

Research Evaluation in Economic Theory and Policy: Identifying and Overcoming Institutional Dysfunctions

Thomas Ferguson (Institute for New Economic Thinking),

Robert Johnson (Institute for New Economic Thinking)

May 25, 2018 | Last updated: June 19, 2018

Proposal

The problem this essay addresses can be framed in terms of two quotations from Alexis de Tocqueville. The first comes from his famous speech in the French Chamber of Deputies just prior to the outbreak of the Revolution of 1848: "We are sleeping on a volcano....do you not see that the earth is beginning to tremble. The wind of revolt rises; the tempest is on the horizon." The second is from *Democracy in America*: "When the past no longer illuminates the future, the spirit walks in darkness."

In 2018, the darkness is all too palpable: A chain of economic reverses that no prominent economists, central bankers, or policymakers anticipated has combined with other shocks from technology, wars, and migrations to produce the political equivalent of the perfect storm. The world financial meltdown of 2008 set the cyclone spinning. As citizens watched helplessly while their livelihoods, savings, and hopes shriveled, states and central banks stepped in to rescue the big financial institutions most responsible for the disaster. But recovery for average citizens arrived only slowly and in some places barely at all, despite a wide variety of policy experiments, especially from central banks.

The cycle of austerity and policy failure has now reached a critical point. Dramatic changes in public opinion and voting behavior are battering long entrenched political parties in many countries. In many of the world's richest countries, more and more citizens are losing faith in the very ideas of science, expertise, and dispassionate judgment – even in medicine, as witness the battles over vaccines in Italy, the US, and elsewhere. The failure of widely heralded predictions of immediate economic disaster when the UK voted to leave the European Union and Donald Trump became President of the United States has only fanned the skepticism.

Placing entire responsibility for this set of plagues on bad economic theory or deficient policy evaluation does not make sense. Power politics, contending interests, ideologies, and other influences all shaped events. But from the earliest days of the financial collapse, reflective economists and policymakers nourished some of the same suspicions as the general public. Like the Queen of England, they asked plaintively, "Why did no one see it coming?"

Answers were not long in arriving. Critics, including more than a few Nobel laureates in economics, pointed to a series of propositions and attitudes that had crystallized in economic theory in the years before the crisis hit.^[1] Economists had closed ranks as though in a phalanx, but the crisis showed how fragile these tenets were. They included:

1. A resolute unwillingness to recognize that fundamental uncertainty shadows economic life in the real world.
2. Neglect of the roles played by money, credit, and financial systems in actual economies and the inherent potential for instability they

create.

3. A fixation on economic models emphasizing full or nearly complete information and tendencies for economies either to be always in equilibrium or heading there, not just in the present but far into the indefinite future.
4. A focus on supply as the key to economic growth and, increasingly after 1980, denials that economies could even in theory suffer from a deficiency of aggregate demand.
5. Supreme confidence in the price system as the critical ordering device in economies and the conviction that getting governments and artificial barriers to their working out of the way was the royal road to economic success both domestically and internationally.

Initially, debates over this interlocking system of beliefs mostly sparked arguments about the usefulness of particular tools and analytical simplifications that embodied the conventional wisdom: Dynamic stochastic general equilibrium models; notions of a “representative agent” in macroeconomics and the long run neutrality of money; icy silence about interactions between monetary rates of interest and ruling rates of profit, or the failure of labor markets to clear.

Increasingly, however, skeptics wondered if the real problems with economics did not run deeper than that. They began to ask if something was not radically wrong with the structure of the discipline itself that conduced to the maintenance of a narrow belief system by imposing orthodoxies and throwing up barriers to better arguments and dissenting evidence.

The empirical evidence now seems conclusive: Yes.

1. “Top 5” Dominance for Promotion and Tenure

Studies by James Heckman demonstrate the critical gatekeeping role of five so-called “top journals” in recruitment and promotions within economics as a field.^[2] Four of the journals – the *American Economic Review*, the *Journal of Political Economy*, the *Quarterly Journal of Economics*, and the *Review of Economic Studies* – are Anglo-American centered and published in the US or the UK as is the fifth, *Econometrica*, though it is sponsored by the Econometric Society, which has long involved scholars from Scandinavia and other countries.

