ROYAL SOCIETY OPEN SCIENCE

royalsocietypublishing.org/journal/rsos

Research

The evolution of scientific credit: when authorship norms impede collaboration

Toby Handfield[1] and Kevin Zollman[2]

[1]SOPHIS, Monash University, Victoria, Australia
[2]Department of Philosophy, Carnegie Mellon University, Pittsburgh, PA, USA

TH, 0000-0003-2995-2067; KZ, 0000-0001-9776-1890

Cite this article: Handfield T, Zollman K. 2026 The evolution of scientific credit: when authorship norms impede collaboration. *R. Soc. Open Sci.* **13**: 251301.
https://doi.org/10.1098/rsos.251301

Received: 10 July 2025
Accepted: 18 November 2025

Subject Category:
Science, society and policy

Subject Areas:
computer modelling and simulation, applied mathematics

Keywords:
meta-science, sociology of science, social epistemology

Author for correspondence:
Kevin Zollman
e-mail: kzollman@andrew.cmu.edu

Scientific authorship norms vary dramatically across disciplines, from contribution-sensitive systems where first author is the greatest contributor and subsequent author order reflects relative input, to contribution-insensitive conventions like alphabetical ordering or senior-author-last. We develop evolutionary game-theoretic models to examine both how these divergent norms emerge and their subsequent effects on collaborative behaviour. Our first model reveals that contribution-insensitive norms evolve when researchers who sacrifice positional advantage face the strongest adaptive pressure—for example, senior authors managing larger collaboration portfolios or bearing heavier reputational stakes. This 'Red King' dynamic potentially explains why fields in which senior researchers command large laboratories, major grants and extensive collaboration portfolios may paradoxically evolve conventions that favour junior authors. Our second model demonstrates that established norms influence researchers' willingness to collaborate, with contribution-sensitive norms consistently outperforming insensitive alternatives in fostering successful partnerships. Contribution-insensitive norms create systematic coordination failures through two mechanisms: 'main contributor resentment' when exceptional work goes unrecognized; and 'second contributor resentment' when comparable efforts receive unequal credit. These findings suggest that widely adopted practices like senior-last positioning and alphabetical ordering may function as institutional frictions that impede valuable scientific collaborations rather than neutral organizational conventions, potentially reducing overall scientific productivity across affected disciplines.

Supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.8210678.

1. Introduction

Credit functions as the lifeblood of the scientific enterprise—by-lines, citation counts and perceived ownership of ideas drive

THE ROYAL SOCIETY PUBLISHING

hiring, promotion, grant awards and career trajectories, thereby motivating researchers to pursue ambitious questions and invest the effort necessary for rigorous discovery [1]. At the same time, collaboration has become the dominant mode of knowledge production: multi-author papers now represent the lion's share of publications, large consortia tackle complex problems and empirical studies link collaborative work to higher citation impact and greater methodological innovation [2–6]. Yet reducing the rich tapestry of individual contributions—ranging from conceiving the research question and designing experiments to collecting data, performing analyses and drafting prose—into a simple, sequential author list entails loss of information. That compression can skew incentives and rewards, privileging certain roles or career stages over others and may shape both individual trajectories and the collective pace of scientific progress in unexpected or undesirable ways.

To examine how different approaches to this information compression problem shape scientific practice, we focus specifically on author-order norms, which vary considerably across disciplines. Some fields employ contribution-sensitive norms (C-Norms), where author order directly signals relative contribution, typically with first authors having contributed most. Other fields have adopted contribution-insensitive norms (I-Norms), with perhaps the most notable being the alphabetical norm often used in mathematics and economics. In many biological sciences, there is a 'senior-author-last' convention, whereby the last author is understood to have had a supervisory role, and perhaps provided funding or infrastructure, but with no determinate implication regarding this author's contribution to the intellectual content of the paper. The last author might have been the one who had the whole idea, without which the paper would not exist, or the last author might just be coasting on the hard work and innovation of more junior colleagues. Hence at least regarding the final position in the author list, this norm is insensitive to contribution. Often the senior-author-last convention is combined, however, with the presumption that first author position signals principal contributor, while middle authors are listed alphabetically, signalling roughly equal, but lesser contributions than both first and last authors [7]. Other practices include footnoted equal contribution statements, rotation systems where collaborators take turns and complex discipline-specific conventions that blend multiple signals. For analytical clarity, however, our investigation focuses on two-author collaborations, allowing us to employ the simpler C-Norm versus I-Norm framework without confronting these hybrid cases. While we primarily examine a senior-last convention, our model generalizes to other I-Norms such as alphabetical ordering [8].

These relatively explicit, widely endorsed norms of author ordering of course interact with other, less official patterns of credit assignment. The 'Matthew effect', whereby eminent scientists receive disproportionate credit for joint work [9–11], may lead readers to attribute outsized credit to senior authors regardless of their position in the author list. Similarly, there can be a bias towards giving more credit to first authors. Even in fields with alphabetical ordering norms, readers might consciously know this convention carries no information about contribution, yet still unconsciously attribute more credit to earlier-listed authors [12,13]. Indeed, surname-initial effects in disciplines with alphabetical norms create strategic incentives and measurable efficiency losses in co-authorship patterns [12,14–17]. These informal biases operate alongside formal norms, creating tensions between official conventions and unofficial credit attribution that can undermine the informational value of authorship practices.

To disentangle these complex interactions between formal norms and actual scientific practice, we need theoretical frameworks that can account for both the emergence of different authorship conventions and their subsequent effects on collaboration behaviours. To investigate these issues, we develop two connected models. First, we analyse a simple game between two researchers who must agree on an authorship norm before collaborating. This evolutionary model reveals when different norms are likely to become established. Second, we examine how established norms affect researchers' willingness to collaborate at all, identifying conditions under which different norms promote or inhibit scientific partnerships.

Building on the seminal analysis by Engers *et al.* [18], which explored how authorship norms affect effort allocation within collaborations, we shift attention to prior stages: the formation of collaborations and the establishment of norms themselves. Whereas Engers *et al.* analysed how different conventions influence individual effort once collaboration is underway, we ask how collaborations arise in the first place and which norms are likely to prevail. This shift reflects the sequential nature of scientific activity: authors must first decide *whether* to collaborate and under *which* convention, before deciding *how* to allocate effort. By foregrounding these upstream choices, our models illuminate the broader structural conditions under which individual strategic behaviour occurs.

