A Robust Community-Based Credit System to Enhance Peer Review in Scientific Research

Wanpeng Tan

Department of Physics and Astronomy, University of Notre Dame

wtan@nd.edu

June 19, 2023

Abstract

Using an analogy with the capitalist economy, we examine the issues within modern basic science research as innovation drives both evolutionary cycles of the economy and research. In particular, we delve into the topics of peer review, academic monopolies and start-ups, the tenure system, and academic freedom in detail. To improve science research with a mature paradigm, a comprehensive solution is proposed, which involves implementing a credit system within a robust community structure for all scientists. Members can earn credit by contributing to the community through commenting, reviewing, and rating academic activities of submitted manuscripts, grant applications, and up to five achievements from each member. As members accumulate experience and credit, they can progress in their roles within the community, from commenter, reviewer, moderator, up to board member (serving in governing committees). High-achieving individuals are evaluated by the community for the quality, rather than the quantity, of their academic accomplishments. High-risk, high-reward projects from academic start-ups will be properly funded, and a healthy feedback and ecosystem will make the scientific community prosper in future innovative cycles in a self-sustaining way.

Contents

1 Evolution of Science

Galileo famously claimed that the book of nature is written in mathematics, and to take it further, the development of science can be viewed as the progression of natural philosophy from vague and descriptive ideas to rigorous, scientific fields such as physics, chemistry, biology, and so on. This process and its continued evolution may be best elaborated in Thomas Kuhn's great book, "*The Structure of Scientific Revolutions*". In an emergent science field, different ideas and theories compete with each other until one wins out, giving birth to the first paradigm. Its further progress rhymes in revolutionary cycles: starting with normal science research (jigsaw/crossword-puzzle-solving-like) within a paradigm, followed by the discovery of anomalies, eventually leading to a crisis, and finally resulting in a new paradigm by resolving the crisis. Each cycle is a revolution or paradigm shift.

Normal science is actually crucial as it pushes exploration towards the limits of the current paradigm by either examining at the precision limit or reaching the scope boundaries of the paradigm. Scientists often discover anomalies near these limits, triggering innovation in critical steps such as inventing tools with unprecedented precision and clarifying vague concepts within the paradigm. However, new ideas outside the paradigm must emerge to break the limits and present new dimensions beyond the paradigm, leading to a revolution or paradigm shift.

In light of such cycles, we can categorize good scientific achievements into four classes: normal incremental research (A_{+1}), normal innovation (A_{+2}), disruptive innovation (A_{+3}),

and revolutionary innovation (A_{+4}). While the latter two categories represent truly important innovative works that result in paradigm shifts, the first two likely make up the great majority of all research activities. Of course, there are also garbage works (A_0) and sometimes even detrimental ones (A_-). For the healthy advancement of science, we should promote A_+, especially high-level A_+ research, while striving to minimize A_0 and A_- works.

In this essay, we will focus on basic science with a mature paradigm, with examples and detailed discussions primarily limited to the field of physics, although similar arguments could also be extended to other fields in basic research.

2 Analogy to Capitalist Economy

It is revealing to learn about issues in scientific research from its analogy to capitalist economy. Innovation drives both cycles of capitalist economy and scientific research, as shown in Figure 1. Compared with the afore-mentioned cyclic revolutionary progress of science, here the cycles are presented from a financial perspective. In a free and open market, healthy competition has successfully nurtured innovation for the growth of a capitalist economy, at least at its nascent stage. However, due to the direct positive feedback, whoever wins out in the market could grow into giant corporations and even monopolies that inevitably block further innovation and competition for their own benefit.

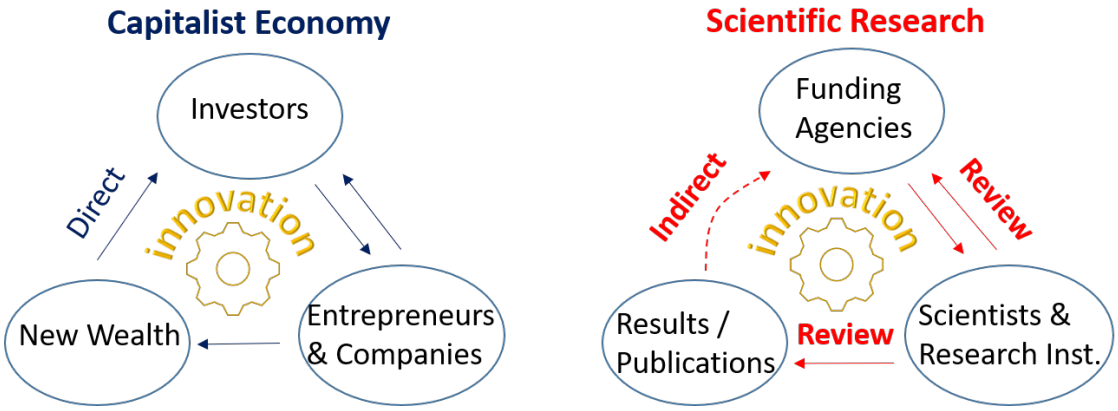

Figure 1: Evolution cycles of capitalist economy and scientific research.

Fortunately, the capital economy has addressed this issue by establishing antitrust laws and regulations to prevent monopolies. Start-up companies are protected for innovation and their growth is rapidly fostered by venture capital and angel investment. Some eventually drive the outburst of the next economic growth with disruptive innovation and become the new giants. Mature companies, especially large ones, may still be the backbone of the economy taking up most investments. However, start-ups, fostered with

roughly 5% of the total investment, are critical as the driving force for innovation and the next level of the economy.

Unfortunately, for research of science, especially basic science, the feedback is often indirect. For example, the benefits of a sponsored research project may not be seen or truly evaluated until many tens of years have passed. It makes the two peer/expert review processes for proposals and publications critical for agencies to make appropriate funding decisions.

The glaring issues in these review processes have long been recognized in science research. However, it is difficult to learn from other human societal activities as scientific research has unique characteristics. One outdated closed review system is still dominant in most fields, though some open yet unsatisfying review practices are emerging. This is arguably the biggest problem facing scientific research, which will be addressed later.

By analogy with the capitalist economy, we can identify more issues in scientific activities. For example, who are the start-ups in the academic world? Where are the venture investments for high-risk projects? Are there monopolies in science? If so, how can we continue promoting scientific innovation while preventing academic monopolies? Or in general, how can we ensure the healthy advancement of science in such funding cycles? Unfortunately, all these problems are closely interconnected and no simple solutions can tackle them separately. A comprehensive approach is needed to solve them as a whole.

2.1 arXiv's Monopoly and Planck's Principle

After hundreds of years of development, monopolies in modern science could emerge just like in a capitalist economy. Dominant scientists and research units may become the obstacles to new ideas as they control allocation of resources and promotion of up-and-coming young researchers. One example below illustrates the current situation.

After decades of ever-increasing dominance since its inception in 1991, arXiv.org has become the largest and most popular preprint archive or eprint service for scientific publications in physics and several other fields. It would have been the most beneficial to the community had arXiv adhered to its original principles for sharing new ideas and works quickly. Sadly, arXiv has increasingly been playing more of a gate-keeping role with obscure moderation and even veiled censorship. A much more lenient, yet much less influential archive viXra.org was established in 2007 as a counter measure for unorthodox articles rejected or unallowed by arXiv, many of which are no doubt crackpottery. But one may wonder about the actual benefit.