Heckman’s research shows that the number of Top 5 (T5) articles published by candidates plays a crucial role in the evaluation of candidates for promotion and tenure. This is true not only in leading departments but more generally in the field, though the influence of the count weakens in lower ranked institutions.

2. The Great Disjunction

Heckman compares citations in Top 5 journals with articles frequently cited by leading specialists in various fields and with publication histories of Nobel laureates and winners of the Clark Medal. He is crystal clear that many important articles appear in non-T5 journals – a finding supported by other studies.^[3] This evidence, he argues, highlights a “fundamental contradiction” within the whole field: “Specialists who themselves publish primarily in field journals defer to generalist journals to screen quality of their colleagues in specialty fields.”

3. Citations as Pernicious Measures of Quality in Economics

Heckman draws attention to the increase in the number of economists over time and the relative stability of the T5. He argues that his findings imply that the discipline’s “reluctance to distribute gatekeeping responsibility to high quality non-T5 Journals is inefficient in the face of increasing growth of numbers of people in the profession and the stagnant number of T5 publications.”

Other scholars who have scrutinized what citations actually measure underscore this conclusion. Like Heckman, they know that citation indices originated from efforts by libraries to decide what journals to buy. They agree that transforming “*journal* impact factors” into measures of the quality of individual *articles* is a grotesque mistake, if only because of quality variation within journals and overlaps in average quality among them. Counts of journal articles also typically miss or undercount books and monographs, with likely serious effects on both individual promotion cases and overall publication trends in the discipline. As Heckman observes, the notion that books are not important vehicles for communication in economics is seriously mistaken.

Analytical efforts to explain who gets cited and why are especially thought provoking. All serious studies converge on the conclusion that

raw counts can hardly be taken at face value.^[4] They distort because they are hopelessly affected by the size of fields (articles in bigger fields get more citations) and bounced around by self-citations, varying numbers of co-authors, "halo effects" leading to over-citation of well-known scholars, and simple failures to distinguish between approving and critical references, etc. One inventory of such problems, not surprisingly by accounting professors, tabulates more than thirty such flaws.^[5]

But cleaning up raw counts only scratches the surface. Heckman's study raised pointed questions about editorial control at top journals and related cronyism issues. Editorial control of many journals turns over only very slowly and those sponsored by major university departments accept disproportionately more papers from their own graduates.^[6] Interlocking boards are also fairly common, especially among leading journals.^[7] Carlo D'Ippoliti's study of empirical citation patterns in Italy also indicates that social factors within academia figure importantly: economists are prone to cite other economists who are their colleagues in the same institutions, independently of the contents of their work, but they are even more likely to cite economists closer to their ideological and political positions.^[8] Other research confirms that Italy is not exceptional and that, for example, the same pattern shows up in the debates over macroeconomics in the US and the UK after 1975.^[9]

Other work by Jakob Kapeller, et al., and D'Ippoliti documents how counting citations triggers a broad set of pathologies that produces major distortions.^[10] Investing counts with such weighty significance, for example, affects how both authors and journal editors behave. Something uncomfortably close to the blockbuster syndrome characteristic of Hollywood movies takes root: Rather than writing one major article that would be harder to assimilate, individual authors have strong incentives to slice and dice along fashionable lines. They mostly strive to produce creative variations on familiar themes. Risk-averse gatekeepers know they can safely wave these products through, while the authors run up their counts. Journal editors have equally powerful incentives: They can drive up their impact factors by snapping up guaranteed blockbusters produced by brand names and articles that embellish conventional themes. Kapeller, et al. suggest that this and several other negative feedback loops they discuss lead to a form of crowding out, which has particularly pernicious effects on potential major contributions since those are placed at a disadvantage by comparison with articles employing safer, more familiar tropes.^[11] The result is a strong impetus to conformism, producing a marked convergence of views and methods.