Our analysis yields two main results. First, I-Norms emerge not through egalitarian ideals but through asymmetric evolutionary pressures: these norms become entrenched when the researchers who sacrifice positional advantage (senior authors in senior-last systems, late-alphabet researchers in alphabetical ordering) face the most intense adaptive pressure. Whether driven by managing larger collaboration portfolios, bearing heavier reputational consequences or confronting steeper opportunity costs from partnership failures, these high-stakes actors evolve towards accommodating conventions that formally disadvantage them. This finding reveals that ostensibly neutral practices like alphabetical ordering or senior-last positioning may actually encode strategic adaptations, not democratic consensus or historical accident.

Second, this evolutionary logic produces systematic inefficiency: C-Norms consistently outperform their insensitive counterparts in fostering successful collaborations. The compression of rich contribution information into positionally arbitrary signals creates predictable coordination failures, suggesting that disciplines employing senior-last or alphabetical conventions may be systematically discouraging valuable scientific partnerships. These widely adopted practices create institutional frictions that impede collaborative science.

These findings connect to broader investigations of how scientific norms shape knowledge production [19–21]. Science operates through an intricate web of formal and informal conventions: statistical thresholds that determine significance, methodological standards for gathering evidence, what constitute interesting or important problems [22], expectations around peer review, criteria for promotion and, of course, protocols governing authorship [7,23–25]. Although these norms often present as arbitrary traditions or accidents of history, they reflect complex interactions between incentive structures, power dynamics and information needs. The norms and customs of scientific practice—including authorship conventions—continuously evolve through competitive adaptation, strategic behaviour and institutional path dependencies.

Viewed through this lens, our results offer insights for both understanding and potentially reforming scientific institutions. Credit attribution systems serve dual roles: they function as incentive mechanisms that motivate effort and information channels that signal contribution. Our findings reveal how tensions between these roles can reduce scientific productivity and potentially skew career trajectories. By treating authorship norms as strategic choices with field-wide consequences, our approach provides a framework for analysing the evolution and effects of scientific conventions more broadly, suggesting that seemingly inconsequential organizational practices may significantly influence both individual careers and collective knowledge generation.

2. Model and results

To explore how author order norms arise and influence scientific collaboration, we develop two connected models. These models analyse simplified games that researchers 'play' when deciding how to collaborate and how to assign authorship credit. We focus on two main types of norms: *contribution-sensitive norms (C-Norms)*, where author order reflects who contributed more; and *contribution-insensitive norms (I-Norms)*, like alphabetical ordering or senior-author-last, where order does not signal contribution.

2.1. How author order norms emerge: a simple game

Our first model examines how C-Norms and I-Norms might become established in scientific fields. We consider a simplified scenario with two researchers—'Junior' and 'Senior'—who are contemplating collaboration. Before they begin working together, they must implicitly agree on which norm to use if they publish together.

Each researcher proposes a norm: either the C-Norm or the I-Norm. The C-Norm places the researcher who contributed more as first author, while the I-Norm always places Junior first. For collaboration to proceed, these proposals must be compatible. 'Compatible' means they do not anticipate any irresolvable disagreement over who gets the *first* author position when publication time arrives. Compatibility arises in three scenarios: (i) both choose the C-Norm; (ii) both choose the I-Norm; or (iii) Senior chooses the I-Norm while Junior chooses the C-Norm. This last case represents a situation where both researchers are effectively yielding priority to the other. Senior is saying 'you should be listed first regardless'; while Junior is saying 'no, if you contribute more, you should go first'. We assume they resolve this amicable standoff by flipping a coin to determine which norm to follow. The only incompatible scenario, preventing collaboration entirely, occurs when Junior insists on the I-Norm (guaranteeing their first authorship) while Senior insists on the C-Norm (where they would be first if they contribute more).

If the researchers successfully agree on a norm and collaborate, the proportion contributed by each is determined by an exogenous distribution. We characterize this distribution using three parameters: the probability that Junior contributed more (w_j), Junior's expected contribution given that they contributed more (b_j) and Senior's expected contribution given that they contributed more (b_s).

When deciding which norm to adopt, researchers consider what maximizes their expected *credit*. This credit is assigned by a third party, representing the scientific community, who observes the published paper and attempts to determine how much each author contributed, based on: the observed author order, their prior expectations about the relative contributions of the authors and their understanding of prevailing norms (i.e. what the author's mixed strategies are in the game).

We assume authors bargain over author order before they know the magnitude of their individual contributions. At time of decision, they merely know the probability distribution over relative contributions (w_j, b_j, b_s), and they know the value that would be added to their ideas if they are able to agree on an author order and publish those ideas together ($\hat{c} > 0$). That is, we assume that collaboration would improve the quality of their joint work, and this is the key incentive that drives them to consider co-authorship.

Formal norms alone do not determine credit assignment; informal biases inevitably intrude. We model this with two parameters. First, $\varepsilon > 0$ represents the probability that the community mistakenly gives all credit to the first author, even under an I-Norm where position ostensibly carries no signal about contribution. This models the stubborn tendency to remember and cite first authors regardless of field conventions. Second, we introduce a parameter to capture the classic 'Matthew effect' [9], whereby with probability $\chi > 0$ the senior author receives all credit regardless of author order or actual contribution. These parameters reflect the cocktail of cognitive biases, information shortcuts and social dynamics that shape scientific recognition beyond formal norms [12,16,17,26,27].

So if the authors agree on compatible norms, they collaborate, and they together realize a paper of value $1 + \hat{c}$, where $\hat{c} > 0$ is a parameter reflecting the additional value generated by collaboration. The scientific community observes the value of the joint paper perfectly, and then makes an estimate, conditional on their beliefs about the players' mixed strategies (they know what norm is likely to be in effect) and on the scientists' expected contributions (they have accurate beliefs about the relative abilities of the individuals), of the individual contributions of each scientist. These credit attributions by the community simply are the pay-offs to the authors.

If the authors do not agree on norms then they must publish their ideas separately. The community is then able to precisely estimate who did what, and the pay-off to each is simply the value of their respective contributions, without the benefit of collaboration.

Throughout our analysis, we model scientists as seeking credit assigned by the scientific community, where 'credit' represents whatever combination of recognition, citations, reputation or perceived contribution the community attributes to each author based on their assessment of individual abilities and contributions. This framing requires only that scientists care—in some way—about how the community judges their work—a motivation extensively documented in studies of scientific behaviour. We do not mean to minimize the importance of what exactly scientists are motivated by. Existing literature has shown that whether scientists pursue truth, glory, priority or some weighted combination thereof shapes effort allocation by scientists, and this can have both good and ill effects on scientific production [19,28–32]. Our models remain agnostic on these deeper motivational questions. We need only assume that researchers have accurate expectations as to how the community assigns them credit, whatever that credit represents. If the community's credit assignment criteria systematically diverge from genuine scientific value—rewarding flashy results over rigorous replication, say, or prioritizing citation counts over conceptual breakthroughs—this introduces additional distortions beyond those we analyse here. For our conclusions about the optimal authorship norms for science, we merely rely on the assumption that collaborations are scientifically valuable, and that the *aggregate* reward for participating in collaborations is commensurate with the additional value that they bring.

