RISK IN SCIENTIFIC QUALITY CONTROL

Kyle Siler
7059 Rotman School of Management
University of Toronto
105 St. George Street
Toronto, ON
M5S 3E6

March 31, 2015

Ksiler <at> gmail <dot> com

RISK IN SCIENTIFIC QUALITY CONTROL

ABSTRACT: This article theorizes risk as a key factor underpinning evaluative cultures in science and knowledge work. It is posited that talent and reward structures in science have opposing distributions. Talent discrepancies decrease moving up the quality gradient, while rewards exponentially increase at the top of the talent distribution. This contributes to a scientific incentive structure that makes avoiding errors of commission easier than avoiding errors of omission. Consequently, most scientific gatekeepers tend to enact risk-averse decision-making. Potential downside risks of science are discussed. Four ideal types of potential innovation outcomes are proffered: The Polarizer, The Specialist, The Lowest Common Denominator and The Overlooked Genius. With these varied innovation outcomes, scholars and gatekeepers are faced with the question of whether increased variance in evaluation of riskier contributions justifies lesser, equal or greater mean quality on the whole. Examples of upside and downside risk are drawn from the social sciences.

KEYWORDS: Peer review, risk, sociology of science, decision-making, academic publishing

RISK IN SCIENTIFIC QUALITY CONTROL

INTRODUCTION

At the competitive cutting-edge of science, where new knowledge is created and disseminated, uncertainty regarding the quality of new innovations is inherent, particularly prior to publication and diffusion. This is a challenge that pervades all scientific disciplines and types of knowledge work. Cole (1983) found similar degrees of dissensus amongst National Science Foundation peer reviewers adjudicating elite grants across natural and social science disciplines. Gatekeepers and scientists do not want to reject or desecrate future seminal articles, but also want to avoid wasting journal space on mediocre contributions, or even worse, publishing badly flawed science. Ergo, risk is a central part of academic gatekeeping and decision-making. This article examines the incentives, tradeoffs and values underpinning different risks in vetting and publishing science, and in knowledge work in general.

TALENT AND REWARD STRUCTURES IN SCIENCE

Attention space from others in social fields is limited, so with academics (Lotka, 1926; Collins, 1998) and rock stars (Rosen, 1982) alike, a select few ‘superstars’ receive a disproportionate amount of attention and resources, based on this hierarchical deference structure. In science, the utility of scientific contributions increases as more scholars become familiar with them. These social incentives can beget a system where perceived quality is influenced to some extent by the opinions of others, which often results in sub-optimal innovations gaining precedence (Salganik et al., 2006). Academia is characterized by a "superstar" market (Rosen, 1981), with scale-free networks. Citations and network ties follow an exponential power law (e.g., Barabási and Albert, 1999), where a small number of scholars and articles receive a disproportionate amount of attention (Lotka, 1926). Since there are benefits to informational co-ordination and shared symbols in science, this also helps produce the 'superstar' model of knowledge production. Even within leading academic journals, a few highly-cited articles account for the majority of citations, while other less-cited articles "free ride" on inflated impact factors skewed by a few articles (Baum; 2011, 2013). Frank and Cook (1995) identified social structures where a few highly ranked entities enjoy disproportionate rewards as "winner-take-all markets." The tendency of academic knowledge to coalesce around superstars and highly-cited landmark articles is indicative of such a market. As Figure 1 shows, this citation structure is remarkably consistent – with slight differences – across a variety of social science journals. Most articles are never or seldom cited, while above the ninetieth percentile, there is exponential growth in citations.

-- Insert Figure 1 about here –

Academia – particularly in the social sciences – works as a winner-take-all market[1] in two main ways. First, getting entry into high-status journals is extremely competitive, as social science journals tend to have very low acceptance rates; often under ten percent (Hargens, 1988). Promotion, tenure and hiring are usually strongly linked to publication in a relatively small proportion of leading journals. Consequently, disproportionate attention and prestige are

[1] Perhaps a better descriptor of such reward structures would be "winners-take-most", as few markets in science are literally winner-take-*all*.

associated with a few journals at the top of the professional hierarchy. In sociology (Burris, 2004) and economics (Colander, 1990), a small number of doctoral programs control hiring markets and enact closure in the labor market over outsiders. Secondly, diffusion processes are filtered through the scale-free networks of academia, meaning a relatively few articles and scholars serve as highly-cited 'hubs' of the network, which other 'spokes' preferentially attach to. In academic fields with less paradigmatic consensus, in an effort to cope with uncertainty, scholars are more likely to cite other highly-cited articles (Lynn, 2012), reflecting another way that status signals are influential in scientific production and consumption.

In general, the talent distribution in science makes minimizing poor articles easier than maximizing the chances of publishing outstanding work. John Maynard Keynes edited *The Economic Journal* for thirty-two years, and remarked, "I feel much clearer about the demerit of the articles I reject than I do about the merit of most of those which are included" (Shepherd, 1994: 26). Einstein was famously quoted, "the difference between genius and stupidity is that genius has limits." While most fields and subfields in the social sciences are characterized by low paradigmatic consensus and low inter-rater reliability among reviewers (Whitley, 1984; Hargens, 1988; Pfeffer, 1993), this low level of consensus is inflated from agreement about the demerit of low-quality contributions. Cicchetti (1991) found that inter-rater reliability in social science journals was much higher for rejected articles than it was for those that were granted a revision attempt. Lindsey (1988) argued that at the upper percentiles of quality in social science, peer review is "little better than a dice roll." Put differently, quality signals at high levels are dampened, as evaluators are less able to strongly identify truly exceptional contributions a priori. As another example of how quality is difficult to discern amongst competitive manuscripts, Smith (2006: 178) was jokingly challenged to release an issue of the *British Medical Journal* comprised entirely of rejected manuscripts. His response: “How do you know I haven’t already done it?” Editors and scholars alike are left with the difficult task of predicting which ripples in the ocean will eventually become tsunamis.

PROPOSITION 1: Talent in science is distributed in a diminishing exponential curve. Marginal returns decline as one moves up the talent gradient.