These papers, and George Akerlof in several presentations, also show that counting schemes acutely disadvantage out-of-favor fields, heterodox scholars, and anyone interested in issues and questions that the dominant Anglo-Saxon journals are not.^[12] This holds true even though, as Kapeller et al. observe, articles that reference some contrary viewpoints actually attract more attention, conditioning on appearance in the same journal – an indication that policing the field, not simply quality control, is an important consideration in editorial judgment. One consequence of this narrowing is its weirdly skewed international impact. Reliance on the current citations system originated in the US and UK, but has now spread to the rest of Europe and even parts of Asia, including China. But T5 journals concentrate on articles that deal with problems that economists in advanced Anglo-Saxon lands perceive to be important; studies of smaller countries or those at different stages of development face higher publication hurdles. The result is a special case of the colonial mind in action: economics departments outside the US and UK that rely on "international" standards advantage scholars who focus their work on issues relevant to other countries rather than their own.

4. Citation Counting and Women Economists

The long, dismal history of women's engagement with economics as a field has finally attracted anguished attention. The kernel of that history is easy to summarize: until recently, talk of glass ceilings was a stretch, because so few women could be found anywhere on the floor.

This peculiar history means that sifting out the role citations play in the bigger picture is inevitably complicated, especially as women enter the field in larger numbers. Giulia Zacchia has compared the publication records and citation patterns of men and women economists. Her conclusion is stark: economics is an "environment in which to reach top academic roles—or sometimes any academic position—women economists are increasingly forced to conform to research activities and publishing habits of their male colleagues ...The tendency to assess research quality based on standardized bibliometrics reveals a dual path of convergence and conformity: i) a progressive reduction in the variety of research interests of women and men economists; ii) a tendency to converge to international standards of perceived research excellence...These phenomena reveal a consistent reduction in diversity in economics, and more broadly the pluralism of research."^[13] Her work and other research by Marcella Corsi and Carlo D'Ippoliti suggest that the T5 system puts up barriers to distinctively feminist expression in economics.^[14] As a result, a significant number of women economists, especially at the top, have react by copying the males as they make their way in the discipline: "Gender homologation was stronger for full and associate professors than it was for Ph.D. students – demonstrating how institutional changes can produce indirect discrimination effects."^[15]

5. What Can Be Done?

All talk of remedies needs to begin by acknowledging two sovereign facts. Firstly, as Alberto Baccini and other analysts have emphasized, systems of evaluation trigger many sorts of schemes to game the system.^[16] The situation is exactly like bank regulation as summarized in "Goodhart's Law": Actions by regulators to rein in banks quickly induce innovations to take advantage of the rules. This does not mean that regulation is hopeless, but that proposed remedies need careful scrutiny and should never be viewed as once and for all fixes.

Secondly, any idea that some formula can completely supplant discussions and assessments of individual work is delusory. The point is powerfully driven home by Kapeller and Steinerberger. They set up an agent-based model of what a perfect refereeing process look likes in a hierarchical system of publication outlets. Then they ask how acknowledging that referees sometimes make mistakes changes things. Their conclusion is eye opening: even small amounts of refereeing error in the system lead to more sweeping changes in the system as a whole than anyone likely suspected. First-rate papers end up at the bottom of the pile; top journals accept substantial numbers of dogs and good papers scatter all around. Their results amount to a yellow caution flag against the idea of taking pecking orders in journals too seriously for any important decisions.

But given the clear evidence that the volcanoes are steaming, it would be foolhardy not to try to do better. We are convinced that economics as a discipline still has much to offer the world and that restoring its standing in public life is much to be desired. But this will be something of an uphill climb. Things cannot go on as they have. The ground beneath our feet is shifting, just as in de Tocqueville's time. Another five years of more of the same is likely to produce far more serious reactions that will make current controversies about expertise look quaint.

Remediation is required at three different levels: The first concerns the way economics as a discipline proceeds; the second relates to how economic theories feed into economic policymaking; and the third concerns the role of women and diversity.