Effectively, our model is simply a 2 × 2 coordination game, but with a distinctive twist: the exact pay-offs depend on the strategies the authors are perceived to be using. The scientific community's assessment of each contributor's work—and thus the credit allocated—hinges on their beliefs about which norms are in operation among scientists. For instance, if the community knows that the authors have settled on the I-Norm, then the community will simply give credit in accordance with their prior beliefs about the relative contributions of each author. If, on the other hand, the authors have adopted the C-Norm, then they will update their credit estimates, conditional on the observed order of the authors. And if the authors have not settled on a particular norm, but are playing mixed strategies, then the community will make the appropriate adjustments to hedge their credit assignments over the various possibilities.

Despite this complexity, we can identify clear preference rankings. Junior authors generally prefer the outcome where both adopt the I-Norm (I, I) over the scenario where both use the C-Norm (C, C). Senior

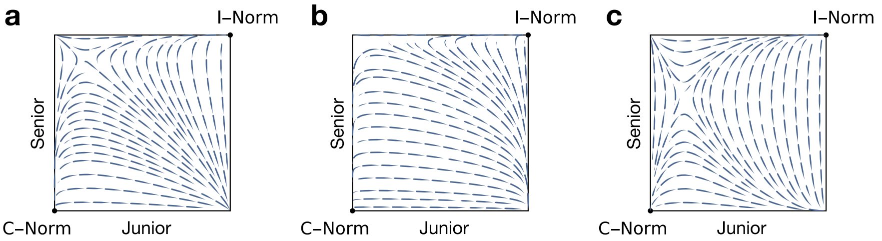

Figure 1. Evolution of authorship norms under different prior expectations of contribution. Phase-plane diagrams showing replicator dynamics for the two-population game. Each streamline indicates the trajectory of norm adoption, with comet heads pointing towards evolutionary outcomes. The horizontal axis represents the proportion of junior researchers adopting the I-Norm. The vertical axis shows the proportion of senior researchers adopting the I-Norm. Solid dots mark stable equilibria, located at opposite corners: bottom-left (all researchers follow C-Norm) and top-right (all researchers follow I-Norm). Community probability of incorrectly attributing credit to first author regardless of norm: $\varepsilon = 0.1$. Community probability of attributing credit to senior author regardless of norm: $\chi = 0.05$. Each stream plot is derived from a different underlying probability distribution for prior expectations about relative contributions of the researchers. (a) Symmetric distribution (equal chances for either to contribute more); (b) Junior-biased distribution; and (c) Senior-biased distribution. When seniors are expected to contribute more (c), the I-Norm's basin of attraction expands, making the senior-last convention more likely to emerge. Conversely, when juniors are expected to contribute more (b), the C-Norm's basin widens, favouring the greatest contributor first convention. With symmetric contribution expectations (a), neither norm has a decisive evolutionary advantage. This pattern remains consistent across various distribution shapes, and for a range of values of ε, χ. See the electronic supplementary material for illustrations.

authors, conversely, prefer (C, C) over (I, I). This preference divergence occurs not because of differences in the community's accurate credit assessment. In both equilibria, when the community correctly identifies the operative norm, authors receive credit matching their expected contributions. Rather, the conflict stems from our parameter ε, which captures first-author bias: under the I-Norm, Junior always enjoys this positional advantage, while under the C-Norm, Senior maximizes their opportunities to be listed first and capture this bonus.

Both authors have a common interest in avoiding the incompatible combination (C, I), where Senior insists on contribution-based ordering while Junior demands the I-Norm. This scenario alone prevents collaboration entirely, making it mutually disadvantageous compared with either coordinated outcome.

Using replicator dynamics to analyse the model, both C-Norms and I-Norms can emerge as stable equilibria, with no other stable equilibria existing under realistic conditions of first-author bias (see the electronic supplementary material for discussion of special cases where first-author bias is extraordinarily large). However, the likelihood of each norm becoming established depends on the distribution of contributions. In figure 1, we show a stream plot of the replicator dynamics for three different initial distributions of expected contributions by the authors.

The most striking finding from this first model is that the I-Norm (senior-author-last) is more likely to evolve when the senior author is expected to make a greater contribution than the junior (Figure 2). Conversely, when the junior author is expected to make the greater contribution, the C-Norm is more likely to evolve. This occurs because when one author is expected to make a greater contribution, they stand to lose more from a failed collaboration: this entails that they are subject to a greater selection pressure, and thus adapt their strategies more rapidly than the other author. For instance, when Senior faces higher stakes in successful collaboration, they evolve relatively fast towards strategies that maximize the probability of collaboration occurring at all—in this case, choosing the I-Norm, which is compatible with either of Junior's possible strategies. Conversely, when Junior faces higher stakes in successful collaboration, they evolve relatively fast towards the C-Norm, which is compatible with either of Senior's possible strategies.

This dynamic extends beyond asymmetries in expected contributions. The key determinant is which party faces greater evolutionary pressure in aggregate. While junior academics are typically in less secure employment, and this is a very salient reason to think of them as under greater adaptive pressure, this is ultimately an empirical question that may well vary across fields. Several considerations might counterbalance the employment security factor: (i) senior researchers typically manage a larger portfolio of simultaneous collaborations, exposing them to selection pressures across multiple fronts; (ii) each publication often carries heavier consequences for senior researchers, whose reputations, grant renewals and laboratory sustainability depend on consistent output; (iii) junior researchers paradoxically benefit from more viable outside options and higher mobility between career paths, potentially reducing

