Rational Randomization by Journal Editors: A Mathematical Derivation

July 2019

Andrew J. Oswald
University of Warwick, UK and IZA, Bonn Germany
Email: andrew.oswald@warwick.ac.uk

Acknowledgements

I would like to acknowledge helpful discussions with, and ideas from, Amanda Goodall. I also thank the CAGE research group, which is supported by the ESRC, at the University of Warwick. Thanks also go to Ben Martin for helpful comments.

Abstract

Should journal editors and grant-giving bodies ever make use of random-draw mechanisms to make their final decisions? This Note argues that the answer is yes. It describes a mathematical rationale for such randomization. Put intuitively, random draw should be used when the gains from the acceptance of unorthodox path-breaking papers outweigh the costs of publishing the unorthodox poor papers. The mathematical case for this rests on an averaging argument and requires 'convexity' (of scientific influence) in a sense explained in the Note. The long-observed skewness of citations distributions is consistent with such convexity. Hence this Note attempts to offer a conceptual and practical addendum to the potentially important randomization proposal of Margit Osterloh and Bruno Frey.

Word count excluding abstract and references: 1800 approx.
Keywords: Citations; skewness; science; Jensen's inequality; academic journals.
Corresponding author: andrew.oswald@warwick.ac.uk.
Address: Professor of Economics and Behavioural Science, University of Warwick, Coventry CV4 7AL, United Kingdom.
Telephone: (+44) 02476 523510

Rational Randomization by Journal Editors: A Mathematical Derivation

Introduction

The idea that journal editors might usefully rely on random-draw methods has recently been proposed by Margit Osterloh and Bruno Frey (this issue of *Research Policy*). The authors propose that 'focal randomization' could be used in the selection of unorthodox papers that are inherently difficult to assess. Frey and Osterloh argue that such a step is now needed in academia. They suggest that conservatism tends to block important papers.

This Note provides a mathematical argument for such randomization. To understand the argument's logic does not require formal mathematical knowledge because it can be conveyed with the intuitive concept of averaging. Consider unconventional hard-to-evaluate papers that will eventually turn out to be either (i) intellectual breakthroughs or (ii) valueless. *If the path-breaking papers are many times more valuable than the poor papers are valueless, then averaging across them will lead to a net gain for society*. The plusses can be so large that the losses do not matter. This is a kind of convexity (of scientific influence). Averaging across the two kinds of papers, by drawing them randomly from a journal editor's statistical urn, can then be optimal.

It might be thought that this argument would run into a fatal counter-objection. Perhaps 'bad' papers are far more numerous than good ones, in which case it would be a mistake to randomize across the whole set of papers. Such a concern sounds sensible. But it misses an important point. The argument in this Note, reflecting the case put forward by Margit Osterloh and Bruno Frey, is that randomization would be used only for drawing articles from a final set of pre-selected and potentially strong papers that have been chosen for inclusion in a hard-to-evaluate 'focal' pool. Randomization would not be used across the entire set of submissions to a journal.

The current Note relies on an appeal to convexity and a 100-year-old mathematical theorem due to the late Johan Jensen. The derivation is given later in this Note. In the practical world of academic publishing, there is evidence consistent with the key empirical requirement -- convexity of scientific influence -- for the rationale for randomization. It is known that there is convexity in citations data. Path-breaking papers produce hugely disproportionate numbers of citations (see e.g. Nederhof and Van Raan, 1993; Goodall, 2009), while many regular papers are hardly mentioned again in the subsequent literature.