There are a total of 40,587 articles posted in viXra in contrast to 1,850,470 posted in arXiv during the same period. The total rejection rate is merely 2%. Physics, especially in subfields such as high-energy particle physics (HEP) and cosmology, is considered to be one of the most attractive fields to crackpots. Even with diligent effort of moderation,

undesired submissions still get through from time to time, and if caught later, the papers would be reclassified into the infamous "crackpot" category – *physics.gen-ph*. What is the reclassification or rejection rate within arXiv for such a field prone to crackpottery?

arXiv	category	hep-th	hep-ph	gr-qc	astro-ph.CO	sum	physics.gen-ph
	article#	1798	1774	1701	1035	6308	77

viXra	category	HEP	Quantum Gravity/Sting Theory	Relativity/Cosmology	sum
	article#	34	39	71	144

Table 1: Numbers of articles posted this year (before 4/16/23) in major categories related to HEP and Cosmology in both eprint archives of arXiv.org and viXra.org

Table 1 appears to show some interesting statistics. Assuming that all articles from *physics.gen-ph* were treated as potential crackpot papers and reclassified from one category (say, *hep-th*), we obtain a reclassification rate of $77/1798 \sim 4\%$. Assuming that they were reclassified from four major HEP and Cosmology categories (*hep-th, hep-ph, gr-qc, astro-ph.CO*), we get a much lower rate of $77/6308 \sim 1\%$. Considering that some articles are cross-listed in multiple categories while there may be articles that genuinely belong to *physics.gen-th*, the actual disapproval rate is probably somewhere in between 1% and 4%. Nevertheless, this is still a very low rate. Similar articles, taken by viXra.org but typically rejected or not allowed without endorsement by arXiv.org, have a similar low rejection rate of $144/6308 \sim 2\%$, which could be even lower as some papers were posted in both archives, possibly as a way to protest or due to arXiv's moderation decay.

It is clear that arXiv's gate-keeping policy is not very efficient: in order to eliminate a small percentage of potential crackpot works, some articles with genuinely disruptive ideas could get thrown out as well. An high-profile example of arXiv over-moderation was reported in the Nature news article "*ArXiv rejections lead to spat over screening process*".

The monopolistic practices of arXiv are just a reflection of the more general Planck's principle in the sociology of scientific knowledge, named after Max Planck, one of the best-known physicists in early 20th century. He once said,

> *"A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it. "*

Unfortunately, Planck's statement has too often been verified in the relatively short history of science. For example, in a National Bureau of Economic Research article, "*Does Science Advance One Funeral at a Time?*", the authors quantitatively explored how established star scientists shape the vitality of new ideas in their fields by examining what happens to the fields when dominating scientists pass away prematurely.

It seems that established scientists may be more resistant to new ideas and may never accept them during their lifetime. However, their persistence may be what initially helped them become a dominating figure in their field. This is why many scientific discoveries (and social revolutions) are often led by young people, before they become well-established themselves. Before the invention of written language, human societies developed very slowly, as brilliant ideas could be stifled by authoritative elders and easily lost between generations. As a result, rediscovering ideas or reinventing the wheel was a persistent phenomenon throughout ancient human history.

Nowadays, all scientific advancements are well-documented, including even much more peculiar ideas. However, the organization of the science community remains authoritative in nature and lacks democracy and diversity for high-risk, high-gain ideas. As a result, truly innovative ideas are still difficult to thrive and can easily be buried in the exponentially growing literature.

Like giant corporations in the economy, science has its elite circles of authority that dominate ideas and resources. Unlike in the business world, we don't have real support for scientific "start-ups". Disruptive ideas, therefore, are hard to find suitable soil in science to germinate and grow. To combat such monopolistic phenomena in science research, we need to apply open science principles, truly support diversity in ideas and projects (especially risky ones), and most importantly establish a more open and democratic community structure involving all scientists and researchers.

2.2 Research Start-ups and High-Risk Investments

At first sight, government funding agencies and private foundations may seem to offer similar funding programs that claim to support high-risk, high-reward projects. They may even appear to serve similar roles as venture capitalists and angel investors. However, in reality, there are no true scientific start-ups receiving such support. Additionally, these organizations lack the knowledge or expertise to do so effectively even if they have the chance to support high-risk projects.

The scientific community seems to be very good at forming giant and medium-sized research groups with the current funding/tenure system. For example, in physics, we see the formation of giant collaborations (up to several thousand scientists) at large facilities like LHC and LIGO. We also see the formation of numerous medium-sized multiple-PI (Principal Investigator) groups in scientific research. These two categories take up most of the resources, and are leading the research in hot-topic frontiers and also responsible in training the next generation mainstream researchers.

In scientific enterprises, there is currently no mechanism in place for cultivating true scientific start-ups, as is the case in the business world. A true start-up in science needs to be independent from the manipulation of large groups and requires long term support

from venture investments in order to survive and grow.

At first glance, a new starting PI's group should play the role of a start-up, or tenure-track professors in universities should. Unfortunately, it doesn't match reality. In their early career, they are under an enormous amount of pressure to get tenured within a limited time frame, typically six years plus possibly many years spent as a postdoc, which is very common in physics. Therefore, they are most likely to take a safe route of pursuing hot-topic research and following more authoritative figures' footsteps in their field. Under such an ordeal, they are no longer able to do any truly disruptive research once they become tenured or truly independent. Those who dare to pursue a different path are often eliminated in the process.

While some may argue that postdoctoral positions could serve as a start-up role in the academic world, the reality is that few of these positions are truly independent, and many are associated with large research groups. In addition, the terms are typically too short (1-3 years). Much fewer tenure-track professorships offer a better position for scientific start-ups, but still pressured too much as mentioned above. After 10-20 years of conformity to mainstream research, eventually tenured scientists may no longer be in their prime for conducting transformative research. They may have become accustomed to safe approaches and even obstructive to the acceptance of disruptive ideas by next-generation researchers.

So where can we find academic start-ups, and how can we support and foster them? One potential solution is to make most postdoctoral researchers independent with long term support (possibly starting with a low-paying position lasting from 5-10 years, renewable, or maybe even indefinitely). In colleges, they could shoulder most of the teaching duties to sponsor their positions. In research laboratories, they could oversee various small-scale research projects or maintain the operation of scientific equipment and the laboratory. In addition, they should be provided with ample opportunities to collaborate freely with each other and pursue or participate in larger-scale projects.

However, the key issue lies in developing a structure or mechanism that allows for the academic start-ups and their disruptive ideas to ascend to the top of their field in a sustainable way, rather than just once.

2.3 Tenure System and Academic Freedom

Just as the survival of a life species in evolution depends on its diversity to meet volatile challenges along the evolutionary path, the prosperity of science also invokes the diversity principle for the sake of innovation. We can't support only mainstream research, and instead we need to protect the rights and freedom of scientific minorities. In the end, it is all about academic freedom and the free expressions of different ideas and views, especially unorthodox ones that might be the disruptive force for the next scientific revolution. In

essence, this is the same spirit of free market and competition that we pursue in a capitalist economy.

The tenure systems adopted in colleges are meant to protect academic freedom. However, the issue is that tenured professors who can enjoy such protection are typically at least ten to twenty years into their careers, especially in basic science fields like physics in North America. Are we content with the situation where academic freedom is more protected for older tenured professors than other researchers, especially young untenured ones? Shouldn't academic freedom be protected universally for all researchers regardless of their status, age, and prestige?

