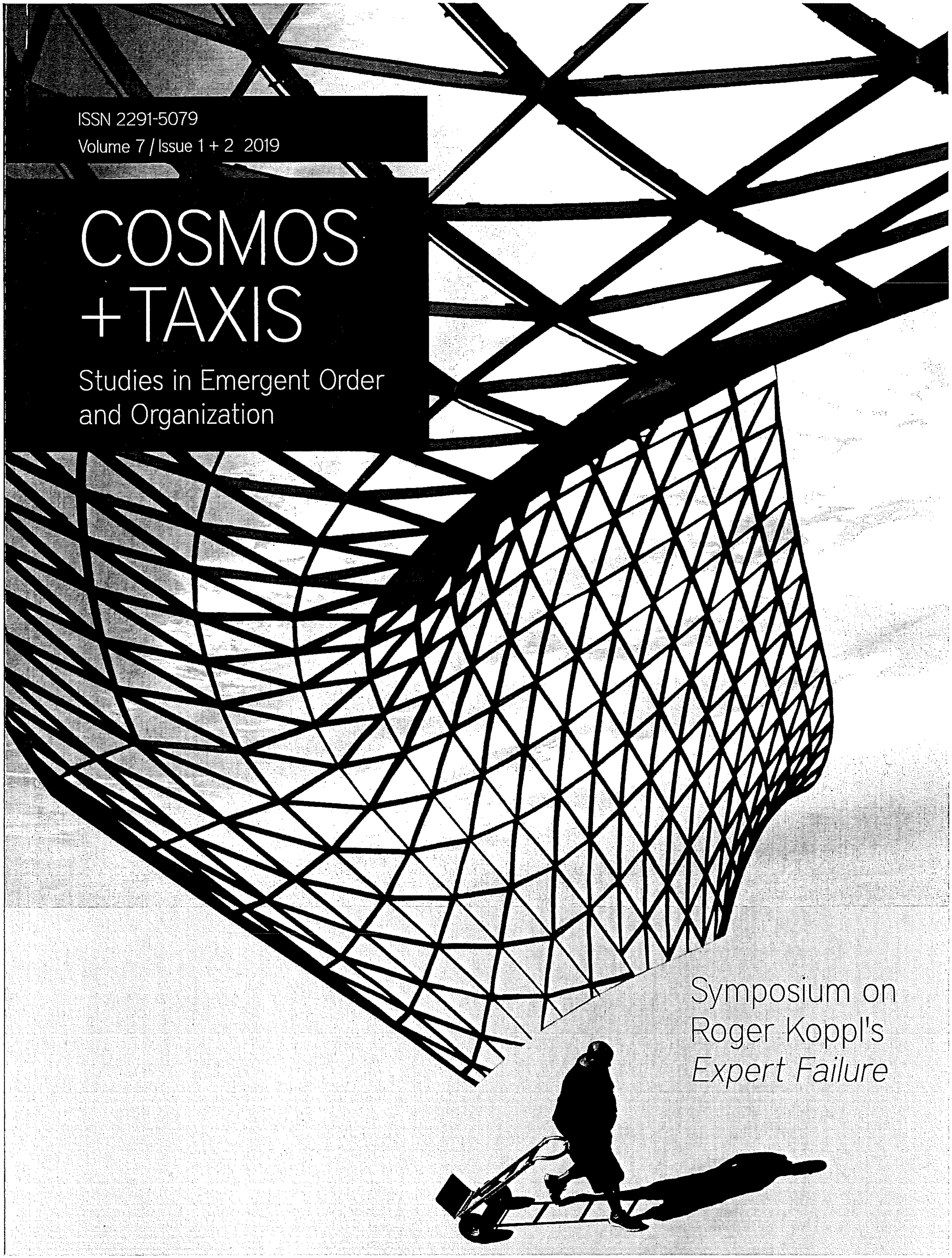
ISSN 2291-5079
Volume 7 / Issue 1 + 2 2019

