Peer review in competitive funding programs: promotion of mediocrity and clique formation

Pawel Sobkowicz

May 15, 2014

Introduction

Peer review is the cornerstone of modern science: from the publication process to the evaluation of funding applications. While there is long tradition and many arguments why peer review is beneficial and necessary component of the scientific processes, the exponential growth of the research community, the ‘publish or perish’ and ‘funding or famine’ pressures as well as the availability of better datamining tools have led to increasing number of voices describing the weaknesses of the system.

Some quotes

At first glance the notion of "excellence through competition" seems reasonable. The idea is relatively easy to sell to politicians and the general public. [. . .] On the practical side, **the net result of the heavy-duty "expert-based" peer review system is that more often than not truly innovative research is suppressed.** *Furthermore, the secretive nature of* **the funding system efficiently turns it into a self-serving network operating on the principle of an "old boys' club."**
A Berezin, The perils of centralized research funding systems, 1998

Some quotes

Diversity – which is essential, since experts cannot know the source of the next major discovery – is not encouraged. [...] **The projects funded will not be risky, brilliant, and highly innovative since such applications would inevitably arouse broad opposition from the administrators, the reviewers, or some committee members.** *[...] In the UK (and probably elsewhere), we are not funding worthless research. But we are funding research that is fundamentally pedestrian, fashionable, uniform, and second-league.*

D F Horrobin, Peer review of grant applications: a harbinger for mediocrity in clinical research?, 1996

Some quotes

Further cohort studies of unfunded proposals are needed. Such studies will, however, always be difficult to interpret – **do they show how peer review prevents resources from being wasted on bad science, or do they reveal the blinkered conservative preferences of senior reviewers who stifle innovation and destroy the morale of promising younger scientists?** *We cannot say.*

S Wessely, Peer review of grant applications: what do we know?, 1998

Some quotes

Only false objectivity is offered by evaluating real people using unreal calculations with numbers of papers, citations, and journal impact factors.
P A Lawrence, Real lives and white lies in the funding of scientific research, 2009

The allegation of cronyism – that the peer review system consists of a group of committees whose members hand out grants to each other and to their friends – is one that is almost impossible to substantiate or refute.
N Wade, Peer review system: How to hand out money fairly, 1973

Some quotes

Only false objectivity is offered by evaluating real people using unreal calculations with numbers of papers, citations, and journal impact factors.
P A Lawrence, Real lives and white lies in the funding of scientific research, 2009

The allegation of cronyism – that the peer review system consists of a group of committees whose members hand out grants to each other and to their friends – is one that is almost impossible to substantiate or refute.
N Wade, Peer review system: How to hand out money fairly, 1973

Recently, in addition to enhanced capacity of studying real world data, an interest in agent based models of peer review processes has grown.

Model assumptions

N_P proposals are submitted each year, with starting $N_P = 2000$ and 2% growth each year.

Model assumptions

N_P proposals are submitted each year, with starting $N_P = 2000$ and 2% growth each year.
Lognormal distribution of innovation value $V(P)$ of proposals P.
Units are, of course, arbitrary...

$$\frac{1}{V(P)\sqrt{2\pi}\sigma_I} \exp \frac{-(\ln V(P) - \mu)^2}{2\sigma_I^2},$$

with $\mu = 0$ and $\sigma_I = 0.4$ used in the simulations.

Model assumptions

N_P proposals are submitted each year, with starting $N_P = 2000$ and 2% growth each year.
Lognormal distribution of innovation value $V(P)$ of proposals P.
Units are, of course, arbitrary...

$$\frac{1}{V(P)\sqrt{2\pi}\sigma_I} \exp \frac{-(\ln V(P) - \mu)^2}{2\sigma_I^2},$$

with $\mu = 0$ and $\sigma_I = 0.4$ used in the simulations.
Only a small fraction (30%) of the proposals get funded.

Model assumptions

N_P proposals are submitted each year, with starting $N_P = 2000$ and 2% growth each year.
Lognormal distribution of innovation value $V(P)$ of proposals P.
Units are, of course, arbitrary...

$$\frac{1}{V(P)\sqrt{2\pi}\sigma_I} \exp \frac{-(\ln V(P) - \mu)^2}{2\sigma_I^2},$$

with $\mu = 0$ and $\sigma_I = 0.4$ used in the simulations.
Only a small fraction (30%) of the proposals get funded.
Out of the rejected ones, 60% are resubmitted with the same innovativeness value, 40% drop out, and are replaced by new proposals/researchers.