Who are in more urgent need of protection for academic freedom? Aren't scientists more innovative at their younger age? Do the young, untenured scientists with unorthodox ideas feel unsafe or afraid of losing their academic freedom when facing intolerant academic administrators who, ironically, are often tenured professors?

To make things worse, tenure is typically determined within one academic unit, often by very few tenured influential figures, taking in consideration many factors other than simply the candidate's academic achievements. Modern science has become increasingly specialized today, and few experts within a single department can truly evaluate a candidate's achievements. In such small elite circles, politics and other unspoken factors may play a much bigger role.

Ideally, it is better to rely on an award system where an individual's scholarly achievements are judged by all peers/experts in a given field. In such an award system, the main emphasis should be on quality rather than quantity of one's achievements. That is, no more than a certain number of (say, five) achievements from one individual should be submitted for evaluation. Better yet, the academic community, instead of individual institutions, should take charge in the evaluation process.

The "Chicago Trifecta", which is based on the principles of freedom of expression, institutional neutrality, and academic achievements as the basis for hiring and promotion, should be adopted to protect academic freedom for all researchers. Universities may assess the quality of additional performance in teaching and service in their hiring and promotion decisions, but the academic achievements of a candidate should be assessed by the academic community instead of each individual institution.

3 Peer/Expert Review

The afore-mentioned over-moderation at arXiv also shows an example of poor review aggravated with obscurity, bias, prejudice, and gate-keeping. One immediate reaction would be that we should apply the principles of open peer review, which is actually a quite complicated concept involving transparency in various aspects of the process. In addition, we

must consider two different kinds of review processes: one is for reviewing manuscripts/ publications as a result of a research project; the other is for refereeing proposals for funding support before a project starts.

First, let's look at the issues in the review of publications. The knowledge explosion in modern times has caused the number of publications to grow exponentially, but too much of this growth has been of little value, largely due to the "publish or perish" pressure in the academic world. As a result, it is hard even for prestigious journals to find enough reviewers in time, which often results in long decays of several months up to even years before publication. Worse yet, authors would typically keep shopping for publication in other journals after their articles are rejected by one journal, and exhausted review resources would be wasted again and again as most journals do not share their review results.

There's no real incentive for reviewers to carry out such burdensome works. Reviewers are seldom paid and yet do the most important works for journal publication, while all the profits from either subscription paywalls or article processing charges go to the publishers, which benefits neither the public nor the research community. An article titled "*The rise and fall of peer review*" argued about the failure of the current peer review system, particularly its inability to prevent fraudulent papers from being published.

When it comes to openness in peer review, it may seem natural to assume that anonymity provides referees with a sense of safety, allowing them to make biased/inappropriate comments, or follow their own agendas. Conversely, open identities may cause referees to feel uncomfortable about writing critical or negative remarks. While this may be true in some cases, in reality, just like the arXiv example discussed above, gate-keeping for orthodoxy can be a stronger factor for referees, particularly when evaluating unorthodox works. In open review practices, it indeed occurs as follows: referees who give positive and constructive feedback choose to remain anonymous, while those who make negative or inappropriate comments choose to sign their reports when reviewing the same unorthodox manuscript. This behavior can be explained by referees' fear of pressure from main-stream peers, rather than retaliation from the authors, unless the authors are much more influential figures in their fields.

In recent years, many journals and publishers, especially in the life sciences, have started to adopt some type of open peer review including *Nature*, *MDPI*, *PLOS*, and many others. In late 2022, *eLife* became the first journal to change its peer review model by removing the "gate-keeping" function for publication.

Several open review platforms for preprints have emerged since 2019 as well: *PREreview*, *Qeios*, *ScienceOpen*. Such practices have evolved very rapidly in the life sciences possibly due to urgent needs since the COVID-19 pandemic started. Although arXiv started much earlier and has become much more dominant, bioRxiv (started in 2013) and medRxiv (started in 2019) have progressed much better, in particular, by incorporating the comment/review system. The independent service *Review Commons* has even started a

central open review platform for preprints and journals in the life sciences.

It is noteworthy that two statisticians, Harry Crane and Ryan Martin, have founded a new type of decentralized research platform *Researchers.One* for peer review and scholarly publication. In particular, their founding mission statement paper published in 2018 also presents a very comprehensive review of the issues surrounding today's peer review. Readers are recommended to refer to this paper and the references therein for more detailed discussions regarding peer review.

Unfortunately, all of these practices and services have not really attracted sufficient review activities, nor have they dramatically reduced the number of meaningless or fraudulent works. The problems may largely be due to a lack of appropriate incentives or rewarding systems for reviewers. A solution may lie in making amends for the distinct separation and conflict between different roles of author, reviewer, editor, etc., played by a researcher.

Secondly, peer review in research grant proposals shares similar problems but also has its own glaring issues. Both review processes favor mainstream research, but due to intense competition for funding, especially from major federal funding agencies, a proposal would hardly stand a chance of being funded if it does not receive excellent reports from all (typically five) expert reviewers. As such, so-called high-risk, high-reward projects have nearly no chance of being funded in most programs. In contrast to peer review for publications, peer review for grant proposals is still mostly closed.

Federal funding agencies are indeed risk-averse, though not completely due to their intentions. These agencies have acknowledged their shortcomings and initiated separate programs to support high-risk projects such as NSF's EAGER and DOE's USP. However, despite these efforts, the situation has not significantly changed due to the limited scales and practices of these programs. In contrast, private foundations are more inclined to fund high-risk high-reward projects, but their actual practice often involves similar measures used for funding mainstream or low-risk projects.

The problem is that neither program officers nor experts in the same field can independently or even jointly decide which high-risk high-reward projects to fund without concerns of bias and fairness. A completely revamped review procedure may be necessary to address this issue.

4 A Proposed Solution

A tentative, comprehensive solution to the aforementioned issues in basic science research is proposed as follows. This solution is not yet mature and requires concrete implementations through trial and error in the real world to improve it. Therefore, the proposed quantitative measures and schemes are primarily presented as examples, and

the details are still subject to debate.

4.1 Principled Considerations

Scientific contributions can be roughly divided into two distinct types: one is original research that directly contributes to the advancement of science; and the other is the evaluation of research results (including the achievements of individual researchers) that determines the direction of scientific research. While these two types of contributions are strongly correlated, they do not always overlap exactly. Therefore, it is necessary to have two systems to quantify both types of contributions. For original research and contributors of original ideas, an achievement level system that emphasizes quality rather than quantity would be appropriate. For evaluation-related service efforts, a quantitative credit/role system would be the most suitable.

There are several well-known quantitative metrics for scientific publications. For example, bibliometric indicators such as CiteScore from Scopus, SJR from SCImago, and Journal Impact Factor (IF) from Clarivate, focus on evaluating the quality of journals in general. However, these metrics are too simplistic and naive to be used to evaluate individual papers. Even citation metrics at the article level are too crude and prone to manipulation. A more genuine approach would involve a properly-designed, more thorough, and more sophisticated peer review system with quantitative ratings from the entire scientific community, as it provides the best means of evaluating individual papers.

Author-level citation metrics like the H-index, while widely used, are also too coarse to evaluate individual achievements. They can be easily skewed by the volume of publications or the size of collaborations, and they are also susceptible to gaming the system. The collective review of the entire community, accompanied by quantitative ratings, probably remains the best approach for evaluating individual achievements.

To ensure the most effective peer review processes that are so critical to scientific research, we need to consider appropriate incentives. Perhaps, the most effective incentive for working scientists to engage more in high-quality peer review is rewarding recognition and increased roles within the scientific community. A quantitative credit/role system can best account for such a mechanism.