COSMOS + TAXIS

Studies in Emergent Order and Organization



Symposium on Roger Koppl's *Expert Failure*

Massive Error

STEPHEN TURNER

Email: turner@usf.edu
Web: http://faculty.cas.usf.edu/sturner5/

Roger Koppl's book points to the big conundrum in dealing with expertise: what can we say about massive expert error? With individual errors, we can use the conventional, consensus, standards to judge something to be erroneous, knowing that the conventional standards may be wrong. We can poll experts. We can wait for claims to be refuted, and for the refutation to be generally accepted. Or we can wait for history to show that the consensus was wrong—perhaps a very long wait. Or we can make a personal judgment, against the consensus. There might be many grounds for doing so, involving something we know or think we know, either about the subject matter or about the mode in which it was produced, which we may regard as flawed. Massive error is more difficult to handle. Some of the same considerations apply, but in a more problematic way. In the case of massive error, we can't consult the consensus, or the conventional standards, because they are going to be part of the error. We can't say the mode of production was problematic unless we have some idea of what the right mode is.

Koppl provides a kind of model of the circumstances in which expert error is likely to be reduced, and, on the other side, an account of the dangers of certain kinds of institutional arrangements that elevate the possibility of error. The model of the best circumstances has some characteristics of an open market; the worst has characteristics of markets that are closed because there is only one seller or buyer, or which has insurmountable barriers to entry. This is a model with clear antecedents in Michael Polanyi ([1951] 1980, pp. 32-48, 49-67; Turner 2005), who stressed the importance of scientists having powers of autonomous decision-making about what topics, methods, and lines of inquiry to pursue. So it is instructive to think about some of the issues with this account, and some of the problems with assessing what actually happened. Polanyi's model was one in which all qualified investigators would receive support, but without direction, so that they are allowed to follow their best scientific instincts.

Polanyi ignored rewards, the problem of recognition of achievements—which Polanyi was nevertheless aware of—and the problem of deciding who was qualified. His policy was nevertheless grounded in experience, Polanyi's own experience in running a Kaiser Wilhelm Gesellschaft lab in Germany in the 1930s, and his later experience in Manchester, where the university itself provided support. It was precisely this model that was obliterated in the post-war period by the rise of big science and the predominance of the project grant system. This made "autonomy" into a fiction, a fiction that had to be maintained by beliefs along the lines that funding and other academic decisions were based solely on merit, omniscience about the future pay-offs of research, and so forth—that the omnipresent decision-making that goes on in every nook and cranny of science had no effect on the content of science or the path that scientific discovery took.

In Koppl's terms, the restriction to qualified investigators is a barrier to entry. The selective allocation of funds, similarly, is a barrier to entry for those who were not given funds. Polanyi struggled to find a way to explain the role of granting agencies—at which involved what he called the Influentials ([1951] 1980, p. 54). But unless the Influentials are doing what the market would do, which is to say serve as something like the omniscient planners of socialist theory, even if they limit their role to talent-spotting, their existence undermines the market analogy. Moreover, we know that the decision makers are not omniscient planners. Bernard Lovell, the radio-astronomer and Nobel Laureate, campaigned against funding for DNA research, because he thought it was oversold and not as important as his own work. Conflicts of interest and patches of simple ignorance abound in these decision-making processes. Even talent-spotting is subject to these issues. This is a good reason not to intervene, as Polanyi understood. But it is not clear that the problems of decision-making are better resolved by creating one big barrier to entry, the barrier of becoming a "qualified" scientist, rather than having every idea subject to peer-review and the deliberations of a grant-making committee.