Model assumptions

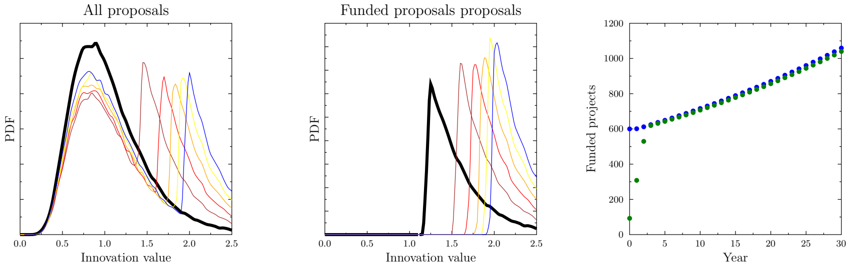
N_P proposals are submitted each year, with starting $N_P = 2000$ and 2% growth each year.
Lognormal distribution of innovation value $V(P)$ of proposals P. Units are, of course, arbitrary...

$$\frac{1}{V(P)\sqrt{2\pi}\sigma_I} \exp \frac{-(\ln V(P) - \mu)^2}{2\sigma_I^2},$$

with $\mu = 0$ and $\sigma_I = 0.4$ used in the simulations.
Only a small fraction (30%) of the proposals get funded.
Out of the rejected ones, 60% are resubmitted with the same innovativeness value, 40% drop out, and are replaced by new proposals/researchers.
Selection is done by groups of N_E (5) evaluators, drawn randomly from a pool of experts $\mathcal{R}$ of size N_X (300).

Ideal world case

In the ideal world case every evaluator would assign the proposal a score equal to its innovation value $S(P, E) = V(P)$ and only the propopsals with topmost scores get funded.



(a) (b) (c)

Figure: (a,b) Distribution of innovativeness of all proposals and of funded proposals ; (c) number of funded proposals (all – blue, inventors – green.

Non-ideal world

Every evaluator suffers from limitations of his/her own innovativeness and would assign to a proposal a score equal to its innovation value modified by the difference between his/her own innovativeness

$$S(P, E) = V(P) \exp \frac{-(V(P) - V(E))^2}{2\sigma_T^2}$$

Evaluator's own innovativeness acts thus as a tolerance filter for the evaluated proposals. In the case of multiple evaluators, the average score is $\bar{S}(P) = \sum_{i \in \mathcal{R}} S(P, E_i)/N_E$.

Tolerance filter

Graphical representation of an evaluator tolerance filter for $\sigma_T = 0.1$.

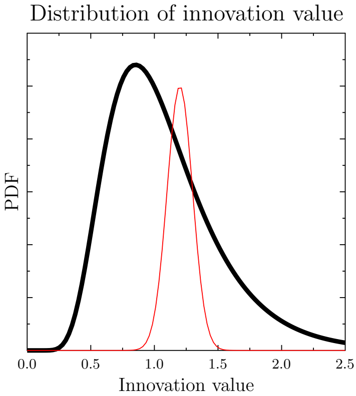


Figure: Black line – distribution of innovation value, red line example of tolerance filter for evaluator with innovativeness equal to 1.25 and tolerance $\sigma_T = 0.1$.

Non-ideal world case

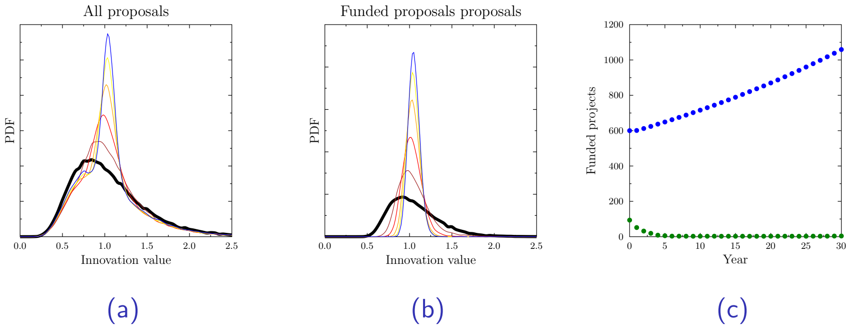


Figure: (a,b) Distribution of innovativeness of all proposals and of funded proposals ; (c) number of funded proposals (all – blue, inventors – green.

Clique formation

Clique recruits members preferentially from the expert pool $\mathcal{R}$.
Three possible models can be considered:

- "Fair" clique: clique evaluators give clique proposals their true value as the score $S(P^C, E^C) = V(P^C)$

Clique formation

Clique recruits members preferentially from the expert pool $\mathcal{R}$.
Three possible models can be considered:

- "Fair" clique: clique evaluators give clique proposals their true value as the score $S(P^C, E^C) = V(P^C)$
- "Preferential scenario" – clique evaluators add a bonus to the clique proposal score $S(P^C, E^C) = V(P^C) + \Delta_1$

Clique formation

Clique recruits members preferentially from the expert pool $\mathcal{R}$.
Three possible models can be considered:

- "Fair" clique: clique evaluators give clique proposals their true value as the score $S(P^C, E^C) = V(P^C)$
- "Preferential scenario" – clique evaluators add a bonus to the clique proposal score $S(P^C, E^C) = V(P^C) + \Delta_1$
- "Jealousy" model – for more than one clique: same as preferential, but clique 1 evaluators not only use the tolerance filter but also penalize proposals from the other cliques $S(P^{C1}, E^{C2}) = V(P^{C1}) \exp(-(V(P^{C1} - V(E^{C2}))^2/2\sigma_T^2) - \Delta_2$

Clique formation

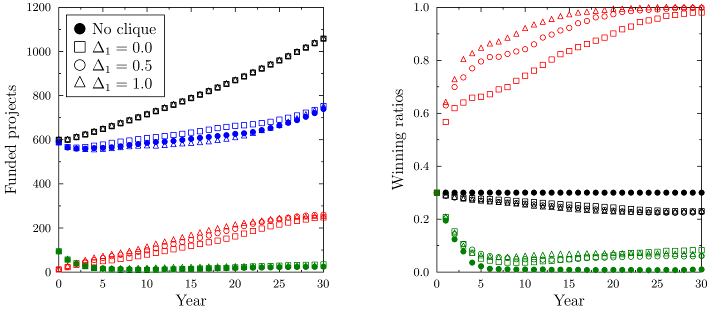


Figure: Evolution of the funded projects in the presence of one clique. Black symbols – all funded projects. Blue symbols – non-clique proposals. Red symbols – clique proposals. Green symbols – innovative proposals.

Two competing cliques

Second clique forms 10 years after the first one.

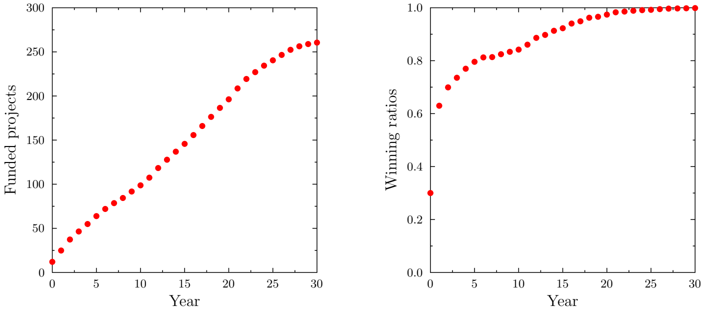


Figure: Evolution of the funded projects in the presence of one clique. Red symbols – single clique. Green symbols – non-conflicted cliques. Blue symbols – jealousy model.

Two competing cliques

Second clique forms 10 years after the first one.

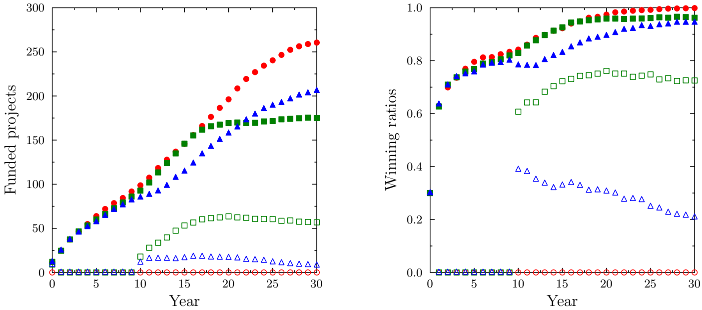


Figure: Evolution of the funded projects in the presence of one clique. Red symbols – single clique. Green symbols – non-conflicted cliques. Blue symbols – jealousy model.

Summary

- Unless the reviewers are very open-minded, peer review may indeed favor regression towards mediocrity

Summary

- Unless the reviewers are very open-minded, peer review may indeed favor regression towards mediocrity
- Even a relatively weak preference within a group of reviewers may lead to disproportionate advantages and biasing the selection process

Summary

- Unless the reviewers are very open-minded, peer review may indeed favor regression towards mediocrity
- Even a relatively weak preference within a group of reviewers may lead to disproportionate advantages and biasing the selection process
- The two processes only weakly influence each other. Separate corrective mechanisms are needed in both cases.

Parting question

If we want to explore the unknown, to aim for true innovations, we must accept the risk of failure. This applies – in particular – to research.

Parting question

If we want to explore the unknown, to aim for true innovations, we must accept the risk of failure. This applies – in particular – to research.
Historically, the rule of thumb is that 90% of truly audacious efforts end in failure, but the remaining 10% pay off the costs and generate true growth.

Parting question

If we want to explore the unknown, to aim for true innovations, we must accept the risk of failure. This applies – in particular – to research.

Historically, the rule of thumb is that 90% of truly audacious efforts end in failure, but the remaining 10% pay off the costs and generate true growth.

Then let me ask the question: do you know any funding agency that boasts the fact that 90% of the research they funded ended in failure?