Such a quantitative solution should consider measures to address the issues of gaming the system. In particular, it should take into account the dynamic nature of all evolutionary systems by implementing dynamic calibration and continuous tuning procedures for the recalculation of ratings and credits. Much like the ever-improving progress of science itself, this constantly updating feature will ensure the most desirable behaviors in scientific publication and peer review.

In essence, a complete and comprehensive solution would require a revolution in the entire structure of scientific research. Based on the above analysis and discussions in

previous sections, it becomes clear what characteristics such a revolutionary solution requires:

- **Principle of Democracy** – An all-participating community is for all properly trained researchers.
- **Principle of Diversity** – Scientific start-ups and high-risk high-reward projects must be adequately funded.
- **Quantitative Credit System** – Rigorous science requires a rigorous credit system for scientific evaluation.
- **Quality-Based Achievement Rewards** – Quality, rather than quantity, should be emphasized in rewards for individual achievement.
- **Contribution-Based Role System** – Members' service contributions to the community should determine their roles in the community.
- **Healthy Self-Sustaining Ecosystem** – If all the above requirements are met, this is automatic.

First and foremost, we need to establish a robust community structure for all members with proper scientific trainings in a given scientific field. Each registered member should be identifiable with a unique research ID for connecting researchers with their works, such as ORCID, which is the most widely adopted one today. A large e-print service like arXiv.org, which already has the largest user base, would be the best starting point for physics.

4.2 Quantitative Credit System

The most critical element for such a community structure to be healthy and sustainable is a well-designed credit system. All members can participate, gain credit, and hence play an ever-increasing role in the evaluation of three different activities in the community: preprints/publications, grant proposals, and individual achievements. Initially, a new member starts with a certain amount of credit (e.g., 10 points) and can earn credit points as a commenter with informal comments. Every member can post a preprint as an author if holding enough credit points (say, 10 points per manuscript), which can be earned by commenting, reviewing, and rating activities of other members. In other words, the credit system encourages and requires that members contribute to the community.

The format of preprints, comments, and reviews can be anonymous or open-ID for maximum flexibility, in particular, in early transition to the new system. To encourage openness, we could double the earned points for members who choose the open format. We hope that it will become mostly open in the end once the healthy cycle starts. Members can post preprints and informal comments in any subfield/category they wish. But reviewing and rating activities require a higher role such as reviewer and above, which must be acquired in the appropriate subfield.

Once members gain enough experience and credit (e.g., posted ≥ 3 solid or higher-

quality papers, with > 1 yr membership, and holding > 100 credit points) in a subfield, they will be promoted to the rank of reviewer who can earn more points by contributing or getting invited to rate and/or write official review reports about a preprint in that subfield if there is no conflict of interest. The next step in the role would be moderators (e.g., ≥ 10 papers, > 5 yr, and > 1000 points) who can invite reviewers and coordinate other efforts while earning even more points. Publishers may hire some of the moderators as academic editors for publishing some of the highly-rated preprints in overlay journals.

Reviewers can rate a preprint with a score A in the range $-1.0 \leq A \leq 5.0$. Assigning $A = -1.0$ means that the preprint is fraudulent, plagiaristic, or otherwise extremely detrimental to science. Assigning $A = 0.0$ is reserved for completely useless papers. $A \geq 1.0$ is for solid incremental pieces of work which should represent most of the publications in today's academic world. $A \geq 2.0$ is for normal innovative works, or more pragmatically, top papers on a given topic in recent years, typically worthy of publication in top journals. $A \geq 3.0$ is for disruptive studies, or top papers in a subfield, typically worthy of good prizes within the subfield. $A \geq 4.0$ for revolutionary works, or top papers in a field, typically worthy of the top prizes of the field. The average score $\bar{A}$ for a given preprint will only show up when the number of ratings N reaches a certain threshold (e.g., $N \geq 5$). The general public can only see preprints with $\bar{A} \geq 1.0$. While non-reviewer members (typically students) can see preprints with $\bar{A} > 0.0$ or A_+ and unrated preprints (i.e., $N < 5$), the rest are visible only to reviewers and above.

When members post an informal comment, write an official review, rate the manuscript with a score of A, or rank a comment/review with a score of S ($-1.0 \leq S \leq 5.0$), as the N-th contributor in order, they will receive credit points as follows,

$$\text{Earned Credit Points} = f_a(\bar{A})/f_{eb}(N) \times \text{SCORE} \tag{1}$$

where the attention factor $f_a(\bar{A}) = 2^{|\bar{A}|}$ is designed to exponentially attract more activities for higher quality works; the early bird factor $f_{eb}(N) = 1 + 2^{N-5}$ for writing a review is intended to entice the first five reviews and suppress too many reviews after about ten, or $f_{eb}(N) = \sqrt{N}$ for other activities is meant to encourage early contributions. For a comment, SCORE = $\bar{S}$, is the average rating score it has received; for an invited review, SCORE = $20 \times \bar{S}$ (inviting moderator may receive one fifth of it); for a contributed review, SCORE = $10 \times \bar{S}$; for rating the preprint, SCORE = $3 - 2 \times |A - \bar{A}|$; for rating a review/comment, SCORE = $(3 - 2 \times |S - \bar{S}|) \times 2^{\bar{S}-4}$, which encourages more ratings on better reviews/ comments. Note that both $\bar{A}$ and $\bar{S}$ only appear when N is above the threshold (e.g., $N \geq 5$). Therefore, earned points may be credited later or not at all, and it could fluctuate as $\bar{A}$ and $\bar{S}$ vary over time.

The main merits of the credit system are summarized as follows:

- **Early-Bird Encouragement** – earlier contributions are credited with more points.

- **High-Quality Attention** – more activities are attracted to higher-quality papers/reviews/comments.
- **Robust Against Gaming** – careless or irresponsible behaviors are hard to gain credit and may result in losing it instead.
- **All Member Participation** – democracy and diversity are ensured by the participation of the entire community.
- **Rewarding Positive Activities** – a reliable role-increasing mechanism is integrated into the self-regulating ecosystem.

4.3 Funding High-Risk Projects

Now we turn to grant proposals. A proposal can be submitted by a member who is in good credit standing (e.g., 50 points per proposal). Criteria for grant reviewers/moderators may be set higher than those for preprint reviewers/moderators. Program officers from external funding agencies may hire or consult with moderators to get proposals reviewed. The usual review approach is sufficient for funding main-stream research projects.

However, a different approach must be adopted for the review of high-risk, high-reward projects in dedicated funding programs (which should ideally represent 5% of the total investment). Essentially, minimum scientific standards should be applied, such as requiring 1-3 positive consultative/applicant-selected reviews and/or 1-3 highly-rated ($\bar{A} > 2$) relevant papers. Concrete numbers depend on how much risk a funder is willing to take. Varied opinions of other reviews from randomly-selected experts in the same subfield could also indicate the level of risks involved. If possible, the funder should include non-specialists from adjacent subfields and/or even completely different fields to evaluate the proposal's potential.

The reason minimum scientific standards are necessary is that transformative ideas, in their nascent forms or times, are often misidentified as pseudoscience by mainstream scientists, sometimes even by the overwhelming majority of scientists. According to Kuhn, experts who are fully immersed in the old paradigm are typically the fiercest critics of the emerging new paradigm. Therefore, few experts would support truly paradigm-shifting proposals in their own field. On the other hand, minimum scientific standards must be applied to exclude the competition of pseudoscience. The critical point is that we should not have a blanket exclusion of all fringe science ideas.

