Lorenz curves in a new science-funding model

Ding-wei Huang
Department of Physics, Chung Yuan Christian University, Chung-li, Taiwan
E-mail: dwhuang@cycu.edu.tw

Abstract. We propose an agent-based model to theoretically and systematically explore the implications of a new approach to fund science, which has been suggested recently by J. Bollen et al.[1] We introduce various parameters and examine their effects. The concentration of funding is shown by the Lorenz curve and the Gini coefficient. In this model, all scientists are treated equally and follow the well-intended regulations. All scientists give a fixed ratio of their funding to others. The fixed ratio becomes an upper bound for the Gini coefficient. We observe two distinct regimes in the parameter space: valley and plateau. In the valley regime, the fluidity of funding is significant. The Lorenz curve is smooth. The Gini coefficient is well below the upper bound. The funding distribution is the desired result. In the plateau regime, the cumulative advantage is significant. The Lorenz curve has a sharp turn. The Gini coefficient saturates to the upper bound. The undue concentration of funding happens swiftly. The funding distribution is the undesired results, where a minority of scientists take the majority of funding. Phase transitions between these two regimes are discussed.

1. Introduction
The funding distribution is relevant to both research policy and sociophysics, where social interactions are investigated by mathematical tools inspired from physics. Recently, Johan Bollen et al. have suggested a new approach to fund science [1, 2]. The scientific community’s responses are favorable. Small-scale experiments are in progress [3, 4]. The idea is to fund every scientist equally and to ask all scientists to give a fraction of their funding to other scientists. The more one receives funding from others, the more one gives away funding to others. Every scientist serves as a funding agency to distribute a small amount of funding. As the funding circulates fluidly in the community, the system can potentially be fairer and more efficient. The funding distribution is expected to converge toward an optimal result favored by the entire community of scientists.

Along this line, we propose an agent-based model to theoretically and systematically explore the implications of this approach. The model is presented in Section 2. The phase transition is addressed in Section 3. Discussions are in Section 4.

2. Model
Consider a community consists of $10^3$ scientists. Assume that the annual research budget is also $10^3$, i.e., the average budget per scientist is taken as the arbitrary unit. The traditional system is to set up a small committee to allocation $10^3$ budget to $10^3$ scientists. Quite a few scientists would expect to receive nothing. In the new system, all scientists receive an equal share of 1, in the arbitrary unit. All scientists are also required to give a fixed ratio $F$ to support others, where
$0 \leq F \leq 1$. More specifically, the funding of each scientist is denoted as $u_i$, where $i = 1, 2 \cdots 10^3$. Every scientist is asked to donate $(F \cdot u_i)$ to five other scientists. We introduce two parameters $\alpha$ and $\beta$ to specify the random process of how to choose these five scientists and how to share the donation among the chosen. The scientist $i$ is chosen with a probability proportional to $(u_i)^\alpha$; once chosen, the share of donation is proportional to $(u_i)^\beta$. When $\alpha = \beta = 0$, the scientists are chosen randomly and the donation are shared equally. When $\alpha > 0$ and $\beta > 0$, the cumulative advantage is introduced into the model. The Matthew effect becomes prominent when $\alpha$ and $\beta$ are large.

![Figure 1](image1.png)

**Figure 1.** (LEFT) Lorenz curves at various $F$, where $\alpha = \beta = 0.5$. Solid lines show $F = 0.1, 0.2, \cdots 0.9$, where $\Delta F = 0.1$. Grey dotted line shows $F = 0$. Grey dashed line shows $F = 0.99$. (RIGHT) Gini coefficient $G$ as a function of $F$ at various $\alpha$ and $\beta$.

The concentration of funding is mainly controlled by the parameter $F$. Typical results of the Lorenz curves are shown in Fig. 1 (LEFT), where the cumulative share of funding ($y$) is plotted on the cumulative share of scientists ($x$) from lowest to highest funding. Line of equality is $y = x$, where all scientists receive the same amount of funding. The inequality increases with the increase of $F$. The inequality can be conveniently measured by the Gini coefficient, which is defined as the ratio between two areas in the plot, i.e., $G \equiv A/B$. The area $A$ is between the Lorenz curve and line of equality; the area $B$ is between $x$-axis and the line of equality and $B = 1/2$. Fig. 1 (RIGHT) shows the typical results of Gini coefficient as $F$ varies. The Gini coefficient increases monotonically with the increase of $F$. In this model, every scientist keeps the minimum funding of $(1 - F)$. The Lorenz curve always lies above the line $y = (1 - F)x$. As a result, the area $A$ is less than $F/2$ and the Gini coefficient is less than $F$. The parameter $F$ becomes the upper bound of the Gini coefficient. When $\alpha$ and $\beta$ are small, this upper bound is reached only when $F$ approaches 1. When $\alpha$ and $\beta$ are large, this upper bound can be reached by a small $F$.