The reward structure of academia is usually shaped differently than the talent structure. Due to the scale-free networks and exponential payoffs in the academic reward structure, academia functions as a winner-take-all market. Prestige is disproportionately loaded onto elite journals and programs, plus Matthew Effects (Merton, 1968) emerge where high-status scientists can further consolidate rewards merely based on their own status. Relatedly, there are also “halo effects”, where people tend to defer and ascribe competence to those associated with high-status scholars and institutions (Peters and Ceci, 1982; Leahey, 2004).

PROPOSITION 2: Rewards in science are distributed in a rising exponential curve. Marginal returns accelerate as one moves up the talent hierarchy.

Thus, combining Propositions 1 and 2, academia is a winner-take-all reward structure with an opposite diminishing talent distribution. As a result, important professional and intellectual decisions tend to occur on the right tail with the highest stakes and the most noise. On the right tail, rewards spike upward as one enters the highest percentiles, but talent differentials become more minuscule. In other words, on the left-tail of the talent distribution, where reward differentials are highest, talent differentials are smallest; where reward differentials are low, on the right-tail, talent differentials are largest. This is a risk-averse strategy that is effective for filtering out bad science, but much less effective in distinguishing the best contributions. With a high-rejection journal in a field where talent is not strongly demarcated in upper echelons of ability, this leads to implicit value judgments that the benefits of avoiding errors of commission outweigh the benefits of avoiding errors of omission.[2]

-- Insert Figure 2 about here --

Figure 2 provides a simple graphic illustration of the hypothesized relationship between talent and rewards in science. With mild exponential functions, reward outpaces ability starting roughly at above the sixty-third percentile, while ability exceeds reward below that threshold. Shifting the curves with either steeper or flatter reward or talent structures will change the payoffs and equilibria. For example, since journals decrease their acceptance rates as they increase in prestige, this creates a more drastic winner-take-all market and a shifted reward curve. Winner-take-all markets disproportionately reward the right tail of performers, yet often are based on tests calibrated to distinguish near the mean. This creates challenges, for example with elite colleges relying on fine gradations of SAT scores at extreme values – with relatively high standard errors – to make admission decisions.[3] As winner-take-all markets become more extreme, the intersection of the talent and reward curves occurs at higher percentiles, and greater

[2] This lies in contrast to other cultural/knowledge fields, such as pharmaceuticals, music, hedge funds, and angel investing, where failure is the modal outcome, yet the exponential payoffs for a single successful innovation justify repeated previous failures.

[3] In contrast, the Putnam Mathematics Competition for undergraduates - an extremely difficult test which usually has a median score of one out of a possible 120 points - is an example of a way to focus evaluation in fine gradations at the extreme right tail of a talent distribution, although it does not meaningfully distinguish amongst most other test-takers.

surpluses of reward relative to talent accrue to the fortunate few above that threshold. To reduce this asymmetry between talent and reward, either the reward structure has to be flattened, or tasks have to be devised to distinguish talent in finer gradations. For example, economics programs are known for using difficult and arcane mathematics courses to distinguish talent levels between students (Colander, 1990; Athey, 2014), albeit along a tangential or marginally relevant dimension. This may be a rational strategy in part because high-ability students benefit from – and presumably are attracted to – difficult courses with low standard errors in evaluation (March, 1994: 44).

TRADEOFFS BETWEEN UPSIDE AND DOWNSIDE RISK

The combination of opposing exponential functions of talent and reward means that small – and perhaps arbitrary – differences in talent distinguishes scholars from one another. Mauboussin (2012) argued that a *paradox of skill* exists in competitive environments with highly skilled people, where luck plays a substantial role in determining outcomes, because there are relatively scant skill differentials at upper bounds for people to distinguish themselves with. Based on analysis of granting outcomes at the National Science Foundation, Cole et al. (1981) argued that roughly one-half of the granting process in this case was a matter of luck. In her own study on National Institutes of Health medical grants, Li (2013) found that in regards to citations eventually received, there were no substantial differences between funded and unfunded proposals, despite halo and Matthew effects associated with holding a prestigious grant.

Accordingly, flattening the reward structure of science (through for example, more egalitarian distribution of grants, lower journal rejection rates) should result in an increase in errors of commission, but at a slower rate than a concomitant decline in errors of omission. At first blush, this seems to be a desirable tradeoff. However, it is not entirely certain that errors of commission and errors of omission should be treated equally. Watts (2014: 17) remarked that most academic journal editors reported being more troubled by publishing unworthy articles as opposed to rejecting good ones. McAfee (2014) argued that rarely – with a few exceptions – will an editor express regret about a rejection. While confirmation bias may explain part of the tendency of editors to feel positively about their previous decisions, the incentive structure of the social sciences also may contribute to the tendency to be cavalier about rejections. Zuckerman and Merton (1971) observed that in the social sciences there was a much greater emphasis on avoiding Type I errors (publishing an unworthy article) vis-à-vis Type II errors (rejecting a worthy article). In contrast, the natural sciences were more concerned about avoiding Type II errors. When acceptance rates in leading journals in the social sciences remain below ten percent, it appears that rejecting worthy articles occasionally (or more than occasionally) is a culturally accepted outcome, in exchange for increased vigilance against potentially mediocre or poor articles. Granted, this tradeoff is not necessarily linear, but risk-averse thinking entails trading a chance at a large gain in exchange for a reduced chance of loss. Even if such decision making sacrifices value, risk-aversion may be rational if value judgments are made that mistakes are potentially disproportionately damaging relative to the benefits of successes.