Regarding the progress of the discipline, an obvious first step is required: The G20 should insist that individual national academies of science place the problem of evaluation of research and personnel high on their respective agendas. The problems highlighted here are not necessarily limited to economics, and some national academies have become concerned with methods used in research evaluation. But within economics the issues appear especially clear cut and pressing, because of the link with policymaking and ultimately the wellbeing of the population. National academies should be adjured to hold *public* hearings and conferences on what went wrong, with both theory and policymaking. They should solicit comments from a wide range of experts, scholars, and even the general public. The results of these assessments should be made public and exchanged among the academies. The end product should be convergence of agreement on specific policy failures and clear inventories of the (mistaken) economic theories that were enlisted to support them. The national academies should take the lead in communicating these results to the media and large funders of research in each country. If these constituencies better understood how shaky the evidence about citations is and the relatively low effectiveness of existing professional screening protocols, pressure for improvement would intensify. As it becomes clear that methods of evaluation can be improved relatively easily, the will to improve the system would intensify.

Possible reforms of economics have stimulated widespread discussion, but produced a wide dispersion of views. What to do about journals poses an especially thorny problem. The T5 system bears an uncomfortable likeness to a classic cartel system, which several US and UK university graduate departments quite obviously support to their own advantage and which provide convenient places for outside interests to wield influence.

More discussions of the best way to open up the system are needed. Some analysts are impressed with the success of the open source publication movement in some of the physical sciences. They aspire to replace the T5 system with a broad set of open access journals on the internet. In some proposals, comments and reactions to articles would be published alongside original contributions in real time.

Not everyone regards this as feasible or desirable nor is it clear how much would materially change were it implemented. Some analysts take the ability of a few mathematicians to display their prowess by solving publicly posted problems on the internet with no journal intervention at all as evidence that journals could be downgraded tout court. But the number of scholars who succeed in solving these problems is few and all of them together are insufficient to structure the kinds of real life scientific communities required for modern countries to flourish. In practice, these open challenge schemes thrive inside a larger ecology of expertise that maintains itself in universities, government laboratories, research institutes, etc. Skeptics also shudder to recall Thomas Hobbes' famous comment that $2+2=4$ would be disputed if it affected someone's interest. They see no reason why on-line mobbing of the type now familiar in socially contentious discussions on the

affected someone's interest. They see no reason why on-line mobbing of the type now familiar in socially contentious discussions on the internet would not occur within science practiced along such lines. At the turn of the twentieth century, national rivalries in medicine, physics, and even mathematics were sometimes intense. Today, though, the number of interested private actors has likely increased. The global warming debate might be considered a kind of warning. That under pressure macroeconomics tends to dissolve into extensively solipsistic communities that cite communicants relatively heavily hardly dampens such fears.

By contrast, practical improvements to personnel review are perhaps easier to imagine and might effectively go more quickly to the roots of the problem. One reason for the popularity of the T5 system is that it requires review committees and deans to make little more than a pretense of reading candidates' research. One simply totes up the number of articles, weights each by journal impact figures, and arrives at a formal or informal total score. End of discussion. Fearful defenders of the status quo cite the bulge in candidates and streaming proliferation of articles as reasons to stick with this procedure, despite its obvious lack of intellectual merit.

An approach that is viable even with relatively large numbers of candidates might be one based on the system used for many years in the MIT Economics Department. Likely rooted in an old aphorism of Paul Samuelson's that in science one honors peak, not average, performance, this approach functions by having candidates put forward their three best publications for evaluation. Members of departments who would never be able to review the entire oeuvre of candidates can fully participate. The process could be improved by contriving some auxiliary safeguards for special circumstances. Such processes could also employ a set of different citation indices, sifted to remove obvious foolishness (such as the failure to control for field size), as one form of evidence among others.

Another easy improvement could be a flat ban on the use of "point scales" or scorecard tallies of minimum numbers of articles in journals of different types as conditions for promotion. A number of countries now require something like this, despite the evidence brought forward by Heckman, Baccini, Corsi, et al., and others.

Some reforms are similarly easy to envisage with respect to the application of economic theories in economic policy. The case for more pluralism is compelling. There is no reason why the US defense establishment should be the only organization that sets up formally constituted "Team B's that deliver alternative, officially reported assessments. It is obvious that since 2008 both governments and international organizations have been carried away with a mania for secrecy about their operating economic projections and the economic theories they are trying to implement. These should be opened up to formal discussions, similar to what used to happen in testimony in the U.S. Congress and regulatory commissions.^[17]

We venture to suggest that a close reading of the debates over financial regulation in the run up to the 2008 collapse will readily reveal the lengths to which authorities and collaborating economists went to stifle public debate. The guiding idea of efforts to reverse this state of affairs would not be to say that particular economic theories are right and or wrong but to assess whether the economic theories being applied are biased because they have never seriously been questioned. To win back public confidence, economists need to justify and support their ideas. That can happen only if we guarantee pluralism, so that they all are forced to argue with one another, instead of retreating to silos.