The biggest pitfall to avoid is inadvertently supporting mostly low-risk projects. If most randomly-selected experts give excellent ratings, there is near-consensus support from the community in the specialized field, and/or there are a huge number of citations in relevant publications, then such projects should not be considered for funding programs designed to support high-risk efforts.

In assessing the high-reward factor, funders cannot rely, at least not entirely, on the

positive reviews of experts selected by the applicant. However, funders cannot rely on the opinions of randomly-selected reviewers either, as they are also likely to be biased in favor of the old paradigm. As one can imagine, the most unbiased reviews on the impact factor are likely to come from non-expert scientists in other fields. The best option would be scientists from immediately adjacent (sub)fields where the proposed new paradigm does not change much. These non-expert scientists may not be able to evaluate the technical details of the proposal, but they can probably tell how impactful it will be if successful.

The final factor to consider is testability. A testable idea could have an immediate impact after the successful execution of the project. In short, funding programs that aim to support high risk, high reward proposals should set a low bar for scientific standards to filter out pseudoscience projects, and then focus on funding those highly-testable projects with great potential.

4.4 Achievement Level System

Level	≈Position	Achievements	Funding	≈Role
L0	Student	N/A	N/A	commenters
L1	Postdoc	A_{+1}	$\leq$\$3k/yr	reviewers
L2.1	Fellow	$1 \times A_{+2}$	$\leq$\$10k/yr	
L2.2	Assist. Prof.	$2 \times A_{+2}$	$\leq$\$15k/yr	moderators
L2.3	Assoc. Prof.	$3 \times A_{+2}$	$\leq$\$20k/yr	leaders
L3	Prof.	A_{+3}	$\leq$\$50k/yr	
L4	Chair Prof.	A_{+4}	$\leq$\$250k/yr	

Table 2: Achievement class levels (L0-L4) of researchers are aligned with suggested basic annual funding levels, and roughly matched with their positions and roles in the community.

Lastly, the most difficult is to replace tenure with an achievement class level system as shown in Table 2. Each member of the community can submit up to five scholarly achievements for evaluation, and such a limit will greatly reduce meaningless works that are prevalent today. Again, good credit standing is required (e.g., 100 points per achievement submission). High-level achievements with the rating ranges shown in Table 3 may also be considered significant contributions to the community and therefore worth credits to members. For example, each A_{+2} achievement is worth 100 points, each A_{+3} is 500 points, and each A_{+4} is 2500 points.

If a submitted achievement is related only to one single preprint/publication, then its evaluation is simple and straightforward as it is determined by the average rating $\bar{A}$ of that paper. If the submission synthesizes multiple papers into a systematic study for evaluation

as a whole, then it will be reviewed for eligibility of one level above the highest rating of the papers (e.g., from A_{+2} to A_{+3}). Achievement reviewers can only rate and review achievements with a target level at or below their own level (e.g., an L2 reviewer can review for A_{+2}-level achievements but not A_{+3}).

L1 and above members can apply for basic support from government funding agencies as shown in Table 2 if they remain in academia. Their possible positions in a university and roles in the community are roughly matched in Table 2. Note that such basic funding support is intended to give them some degree of independence and protect their academic freedom, e.g., as seed funding for high-risk or free exploratory research. In addition, they can also apply for larger project-based funding as discussed above. However, their salaries will continue to be paid through the teaching, research, and service they perform in their positions. The new level system will relieve individual institutions of the burden of achievement evaluation in their hiring and promotion decisions, which the community can certainly do much better.

The new achievement reward system does not mean that we will immediately abandon the tenure system. It could work in parallel, at least initially. Once most researchers have been awarded the appropriate L1-4 levels, community evaluation will eventually replace individual institution evaluation. Positions and the tenure system will gradually be either replaced or tied to the achievement levels of scholars and their roles in the community.

Moderators who are among the top credit holders in their class and subfield may assume the highest leadership role and volunteer to serve as board members on various committees. Besides possible representatives from other agencies, the leadership team should include equal numbers of L2, L3, and L4 moderators and be representative of all subfields to ensure diversity. Each position will be re-selected from other top credit holders every four years to avoid bias.

In summary, the major advantages of this solution are: it is driven by the entire scientific community, rather than by elite circles or monopolies; the credit system encourages healthy role-playing and positive feedback in a self-regulating ecosystem, which operates in a self-sustaining way; limiting the number of achievements an individual can submit for evaluation to five significantly reduces the production of low-quality works, resulting in a much cleaner field for all scientists; and high-risk, high-reward projects from academic start-ups can finally receive proper funding.

Notes: many of the pieces and ideas discussed in this essay can be found in their prototypical versions in https://www.wanpengtan.com/category/open-science/.

Detrimental	Useless	Positive	Incremental	Innovative	Disruptive	Revolutionary
A_-	A_0	A_+	A_{+1}	A_{+2}	A_{+3}	A_{+4}
$-1 \le \bar{A} < 0$	$\bar{A} = 0$	$0 < \bar{A} \le 5$	$1 \le \bar{A} < 2$	$2 \le \bar{A} < 3$	$3 \le \bar{A} < 4$	$4 \le \bar{A} < 5$

Table 3: Achievement levels are shown with their corresponding score ranges.

Appendix A: More Details of the Proposed Solution

A.1 Earned Credit Points

A more elaborate formula for Earned Credit Points can be written as,

$$f_o f_i \frac{f_a(\bar{A})}{f_{eb}(N)} \times \text{SCORE} = f_o f_i \times \begin{cases} \dfrac{2^{|\bar{A}|}}{1+2^{N-5}} \times \bar{S} \times \begin{cases} 10, & \text{for contributed review.} \\ 20, & \text{for invited review.} \\ 20/5, & \text{for inviting moderator.} \end{cases} \\ \dfrac{2^{|\bar{A}|}}{\sqrt{N}} \times \bar{S}, & \text{for informal comment.} \\ \dfrac{2^{|\bar{A}|}}{\sqrt{N}} \times (3 - 2|A - \bar{A}|), & \text{for rating manuscript.} \\ \dfrac{2^{|\bar{A}|}}{\sqrt{N}} \times (3 - 2|S - \bar{S}|) \times 2^{\bar{S}-4}, & \text{for rating review/comment.} \end{cases} \tag{2}$$

where

- A is the submitted rating score for the research article in the range $[-1.0, 5.0]$,
- S is the submitted rating score for the review/comment also in the range $[-1.0, 5.0]$,
- $\bar{A}$ is the average rating score of the research article,
- $\bar{S}$ is the average rating score of the review/comment,
- N indicates that this is the N-th contribution submitted in the category (in particular, contributed/invited reviews count in different categories).

All numerical details of the proposed solution may be adjusted for different scientific fields. Additional factors may be considered. The openness factor $f_o = 2$ doubles the earned points for submitters who disclose their identities. The inflation factor f_i could be applied over time or to compensate for different fields. The attention factor f_a can be very effective in reducing attention to low-quality and harmful works while increasing attention to higher-quality papers; likewise, the factor of $2^{\bar{S}-4}$ encourages more ratings on higher-quality comments/reviews.