Editors and journals must make decisions regarding potential tradeoffs between the risks of committing Type I (accepting an unworthy article) and Type II (rejecting a worthy article) errors respectively.[4] This raises the question of whether a quality submission with a high degree of variance in evaluation is preferable to a similar submission – perhaps of slightly lower mean quality – with less potential to be polarizing. In a study of peer review in a large national granting agency, Langfeldt (2001: 834) showed that different systems of aggregating scores of individual reviewers can yield divergent funding outcomes. Some systems favor manuscripts with high consensus (in particular, without a strong negative appraisal), while others are more inclined to promote a polarizing article with one or two strong advocates. Tradeoffs vary between disciplines, with editors in the physical sciences and journals with lower rejection rates being more willing to risk publishing questionable articles in search of 'exceptional' contributions (Zuckerman and Merton, 1971; Cole, 1991). Further, editors of generalist journals tend to focus on avoiding Type I errors, while specialist editors tend to emphasize avoiding Type II errors (Cicchetti, 1991). Costs of publishing a controversial article can be significant, in terms of shepherding such work through often polarized reviewers, opportunity costs of not publishing a more worthy article and potential legitimacy losses for the editor and journal as a whole. With survivor biases, costs of rejecting an excellent article are less tangible, especially if those articles end up at seldom-read journals or are never published at all. Gans and Shepherd (1994) chronicled reports of rejected classic economics articles, serving as strong and rare negative feedback against errors of omission in academia. Taking risks with controversial articles can be

[4] In Fall 2013, a new academic journal, *Sociological Science* debuted, featuring a stripped-down peer review system emphasizing efficiency and authorial autonomy, while de-emphasizing the gestational function of peer review. This new system is underpinned by an unconventional philosophy towards publishing risks. The editorial board (2013) states, "*Sociological Science* prefers errors of commission (accepting papers that should have been rejected) over errors of omission (rejecting papers that should have been accepted)."As an online journal without costs of printing and distributing hard copies of articles, the luxury of committing errors of commission may in part be enabled by operating solely online. The costs and benefits of this new model remain open questions that will begin to be answered as the journal starts publishing articles. What will happen to the journal after the inevitable publication of the first controversial or widely panned article, especially because this is a cost they are theoretically willing to incur, based on their cost/benefit analysis? If errors of commission occur in traditional peer review settings, are journals often inured from a reputation hit either due to prestige (if high-status) or anonymity (if low status)?

difficult for leading journals, who are under close scrutiny and face criticism for a mistake. During his term as editor of *Science*, Kennedy (2002: 1793) published an editorial explaining his decision to publish a controversial article. Defending his decision, he wrote:

"[W]e are prepared for occasional disappointment when our internal judgments and our processes of external review turn out to be wrong, and a provocative result is not fully confirmed. What we ARE very sure of is that publication is the right option, even – and perhaps especially – when there is some controversy."

Correctly identifying the costs and benefits of publishing risky articles – particularly when there is a substantial cache of high-quality less controversial manuscripts – is difficult. Even if such assessments can be made accurately, the normative question of to what degree journals, editors and scientists should incur such risks remains an open question.

RISK AND INNOVATION

Academic journals experience a tension between countervailing forces of innovation and orthodoxy. The general model of science is that leading journals innovate and disseminate cutting edge research, which is then applied, replicated and refined by scholars, institutions and journals downstream in the professional hierarchy (Kuhn, 1962; Stinchcombe, 1994). Hackett (2005: 815) posits that scholars assume riskiness in their research through choosing projects of varying importance and viability. Innovation is a vaunted goal in most organizations; particularly in science, where *priority* – being the first to claim an innovation or research finding – is how scholars receive credit and status (Merton, 1968; Latour and Woolgar, 1979). However, organizations often publicly pay lip service to innovation, but then for bureaucratic and/or cultural reasons, act in ways that squelch innovation and creativity (Flynn and Chatman, 2001). In turn, scientific innovation often emerges via atheoretical work with links outside of academia, where inquiry is not necessarily bound by prevailing scientific cultures and orthodoxies (Evans, 2010).

Gans and Shepherd (1994) reported numerous cases of famous economists and articles receiving rejections from leading journals. The rejection of Akerlof's (1970) "The Market for Lemons" is among the most famous of those rejections. Harvey (2014: 76-77) confessed that as editor of the *Journal of Finance* – a prestigious journal that had been notoriously dismissive of Akerlof's work in peer review – the potential of making a similar mistake in the future influenced his own editorial decision-making. While rejecting a seminal article is obviously a regrettable outcome, McAfee (2014: 57) argued that a journal that does not reject innovative articles like "The Market for Lemons" will publish articles with lower mean quality on the whole. This raises the important issue of what the implied quality structure of articles is, particularly relative to the innovativeness or unconventionality of the contribution. Levinthal and March (1993: 106) posited that most new ideas are bad ideas, so statistical discrimination against innovative work could be efficient, if not also desirable.

Three main questions underpin decisions regarding risky scientific valuation. First, assuming that unconventional scientific contributions will increase variance and uncertainty, what should be the necessary risk premium – if any – to take chances on such innovations? Second, what is the distribution of quality amongst potential choices, either as perceived through some measure of true quality and/or filtered through evaluative schemas? Third, should scientific gatekeepers attempt to maximize the mean quality of articles? Given that a few exceptional articles tend to exert a disproportionate influence on science (Horrobin, 1990), do a few exceptional results justify numerous other mediocre outcomes? Alternatively, even if unconventionality pays off on the whole, the downside risk of publishing mediocre – or outright false – science may be restrictive, and justify greater risk-aversion.

DOWNSIDE RISKS WITH ERRORS OF COMISSION

Two recent examples show the downside risks associated with Type I errors faced by journal gatekeepers. Regnerus' (2012) *Social Science Research* article about LGBT parenting outcomes and Bem's (2011) *Journal of Personality and Social Psychology* article on extrasensory perception both received sharp and widespread criticism after publication. Intentionally or not, journal editors managed to send methodologically – if not also ethically – dubious manuscripts to sympathetic reviewers. This could have been due to fluke, random draws in vast reviewer pools and/or poor selection of reviewers. Regnerus (2012) claimed to cast doubt on the conventional sociological wisdom that there is no difference in outcomes for children raised in households same-sex marriages vis-à-vis those with opposite-sex parents. Almost immediately, Regnerus' article was emphatically rebuked by two-hundred leading researchers in the subfield (Gates et al., 2012), an *SSR* auditor (Sherkat, 2012), an Amicus Brief submitted by the American Sociological Association (2013) and a Michigan District Court (2014).[5]