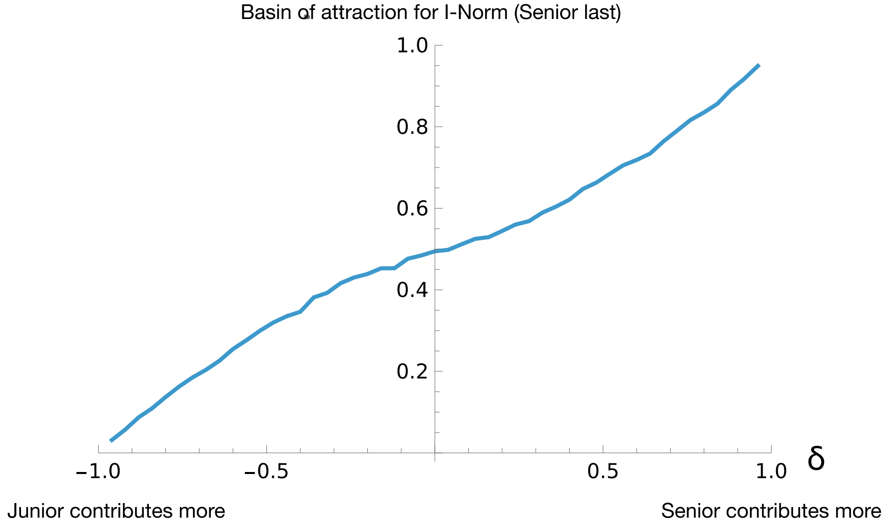

Figure 2. Basin of attraction for the I-Norm, as a function of difference in expected contributions between senior and junior author. $\varepsilon = 0.1$, $\chi = 0.05$. The qualitative result is robust for various levels of $\hat{c}$, ε, χ (electronic supplementary material).

their sensitivity to academic selection; and (iv) senior authors might be expected to make the greatest contribution to any collaboration, whether from accumulated expertise, a reputation for brilliance or the gravitational pull of the Matthew effect, and this makes the opportunity cost of a failed collaboration particularly costly for senior researchers. The relative strength of these countervailing forces presumably varies across disciplines and institutions, and their empirical weight remains an open question.

This evolutionary mechanism, through these channels or others, might explain some of the diversity of authorship norms across scientific disciplines. Fields where senior authors typically contribute more or collaborate at a higher rate should, according to our model, more frequently develop senior-author-last conventions. This prediction aligns with observed patterns in laboratory sciences where principal investigators oversee multiple research teams simultaneously, and where senior-author-last appears to be a more common norm. Conversely, fields where junior and senior researchers maintain more equal collaboration loads and contribution expectations should more frequently develop C-Norms.

This finding is an example of what evolutionary game theorists call the 'Red King effect' [33]. Contrary to the more familiar Red Queen dynamics where faster evolution is advantageous, the Red King effect describes situations where slower-evolving populations can achieve more favourable outcomes in mutualistic interactions. When Senior is expected to contribute more, they have more to lose from failed collaborations, creating stronger selective pressure on their strategies. This accelerated evolution paradoxically leads Senior to adopt the more accommodating I-Norm strategy that guarantees collaboration but sacrifices potentially deserved first-author credit. Junior, evolving more slowly due to lower stakes, maintains the C-Norm preference longer, eventually pulling the system towards that equilibrium. This aligns with Bruner's [34] findings on bargaining norm evolution, where faster-adapting groups often settle for less favourable divisions. Our model thus reveals how asymmetric evolutionary speeds—driven by differential stakes in scientific collaboration—can shape authorship conventions across disciplines, potentially explaining why fields with powerful senior researchers might counterintuitively develop norms that favour junior authors in author ordering.

All of our conclusions about this first model carry over unchanged if we reinterpret the I-Norm as simply alphabetical ordering, treating our two researchers, Junior and Senior as 'A' and 'B', respectively. In this guise, the replicator-dynamic equations and basins of attraction are identical: alphabetical ordering (the I-Norm) becomes more likely to evolve whenever authors with later-alphabet names face greater selective pressure than those earlier in the alphabet. However, unlike seniority, which correlates with systematic differences in collaboration patterns, career stakes and research portfolios, alphabetical position is essentially random and unlikely to create the persistent asymmetries in adaptive pressure that our model requires. Authors named 'Smith' or 'Zhang' do not systematically face higher collaboration costs or reputational risks than those named 'Ahmed' or 'Brown'. This suggests that while our framework demonstrates the theoretical possibility of alphabetical norms emerging through evolutionary dynamics, their actual adoption probably stems from different mechanisms—perhaps coordination convenience or historical accident—rather than strategic adaptation. What does persist, however, is the

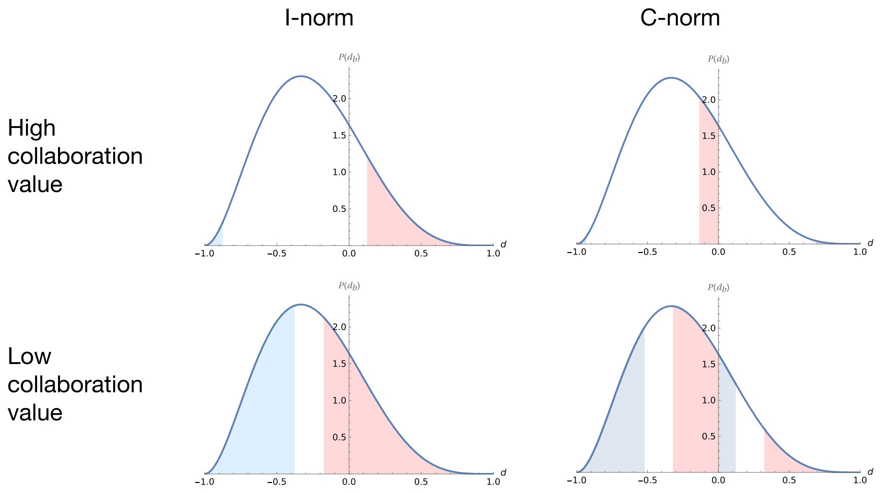

Figure 3. Collaboration failure regions under different authorship norms. The horizontal axis represents the ex-post difference in contributions between Junior and Senior ($d = c_S - c_J$), while coloured regions indicate where collaboration breaks down (red: Senior refuses; blue: Junior refuses). The bell-shaped curves illustrate probability distributions of contributions. Under the I-Norm (left panels), collaborations primarily fail at the extremes due to 'main contributor resentment'—when one author contributes substantially more than expected but receives insufficient credit. Under the C-Norm (right panels), failures occur both at extremes and in the middle zone of near-equal contributions, where 'second contributor resentment' emerges because the norm forces a stark credit distinction despite similar work. Upper panels show results with high collaboration value ($\hat{c} = 0.05$), while lower panels show low collaboration value ($\hat{c} = 0.01$).

key takeaway of the first model: both C-Norms and I-Norms are stable evolutionary equilibria, and so the C-Norm is far from an inevitable convention.

2.2. How norms affect collaboration: will they even work together?

Our second model shifts focus to the consequences of having either a C-Norm or an I-Norm already in place. We assume a norm is established in a field and want to know how these different norms affect researchers' willingness to collaborate.

Again considering Junior and Senior, they now know what the authorship norm in their field is. After collaboration, they will know exactly how much each contributed. The question is: will they choose to publish jointly or would one of them rather 'go it alone' and publish their part separately?

