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-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 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 majority of 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 the economy and research as shown in Figure 1. Compared with the 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. However, due to the direct positive feedback, whoever wins out could grow into giant corporations and even monopolies that inevitably block further innovation and competition for their own benefit.

Fortunately, the capital economy has addressed this issue by establishing antitrust laws and regulations to prevent monopolies. Startup companies are protected for innovation, and their growth is 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.

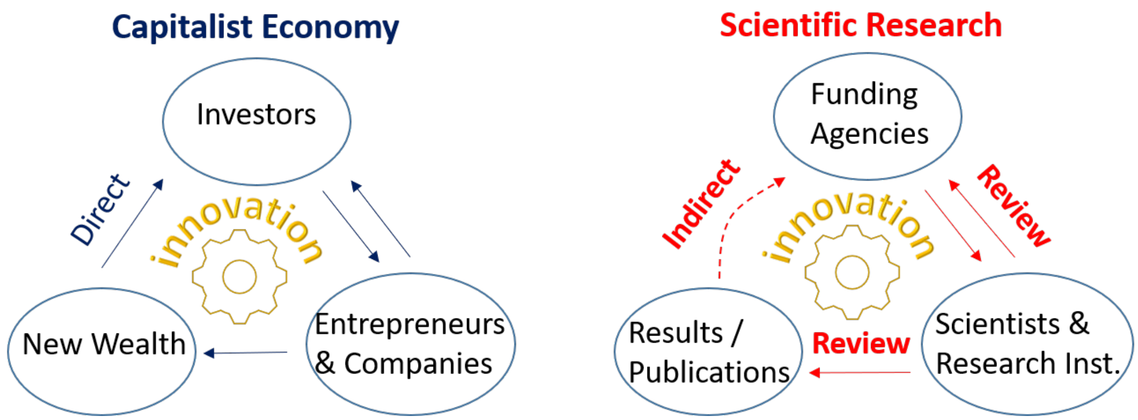

Figure 1: Evolution cycles of capitalist economy and scientific research.

Mature companies, especially large ones, may still be the backbone of the economy taking up most investments. However, startups, 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 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. 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 startups 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 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 but typically rejected or not allowed without endorsement by arXiv, 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, who 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.

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 "startups". 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 truly support diversity in ideas and projects (especially risky ones), and most importantly establish a more open and democratic community structure involving all scientists.

2.2 Research Startups and High-Risk Investments

At first glance, 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 startups 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.

In scientific enterprises, it is common to see the formation of large research groups, including giant collaborations in physics (up to several thousand scientists) at facilities like LHC and LIGO. These groups often lead the research in hot-topic frontiers. However, there is currently no mechanism in place for cultivating scientific startups, as is the case in the business world. A true startup in science needs to be independent from the manipulation of large groups and requires long-term support in order to survive and grow.

While some may argue that postdoctoral positions could serve as a startup role, 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 startups, but the enormous pressure to get tenured within a limited time frame, typically six years plus possibly many years spent as a postdoc, often leads individuals to pursue hot-topic research and follow the footsteps of more authoritative figures in their field. Those who dare to pursue a different path are often eliminated in the process. 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 startups, and how can we support and foster them? One potential solution is to make most postdoctoral researchers independent with long term support (5-10 years, renewable indefinitely). However, the key issue lies in developing a structure or mechanism that allows for the academic startups 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 species 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 10 to 20 years into their careers, especially in basic fields like physics. 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 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 determined within one academic unit, often by a few 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, an individual's scholarly achievements should be 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 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.

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: manuscript/publication review and grant application review.

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 academe. 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, 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. The independent service *Review Commons* has even started a central open review platform for preprints and journals in the life sciences.

Unfortunately, all of these practices and services have not really attracted sufficient review activities, nor have they dramatically reduced 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 the 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, a proposal would hardly stand a chance of being funded if it does not receive excellent reports from all 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 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 chagned 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.

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.

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) as a commenter and earns credit points 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 can be anonymous or open-ID for maximum flexibility, and to encourage openness, we could double the earned points for members who choose the open format.

Once members gain enough experience and credit (e.g., posted ≥ 3 papers, > 1 yr membership, and > 100 credit points), they will be promoted to the rank of reviewer who can earn more points by contributing or getting invited to rate or write official review reports about a preprint 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 posted in their qualified subfield with a score A in the range $-1.0 \leq A \leq 5.0$. Assigning $A = -1.0$ means that it 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 works 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. $A \geq 4.0$ is 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 can see preprints with $\bar{A} > 0.0$ or A_+ and unrated preprints (i.e., $N < 5$).

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 for encouraging 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 = $3 - 2|A - \bar{A}|$; for rating a review/comment, SCORE = $(3 - 2|S - \bar{S}|) \times 2^{\bar{S}-4}$, which encourages more ratings on better reviews/comments. Note that both $\bar{A}$ and $\bar{S}$ only appear when N is above the threshold ($N \geq 5$). Therefore, earned points may be credited later or not at all, and it could fluctuate as $\bar{A}$ and $\bar{S}$ vary over time.

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 (ideally 5% of all investments). 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. Varied opinions of other reviews from randomly-selected experts in the same subfield could indicate the level of risks involved. The funder should also have non-expert scientists from adjacent subfields and/or even from totally different fields to evaluate the potential of the proposal.

Lastly, the most difficult is to replace tenure with an achievement class level system as shown in Table 2. Each member 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). If a

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

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

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.

submitted achievement is related only to one single paper, 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.

L1 and above members can apply for basic funding support from government funding agencies as shown in Table 2 if they stay in academia. Note that such support is 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. The new level system will make individual institutions relieved of the burden on achievement evaluation for their hiring and promotion decisions, which the community can certainly do much better.

Moderators who are among the top credit holders within their class and subfield may assume the highest leadership roles and serve as board members for 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 every four years to prevent bias.

In summary, the major advantages of this solution are: it is driven by the entire scientific community, rather than by elite circles; 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 for evaluation to five significantly reduces the production of low-value works, resulting in a much cleaner field for all scientists; and high-risk, high-reward projects from academic startups can finally receive proper funding.