Part of the failure of peer review with the Regnerus article was due to a lack of differentiation among peer reviewers. The *SSR* audit uncovered that at least two of the three reviewers had professional links to Regnerus and had gone on record as politically opposing LGBT rights (Sherkat, 2012). In contrast, despite the low degree of inter-rater consensus in most

[5] Judge Bernard A. Friedman of Federal District Court in Detroit, appointed by Ronald Reagan, evaluated Regnerus' work very harshly. A couple of withering quotes: "The Court finds Regnerus's testimony entirely unbelievable and not worthy of serious consideration.", deciding that the research was "hastily concocted at the behest of a third-party funder." (Frank, 2014)

social sciences (Hargens, 1988), editors blatantly choose reviewers with diverse interests and expertise in a way that decreases the likelihood of consensus (Kraemer, 1991). This raises the question of what degree to which gatekeepers want to differentiate reviewers. Too much differentiation would result in three relatively random evaluations; too little differentiation is prone to insularity. What is the optimal mix of differentiation and similarity; of breadth and depth? Would this optimal mix vary between fields and institutions? The Regnerus case was rare not in that it was based on sloppy or dubious science, but that it took a stance contrary to the mainstream literature in the field, and was also couched in a political position that most sociologists find personally objectionable.

There is a small community of scholars sympathetic to Regnerus' work, who mobilized a small Social/Intellectual Movement (Frickel and Gross, 2004) to attempt to counter the critical onslaught Regnerus faced after the publication of his *SSR* article (Baylor ISR, 2012). Given that these contrary-minded sociologists are in the reviewer pool, a random draw will theoretically occasionally yield three reviewers who are amenable to research that would be rejected by most of the rest of the discipline. However, gatekeepers usually take steps to combat this when the ideological priors of reviewers are reasonably known. Kraemer (1991: 153) argued that editors should and often do make decisions that decrease inter-rater reliability. Editors tap reviewers from varying subfields to "supplement and complement" each other, and help orient the article vis-à-vis a broader variety of perspectives across complex and intellectually diverse fields. At best, this diversifies the reviewing portfolio and hedges against risk of groupthink among coincidentally similar reviewers. On the downside, diversification may force editors to adjudicate between disparate and incommensurable reviewers down to the lowest common denominator, while forgoing the benefits of specialization. Since social similarity begets fact complexity (Carley, 1991; Mark, 1998), science is strengthened by some degree of consensus and shared ideology. Kuhn (1962) famously posited that the high-consensus culture of normal science is extremely productive. In turn, it can be a fine line between garnering the benefits of specialization and leaving the evaluation prone to groupthink. Gatekeepers are faced with a strategic optimization problem regarding how to most appropriately balance the risks and benefits of tapping similar reviewers for depth, while hedging their bets by tapping dissimilar reviewers for breadth.

The Bem (2012) article, which purported to provide evidence in support of the existence of extra-sensory perception (ESP), was very controversial amongst psychologists (Zimmer, 2011). The controversy was exacerbated via the fact that it was published in the *Journal of Personality and Social Psychology*; a leading journal in the field. As Carey (2011) reported, "the decision [to publish Bem's paper] may delight believers in so-called paranormal events, but it is already mortifying scientists."[6] The methodology and findings of the article were rebuked (Wagenmakers et al., 2011) and attempts by researchers to replicate Bem's controversial results have failed (Ritchie et al., 2012). *JPSP* editor Charles Judd defended the decision to publish the article, arguing that the manuscript passed muster with four "very trusted" reviewers in the field (Carey, 2011).

The Bem publication presents a similar potential problem to the Regnerus case; that the editor may not have adequately intellectually diversified his choices of peer reviewers to account for the skepticism such research faces in the field of psychology. Alternatively, it could merely have been a matter of luck. As with any polarizing research, if there is a minority of sympathetic readers in the population of reviewers, it is inevitable that with hundreds of manuscripts

[6] ESP research critic Ray Hyman responded with a couple of particularly blunt quotes: "It's craziness, pure craziness. I can't believe a major journal is allowing this work in….I think it's just an embarrassment for the entire field." "[Bem's] got a great sense of humor. I wouldn't rule out that this is an elaborate joke." (Carey, 2011)

processed annually, on some occasions, the editor will coincidentally tap the few sympathetic readers in the field to be reviewers. Exceptionally good manuscripts will occasionally face an analogous fate, except with having the paper sent to a few idiosyncratically critical reviewers. A problem for science is that while both of these errors are inevitable, the lucky bad article gets published, while the unlucky good article may never be published. Schrodt (2012) argued that the field of psychology is anomalous in the social sciences in that it has problems with excessively favoring the publication of work with unexpected or counterintuitive results, as exemplified by Bem's (2012) article. More broadly, Ioannidis (2005) argued that most published research findings are false, and are seldom verified through replication. In response to these concerns, psychologists are developing institutions to encourage and share "file drawer" null results as well as replication of controversial studies (Carpenter, 2012).

Stinchcombe and Ofshe (1969) offered a probabilistic model of peer review, where elements of luck underpin the consensus – or lack thereof – amongst reviewers. Better articles are more likely to be published, but with low inter-rater reliability in the social sciences, it is inevitable that some mediocre articles will be published, and some excellent articles will be rejected (Starbuck, 2003). This raises the issue of what degree "mistakes" are merely a high-status journal publishing an article that based on its true quality should have been published in a slightly lower-ranked journal, or if profoundly flawed articles are making the cut. This latter risk is especially prevalent, since inferior strategies can beget increased variance in evaluation, which inevitably contributes to a few positive outliers, especially in large populations (Denrell and March, 2001). When academic journals accept less than the (perceived) top ten percent of articles, this raises the specter of overselecting bad ideas with high variance.

Filtering out human error is another challenge of scientific gatekeeping and a source of potential downside risk, particularly when author(s) usually have privileged access to their own data. An infamous case of scientific error in the social sciences is Jasso's (1985) work on the relationship between age and coital frequency. Kahn and Udry (1986) re-analyzed the dataset Jasso used for her work and concluded that her findings were driven by a simple coding error. With the exclusion of a few problematic data points (allegedly extremely sexually active octogenerians), the central – and seemingly unique – findings of the article were reversed. A similar case occurred in economics, where graduate student Thomas Herndon uncovered that an influential *American Economic Review* article by Reinhart and Rogoff (2010) on nation debt and economic growth was based upon a number of errors, including mistakes with elementary spreadsheet formulas (Herndon, Ash and Pollin, 2013).[7] In other cases, data errors may not even be the fault of the researchers and even when corrected, may not mortally flaw an article (e.g., see Fischer's (2009) reanalysis of McPherson et al.'s (2006) work on isolation in social networks).