If there were no benefit to collaboration, then the decision to collaborate would be strictly zero sum: if one author gets assigned more credit by the community than their actual contribution, that author benefits from collaboration, while the other will lose, making collaboration inevitably fail. However, because we assume that collaboration improves the quality of the work—either by boosting visibility or by generating better ideas [5,6]—the game is non-zero sum and there is the potential for both authors to prefer collaboration, even though the community will almost inevitably give one author less credit than their actual share.

Credit is assigned by the same process as in model 1: the community has priors on the contributions of both authors, and also knows the norm in place. The community observes the quality of the paper and, conditional on its priors, assigns credit to each author.

Authors anticipate being judged for their contribution by the community and can thus estimate whether their pay-off will be greater if they 'go it alone' or if they collaborate. The collaboration will only proceed if both authors prefer that option. The important question, then, is when authors will refuse to collaborate under each norm.

Collaboration Under the I-Norm. With the I-Norm, collaborations will break down when one author contributes substantially more than was expected ('main contributor resentment'). This is because the I-Norm will cap the credit a scientist receives at whatever the community antecedently expected. If either scientist makes a surprisingly large contribution, the I-Norm cannot reflect this, and this can lead the author to prefer sole-authorship (figure 3, left column).

Collaboration Under the C-Norm. The C-Norm presents different challenges. C-Norms can discourage collaborations in two situations (figure 3, right column):

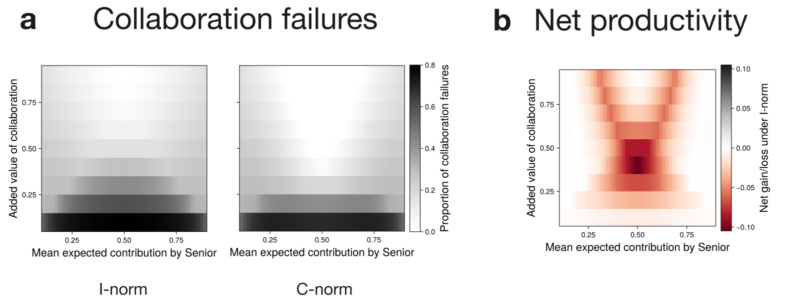

Figure 4. Comparative efficiency of C-Norms versus I-Norms. (a) Collaboration failure rates for the C-Norm and the I-Norm across parameter space. Darker regions indicate higher failure rates. (b) Net efficiency comparison between norms (difference between the norms in collaboration failures, weighted by collaboration value). Negative values (red) indicate the C-Norm is more efficient; positive values (grey) indicate the I-Norm is more efficient. The C-Norm demonstrates superior efficiency across most parameter combinations, with the I-Norm showing only marginal advantages in limited scenarios (faint grey regions). The underlying probability distributions are all beta distributions with $\alpha + \beta = 7$. The horizontal axis represents Senior's expected contribution, while the vertical axis shows collaboration value ($\hat{c}$). Additional distribution scenarios are illustrated in electronic supplementary material.

(1) *When contributions are very unequal*. Being first author sends only a very broad signal that may not give sufficient reward when one author has made an outsize contribution. If one person does almost all the work, they will receive more credit by publishing alone than they will receive from being first author under a C-Norm. This is similar to the way collaborations fail under the I-Norm ('main contributor resentment' again), but because the C-Norm sends a stronger signal of who contributed the most, it generally enables collaboration under a wider range of unequal contributions.

(2) *When contributions are very similar*. Under a C-Norm, if contributions are almost equal, whoever ends up listed second might be under-credited compared with what they could achieve by publishing independently ('second contributor resentment'). This occurs because the C-Norm forces a fixed distinction in credit, even when contributions are nearly equivalent.

In general, collaborations fail less often when the value of collaboration ($\hat{c}$) is high. This is straightforward to explain: when the benefits of cooperating are greater, it is easier to achieve cooperation, because it is more likely to be in the interests of both.

Comparing the Norms. Neither norm is perfect. Both can lead to collaboration failures. However, when comparing them across various contribution scenarios, we find that the I-Norm tends to lead to more collaboration failures and a greater loss of scientific value from collaboration. The C-Norm, while imperfect, generally fostered more successful collaborations in our model (figure 4). There are some cases where the I-Norm is superior—and this is generally where the collaborations are of moderate to high value, and the expected contributions are highly unequal. Unlike figure 3, which depicts failure regions as a function of realized contribution differences under a given distribution, figure 4 generalizes across ex ante distributions. When Senior is expected to contribute substantially more or less than Junior (horizontal axis extremes), both norms generate similar failure rates.

As with the first model, while our primary analysis has focused on the senior-author-last convention, these results generalize meaningfully to other I-Norms, most notably alphabetical ordering. Just as with the senior-last norm, alphabetical ordering fails to precisely communicate actual contributions, creating similar inefficiencies in collaboration decisions. When author order is determined alphabetically, researchers with greater contributions—regardless of their surname's position in the alphabet—may find their work undervalued and choose to publish independently rather than collaborate. This suggests that disciplines with alphabetical ordering conventions like economics, mathematics and theoretical computer science may experience collaboration failures similar to those in fields using senior-last conventions. Recent empirical work has documented strategic behaviours and publication patterns consistent with these predicted inefficiencies, including a greater reluctance to collaborate on the part of authors with surnames late in the alphabet [12,14]. The potential lost scientific value in fields using alphabetical norms could be significant, particularly for collaborative projects where the benefits of teamwork are modest. Just as our model predicts for senior-last norms, contribution-sensitive alternatives might foster more effective collaboration across disciplines currently relying on alphabetical conventions.

3. Discussion

Our analysis demonstrates two key findings about scientific authorship norms. First, both C-Norms and I-Norms can emerge as stable equilibria in scientific communities, with the likelihood of each norm becoming established affected by field-specific factors. Critically, we found that I-Norms like the senior-author-last convention are more likely to evolve when senior researchers face greater evolutionary pressure, such as when they engage in multiple simultaneous collaborations or when they have higher stakes for failed collaborations.

Second, and perhaps more importantly, we found that different authorship norms have significant implications for scientific collaboration. While both C-Norms and I-Norms can lead to collaboration failures under certain conditions, C-Norms generally foster more successful collaborations across a wider range of scenarios. When researchers know their actual contributions will be reflected in author ordering, they are more likely to pursue valuable joint work.

These findings have important implications for understanding scientific productivity and potentially for reforming scientific institutions. The norm dynamics we identify may help explain the puzzling diversity of authorship conventions across scientific disciplines. Moreover, our results suggest that fields employing I-Norms might be inadvertently discouraging collaborations that would produce valuable scientific knowledge.