Although almost any scientific journal or period would suffice to show that convexity, I have collected, for illustrative purposes, some data on Research Policy articles published in the year 2000. (The year 2000 was chosen as a round number and to give papers a reasonable time to be recognized and cited.) The six most-cited articles in that year have citations totals as follows: 2152, 846, 577, 544, 521, 521.[1] The four least-cited articles (ignoring letters, notes, and book reviews as best I can) have citations totals of 5, 5, 3, 2.[2] The non-linear convexity captured in these different numbers could be depicted as a curved power-law on a graph and would have the standard characteristics of a Lotka's Law distribution (Lotka 1926). The disproportionality here in the citations numbers is marked. My examination of the data suggests that in a citations sense – and admittedly that is just one criterion – the most-cited Research Policy article in 2000 was worth the same as the combination of the lower half of the journal's remaining articles in that year. That is an extreme case, of course, but it serves as an illustration of the non-linearity of 'influence', one might say, in the world of academic research.

The formal argument

Assume that a journal editor wants to attract subscribers for next year and beyond. Let the representative subscriber have an increasing and twice-differentiable utility function $v = v(q)$ where q is the quality of the articles published[3]. In principle, q is a vector, because there will be many articles published in the journal's future, but without loss of generality the variable q will be treated here as a scalar.

There are then two situations to consider. One is where the journal editor randomizes the quality of articles, so that q is a random variable. A second is where the journal editor sticks to a reliable known quality, so that q is fixed at, say, q^*, and any hard-to-evaluate articles are summarily rejected. Both of these are stylized extremes (neither would probably be implementable in a strict sense when compared to the complications of the real world), but they are helpful simplifications to set out the argument.

As there is potential uncertainty, assume that decision-makers care about their expected utility. Let E be the expectations operator. Then the cases are respectively:

[1] Search dated 5 May 2019 on the Web of Science.
[2] This includes articles that earned zero citations.
[3] A reasonable question to ask would be: what does utility actually mean here? One answer would be: the scientific influence, and thus ultimately the lasting scientific value, of a journal article.

Ev(q) = expected utility of the journal subscriber when article quality has a random component

v(q) = expected utility of the journal subscriber when article quality is fixed at q*.*

It is necessary to do a fair-minded comparison between the two journal systems. One natural approach is to assume that q^*, the fixed level of journal quality in a deterministic world, is simply equal to the mean of the quality distribution, Eq, in the other world that has randomness. Thus assume $q^* = Eq$. This ensures a kind of like-for-like comparison between the underlying quality of submissions in the two worlds.

A translation of this mathematical equality into simple English would go as follows. Assume, as Osterloh and Frey effectively do, that editors and referees are capable of rejecting a large number of demonstrably bad papers, as well as identifying some demonstrably good papers that should definitely be published. The difficulty for an academic journal comes in the grey zone in between – that is, in the case of the remaining interesting submissions that are hard to evaluate but might (and might not) go on to be important contributions to science. Randomization is a way of avoiding the need to reject all the submissions that are hard to evaluate. According to the randomization notion, the hard-to-evaluate submissions could be put into a pool, and then a random draw could be made from that pool. The equality assumption of $q^* = Eq$ formalizes the notion that, *ex ante*, the papers in the pool are known to be of the same average quality as those identifiable as of publishable quality q^*, but that it is not possible in advance to tell which paper is which. That equality is what it means to make a fair-minded comparison[4] between a world with and without randomization.

However, when would randomization actually be optimal for a journal? If the equality assumption of $q^* = Eq$ is made, we have, after substitution, the following utility difference between the two worlds:

The utility gain from randomness $= Ev(q) - v(q^*) = Ev(q) - v(Eq)$

This allows Jensen's inequality theorem to be applied.

Jensen's theorem states that $Ev(q) - v(Eq)$ is positive when $v(.)$ is strictly convex, is negative when $v(.)$ is strictly concave, and is zero when at the border between the two. The proof is given in Jensen (1906) and in standard mathematics textbooks; it is not repeated here.

[4] If lots of bad papers were deliberately included in the pool, then obviously that would not be a fair-minded comparison, and random draw would in that case be less desirable.

Graphically, Jensen's result can be thought of as arising from the fact that in a graph with a convex curve any linear segment between two points has to lie above the curve. Geometrically, this is the notion that randomized mixtures are beneficial. They average across 'good' and 'bad' outcomes, and the good ones are disproportionately good.