Kuhn (1977) characterized science as being defined by an "essential tension" between tradition and innovation (also see Hackett, 1997; Lamont, 2009). The Jasso case led Kahn and Udry (1986: 49) to conclude with the cautionary tale that scholarly gatekeepers need to be more skeptical of counterintuitive results. The allure of publishing counterintuitive results can often be substantial, as journals and scholars alike pride themselves on being cutting-edge, and benefit from being the first to report new and innovative findings (Mone and McKinley, 1993). On the

---

[7] Like the proverbial child laughing at the naked emperor, the flaws in the Reinhart and Rogoff (2010) article were uncovered by a graduate student undertaking seemingly mundane replication work as part of a class assignment. Thomas Herndon graduated from the notoriously countercultural Evergreen State College, then pursued graduate work at UMass-Amherst; a heterodox economics department in a high-consensus field that seldom takes such approaches seriously. From this 'outsider' vantage point, Herndon shook up the existing consensus view on nation debt and growth in mainstream economics.

other hand, concerns abound that peer review tends to be inimical to unconventional approaches and results (Horrobin, 1990; Staw, 1995; Smith, 1996; Redacted a). Editors often argue that innovative work should be granted slack because new methods and problems lack benefits of cumulation, replication and consensus (Friberg, 2014: 49: Beyer et al., 1995). Science values innovation and priority (Merton, 1968), and journals and scholars alike pride themselves on being – or at least perceived as being – at the frontier of knowledge. In turn, gatekeepers must strike an appropriate balance of engaging the uncertain frontier of knowledge, while trying to minimize the risk of publishing something that will be poorly received.

Having discussed potential downsides – if not worst-case scenarios – of scientific gatekeeping, it is also important to understand upside risks. Particularly in the social sciences, where editors have the benefits and burdens of low acceptance rates, leading journals have a substantial pool of quality submissions to choose from. Within this pool of new scientific research, there are different types of innovations and risks posed for scientists and scholars alike.

POTENTIAL INNOVATION OUTCOMES

Innovations have varying diffusion patterns and occupy different niches in the social structure. In turn, even high-quality scientific contributions may be perceived in different ways across academic fields. As follows are four ideal types of scientific innovations, based on varying audiences and niches.

The Polarizer: Such an innovation is not a matter of one arbiter catching a grievous error that another misses. Instead, there exists a fundamental disagreement in regards to the perceived quality and/or significance of the work based on personal opinion. Some very successful polarizing articles will be based on the strong support of a critical mass of scholars that highly regard the contribution, despite the fact that there are others who view the innovation negatively. In other words, the mean perceived quality of the article may be relatively modest, but there is substantial variance in appraisals. Differences in opinions are often delineated by academic subfields, particularly in a low-consensus scientific field like sociology.

Udry's (2000) *American Sociological Review* article, which linked gendered behaviors to biological roots, is an example of a polarizing article. In a rare candid look inside the editorial process, *ASR* editor Glenn Firebaugh (2001) decided that the publication and the peer review process it survived were contentious enough to warrant an explanation. Firebaugh acknowledged that the article required numerous rounds of revision before publication and faced six different peer reviewers. Throughout the peer review process, the manuscript yielded a variety of recommendations from reviewers: vehement rejection, speculative willingness to consider a substantially revised manuscript and vehement support. Firebaugh (620) acknowledged that publishing such a polarizing manuscript was a risk, but was willing to do so, arguing that always excluding controversial articles would not be in the best interests of the journal or the discipline.

Udry's (2000) article faced strong criticism – although perhaps not exclusively – from sociologists identifying as feminist and gender scholars, resulting in rejoinders being published in a subsequent issue of *ASR* (Miller and Costello, 2001; Kennelly et al., 2001; Risman, 2001). In this case, sociologists interested in integrating the relatively new and unexplored area of genetics and society were likely pleased by the inclusion of Udry's article. On the other hand, while the article appeared to particularly run afoul of sociologists of gender, it is arguable that the trenchant critiques of Udry's work could have advanced knowledge in that subfield as well.

Udry's arguments may have served as inoculations for others to strengthen their own beliefs and solidarity within their subfield. In a diverse discipline like sociology, this might be an acceptable – if not also desirable – outcome, particularly for a generalist journal. Another explanation for why and how Udry's work faced opposition, is that the notion that gendered behaviors can be explained in part by biological factors challenged the professional and intellectual turf of scholars who believe that gendered behaviors tend to be influenced by social – as opposed to biological – factors. Clashes over knowledge domains distinguish professions (Abbott, 1988) and divide up professional rewards. These turf wars can be anywhere along the continuum between symbiotic and internecine for both warring fields.

Polarizing articles raise the issue of if – and to what extent – increased variance in evaluations is tolerable, assuming the mean quality of such polarizing contributions are equal to or even greater than less controversial manuscripts. As a rule, people tend to be risk-averse; preferring a smaller mean in exchange for lower variance (Rabin and Thaler, 2001). People also tend to find states of uncertainty[8] to be aversive and generally desire to avoid such situations (Mueller et al., 2011). This is particularly true of people in leadership situations; managers tend to perceive risk more as a matter of potential losses than potential gains (March and Shapira, 1987). Since the frontier of science is characterized by some degree of uncertainty (Cole, 1983), attitudes towards risk and uncertainty are particularly important. Cultural and psychological frames can alter attitudes towards risk. For example, when propositions are framed as losses, as opposed to gains, people tend to become risk-seeking (Kahneman and Tversky, 1979). Denrell and March (2001: 535) argued that risk-aversion tends to emerge in organizations, since

[8] I am distinguishing risk and uncertainty as per Knight's (1921) definitions. Risk is defined when the outcome is unknown, but numerical probabilities can be assigned to the potential outcomes. Uncertainty is defined as when precise probabilities of outcomes cannot be calculated. When people have to make decisions based on uncertain parameters, they need to commensurate (Espeland and Stevens, 1998) uncertainty into estimated risk in order to make a decision.

successful choices are repeated and fast learners eliminate slower learners. Risky and new alternatives are inefficient choices, particularly in the short-run. The Udry article is a good example of the potential costs of a polarizing article; it took six reviewers and multiple rounds of revision to publish the article, then the journal had to devote space and time to justifying the review process to green-light the manuscript and allow for rebuttals. Such resources cannot be reasonably expended for every article a journal considers publishing.