3.1. Comparison with existing literature

Our evolutionary framework for analysing authorship norms provides several distinct contributions over previous approaches. Engers *et al.* [18] pioneered formal modelling of authorship conventions, examining a scenario where researchers negotiate author order after observing their contributions. Their model employs Nash bargaining to resolve authorship disputes and yields a striking conclusion: I-Norms (alphabetical ordering in their preferred interpretation) constitute stable equilibria, while C-Norms never achieve equilibrium status. This prediction, however, creates an immediate puzzle, given the widespread prevalence of C-Norms across numerous scientific disciplines.

This predictive limitation stems partly from their methodological choices. The axiomatic bargaining approach, while elegant, assumes complete information, simultaneous choice and perfectly rational actors operating with common knowledge of the bargaining structure. Scientific norm evolution, in contrast, emerges through iterative interactions among heterogeneous agents (from graduate students to senior researchers) with varying information access and bargaining power, operating in environments where conventions develop gradually through path-dependent processes rather than clean-slate negotiations. Our evolutionary dynamics framework better captures these messy realities of scientific practice, explaining how both norm types can emerge as stable equilibria depending on field-specific conditions and selection pressures.

The models also differ in their analytical focus. Engers *et al.* [18] endogenize effort decisions while treating collaboration as inevitable, investigating how different norms influence researchers' incentives to contribute optimally to joint work. We take the complementary approach of making contribution levels exogenous while endogenizing the collaboration decision itself. This shift in emphasis allows us to explore when valuable scientific partnerships form at all, rather than assuming they always materialize. Both perspectives offer valuable insights into different aspects of scientific practice—one addressing effort allocation within established collaborations, the other examining the upstream decision to collaborate in the first place.

Notably, these distinct approaches converge towards similar conclusions about norm efficiency. Engers *et al.* find that C-Norms provide superior incentives for optimal effort allocation compared with contribution-insensitive alternatives, because there is in effect a race for first author position. Our analysis reveals that I-Norms also create more frequent collaboration failures than their contribution-sensitive counterparts. This convergence from different analytical angles strengthens the case that C-Norms may better serve both individual researchers and scientific productivity as a whole.

Bikard *et al.* [5] offer another important comparison point, investigating how credit allocation shapes collaboration choices under fixed attribution rules. Their empirical finding that collaboration yields individual benefits primarily when credit is disproportionately allocated reinforces our theoretical insight that authorship norms function simultaneously as incentive mechanisms and information channels. Our work extends their analysis by endogenizing the conventions themselves, demonstrating how norms emerge as equilibrium outcomes of strategic interactions among researchers with divergent interests.

This evolutionary perspective adds theoretical foundation to empirical observations about scientific collaboration patterns, while suggesting new avenues for empirical research on how credit-allocation practices shape scientists' collaboration decisions across different fields.

3.2. Limitations and future directions

Our models have several limitations that might be addressed in future work.

First, our model of collaboration failures focuses specifically on credit attribution issues, but there are other reasons why scientific collaborations might fail. Coordination challenges, differing research priorities and resource constraints can all impede joint work. However, these factors can be conceptually incorporated into the benefit of collaboration parameter ($\hat{c}$). When these additional obstacles are significant, the effective value of collaboration decreases, potentially leading to the patterns of collaboration failure we identify.

Second, our analysis considers only two researchers and two possible norms. Extending the model to include larger collaborations and a broader range of potential norms would provide additional insights. One obvious consideration to include is that with growing team sizes, reaching agreement on relative contributions of all authors becomes increasingly difficult, favouring I-Norms as a less conflictual alternative [11]. This extension would be particularly valuable given empirical findings that authorship patterns become more complex with more authors [35], and that different disciplines exhibit distinct patterns of labour division across authorship positions [36]. Including contribution statements or equal contribution designations in our models could capture more nuanced aspects of modern scientific collaboration.

Third, while our model demonstrates the theoretical efficiency advantages of C-Norms, it does not address the practical challenges of implementing such norms in fields that have established conventions. Norm transitions involve coordination problems and potential resistance from researchers who benefit from existing arrangements.

One interesting suggestion from an anonymous reviewer is that we might model the senior–junior asymmetry as entailing additional factors that affect the collaboration. For instance, we might model senior authors as being more expert in updating their strategy to reflect the relative pay-offs (e.g. we might use the replicator–mutator dynamics, but with a lower mutation rate (error rate) for senior authors), or perhaps as better able to accurately estimate the relative contributions of the authors. While we have not undertaken such further analyses, we think it is plausible that they will have a general tendency to introduce more noise in the evolutionary process affecting junior authors, and that this in turn might increase the intensity of selection on senior authors, further reinforcing the very dynamic that favours a senior-author-last norm.

Finally, our model assumes the scientific community instantly updates its beliefs about prevailing norms—inferring author strategies and adjusting credit assignments in real time. Realistically, community beliefs might evolve on timescales comparable to the authors' strategic adaptation, introducing coevolutionary dynamics that could shift equilibria. Whether such lagged updating would stabilize or destabilize the norms we identify remains an open question for future work.

3.3. Implications for scientific practice

Our analysis suggests potential benefits from C-Norms, though translating these theoretical insights into practical recommendations requires careful consideration. While our models provide a coherent framework for understanding authorship conventions, they represent just one—very preliminary—approach to a complex social phenomenon. Any proposed reforms to established practices should be supported by corroborating evidence—both empirical studies across disciplines and complementary theoretical models—before we can begin to estimate the likely effects of interventions. The limitations of our approach highlight the importance of triangulating findings through multiple methodologies before advocating for systematic changes to scientific credit attribution systems.

In real scientific communities, senior figures derive authority and reputational capital from existing by-line rituals and are unlikely to cede status absent compelling incentives. Moreover, mandating detailed contribution statements risks devolving into perfunctory bureaucracy—boxes ticked for compliance, not clarity—unless accompanied by robust enforcement and accountability structures. Any credible effort to reform authorship conventions must therefore rest on realistic assumptions about strategic behaviour, clear protocols that deter token compliance and a foundation of evidence extending well beyond a single model.

3.4. Conclusion

Scientific authorship norms are not merely conventional practices but structural features of scientific communities that shape collaboration patterns and knowledge production. Our evolutionary approach to understanding these norms reveals how seemingly arbitrary conventions can arise from asymmetric selection pressures and how they subsequently affect scientific collaboration.