An improvement on the early bird factor of f_{eb} for comment and rating could be,

$$f_{eb}(N) = \begin{cases} 1, & \text{for } N \le N_{\text{thr}} \\ \sqrt{N}, & \text{for } N > N_{\text{thr}} \end{cases} \tag{3}$$

where the threshold could be set to $N_{\text{thr}} = 5$ when the average rating of the comment or review would appear to members. A category's activities will not be credited until its N reaches the threshold. For unrated($N < 5$) preprints, no credit will be calculated yet for any related activities.

Another time factor f_t (in addition to f_o and f_i) could be multiplied to reduce the credit value exponentially over time,

$$f_t(T) = \begin{cases} 1, & \text{for } T \leq 1 \\ 2^{-T}, & \text{for } T > 1 \end{cases} \tag{4}$$

where T is the time elapsed since the article was posted, in units of, say, two weeks or 14 days. However, from a historical point of view, our understanding of science, especially when it comes to disruptive ideas, may zigzag, in other words, we may occasionally rediscover, at a much later date, the significance of an early idea that we have mostly overlooked collectively. On the other hand, fraudulent or plagiaristic works may take a long time to be exposed. To encourage researchers to be the first to uncover or recognize such works (hopefully not very often), we may need staged ratings for some unusual papers, i.e., resetting T and N at different stages of our consensus on a paper. For example, for a potential new stage period of at least three months, we should do the reset if the average new rating is much larger than the average rating of the previous stage (e.g., $\Delta\bar{A} > 1$). In this case, whoever initiated such changes would get more credit for their courageous game-changing comment/review/rating. All staged and the overall average ratings would be kept for such papers, but the average rating at the current stage should be used to calculate earned credit points.

A.2 Further Clarifications

The penalty and reward are reflected in the credit points earned for each comment/review/rating action. Nevertheless, the average rating of each preprint/comment/review should be the simple average (i.e., unweighted) to protect the principle of democracy, that is, we should not diminish any single vote.

The credit formula in Eq. 2 should be sufficient for dealing with most meaningless or even harmful works. However, we need to single out one particular category for clarification: potential crackpot works. In general, crackpottery should not be regarded offensive, and can usually be rated zero or better. Only in extreme cases can some clearly unscientific articles be given a score as low as $A = -0.1$. The idea is to protect disruptive ideas that may not yet be fully understood by the community. We hope that the rating range of $0 < \bar{A} < 1$ will be predominantly reserved for such fringe science ideas. The proposed system should greatly reduce, if not eliminate, the large amount of subpar work produced by authors under the publish-or-perish pressure.

In contrast, for the two most intolerable categories of plagiarism and fraud (involving cheating, data fabrication, etc.), in addition to assigning a score of -1, further actions on such bad works are necessary as discussed below regarding moderation duties.

Detrimental	Useless	Trivial	Useful	Good	Excellent
$-1 \le S < 0$	$0 \le S \le 1$	$1 \le S < 2$	$2 \le S < 3$	$3 \le S < 4$	$4 \le S \le 5$

Table 4: Standards for rating reviews/comments are shown with their corresponding rating ranges.

The credit points required for submitting a preprint, proposal, or synthetic achievement are not expended, but represent an accumulated level of credit needed for submission. For example, a member with 100 credit points can submit up to 10 preprints, two proposals, and one synthetic achievement. For accountability purposes, all co-authors should be in good credit standing for the submission. The standards for rating reviews/comments (S) are completely different from those for preprints (A) as we only consider whether the review or comment is relevant and useful for the rating S as shown in Table 4. A bad comment or review could receive negative credit (as low as $\bar{S} = -1$). The rating score ($3 - 2|A - \bar{A}|$ or $3 - 2|S - \bar{S}|$), ranging from -9 to 3, is designed to discourage abusive activities as it tends to reduce credit if not done carefully. Only reviewers and above can rate in order to prevent abuse and immature activities.

In general, reviews and comments should appear immediately, except for a possible grace period, such as five minutes, for the submitter to withdraw them or correct any inadvertent errors. However, the average rating scores for preprints/reviews/comments only appear when N reaches a threshold like $N \ge 5$. A possible requirement might be that ratings submitted after $N > 5$ must be accompanied by a comment or review to be credited, to prevent gaming the system. We might also consider closing ratings and comments when N reaches 1024. Both rating and review submissions require a role in the appropriate subfield and no conflict of interest, which should exclude personal/family/ business relations, advisees/advisors, close collaborators, and colleagues in the same institution as typically required by NSF.

To encourage members to rate more of their more familiar topics, we could add an additional skill factor f_s to their earned credit points in ratings (the last two cases in Eq. 2),

$$f_s = 2^{S_{\text{max}}-3} \tag{5}$$

where S_{max} is the maximum average rating of the best review/comment the member has received for the concerned article.

The ratings of a review/comment are typically not as important as those of an article, and the factor of $2^{\bar{S}-4}$ in Eq. 2 takes some of that into account. But more may be needed.

For example, we could use a suppression factor such as 1/2 for an invited review, 1/3 for a contributed review, and 1/10 for a comment, in the last case of Eq. 2.

Not only preprints and their revisions, but also reviews and comments are assigned citable DOIs to ensure that all scientific contributions will be preserved and properly referenced in the future. Each revision of a preprint submission should be attached with a change log from the submitter. Major revisions may reset the rating. It is possible to consider creating a separate high-risk category for preprints if the variance of their ratings is very large (e.g., $\sigma > 1$), which could be valuable for funding programs designed specifically for high-risk projects.

Three levels of anonymity could be implemented. The most desired one is undoubtedly open-ID. The second level is semi-anonymous, meaning that there is no public association with a specific ID. Instead, either a pseudonym chosen by the member or a randomly generated name is publicly displayed while the activity remains internally linked to the actual member ID. The same regulatory criteria apply and credits can also be earned similarly, except with a reduction by a predetermined factor (e.g., 1/2 of those earned with open-ID). Members have the option to reveal their open-ID later, but the credits earned during the semi-anonymous phase remain reduced as before.

The third level is complete anonymity, where there is no internal link to any open-ID, making it essentially a completely separate account. Access to this account is only possible using the passcode initially set up by the member. Strict moderation is necessary for completely anonymous activities, allowing only comments, but not ratings or reviews. Separate comment credits could still be earned, albeit at a significantly reduced rate (e.g., 1/10 of those earned with open-ID). However, these credits will not be added to the member's account until they decide to merge and reveal their open-ID. The intent is to give the most credits to open-ID activities and the least to complete anonymity, which should be the last resort for members who fear retaliation or similar concerns, and allow them comment unofficially.

The assignment of article ratings according to authors' contributions requires further clarification, especially with respect to the achievement level system. To prevent abuse and unfair advantages in large collaborations, it is not appropriate to simply assign the same achievement rating to all authors of an article solely based on the article's rating. Therefore, for each article, the submitting author, with the consent of all authors and considering their individual contributions, should divide the authors into four groups: first authors, second authors, third authors, and general authors. Each of the first three groups should have a limited number of authors (e.g., up to five authors). The last three groups are optional and can be empty. The aim of this categorization is to assign credits and responsibilities based on the specific contributions of each author, considering their relative importance and involvement in the research.