Making economics safe for diversity, of course, also implies improving the situation of women in economics. Many of the most relevant reforms, however, concern the structure of the discipline as a whole, not better uses of citations per se. It is obviously fanciful to imagine that much can be done without substantially increasing the number of women at all levels of the profession and securing their representation on all journal boards and committees that pass on research. That seems patent. It would also help if the discipline routinely calculated indices of gender equity in the economics profession (e.g., a glass ceiling index) to monitor and ensure gender-neutrality of research institutions by looking at hiring procedures, research assessments, and other strategic venues for the advancement of women, and increasing awareness of unconscious bias. All this, of course, presupposes workplaces in which sexual harassment is not tolerated and in which not simply talk, but action occurs to remedy the problems as they come to light.

[1] Nobel Prize winning economists who have recently appeared on panels critical of the way work in economics is now evaluated include George Akerlof, Angus Deaton, Lars Hansen, and James Heckman. See, e.g., "Taking the Con Out of Economics: The Limits of Negative Darwinism," Panel Presented at the Institute for New Economic Thinking Conference in Edinburgh, Scotland, October 2017, available on the web at <https://www.ineteconomics.org/conference-session/taking-the-con-out-of-economics-the-limits-of-negative-darwinism> and "Publishing and Promotion in Economics: The Curse of the Top 5," Panel Presented at the American Economic Association Annual Meeting, Philadelphia, January 2017; on the web at <https://www.aeaweb.org/webcasts/2017/curse>

[2] James J. Heckman and Sidharth Muktan, "Publishing and Promotion in Economics: The Curse of the Top 5," Paper Presented at the Institute for New Economic Thinking Plenary Conference in Edinburgh, Scotland, October, 2017; on the web at <https://www.ineteconomics.org/uploads/downloads/Heckman-Presentation-Publishing.pdf>

See also the discussion of concentration over time in Florentin Glotzel and Ernest Aigner, "Six Dimensions of Concentration in Economics: Scientometric Evidence from a Large Scale Data Set," Institute for Ecological Economics, Vienna University of Economics and Business, Working Paper No. 15, Year 3, 2017; on the web at http://epub.wu.ac.at/5488/1/EcolEcon_WorkingPaper_2017_15.pdf

[3] We are grateful to Alberto Baccini, who shared with us some counts from his own data for the period 2009 to 20018; the less than commanding position of the T5 is apparent.

[4] See, e.g., Todeschini, R. and A. Baccini, *Handbook of Bibliometric Indicators. Quantitative Tools for Studying and Evaluating Research*. (Weinheim, Germany: Wiley-VCH, 2016).

[5] Alan Reinstein, James R. Hasselback, Mark E. Riley, David H. Sinason, "Pitfalls of Using Citation Indices for Making Academic Accounting Promotion, Tenure, Teaching Load, and Merit Pay Decisions," *Issues in Accounting Education* 26, No. 1, pp. 99-131.

Studies also show low rates of agreement between citation counts and the judgments of the relevance of papers by peer reviewers. See the discussion in Wouters, P. et al. (2015), *The Metric Tide: Literature Review (Supplementary Report I to the Independent Review of the Role of Metrics in Research Assessment and Management)*, HEFCE. DOI:

10.13140/RG.2.1.5066.3520; on the web at http://www.dcsience.net/2015_metrictideS1.pdf

[6] Cf. also Colussi T., "Social Ties in Academia: A Friend is a Treasure", *The Review of Economics and Statistics*, Vol 100, Issue 1 (2017), pp. 45-50.

[7] Alberto Baccini and Lucio Barabesi, "Gatekeepers of Economics: the Network of Editorial Boards in Economic Journals," in Lanteri, A. Vromen, J. eds., *The Economics of Economists* (Cambridge: Cambridge University Press, 2014), pp. 104-150.