The efficiency advantages of C-Norms identified in our analysis suggest that scientific communities should critically examine their authorship practices. By aligning credit attribution more closely with actual contributions, fields might foster more valuable collaborations and ultimately enhance scientific progress. While changing established norms presents significant challenges, the potential benefits for scientific productivity make such efforts worthy of consideration.

Future work should extend our models to more complex collaboration scenarios, investigate empirical patterns of authorship and collaboration across different fields and explore practical interventions that might shift credit attribution practices towards more efficient arrangements. By treating authorship norms as evolving social conventions rather than fixed traditions, we open new possibilities for understanding and potentially improving the social structure of science.

4. Methods

4.1. Model 1: the evolution of authorship norms

To analyse how different authorship norms emerge, we developed a game-theoretic model with two researchers whom we call 'Junior' and 'Senior'. Before collaboration, each researcher proposes either a C-Norm, where author order reflects relative contribution, or an I-Norm, where Junior is always listed first regardless of contribution.

Collaboration proceeds if their norm proposals are compatible. Compatibility occurs when: (i) both choose the C-Norm; (ii) both choose the I-Norm; or (iii) Senior chooses the I-Norm while Junior chooses the C-Norm. In the last case, we assume they determine which norm to use by flipping a coin. The only incompatible scenario occurs when Junior insists on the I-Norm while Senior demands the C-Norm, leading to no collaboration.

If collaboration proceeds, the value of their joint paper is $1 + \hat{c}$, where $\hat{c} > 0$ represents the added value from collaboration. Nature determines each researcher's contribution, with Junior's contribution denoted as c_j and Senior's as $1 - c_j$. The probability that Junior is the greater contributor is $w_j \in (0, 1)$, with expected contribution $b_j \in (0.5, 1)$ if this is the case. Similarly, Senior's expected contribution when they contribute more is $b_s \in (0.5, 1)$.

If researchers fail to agree on a norm, they publish independently, with papers of value c_j and $1 - c_j$, respectively. In this case, the scientific community (modelled as an unbiased third party) assigns credit perfectly aligned with actual contributions.

When researchers collaborate, credit assignment becomes more complex. The scientific community observes author order and, knowing the researchers' strategies and the underlying probability distribution, estimates each researcher's contribution. We model a small probability $\varepsilon > 0$ that the community mistakenly gives all of the credit to the first author, regardless of the community's norms. We also include a small probability $\chi > 0$ that the community mistakenly gives all the credit to the senior author.

For example, under the I-Norm (Junior always first), if the community observes Junior listed first, they calculate the probability that Junior actually contributed more as

$$m_j = \frac{(1 - p_j(1 - p_s))w_j}{(1 - p_j(1 - p_s))w_j + (p_j p_s + 0.5(1 - p_j)p_s)(1 - w_j)},$$

where p_j and p_s are the probabilities that Junior and Senior play the I-Norm, respectively. Junior's expected credit is then $m_j b_j + (1 - m_j)(1 - b_s)$.

We analyse this game using replicator dynamics to model how strategies evolve in response to pay-offs. The replicator dynamics for our two-population game are

$$\dot{p}_j = p_j(1 - p_j)(\pi_j(\text{I-Norm}) - \pi_j(\text{C-Norm})),$$

$$\dot{p}_s = p_s(1 - p_s)(\pi_s(\text{I-Norm}) - \pi_s(\text{C-Norm})),$$

where π_j and π_s are the expected pay-offs for Junior and Senior when playing each strategy.

4.2. Model 2: collaboration decisions under established norms

Our second model examines how established norms affect researchers' willingness to collaborate. We assume a norm is already in place and researchers know their exact contributions before deciding whether to publish jointly or independently.

If they publish independently, each receives credit equal to their contribution (c_j and $1 - c_j$). If they collaborate, their joint paper has value $1 + \hat{c}$. The scientific community assigns credit based on author order and the established norm. Researchers collaborate only if both weakly prefer this option to publishing alone.

Under the I-Norm, collaboration fails when one researcher contributes substantially more than expected. This occurs when:

— Junior refuses if $c_j > \mu_j(1 + \hat{c})$, where μ_j is the community's prior expectation of Junior's contribution;
— Senior refuses if $1 - c_j > (1 - \mu_j)(1 + \hat{c})$.

Under the C-Norm, collaboration can fail in two distinct regions:

(1) When contributions are highly unequal: the main contributor refuses if they did substantially more work than would be signalled by first authorship.
(2) When contributions are nearly equal: the second-listed author refuses if their contribution is close to the first author's but they would receive disproportionately less credit.

For each norm, we calculate the ex ante probability of collaboration failure and the expected loss of scientific value from failed collaborations across different contribution distributions. This allows us to compare the efficiency of each norm, revealing that C-Norms generally foster more successful collaborations across a wider range of scenarios.

Our analyses include robustness checks with varying parameter values for $\hat{c}$, different probability distributions of contributions and different values of χ, ε. These extensions are detailed in the electronic supplementary material.

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. Data and relevant code for this research work are stored in GitHub [37] and have been archived within the Zenodo repository [38].

Supplementary material is available online [39].

Declaration of AI use. Generative AI was used in the preparation of this manuscript, chiefly for editing purposes. We note that conversations with generative language models, like conversations with colleagues, can clarify one's thinking without transferring authorship. The ideas, claims and any errors in this work are ours alone.

Authors' contributions. T.H.: conceptualization, formal analysis, investigation, methodology, visualization, writing—original draft, writing—review and editing; K.Z.: conceptualization, formal analysis, investigation, methodology, visualization, writing—original draft, writing—review and editing.

Both authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Authors' Notes. Authors may or may not be listed in alphabetical order.

Conflict of interest declaration. We declare we have no competing interests.

Funding. This study was supported by Australian Research Council, DP190100041.