The idea of barriers to entry is a nicely general concept, which obviously applies far beyond examples like the Polanyian approach to the organization of science. And it has special relevance to expertise, because it is difficult to separate from the concept of expertise itself: if there are no barriers to calling oneself an expert, if anyone who can get someone to pay for an opinion is an expert, one wonders how the putative expert is selling his or her opinion? Koppl's example of forensics, it is true, does provide an extreme case of salesmanship. I recall passing the door of a lawyer's office in Miami years ago, and seeing a cascade of flyers and cards on the floor. Curious, I looked down to see what they were. They were advertisements for expert witness services. So this really looked like a market. But even here the sales pitch included qualifications. This may only have been relevant because qualifications made the expert a more persuasive witness. But they were part of the marketing process, and no accident.

Qualifications have a Janus faced character. The lack of a given qualification is a barrier to entry. But the possession of a qualification is a kind of certification that the market finds desirable. This is the phenomenon I tried to account for in "Scientists as Agents" (2002). It was inspired by another pile—of documents found outside a door, in this case a door of a hotel conference room. The room had evidently been occupied by a grant or hiring committee in the area of medical research. The documents were vitae of applicants. I glanced at them out of curiosity, and was astonished at one feature: most of them listed a full page or more of prizes. These began as early as third grade science fair prizes. My question: why were there so many prizes in medical science?

Prizes aren't like one-shot certifications. But they too serve to certify. And indeed, the articles listed in the rest of the vitae, along with the degrees and other qualifications, also represented certifications. They were each the product of some sort of evaluative process in which someone blessed the person or what the person produced. And they have value. Not as much value as a Nobel Prize, which Paula Stephan and Sharon Levin have quantified (1993), but enough value to put on a vita. Now things get interesting: if it has value, a market can't be far behind. And indeed, prize givers, degree givers, article accepters, letter of recommendation writers, and so on and so forth, are all competing in some sort of market in which their product has a value relative to the other products, to other certifications. They may not be paid for their opinion directly, but they are paid indirectly in various ways. So they are experts, by Koppl's definition. But they are second order experts, certifiers of the expertise of others.

Why is there so much demand for this kind of certification? For the same reason that we want our termite killing service to be certified: there is a principle-agent problem of trust, and we will pay, or make sacrifices, to diminish. And there are people who will come up with a certifying product—a seal of approval. Koppl cites but does not quote the Granny in the supposed exchange with Bertrand Russell, who says, "It's all rubbish! The world is flat and rests on the back of a turtle!" when Russell says, "If that is so, what does the turtle stand on?" the granny says, "You can't fool me, sonny, it's turtles all the way down!" With this we can go full Austrian and say that it is markets for opinions, including opinions about opinions, all the way down, or at least until the costs of going to the next turtle down exceeds the benefits—in our opinion.

But there is a problem with the market analogy, which Koppl is very sensitive to. Opinions are not, like commodities, more or less discrete and independent dry goods, made by independent and competing producers. They depend on a division of knowledge and "knowledge labor," or discoveries and ideas that are communicated. Things may be only partially understood outside the narrow limits of certain roles in the division of knowledge. But this knowledge, acquired second hand, is the basis of opinion. In a certain sense, scientific knowledge at least is "collective" or at least distributed and held by no one in the sense that there is someone who possesses all of it. There is a communal character to scientific knowledge that is antithetical to the market analogy.

In contrast to the image of the autonomous scientist, this dependence on others is pervasive, because massive ignorance is the basic condition of all scientific inquirers (Schwartz 2008). Only a small domain of specialized knowledge can ever be fully accessed by any individual. The beliefs on which they depend, and which are the basis of their opinions come from elsewhere in the division of knowledge, and are normally only very imperfectly assimilated, more or less on a need to use basis. What about raw "evidence"? Evidence is the basis for inferences that themselves, for any meaningful purpose, depend on opinions—about what the right methods are, about what should count as evidence, and so forth. And in most domains this depends on tacit knowledge acquired by personal contact, meaning that a division of labor is being mapped on to the division of knowledge.