### 3. Phase Transition

The two parameters $\alpha$ and $\beta$ control the effect of cumulative advantage. The typical 3D profiles are shown in Fig. 2. Two distinct regimes can be observed as valley and plateau. In the valley regime, the Gini coefficient is well below the upper bound. When $\alpha$ and $\beta$ increase, the Gini coefficient has a gentle rise. In the plateau regime, the Gini coefficient saturates to the upper bound. The parameter $F$ sets the height of plateau. As $F$ increases, the plateau is expanding and the valley is shrinking. In general, parameter $\alpha$ has more influence than parameter $\beta$. The
valley regime can be associated with a small $\alpha$. The plateau regime can be associated with a large $\alpha$. When $\beta$ is large, the transition between valley and plateau is smooth and along the line $\alpha = \text{const}$. The Matthew effect is controlled by $\alpha$ alone. When $\beta$ is small, the transition is sharp and along the line $\alpha + \beta = \text{const}$. Parameters $\alpha$ and $\beta$ are additive to the Matthew effect. As $\beta$ becomes smaller, the transition appears at a larger $\alpha$.

Figure 2. Gini coefficient $G$ as a function of $\alpha$ and $\beta$. (LEFT) $F = 0.3$. (MIDDLE) $F = 0.5$. (RIGHT) $F = 0.7$.

The phase transition can also be revealed by the variation of Lorenz curves. Fig. 3 (LEFT) shows the sharp transition at $\beta = 0$ ($F = 0.5$), where eleven curves are plotted at an equal-distanced $\Delta \alpha = 0.2$. Basically the curves are indistinguishable in the range $0 \leq \alpha \leq 1.2$. When $\beta \geq 1.6$, the curves collapse to the asymptotic line of $y = (1-F)x$. In contrast, Fig. 3 (MIDDLE) shows the smooth transition at $\beta = 2$. The curves collapse to the asymptotic result when $\alpha \geq 1$. Gradual changes can be observed in the range $0 \leq \alpha \leq 0.8$.

Figure 3. (LEFT) Sharp transition as $\alpha$ increases and fixed $\beta = 0$, where $\Delta \alpha = 0.2$ and $F = 0.5$. (MIDDLE) Smooth transition as $\alpha$ increases and fixed $\beta = 2$, where $\Delta \alpha = 0.2$ and $F = 0.5$. (RIGHT) Lorenz curves with various levels of unfunded scientists, where $F = 0.9$ and $\alpha = \beta = 0.55$.

4. Discussions
The new model has three parameters: $0 \leq F \leq 1$ (ratio of donation), $\alpha \geq 0$ (how to choose), and $\beta \geq 0$ (how to share). Parameter $F$ controls the scale of redistribution and sets an upper
bound to the Gini coefficient. Parameters $\alpha$ and $\beta$ control the strength of Matthew effect and move the system between valley and plateau regimes. It is expected that the concentration of funding can be effectively suppressed. The Gini coefficient can be easily controlled by adjusting the value of $F$. However, the desired results only present in the valley regime. The plateau regime is to be avoided. In the valley regime, the fluidity of funding is significant. The Lorenz curve is smooth. The funding distribution is continuous. All scientists are equal in regards to acquire different amount of funding. In contrast, in the plateau regime, the cumulative advantage is significant. The Lorenz curve collapses to the asymptotic result and has a sharp turn. The undue concentration of funding still happens swiftly, even though all scientists are treated equally and follow the well-intended regulations. The seemingly perfect regulations cannot prevent the undesired results from happening, where a minority of scientists take the majority of funding.

We introduce a stochastic process for the redistribution of funding. The more one has, the more one is asked to give away. With naive expectation, the accumulation of funding is totally discouraged. However, it is well known that funding is positively correlated to the research output. The positive feedback is controlled by the two parameters $\alpha$ and $\beta$. An unselfish scientist will give funding to support the best scientists. And it is also reasonable to expect that the donation is not shared equally, where a larger portion should be given to the better scientist. When the positive feedback is too strong, the undesired result appears. To control the concentration of funding, choosing randomly (small $\alpha$) has more effect than sharing equally (small $\beta$). Sharing biasly (large $\beta$) alone is not enough to trigger the transition. But choosing biasly (large $\alpha$) alone will trigger the transition.

As a final remark, we expect that this model can be used to describe the funding distribution in the current system. The mechanism of current funding system is very different from the presented model, especially in the details of operations. We believe that the model captures the two essential features of funding allocation: cumulative advantage and competition with each other. Either it’s a direction interaction among the scientists (as in the model) or an indirect interaction through a small committee (as in the current system), the parameterization of the model should be able to describe the data. The model assumes that all scientists in the community are funded. In the current system, not all scientists are funded. A recent study indicated that 18%-45% of researchers are unfunded [5]. Varied in disciplines, engineering is the lowest 18% and basic medical sciences is the highest 45%. To compare with empirical data, the model can be further modified to include a finite ratio of scientists without funding. The typical results are shown in Fig. 3 (RIGHT). By adjusting the ratio of unfunded scientists, the results are conform to the proverbial 80-20 rule, i.e., 80% of funds are concentrated in 20% of scientists, which implies a Gini coefficient $G \sim 0.6$. By assuming a large $F$ and locating $(\alpha, \beta)$ in the valley regime but near the transition, the model is expected to give a fair description of empirical data.

5. References

[1] Bollen J, Crandall D, Junk D, Ding Y, and Börner K 2017 Scientometrics 110 521.
[2] Bollen J, Crandall D, Junk D, Ding Y, and Börner K 2014 EMBO Reports 15 131.
[3] Gewin V 2014 Nature 507 265.
[4] http://retractionwatch.com/2016/09/20/what-if-scientists-funded-each-other/
[5] Larivière, V, Macaluso, B, Archambault, É, and Gingras Y 2010 Research Evaluation 19 45.