A final technical point should be noted. It might be that, even with the convexity apparent in citations data, it could be that the function $v(.)$ is not convex. This might occur if the typical subscriber to the journal is highly risk-averse. Say that the subscriber to the journal has a utility function that is given by $v = v\,(u(c(q))$, where $u(.)$ is strictly concave even though a citations function defined as $c(.)$ is strictly convex. Then it is possible that $u''(.)$ is so large and negative that it outweighs the influence of the second derivative of the $c(.)$ function, whereupon $v(.)$ could still, overall, be a concave function. This would correspond to the putative case of a journal that could get enormous bad publicity from ever publishing a poor article, although it is perhaps not obvious that such a case is empirically likely.

Discussion

In concluding, randomization by editors can be rational. The mathematical argument given here is an application of Jensen's Inequality and rests on a particular kind of convexity, namely, the convexity of scientific influence.

We observe evidence of convexity in academia. Major papers in scientific research are far, far more valuable than papers of modest quality. Think of an accelerating upward-sloping curve in a diagram with scientific influence on the vertical axis and quality on the horizontal axis. That accelerating curvature is what make this Note's averaging argument work mathematically. The attraction of randomization is that it automatically averages across brilliant and bad papers, so that it is desirable to randomize when the brilliant papers are disproportionately valuable and when the bad papers do not matter too much[5]. The broad idea here could be used in different scholarly settings (as discussed, for example, in Goodall and Osterloh 2016).

Randomization would not be necessary if editors and referees had perfect foresight; but of course they do not. Some way has to be chosen for deciding how to deal with unusual papers that might turn out to be breakthroughs and might turn out to be worthless.

---

[5] Papers that are poor, or of minor importance, can be discarded by future researchers, as indeed seems to happen if historical citations data are examined. This idea might be thought to be inconsistent with the Ortega hypothesis (Cole and Cole 1972), but that would require a longer discussion than is feasible here.

It may be helpful to end on the reasons why scientific academia is different from most industries. Consider the case of automobile manufacturing compared to research journals. Automobile factories are keen to avoid randomness in their product. They aim for a known and exact level of quality q^*. If I buy a motor car, I may be pleased if I get a perfect one, where there was not a single flaw in the paint on the day that it left the factory, but the difference between a nearly-perfect new car that was made late on a rainy Friday afternoon and a perfect one made after lunch on a sunny Tuesday afternoon is likely to be minor. In the case of automobiles, the gain from quality can therefore be expected to be concave. Jensen's inequality then predicts, for automobiles, that consumers will want constancy not variability. The same goes for purchasers of Kleenex tissues and cans of baked beans. A steady homogeneity of the product is what consumers want.

However, academia is not of that kind. Steadiness and homogeneity are undesirable in the production of new scientific thought. Although often hard to assess *ex ante*, the best journal articles tend to be highly unorthodox and they go on to be many, many times more influential than even reasonably good articles. It is convexity of this sort that provides the conceptual case for randomization by journal editors.

References

Cole, J.R., and Cole, S. (1972). The Ortega hypothesis. *Science*, 178, 368-375.

Goodall, A. (2009). *Socrates in the Boardroom: Why Research Universities Should be Led by Top Scholars*. Princeton University Press, Princeton and Oxford.

Goodall, A., and Osterloh, M. (2016). Random selection as a way to increase the supply of female leaders. Working paper, Cass Business School, City University, London.

Jensen, J. L. W. V. (1906). Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica* 30 (1), 175–193.

Lotka, A.J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences* 16 (12), 317–324.

Nederhof, A.J., and Van Raan, A.F.J. (1993). A bibliometric analysis of 6 economics research groups: A comparison with peer-review. *Research Policy* 22(4), 353-368.

Osterloh, M., and Frey, B. (2019). How to avoid borrowed plumes in academia. *Research Policy* (this issue).