As the Udry case shows, opinions regarding polarizing articles are often demarcated by academic subfields. Science is delineated by cultural and professional boundaries (Gieryn, 1983), so it is not surprising that knowledge is contested across those boundaries. Polarization can occur along a number of cleavages between groups and fields, including professional turfs, institutions and high-brow versus low-brow tastes. People use cultural preferences distinguish themselves and reinforce symbolic boundaries (Bryson, 1996). To paraphrase Bryson's argument about cultural exclusion and distinction with heavy metal music, science often begets pleas like "Anything but postmodernism", or "Freud", or "critical realism", or [fill in the blank]. Polarizing articles, theories and research become focal points for this distinction in science. While some ideas and methods and science – like heavy metal – may be reviled by many, this in part can be the appeal of a polarizing idea. Some ideas garner apathy from non-adherents, the polarizing idea at least garners attention. In some cases – like with Socrates and Galileo – a subversive, polarizing idea may inspire enmity because it is powerful and important enough to be threatening to entrenched interests and the status quo. Emotional energy (Collins, 1993) and strengthened professional boundaries can be borne from this antagonism for one or both parties.

The volatility and polarizing nature of peer evaluations can also vary by field. Cole et al. (1981) found that upon re-evaluation by a new team of experts, NSF proposals initially scored in the bottom quartile were much more likely to be chosen for funding in economics (and top quartile papers were more likely to be denied) vis-à-vis proposals in chemical dynamics or solid state physics, where most changes from "funded" to "not funded" occurred among articles near the cut line. This raises the question of whether increased variance is a desirable thing or not. In economics, it appears that polarizing articles were more likely to have the potential to be positively evaluated. In some cases, a proposal that earned disapproval from some peer reviewers may be more likely to be vaunted by others. In contrast, perceptions of quality in chemical dynamics and solid state physics were more finely gradated, where a proposal that received poor grades likely did so due to poor quality, as opposed to being a polarizing proposal that happened to get an unlucky draw in peer review.

The Specialist: The Specialist contribution is similar to The Polarizer in that there is a community of scientists who think highly of the work and find it useful. However, unlike the polarizer, it is a relatively low variance proposition, and is embedded in what Kuhn (1962) dubbed as normal science. The Specialist also differs from the Polarizer in that those outside of this community are blasé or uninterested in regards to the contribution, as opposed to derisive or hostile. Scientific and professional boundaries may demarcate specialist articles, but these boundaries are not sites of contestation, unlike with contentious polarizing contributions. Specialist articles tend to be esoteric, meaning that there is an inherent ceiling on interest and diffusion for a given article or innovation. Like the proverbial Inuit who can make fine distinctions between different types of snow that appear identical to others, scientists strongly socialized in a field of knowledge can perceive innovations differently than others. A banal finding to one scholar can be groundbreaking to another, particularly when they are embedded in a strong paradigm. Sometimes the esoteric nature of a highly specialized contribution is perceived as related to quality. As the supervisor of Theodore Kaczynski's[9] mathematics dissertation glowingly remarked, "I would guess that maybe 10 or 12 [mathematicians] in the country understood or appreciated it."

Esoteric work raises the question of how effective a scientific innovation with such a small constituency can be. If attached to high-status alters capable of disseminating the research to broader audiences, esoteric work has the potential to diffuse and be impactful. Benefits are often associated with eigenvector centrality and attachment to high-status nodes in knowledge networks (Bonacich, 1987; Podolny, 1993). Esoteric scholars and work can receive deference from others, particularly when paired with high status. This can be a problem, as people are more likely to excuse sloppy research if they believe it was from a high-status researcher (Peters and Ceci, 1982; Leahey, 2004). On the other hand, Abbott (1988) argued that professions – characterized by refined, specialized knowledge – require deference from the public in order to establish their knowledge domains as legitimate. Competition for this deference and legitimacy from relevant publics underpins the hierarchy of professions and knowledge. Specialists benefit from scientific boundaries and professional closure, as long as their turf is seen as legitimate and important by other stakeholders in the profession. Specialist work may also benefit from boundary spanners to 'sell' their work to those who may otherwise be indifferent to or incapable of understanding the work, but selling work to non-specialists comes at the expense of complexity.

[9] Kaczynski – also known as the Unabomber – was known as a brilliant mathematician and the youngest tenure-track professor in the history of UC-Berkeley at the time, before he left the university and embarked upon a deadly ecological terror campaign.

<u>The Lowest Common Denominator</u>: This low-variance innovation is the opposite of the Polarizer. While the Lowest Common Denominator is viewed positively by some, as an expense of this popularity, it is not necessarily seen as profound or brilliant work. In contrast, the Lowest Common Denominator innovation is viewed similarly by all or most, but at the cost of simplifying the idea to facilitate mass diffusion. The mean quality of the innovation is perceived as positive, although rarely as exceptional. The relative simplicity of such manuscripts enables them to be easily cited and serve as hubs linking disparate communities in knowledge networks. In some cases, the malleability of interpreting a scholarly innovation can aid diffusion. Scholars and ideas can diffuse through different fields at different times with different interpretations (e.g., Lamont, 1987; McCain and Salvucci, 2006). The potential complexity of mass-appeal articles varies. Less cohesive fields are comprised of people who share less information. Since shared information begets fact complexity (Carley, 1991), low consensus fields may have to "water down" mass appeal articles to a greater extent than fields with tightly bound paradigms. At the same time, fields with low paradigmatic consensus such as sociology may also rely upon mass-appeal generalist articles in order to retain some semblance of cohesive identity and intellectual solidarity in the field. In the social sciences – where most fields are diffuse and low consensus – creating work that builds bridges between diffuse fields and professions is often perceived as an exemplar of quality (Guetzkow et al., 2004).