All this is banal, and Koppl gives a much more elaborate account of these considerations, which I will not repeat here. He suggests some ways in which they become a problem, for example in the entangled state, where authority, expertise, politics, and mix. He also gives the example of the Flexner report (Brown 1979). This is a crucial example, for the simple reason that the pattern of institutional development that Flexner pioneered for medicine in the US was applied elsewhere (famously to China[1]) and more importantly to other fields (Buxton and Turner 1992), thus creating a template for the categories of expert and professional themselves. Why did this anti-market template get applied so widely?

A brief history of this template and its applications are useful here. One example is particularly relevant. Charles Merriam, the powerful advisor for social science to the Laura Spelman Rockefeller Memorial/Rockefeller Foundation, did the same thing Flexner did for Medicine, but for public administration.[2] The template was the same—certification, professional education, continuing education, connection to a university as well as an ongoing process of doing research (complete with an academic discipline and hierarchy) and training people in light of the findings, and creating a hierarchy from experts to practitioners. This was the opposite of spontaneous order: it was instead the self-conscious creation of an order, with experts buttressed and confirmed by an institutional structure, with incentives for following the experts, and with barriers to entry that provided advantages for those who were licensed or certified, and so forth.

So why did this happen, and why did the template succeed so well that "professionalization" became a major theme of the twentieth century, and sociologists began to talk about "the professionalization of everyone" (Wilensky 1964). We can get some "market" answers by looking at these two cases: real estate and public administration. In the real estate case, there were always market alternatives—"for sale by owner," having a lawyer handle the transaction, and so forth. As Koppl notes, market barriers lead to exclusions that are unrelated to value: this was especially evident with real estate. When real estate professionalized and created the category of Realtors, in a process led by none other than Richard T. Ely (Malpezzi 2009, pp. 2, 5-7, 10-12, 22), it involved race. And this produced a response: Black real estate agents created their own certification and professional body in 1947, which pursued different policy aims, and called themselves "Realtists."[3] In the public administration case it was a matter of public choice: cities chose democratically to be professionally administered rather than leave decisions to politics or hand out positions to political appointees. State agencies, for their own reasons, used certifications as job requirements. And there was entanglement with the state. In many places there was a state associated training and continuing education program, and therefore a de facto monopoly.

Flexner may have created the model, but clearly there was a big market for expertise organized according to the template. There were obviously many reasons for this, but trust was one of them: if you need a service and lack the knowledge to find the right person or method, or there are other complications, going for the standard certified model is a safe bet. One might think of it this way: people want experts who are certified, but they also want choice; experts in organized bodies, not surprisingly, want to exclude "fake" experts from competition, and create a monopoly, at least of certification, the effect of which, together with the forms of discipline that go along with certification, is intellectual uniformity. But of course this is a recipe for massive error. The exact same mechanisms for creating some sort of uniform expertise, which are incentives, sanctions, acts of authority, hierarchical forms of education, and so forth, are of course contrary to the mythic idea of autonomous agents making epistemic choices that Polanyi clung to.

So while these structures create intellectual authority, or expertise, they also create the conditions for massive expert error. Is this just a chimera? Is it so rare that it can be ignored, or that we can adopt the heuristic of deferring to "scientific consensus" as the best even if not correct opinion? Of course, we cannot escape the turtles problem: in every case of massive error we can only appeal to more opinion. We can find plenty of cases where opinions change, and the original opinion comes to be regarded universally as error. And we can see some patterns here.

One is this. There are no expert solutions to many of the problems which experts are called upon for expert opinions. The demand for expertise far outstrips the expertise available. So there is a big gap between what buyers want and what sellers can provide. Mental health issues provide numerous examples: people are desperate for answers and help. Autism has long had history of expert failure, from the notorious case of Bruno Bettleheim to the present overuse of the diagnosis. Similarly, Jerome Kagan has recently denounced the ADHD diagnosis (2012), saying it has largely been invented by drug companies. Yet an army of experts, clinicians, and educators are devoted to it, not to mention the parents who are eager for "solutions." The list of failure could go on. Education reform has been on the public agenda for more than a century. Educational research, as Ellen Condliffe Lagemann has shown, has been a succession of fads (Lagemann 2000). This gap never closed.