[8] Carlo d'Ippoliti, "Many Citedness': Citations Measure More Than Just Scientific Impact," Institute for New Economic Thinking Working Paper No. 57, Revised April 6, 2018; on the web at <https://www.ineteconomics.org/research/research-papers/many-citedness>

[9] See Lasse Folke Henriksen, Leonard Seabrooke, and Kevin Young, "Fathers of Neoliberalism: The Academic and Professional Performance of the Chicago School, 1960-85," Paper Presented at Institute for New Economic Thinking Conference, Edinburgh, Scotland, October, 2017; on the web at <https://www.ineteconomics.org/research/research-papers/fathers-of-neoliberalism>

[10] Jakob Kapeller and Stefan Steinerberger, "Emergent Phenomena in Scientific Publishing: A Simulation Exercise," *Research Policy* 45 (2016), pp. 1945-52; Christian Grimm, Stephan Pühringer and Jakob Kapeller, "Paradigms and Policies: The State of Economics in the German-Speaking Countries," Institute for Comprehensive Analysis of the Economy, University Linz, ICAE Working Paper Series, No. 77 – March 2018; but especially Matthias Aistleitner, Jakob Kapeller, and Stefan Steinerberger, "The Power of Scientometrics and the Development of Economics," Institute for Comprehensive Analysis of the Economy, University Linz, ICAE Working Paper Series, No. 46 – March 2016 (Updated: August 2017).

[11] See the papers cited supra.

[12] See, e.g., Akerlof's "Sins of Omission," Paper Presented at the Institute for New Economic Thinking Plenary Conference, Edinburgh, Scotland, on the web at <https://www.ineteconomics.org/uploads/downloads/AKERLOF-Presentation.pdf> and Lee, F. S., X. Pham and G. Gu, "The UK Research Assessment Exercise and the Narrowing of UK Economics." *Cambridge Journal of Economics* 37, 4 (2013) : 693-717.

[13] Giulia Zacchia, "The Dark Side of Discrimination in the Economics Profession," Institute for New Economic Thinking Blog, November 3, 2017; available on the web at <https://www.ineteconomics.org/observatory/blog/the-dark-side-of-discrimination-in-the-economics>

2017; available on the web at <https://www.ineteconomics.org/perspectives/blog/the-dark-side-of-discrimination-in-the-economics-profession> See also her "Diversity in Economics: A Gender Analysis of Italian Economic Production," Institute for New Economic Thinking, Working Paper No. 61; available on the web at <https://www.ineteconomics.org/research/research-papers/diversity-in-economics> For a study strongly suggesting that different standards are applied by reviewers and editors to publications by women economists, see Erin Hengel, "Evidence From Peer Review That Women Are Held to Higher Standards," *VOX*, December 22, 2017; on the web at http://www.dcsience.net/2015_metricideS1.pdf

[14] Marcella Corsi, "Diversity and the Evaluation of Economic Research: the Case of Italian Economics," Paper presented at the Institute for New Economic Thinking Conference, Edinburgh, Scotland, October 2017; available on the web at <https://www.ineteconomics.org/uploads/papers/CORSI-Diversity-and-the-Evaluation-of-Economic-Research.pdf>

Marcella Corsi, Carlo d'Ippoliti, and Giulia Zacchia, "How Academic Conformity Punishes Women – and Restricts the Diversity of Economic Ideas," Institute for New Economic Thinking Blog, December 14, 2017, available on the web at <https://www.ineteconomics.org/perspectives/blog/how-academic-conformity-punishes-women-and-restricts-the-diversity-of-economic-ideas>

[15] Zacchia, "The Dark Side."

[16] See, e.g., Edwards, M. A. and S. Roy, "Academic Research in the 21st Century: Maintaining Scientific Integrity in a Climate of Perverse Incentives and Hypercompetition," *Environmental Engineering Science* 34, 1 (2017) : 51-61; and Smaldino, P. E. and R. McElreath, "The Natural Selection of Bad Science," *Royal Society Open Science* 3, 9 (2016).

[17] Those continue to happen, but they now play little role in most policy formation; the end of official support for many legislative caucuses is a related factor.

Existing Initiatives & Analysis