First authors are meant to be those who have made critical contributions to the work

and bear the most responsibility for the results, typically including co-first authors and corresponding authors/responsible PIs in current journal publication practices. Therefore, they should share the actual rating $\bar{A}$ of the article for the consideration of their achievement levels. The group of second authors should consist of individuals who have made major contributions but not as critical as the first group. As such, they should receive a lower rating to reflect their level of achievements, for example, $\bar{A} - 0.5$ if $\bar{A} > 1$ or $\bar{A}/2$ otherwise (which also implies half responsibility for problematic research). The third group comprises authors who have made distinctive contributions but not as significant as those in the second group, and should hence receive an even lower rating, e.g., $\bar{A} - 1$ if $\bar{A} > 4/3$ or $\bar{A}/4$ otherwise. The last group, general authors, can accommodate the majority of authors in a large collaboration. They should receive the lowest rating (and responsibility), for example, $\bar{A} - 1.5$ if $\bar{A} > 15/9$ (possibly capped at 1.9) or $\bar{A}/10$ otherwise.

A.3 More about the Member Roles

Reviewer and moderator roles are subfield dependent. Members may serve as reviewers or moderators in multiple subfields as long as they meet the minimum requirements (e.g., sufficient number of good papers and earned credit points) of a given subfield. From an achievement point of view, we could require that reviewership should be at least at the L1 level and moderatorship at the L2 level. Moderators who invite a reviewer can earn one fifth of the points earned by the invitee (based on typically five reviews per article, or the limitation that each moderator can invite up to five reviewers). If an invited reviewer declines, the moderator can invite another. However, the maximum number of accepted invitees should be set to 10.

Moderators can take on more responsibilities like closing comments, reviews, and ratings for a preprint if too much activity is deemed unbeneficial, or regulating other harmful behaviors such as gaming the credit system. For example, a moderator can call alert other moderators to extremely harmful (fraudulent, plagiaristic) activities (note that crackpottery is not included in such actions). A temporary rating of -1 may be assigned for two or three days until enough (e.g., ≥ 3) other moderators have responded. Otherwise, the case will be dismissed. If enough responses are received in time and more than half of them agree, the rating will be permanently set to -1; if more than two thirds agree, the alleged member will be suspended pending further action by the ethics committee of board members. After a decision is made, the board will determine how long the suspended membership will be, such as one month, one year, up to lifetime, depending on the severity of the offense. As far as credit is concerned, the initiating moderator will receive points equal to $100 \times (\text{rate of agreement} - 0.5)$ that could be negative, the other participating moderators will get 10 points each regardless of the outcome, and the board members involved will obtain 20 points each.

Another important type of activity needs further clarification. Moderators can invite a member to write a review article on a topic. Unlike submitting a research article, no credit is required, and the invitee can even earn credit points as follows,

$$\text{ECP} = \begin{cases} 1000 \times 2^{\bar{S}-5}, & \text{for } \bar{S} \geq 3.0 \\ 250 \times (\bar{S}-2), & \text{for } \bar{S} < 3.0 \end{cases} \tag{6}$$

where $\bar{S}$ is the average rating it has received. Note that we use the symbol S instead of A because the standards for such ratings should be similar to those for reviews/comments as shown in Table 4, and definitely NOT to those for research papers. The inviting moderator may receive a credit equal to one tenth of what the invitee has obtained. As always, a bad review article can reduce the credit points of both members involved.

In general, commenters can post informal comments in any field; reviewers can review and rate only in the subfields for which they are qualified; moderators can regulate problematic behaviors in their own field, but can invite reviewers only in their qualified subfields. All issues, if appealed or as required, are subject to final resolution by an appropriate committee of board members.

Achievement levels for individual researchers and their works are listed in Tables 2 and 3. More achievement levels could be added such as L2.4, L2.5, L3.1, L3.2, etc. In practice, lower level achievements below A_{+2} do not need to be reviewed.

Given recent advances in artificial intelligence (AI), the implementation of an AI system may be desirable, particularly in detecting behaviors of gaming the credit system and abusive, fraudulent, plagiaristic, or other offensive activities. In addition, AI could be used to detect the patterns of articles that should have staged ratings as discussed above.

Appendix B: Practical Implementation in the Real World

B.1 Motivations

The current incentive system, or lack thereof, in most existing review services and platforms, does not work. Typically, reviewers do this work on a volunteer or honorary basis, and some services have begun to implement more recognizable measures, such as certification for review work. However, these measures do not significantly benefit reviewers in terms of their career or position in the community. More recently, some platforms have attempted to introduce monetary incentives or equivalent tangible credits that can be redeemed in exchange for other publishing/editing services. None of these measures provide adequate incentives.

Perhaps the best incentive is to enhance the role of reviewers in the community in recognition of their high-quality review work. This would motivate scientists to engage in

reviewing each other's work more frequently. The proposed community-based credit/role system for peer review is a promising approach that should be pursued.

In addition to providing the right incentives, such a system could also contribute to the financial self-sustainability of the platform. For instance, the platform could receive donations or fees by participating in the creation of overlay journals for peer-reviewed preprints, by assisting funding agencies, particularly private foundations, in reviewing proposals, and by providing academic institutions with more reliable merit evaluation of candidates for their hiring and promotion decisions.

Therefore, such a credit/role system could be implemented by both non-profit organizations and for-profit companies. The system would ensure that the member community is self-regulating and that the quality of peer review is inherent in the system without external interference from the platform.

B.2 Practical Approaches

There are two components in the proposed system. One is the common core software framework that could be developed on an open-source software development platform such as GitHub or GitLab. This development could be sponsored by open science organizations such as Code for Science and Society. The goal is to establish a centralized location for the software so that different fields and communities do not have to reinvent the wheel. Any field and its community could request new features in the software's development, and they could even initiate a new fork if they have radically different requirements for the system.

Some suggested features related to the coding for comment and review sections would be as follows. Comments and reviews should be able to accommodate various formats: markdown, latex, pdf, and possibly other rich text formats. This would provide flexibility for members to express their thoughts and ideas. Implementing a real-time preview feature would be beneficial, as it would allow members to review their comments or reviews before submitting them. This feature helps catch any typos or mistakes, ensuring that the final submission is accurate and error-free. Readers should be able to choose between two different viewing structures with sorting options: a threaded layout that helps maintain the conversation flow and allows for easier tracking of responses; a flat layout in chronological order of the post, with a back-link if it is a reply to another post.

The second component involves the concrete implementation of the system on an operational platform specific to a given field. It is obvious that different fields may require different implementations of the system, especially when it comes to numerical details. The crucial aspect is that the implementation should be field-oriented rather than institution-oriented. In other words, it should aim to encompass the entire community of scientists in a given field, across geographical boundaries. In the early stages, several platforms may

serve the same field, but the hope is that eventually one of them will prevail, or that they will unite and merge into a single service.

Ideally, the most suitable place to implement the system would be on large preprint service platforms that are widely used within a given field. However, due to its entrenched dominance and inertia, the largest eprint server, arXiv.org, does not allow comments or reviews, let alone a quantitative review system. Although this path would have been the most efficient, it appears to be a long shot.

Conversely, emerging smaller preprint servers like *bioRxiv* and *medRxiv* are more willing to try new ideas and could play a more significant role in the adoption of the proposed system. In addition, newly established dedicated review platforms such as *PREreview.org*, *ReviewCommons.org* (non-profit), and *ReviewerCredits.com* (for-profit) could gain increased recognition and significantly expand their user base by implementing the new system. Interestingly, a for-profit company called *ScienceOpen.com*, which offers both preprint/publishing and peer review services, has already implemented most of the required structures except for the new credit/role system. It may soon demonstrate the desired effect through a relatively straightforward integration of the new quantitative system.

