)
45. Glenn TC, Pierson TW, Bayona-Vásquez NJ, Kieran TJ, Hoffberg SL, Thomas IV JC, et al. Adapterama II: universal amplicon sequencing on Illumina platforms (TaggiMatrix). *PeerJ.* 2019;7:e7786. <https://doi.org/10.7717/peerj.7786> PMID: [31616589](#)
46. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 2011;17(10).
47. Gaspar JM. NGmerge: merging paired-end reads via novel empirically-derived models of sequencing errors. *BMC Bioinformatics.* 2018;19(1):536. <https://doi.org/10.1186/s12859-018-2579-2> PMID: [30572828](#)
48. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ.* 2016;4:e2584. <https://doi.org/10.7717/peerj.2584> PMID: [27781170](#)
49. Zhang P, Spaepen S, Bai Y, Hacquard S, Garrido-Oter R. Rbec: a tool for analysis of amplicon sequencing data from synthetic microbial communities. *ISME Commun.* 2021;1(1):73. <https://doi.org/10.1038/s43705-021-00077-1> PMID: [37938657](#)