Normal academic research, research not driven by a willing buyer with a policy agenda, is not exempt from these problems. As Richard Horton, editor-in-chief of *The Lancet*, writes,

> [M]uch of the scientific literature, perhaps half, may simply be untrue. Afflicted by studies with small sample sizes, tiny effects, invalid exploratory analyses, and flagrant conflicts of interest, together with an obsession for pursuing fashionable trends of dubious importance, science has taken a turn towards darkness. (2015, p. 1380)

Horton adds a comment about markets: "Can bad scientific practices be fixed?" Not without changing the market. "Part of the problem is that no one is incentivised to be right. Instead, scientists are incentivised to be productive and innovative" (2015, p. 1380).

One facilitator of this turn to darkness has been the abuse of statistics, acknowledged by the American Statistical Association (Wasserstein and Lazar 2016—and publicized in recent discussions of p-hacking and in connection with the reproducibility crisis. The issues are very basic. P values are conventionally used to certify a research finding as a fact. This convention, and its abuse, is a major source of the reproducibility crisis in psychology. A recent suggestion (Benjamin et al. 2018) to raise the level of significance from 0.05 to 0.005 would cause whole fields to come close to disappearing—and this would certainly include the fields of evidence-based policy. And the p-value issue just scratches the surface of the problems, which extend to virtually every area in which statistics are used, and in which the small manipulation of assumptions can produce radically different results.

One such problem is this: research subjects and goals are not randomly distributed. People are looking for and attempting to establish particular results. As John Iaonnidis has pointed out, the effect of this is to make the expert consensus little more than a measure of bias (2005). And obviously this bias is often politically motivated bias. The existence of this kind of bias, which often occurs when topics are intentionally under-researched, is admitted even by Brookings, whose reputation for impartiality is itself questionable.

> Psychologists, sociologists, and educational researchers have devoted far less attention to the black-white test score gap over the past quarter-century than they should have. Cowed by the hostile reaction to Daniel Patrick Moynihan's 1965 report on the status of the black family and to Arthur Jensen's 1969 article arguing that racial differences in test performance were likely to be partly innate, most social scientists have chosen safer topics and hoped the problem would go away. (Jencks and Phillips 1998)

There are many other topics that are no-go zones. And there is even a philosophical literature defending the practice of avoiding research on topics that lead in the wrong political direction (Kitcher, 2000, pp. 193-97). This kind of politically motivated sefl-censorship more or less assures that there will be massive error.

Koppl doesn't have a solution to the problem of massive error, and neither do I. "Error" is a problematic notion in this context, because judgments about error also rests, so to speak, on turtles that go all the way down. There is no perch outside of opinion on which we can rest our judgments. It is, as Michael Oakeshott would say, platforms, that go all the way down (1975, pp. 9, 27, 34). Our beliefs about the world rest on research that relies on experimental and statistical conventions. These in turn rest on other opinions, other consensuses. What we take to be true about the world depends on what someone decided to fund. The science, and the expertise we have, is the product of "the world," but it is the world as disclosed by past decisions to disclose it and disclose it in a particular way. The "ways" are necessarily limited in ways that are unknown to us. The path we took could have been different. And had we taken a different path, we might have been in a position to see what the limitations of the path we took were. If we did not invest in that path, we might not ever be in that position. It is pleasing to think that the truth will out, eventually. But turtles can live a long time. And science is as entangled in problematic decision processes as the state.