<u>The Overlooked Genius</u>: Like many competitive industries, science engages in overproduction. Most academic articles are seldom or never cited, yet considerable monetary resources from and universities and knowledge-funding institutions, as well as vast time commitments of scholars are dedicated to the publication system. The value of knowledge in fields is dependent in part to

the degree that others are aware of it, so excellent articles with little prominence will disproportionately suffer, while the value of other manuscripts that diffuse successfully will be buoyed relative to their true quality. The diffusion process can be strongly influenced by social influences and luck (Granovetter and Soong, 1986; Salganik et al., 2006), and in a winner-take-all market like academia, small and arbitrary factors could make the difference between a manuscript going viral and merely being modestly cited. Just as creative destruction occurs in industry, academia leaves “forgotten intellectuals” (McLaughlin, 1998) as science progresses and moves on to new scholars and ideas. Some exceptional scholars and contributions may only be appreciated by a few, which does not change the true quality of the work. However, failed diffusion does diminish the influence of the work in science, where the utility of work is derived in part from how many other people are familiar with and use it.

Figure 3 presents the incentive structures of the ideal types of supply-side innovations. The figure presents a simple population of ten people who rank the merit (or demerit) of a given contribution from 10 (outstanding) to 0 (neutral/apathetic) to -10 (pernicious). Scores are organized by most to least favorable in the population; it is not necessarily the same evaluators ranking each article highly or poorly.

-- Insert Figure 3 about here –

The shapes of the evaluations in Figure 3 are more important than the hypothetical sums of quality presented as examples. Scientific contributions of higher quality will shift the curves in Figure 3 upward, although it may do so at different rates amongst different population members. For example, an improvement in a specialist manuscript may not make any difference in the evaluations of audiences who are apathetic or unaware of the findings, but may significantly bolster evaluations for those who are interested. In contrast, an improvement to a generalist article intended for mass appeal is more likely to result in a linear increase in perceived quality throughout the entire population.

EVALUATIVE CULTURES AND INNOVATION

Since different types of innovations have different payoff functions, it is worth asking which cultural values[10] should guide science in determining how to gestate and publish knowledge. On the demand side, Starbuck et al. (2008: 262) posited that evaluators are faced with a strategic dilemma of striving to either maximize chances of publishing outstanding articles, or to minimize odds of publishing articles of poor quality. This dilemma poses the question of what the tradeoffs are between improving the chances of publishing a great article and the risk of publishing a dud. The various types of innovations and risks described in this article suggest ways in which the payoff structure is influenced by creativity and scientific fields. Put differently, should gatekeepers strive to maximize the mean quality of their decisions, or work towards a *maximin* solution, where the focus is on ensuring decisions meet a certain quality threshold, as opposed to maximizing the mean?

Schrodt (2013: 37) argued that academia tends to be intellectually risk-averse: “[w]ith a few exceptions the peer review process keeps junk out of the journals. But does so at the cost of biasing the system to publishing safe, predictable, incremental, lowest-common-denominator articles.” Similarly, Horrobin (1990: 1439) contended that peer review is a generally effective

[10] Chin (forthcoming) raises the interesting question of what evidentiary burden should be expected of science in the courtroom, as many legal decisions are informed by scientific ‘facts.’ Presumably, this burden should be at least as high as in science, if not higher.

device for quality control but not suited to identify “those rare articles that genuinely offer the possibility of new approaches” (also see Boudreau et al., forthcoming). This weakness is problematic, because it is unconventional, groundbreaking work that exerts the most profound influence on science. The concern that gatekeepers tend to make conservative decisions also raises the normative question of whether – and if so, to what extent – scientific gatekeepers should take more risks on controversial or unconventional articles. As a result, science becomes increasingly narrow and insular, in a Markovian process where scientists build on the published corpus to construct their own innovations. Further, it is worth considering that it is rational for the truly talented to work in areas with relatively low standard errors of evaluation – which tend to be more difficult – while the less talented have less to lose by working in areas with high standard errors (March, 1994: 44). This may be an incentive for the most talented scientists to pursue normal science (Kuhn, 1962), as opposed to paradigm-challenging work. Relatedly, Ellison (2002) argued that scientific gatekeepers tend to fixate on limited, mutable parts of manuscripts, to the detriment of focus on larger ideas. If true, this is an intellectual problem. However, by focusing on the parts of the manuscript with the most consensus, this facilitates the most fair and efficient evaluation possible, even if the criteria disproportionately relied upon are incomplete and/or epistemologically questionable.

In financial markets, increased value can be derived in exchange for a higher risk-of-ruin (Poundstone, 2006). Put differently, the risk aversion of others creates arbitrage opportunities in competitive markets. Does a similar incentive structure in academia exist? Would accepting a higher risk-of-ruin (i.e. the rate at which ‘mistake’ articles are published) result in higher mean quality if the increased risk also yielded more exceptional outcomes? Preston McAfee’s perception of innovation incentives at the *American Economic Review*, is that the mean quality of published articles rises if unconventional articles are more likely to be rejected (2014: 57). Regardless of whether this is an accurate description of the landscape of scientific innovation, such statements are inevitably colored by subjective value judgments. Journals, editors and scholars have brands and reputations; are they more burnished by a home run, or harmed by a strikeout? More generally, it is worth asking whether the occasional home run justifies a few more strikeouts[11], or if a higher batting average comprised only of singles and doubles is preferable. Even with a blatantly – and perhaps optimal – risky strategy, the line between "mistake" and inevitable downside risk can be blurry. Once again, the answers to these questions are informed by cultural attitudes towards publishing risk in science.