References

1. Merton RK. 1974 *The sociology of science: theoretical and empirical investigations*. Chicago, IL, USA: University of Chicago Press.
2. Wuchty S, Jones BF, Uzzi B. 2007 The increasing dominance of teams in production of knowledge. *Science* **316**, 1036–1039. (doi:10.1126/science.1136099)
3. Larivière V, Gingras Y, Sugimoto CR, Tsou A. 2015 Team size matters: collaboration and scientific impact since 1900. *J. Assoc. Inf. Sci. Technol*. **66**, 1323–1332. (doi:10.1002/asi.23266)
4. Franceschet M. 2011 Collaboration in computer science: a network science approach. *J. Am. Soc. Inf. Sci. Technol*. **62**, 1992–2012. (doi:10.1002/asi.21614)
5. Bikard M, Murray F, Gans JS. 2015 Exploring trade-offs in the organization of scientific work: collaboration and scientific reward. *Manage. Sci*. **61**, 1473–1495. (doi:10.1287/mnsc.2014.2052)
6. Thelwall M, Kousha K, Abdoli M, Stuart E, Makita M, Wilson P, Levitt J. 2023 Why are coauthored academic articles more cited: higher quality or larger audience? *J. Assoc. Inf. Sci. Technol*. **74**, 791–810. (doi:10.1002/asi.24755)
7. Mongeon P, Smith E, Joyal B, Larivière V. 2017 The rise of the middle author: investigating collaboration and division of labor in biomedical research using partial alphabetical authorship. *PLoS One* **12**, e0184601. (doi:10.1371/journal.pone.0184601)

8. Waltman L. 2012 An empirical analysis of the use of alphabetical authorship in scientific publishing. *J. Informetr*. **6**, 700–711. (doi:10.1016/j.joi.2012.07.008)
9. Merton RK. 1968 The Matthew effect in science. *Science* **159**, 56–63. (doi:10.1126/science.159.3810.56)
10. Perc M. 2014 The Matthew effect in empirical data. *J. R. Soc. Interface* **11**, 20140378. (doi:10.1098/rsif.2014.0378)
11. Zuckerman HA. 1968 Patterns of name ordering among authors of scientific papers: a study of social symbolism and its ambiguity. *Am. J. Sociol*. **74**, 276–291. (doi:10.1086/224641)
12. Weber M. 2018 The effects of listing authors in alphabetical order: a review of the empirical evidence. *Res. Eval*. **27**, 238–245. (doi:10.1093/reseval/rvy008)
13. Maciejovsky B, Budescu DV, Ariely D. 2009 Research note—the researcher as a consumer of scientific publications: how do name-ordering conventions affect inferences about contribution credits? *Mark. Sci*. **28**, 589–598. (doi:10.1287/mksc.1080.0406)
14. Ong D, Chan HF, Torgler B, Yang YA. 2018 Collaboration incentives: endogenous selection into single and coauthorships by surname initial in economics and management. *J. Econ. Behav. Organ*. **147**, 41–57. (doi:10.2139/ssrn.2557365)
15. Efthyvoulou G. 2008 Alphabet economics: the link between names and reputation. *J. Socio-Econ*. **37**, 1266–1285. (doi:10.1016/j.socec.2007.12.005)
16. Einav L, Yariv L. 2006 What's in a surname? The effects of surname initials on academic success. *J. Econ. Perspect*. **20**, 175–187. (doi:10.1257/089533006776526085)
17. Van Praag CM, Van Praag BMS. 2008 The benefits of being economics professor A (rather than Z). *Economica* **75**, 782–796. (doi:10.2139/ssrn.977540)
18. Engers M, Gans JS, Grant S, King SP. 1999 First author conditions. *J. Polit. Econ*. **107**, 859–883. (doi:10.2139/ssrn.90728)
19. Kitcher P. 1993 *Advancement of science: science without legend, objectivity without illusions*. New York, NY, USA: Oxford University Press.
20. Fortunato S *et al*. 2018 Science of science. *Science* **359**, eaao0185. (doi:10.1126/science.aao0185)
21. Ioannidis JPA, Fanelli D, Dunne DD, Goodman SN. 2015 Meta-research: evaluation and improvement of research methods and practices. *PLoS Biol*. **13**, e1002264. (doi:10.1371/journal.pbio.1002264)
22. Foster JG, Rzhetsky A, Evans JA. 2015 Tradition and innovation in scientists' research strategies. *Am. Sociol. Rev*. **80**, 875–908. (doi:10.1177/0003122415601618)
23. Barabási AL, Wang D. 2021 *The science of science*. Cambridge, UK: Cambridge University Press. (doi:10.1017/9781108610834)
24. Jian D, Xiaoli T. 2013 Perceptions of author order versus contribution among researchers with different professional ranks and the potential of harmonic counts for encouraging ethical co-authorship practices. *Scientometrics* **96**, 277–295. (doi:10.1007/s11192-012-0905-4)
25. Marušić A, Bošnjak L, Jerončić A. 2011 A systematic review of research on the meaning, ethics and practices of authorship across scholarly disciplines. *PLoS One* **6**, e23477. (doi:10.1371/journal.pone.0023477)
26. Sarsons H. 2017 Recognition for group work: gender differences in academia. *Am. Econ. Rev*. **107**, 141–145. (doi:10.1257/aer.p20171126)
27. Barabási AL, Wang D. 2021 Scientific credit. In *The science of science*, pp. 134–146. Cambridge, UK: Cambridge University Press. (doi:10.1017/9781108610834.016)
28. Zollman KJS. 2018 The credit economy and the economic rationality of science. *J. Philos*. **115**, 5–33. (doi:10.5840/jphil201811511)
29. Strevens M. 2003 The role of the priority rule in science. *J. Philos*. **100**, 55–79. (doi:10.5840/jphil2003100224)
30. Strevens M. 2011 Economic approaches to understanding scientific norms. *Episteme* **8**, 184–200. (doi:10.3366/epi.2011.0015)
31. Kitcher P. 1990 The division of cognitive labor. *J. Philos*. **87**, 5–22. (doi:10.2307/2026796)
32. Dasgupta P, David PA. 1994 Toward a new economics of science. *Res. Policy* **23**, 487–521. (doi:10.1016/0048-7333(94)01002-1)
33. Bergstrom CT, Lachmann M. 2003 The Red King effect: when the slowest runner wins the coevolutionary race. *Proc. Natl Acad. Sci. USA* **100**, 593–598. (doi:10.1073/pnas.0134966100)
34. Bruner JP. 2019 Minority (dis)advantage in population games. *Synthese* **196**, 413–427. (doi:10.1007/s11229-017-1487-8)
35. Corrêa EA, Silva FN, Costa LF, Amancio DR. 2017 Patterns of authors contribution in scientific manuscripts. *J. Informetr*. **11**, 498–510. (doi:10.1016/j.joi.2017.03.003)
36. Larivière V, Desrochers N, Macaluso B, Mongeon P, Paul-Hus A, Sugimoto CR. 2016 Contributorship and division of labor in knowledge production. *Soc. Stud. Sci*. **46**, 417–435. (doi:10.1177/0306312716650046)
37. ghostleopold. 2025 author_order. GitHub. See https://github.com/ghostleopold/author_order.
38. ghostleopold. 2025 ghostleopold/author_order: public release (v1.0). Zenodo. (doi:10.5281/zenodo.17604290)
39. Handfield T, Zollman K. 2025 Supplementary material from: The Evolution of Scientific Credit: When Authorship Norms Impede Collaboration (doi:10.6084/m9.figshare.c.8210678)