NOTES

1 Medicine in China. The Rockefeller Foundation. *A Digital History.* https://rockfound.rockarch.org/china-medical-board (accessed 4 June 2018)

2 Rockefeller Foundation. PACH and the Spelman Legacy. Public Administration. *A Digital History.* https://rockfound.rockarch.org/public-administration (accessed 6 June 2018)

3 NAREB Realtist. http://www.nareb.com/faq/what-is-a-realtist/

REFERENCES

Benjamin, Daniel J., et al. (72 authors). 2018. "Redefine Statistical Significance." *Nature Human Behavior* 2: 6-10. DOI: 10.1038/s41562-017-0189-z. https://www.researchgate.net/publication/318642148 (accessed 6 June 2018)

Brown, E. Richard. 1979. *Rockefeller Medicine Men: Medicine and Capitalism in America.* Berkeley: University of California Press.

Buxton, William, and Stephen Turner. 1992. From Education to Expertise: Sociology as a "Profession." In: *Sociology and Its Publics.* Edited by Terence C. Halliday and Morris Janowitz, Chicago: University of Chicago Press. Pp. 373-407.

Ely, Richard T. https://en.wikipedia.org/wiki/Richard_T._Ely

Horton, Richard. 2015. "Offline: What is Medicine's 5 Sigma?" *The Lancet* 385(9976): 1380.

Iaonnidis, John. 2005. "Why Most Published Research Findings Are False." *Chance* 18:4, 40-47, DOI: 10.1080/09332480.2005.10722754. https://doi.org/10.1080/09332480.2005.10722754 (accessed 5 June 2018)

Jencks, Christopher, and Meredith Phillips. 1998. "The Black-White Test Score Gap: Why It Persists and What Can Be Done." Brookings. https://www.brookings.edu/articles/the-black-white-test-score-gap-why-it-persists-and-what-can-be-done/ (accessed 4 June 2018)

Kagan, Jerome. 2012. "What about Tutoring instead of Pills?" Spiegel Interview. Speigel Online. http://www.spiegel.de/international/world/child-psychologist-jerome-kagan-on-overprescibing-drugs-to-children-a-847500.html (accessed 4 June 2018)

Kitcher, Philip (2000) *Science, Truth, and Democracy.* Oxford: Oxford University Press.

Lagemann, Ellen Condliffe. 2000. *An Elusive Science.* Chicago: The University of Chicago Press.

Malpezzi, Stephen. 2009. "The Wisconsin Program in Real Estate and Urban Land Economics: A Century of Tradition and Innovation." The Department of Real Estate and Urban Land Economics & The James A. Graaskamp Center for Real Estate Wisconsin School of Business University of Wisconsin-Madison. https://bus.wisc.edu/centers/james-a-graaskamp-center-for-real-estate/about%20the%20graaskamp%20center/~/media/ef20196ae0394932888b49030d093787.ashx

Oakeshott, Michael. 1975. *On Human Conduct.* Oxford: Clarendon Press.

Polanyi, Michael. [1951] 1980. *Logic of Liberty: Reflections and Rejoinders.* Chicago: The University of Chicago Press.

Schwartz, Martin A. 2008. "The Importance of Stupidity in Scientific Research." *Journal of Cell Science* 121(11): 1771. doi: 10.1242/jcs.033340.

Stephan, Paula E., and Sharon G. Levin. 1993. "Age and the Nobel Prize Revisited." *Scientometrics* 28(3): 387-399.

Turner, Stephen. 2002. Scientists as Agents. In: *Science Bought and Sold*, edited by Philip Mirowski and Miriam Sent. Chicago: University of Chicago Press, pp. 362-284.

Turner, Stephen. 2005. Polanyi's Political Theory of Science. In: *Emotion, Tradition, Reason: Essays on the Social, Economic and Political Thought of Michael Polanyi.* Edited by Struan Jacobs and Richard Allen. Aldershot, UK: Ashgate, pp. 83-97.

Wasserstein, Ronald, and Nicole Lazar. 2016. The ASA's Statement on p-Values: Context, Process, and Purpose. *The American Statistician* 70(2): 129-133. https://doi.org/10.1080/00031305.2016.1154108 (accessed 6 June, 2018)

Wilensky, Harold L. 1964. "The Professionalization of Everyone?" *American Journal of Sociology* 70(2): 137-158.