Lamont (2009) posited that academic fields are distinguished by differing evaluative cultures and philosophies. The perceived importance of parts of scientific contributions, including theory, methods and originality vary between fields, institutions and scholars (Redacted, b). Given that new ideas and innovations can be simultaneously perceived in different ways by different scholars and fields, this poses decision-making challenges for scientific gatekeepers. These challenges are in part defined by the winner-take-all structure of most academic markets, and the opposing diminishing returns to quality as one moves up the talent gradient. Digital distribution of manuscripts also has the potential to change the risks and incentives of scientific gatekeeping. As online-only journals do not have the scarcity of space that a traditional print journal does, this calls into question the utility of the practice of extremely low acceptance rates in most social science journals, determined by the often random decisions of a few randomly chosen reviewers before publication. Can or should determinations of quality

[11] Incidentally, in Major League Baseball from 2009-13, strikeout rates were correlated .423 with home runs for batters (Kendall, 2014). In this case, upside risk and downside risk are positively correlated.

be made by all readers post-publication? How should inevitable publishing mistakes made by journals that are willing to print borderline or polarizing articles be perceived?

The conventional gatekeeping system in the social sciences trades decreased incidences of errors of commission (i.e. publishing an unworthy article) in exchange for a larger proportion of errors of omission (rejecting worthy articles). The damage gatekeeping errors of omission does to science may be mollified in part by the fact that scholars can – and often do – submit articles to other journals following a rejection, albeit usually a lower target in the status hierarchy (Calgagno et al., 2012). Further, this risk-averse culture is also enabled by an enormous consumption of skilled intellectual labor in the form of voluntary peer review – as well as unpaid, underpaid and/or subsidized editing jobs – enabling rejections of articles in multiple highly-selective journals prior to eventual publication. This is another factor that underpins the costs and benefits of the conventional peer review system. These costs and benefits should also be weighed regarding the types of innovations that scientists attempt to publish. Different disciplines, journals, institutions and scientists have different goals, philosophes, as well as varying risk and innovation preferences. In the search for scientific quality, tradeoffs between errors of commission and omission for different types of innovations are an overlooked dilemma worth discussing as the black box of peer review is slowly opened. Innovation is costly not only in terms of labor, but also in the costs of failed attempts and risks. The question remains what costs scientists are willing to incur to optimize the innovation and validity of their intellectual fields.

WORKS CITED

Abbott, Andrew. 1988. *The System of Professions*. Chicago: University of Chicago Press.

Akerlof, George A. 1970. "The Market for Lemons: Quality Uncertainty and the Market Mechanism." *Quarterly Journal of Economics*, Vol. 84, pp. 488-500.

American Sociological Association. 2013. Brief of Amicus Curiae - American Sociological Association in Support of Respondent Kristin M. Perry and Respondent Edith Schlain Windsor. Nos. 12-144, 12-307.

Athey, Susan. 2014. Advice for Applying to Grad School in Economics. Accessible: http://faculty-gsb.stanford.edu/athey/gradadvising.html

Barabási, Albert-László and Réka Albert. 1999. "Emergence of Scaling in Random Networks." *Science*, Vol. 286, No. 5439, pp. 509-512.

Baum, Joel A.C. 2013. "The excess-tail ratio: correcting journal impact factors for citation distributions." *M@n@gement*, 16(5), 697-706.

Baum, Joel A.C. 2011. "Free-Riding on Power Laws: Questioning the validity of the Impact Factor as a measure of research quality in organization studies." *Organization*, Vol. 18(4), pp. 449-466.

Baylor University Institute for Studies of Religion. 2012. "A Social Scientific Response to the Regnerus Controversy." Accessed April 20, 2014 from the World Wide Web: http://www.baylorisr.org/2012/06/a-social-scientific-response-to-the-regnerus-controversy/

Beall, Jeffrey. 2014. "List of Standalone Journals: Potential, Possible, Or Probable Predatory Scholarly Open-Access Journals." Accessible: http://scholarlyoa.com/individual-journals/

Bem, Daryl J. 2011. "Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect." *Journal of Personality and Social Psychology*, Vol. 100(3), pp. 407-425.

Beyer, Janice M., Roland G. Chanove and William B. Fox. 1995. "The Review Process and the Fates of Manuscripts Submitted to *AMJ*." *Academy of Management Journal*, Vol. 38(5), 1219-1260.

Bonacich, Philip. 1987. "Power and Centrality: A Family of Measures." *American Journal of Sociology*, Vol. 92(5), pp. 1170-1182.

Boudreau, Kevin J., Eva Guinan, Karim Lakhani and Chistoph Riedl. Forthcoming. "Looking Across and Looking Beyond the Knowledge Frontier: Intellectual Distance and Resource Allocation in Science." *Management Science*.

Bryson, Bethany. 1996. ""Anything But Heavy Metal": Symbolic Exclusion and Musical Dislikes." *American Sociological Review*, Vol. 61(5), pp. 884-899.

Burris, Val. 2004. "The Academic Caste System: Prestige Hierarchies in PhD Exchange Networks." *American Sociological Review*, Vol. 69(2), pp. 239-264.

Calcagno, V., E. Emoinet, K. Gollner, L Guidi, D. Ruths, C. de Mazancourt. 2012. "Flows of Research Manuscripts Among Scientific Journals Reveal Hidden Submission Patterns." *Science*, Vol. 388, 23 November, pp. 1065-1069.

Carey, Benedict. 2011. "ESP Report Sets Off Debate on Data Analysis." *New York Times*, Jan. 10. Accessible: http://www.nytimes.com/2011/01/11/science/11esp.html

Carley, Kathleen. 1991. "A Theory of Group Stability." *American Sociological Review*, Vol. 56(3), pp. 331-354.

Carpenter, Siri. 2012. "Psychology's Bold Initiative." *Science*, Vol. 335(6076), pp. 1558-1561.

Chin, Jason M. Forthcoming. "Psychological Science's Replicability Crisis and What it Means for Science in the Courtroom." *Journal of Psychology, Public Policy and Law*.

Cicchetti, Domenic V. 1991. "Reflections from the Peer Review Mirror." *Behavioral and Brain Sciences*. Vol. 14(1), pp. 167-186.

Colander, David. 1990. *Becoming an Economist*. Boulder: Westview Press.

Cole, