# Why don't scientists invest all their time in one project?

Sam Zhang[1]

[1]Department of Applied Mathematics, University of Colorado Boulder

Dec 8th, 2020

## Theoretical motivation

### Optimizing for the most impactful project

Suppose a scientist is working on $n$ projects, and the impact of each of the projects is given by an exponential distribution that has mean given by the amount of effort put into it, $\theta$:

$$X_i \sim \exp(\text{mean}{=}\theta).$$

Then how should they allocate their effort across multiple projects to maximize their most successful project $X_{(n)}$? We have through a simple derivation[1] that the expected value of the maximum of i.i.d exponential variables is

$$\mathbb{E}\left[X_{(n)}\right] = \theta H_n,$$

where $H_n$ is the nth Harmonic number, so in particular,

$$\theta \log(n) \leq \mathbb{E}\left[X_{(n)}\right] \leq \theta\left(\log(n) + 1\right).$$

Since the impact of the most successful project increases linearly in $\theta$ but only logarithmically in $n$, if we have some constant pool of effort that we can divide equally across projects $\kappa = \theta n$, then lacking additional constraints, the way to maximize the impact of the most successful project is to concentrate all of the effort in that one project.

### Optimizing for the total impact of projects

This "paradox" is somewhat alleviated if we view scientists as attempting to maximize their *total* impact rather than their biggest impact. Again suppose

$$X_i \sim \exp(\text{mean}{=}\theta)$$

and let

$$Y = \sum_{i=1}^{n} X_i.$$

[1]See, eg, https://mikespivey.wordpress.com/2013/05/13/expectedmaxexponential/

Then $Y \sim \text{Gamma}(n, \theta)$, and the expected total impact is

$$\mathbb{E}[Y] = n\theta.$$

Thus from the perspective of expected total impact, we do not have any immediate preference how effort is allocated across multiple projects. As a tiebreaker, one might consider the impact of the most successful project from the earlier section, and decide to allocate effort into fewer projects. But now one has a valid trade-off, which is that the variance

$$\text{Var}\, Y = n\theta^2$$

increases as the square of $\theta$ but only linearly in $n$. Thus investing one's efforts in more projects diminishes the likelihood of a single blockbuster, but also reduces the risk of having too low of a total impact.

# Discussion

Empirically, most scientists don't put all of their effort into a single large project. We discussed one possibility why this thought experiments fails in the above section on total impact, but here we elaborate on that as well as other possibilities:

1. **Risk aversion around total impact.** As discussed above, if scientists care about total impact, rather than maximum impact, then allocating the same amount of effort across more papers diminishes the variance of the total impact without affecting the mean impact. Although maximum impact may be considered in career evaluations, there are similarly strong disincentives against making too low of a total impact at these career evaluations as well, especially in the early career.

2. **Explicit evaluation of productivity.** Scientists may be explicitly evaluated on their productivity, rather than just maximum or total impact. If this incentive is strong enough, this could tip the objective function in favor of more publications than less, even though it would decrease the expected impact of the most successful project.

3. **Cumulative impacts of publishing.** The impact of a paper grows cumulatively over time from the moment of publishing, and allocating more effort into a project increases the time until publication. Since scientists have career thresholds (e.g, tenure evaluation) where their impact is evaluated, they may not be able to afford to work on single projects for long stretches of time, at least not for their early careers.

4. **Learning through publishing.** Scientists *learn* and explore the topic space as they publish papers, and this may increase the probability of success over time. There is some subtlety here, since research suggests the most impactful papers are distributed uniformly at random across all academic careers. That doesn't imply that different research strategies don't produce different payoffs (though it might), just that aggregating across all strategies, scientists experience roughly a uniform chance of success for their most successful paper at any given time.

5. **Other constraints.** Projects may not be able to absorb unlimited amounts of scientific effort, if some other bottleneck arises. For instance, data collection may occur passively

over a long period of time, during which the researcher would naturally switch to a different project. Some of these constraints may be hard to measure, such as internal psychological constraints, like boredom around working on the same project.