Any of the aforementioned platforms would be suitable for starting experiments with the new system. There is no need to first build a national or international community structure from scratch. Nor is it necessary to implement all aspects simultaneously. However, it is crucial to first establish the basic credit/role mechanism as proposed. Certain existing beneficial practices should be incorporated. For example, member registration should be integrated with *ORCID*, a practice already in place on several existing platforms like *ScienceOpen.com* and *PREreview.org*. Additionally, all reviews should be assigned citable DOIs, as is currently the case on platforms such as ScienceOpen.com and PREreview.org. At a later stage, the option of assigning citable DOIs to highly-rated comments (e.g., $\bar{A} > 3$) could also be implemented.

In the early phases of implementation, all numerical details in the earned credit formulation should be dynamically calibrated, possibly quite frequently. The specific values of the parameters and even the concrete factor formalisms could be considered as the initial reference set of the parametrization. Therefore, it is crucial to completely separate the data sets (such as ratings, comments, and reviews with their timestamps) from the formulation. This means that nothing is hard-coded, so that all members' credits can be easily recalculated if a better parametrization is found. By doing so, this would discourage any abuse of loopholes or attempts to game the system. It would also encourage more attention to higher quality work and prompt, fair reviews, as such desired behaviors will eventually be rewarded, even if not immediately due to temporary system bugs.

We could start the experiment by focusing on the review of preprints/publications first. Once the system has matured, meaning the credit formula has become stabilized, we can

then add the review of grant proposals and embark on a new phase of experimentation. Finally, we can proceed with the community evaluation of individual achievements, primarily for synthetic accomplishments, since single-paper achievements are automatically evaluated in the first step.

By progressively implementing the system starting with paper reviews, followed by proposal reviews, and concluding with achievement evaluations, the credit/role system will eventually establish the implemented platform as the most attractive choice for all scientists and researchers, especially aspiring young individuals. The growth of a platform relies heavily on the size of its user base, and this approach will contribute to its expansion. It is possible that an implemented platform could serve several relevant fields at the same time. Furthermore, a successful platform could expand its scope to include other businesses, for example, organizing conferences (including determining topics and invited speakers), and reviewing proposals for experimental facilities, among others. This would be a dream come true for all scientists.

For platforms intended to serve a much broader community, the community-building procedure could be as follows. Initially, sub-communities could be created based on major sub-fields commonly recognized within the field (e.g., categories of major preprint servers, topics covered by major journals, and major research directions in university departments). Let's take the discipline of physics and astronomy as an example. Five initial sub-communities could be created: Astronomy & Astrophysics (AA); Atomic, Molecular and Optical Physics (AMO); Condensed Matter Physics (CM); Nuclear and Particle Physics (NP); and Miscellaneous Physics & Astronomy (Misc).

Over time, more sub-communities may emerge within each major sub-field or form from several fields for a new cross-disciplinary direction. For example, Cosmology could be a sub-community branching out from AA, Biophysics from Misc and another discipline of Biology, Quantum Information from AMO, Computer Science, and other fields. In addition, experimental and theoretical spin-offs could be formed from their parent communities.

The formation of a new sub-community could be contingent on meeting certain criteria. For example, it may require that they have a minimum number of moderators (e.g., >10) and reviewers (e.g., >100) in their new subfield. Meanwhile, they should not be so dominant in their parent community that their departure could leave the parent community too small. Two sub-communities can also merge into one if the majority of their advanced members (reviewers and above) agree. If a sub-community experiences a decline in size over time (e.g., <6 moderators or <60 reviewers), it must either return to its parent community or merge with another sub-community.

B.3 Concerns and Criticisms

The proposed credit/role system has some similarities to what social media platforms have implemented. However, there are notable differences to highlight: the bar for entrance into the research community is considerably high, as members must be properly trained scientists in their respective fields; the research community operates in a self-regulated manner, similar to certain online forums but unlike the majority of social media platforms; and the proposed system is significantly more quantitative in nature.

Given that researchers' careers (e.g., their role and position in their research community) are at stake, it sets expectations for more responsible and professional behavior. In contrast to the often chaotic behaviors observed on social media platforms, the research community is expected to maintain a higher level of decorum. Furthermore, the dynamically calibrated credit formula will promote the best behaviors from members, based on the belief that the quantitative system will become increasingly trustworthy over time.

Qualitative or semi-quantitative methods, such as thumb-up or down, used in social media platforms have proven their usefulness. However, they cannot be made more quantitative due to the inherent nature of many topics discussed, which are often difficult to quantify and lack scientific rigor. In contrast, for the realm of rigorous science, a more quantitative system is expected to be significantly more effective.

What is the point of the analogy to a capitalist economy? The basis of the analogy is that both systems are driven by cycles of innovation. However, there are fundamental differences between the tangible products of a capitalist economy and the much less tangible outcomes of research. These differences necessitate a complex peer review process for evaluating scientific progress, which is the central focus of this article. Nonetheless, addressing this unique peer review procedure requires considering potential issues associated with start-ups and monopolies that are common to both and that could impede innovation. The point is that capitalist economies have been more successful in dealing with these issues than scientific research, which is both unfortunate and inexplicable.

How can the scientific community do better than the governmental structure that is meant to prevent monopolies in a capitalist economy, which can often be corrupted? While it is true that the capitalist economy and its regulatory government are not flawless, they do have a democratic mechanism in place that promotes fair competition and fosters healthy innovation. In contrast, the scientific community can be perceived as more authoritarian than democratic, i.e., not even up at the level of the capitalist economies. As a result, unorthodox ideas in science often face gatekeeping barriers to recognition and acceptance.

The proposed system advocates the principles of democracy and diversity that can truly preserve the freshness and vitality of the driving force of innovation. Considering the rigorous nature of scientific research, there is good reason to believe that a properly implemented system could significantly enhance the self-regulating structure within the

scientific research community, making it more robust.

In scientific research, ideas embodied in preprints, publications, and proposals can be considered the “products”. But are there more tangible products in science, comparable to the marketable goods produced in the economy? In the realm of basic science research, ideas are undoubtedly the most important output, and the proposed review system is particularly well-suited for evaluating such results. On the contrary, when it comes to research focused on applications and technology, more tangible products emerge, some of which even lead to the creation of start-up companies in the capitalist economy. We argue that applied research progresses more effectively than basic research, precisely because of the existence of such tangible products. The real challenges lie in basic science research, where progress seems to be increasingly stagnant, and cases of meaningless work or fraud have become all too common.

The failure of previous quantitative practices in the scientific community can be attributed to several complex factors, such as lack of sustainability, flawed incentives, inadequate feedback mechanisms, and incomplete metrics. A major concern is the potential for gaming the system. Fortunately, the proposed system has built-in measures to counteract such attempts. In particular, the credit rewarding process is dynamic and can be continuously fine-tuned, rendering any gaming efforts ultimately ineffective. In addition, with the aid of advanced machine learning techniques, we can further enhance the system’s performance and robustness as we accumulate a larger dataset of statistical information.

Despite the potential drawbacks associated with highly quantitative measures, it is important to further defend the criticisms directed towards such a quantitative system. The advancement of science itself is an evolutionary process that continually strives for greater quantification and rigor. If we never attempt to make a field more quantitative and rigorous, then it will stand little chance of becoming part of the science. Why shouldn’t the evaluation of rigorous science be as quantitative as scientific research itself? We dare to propose an initial endeavor to quantify measures that could be applied in peer review, in the hope of establishing a first quantitative paradigm for peer review. It would be a pity if rigorous science could not be assessed in a quantitative manner.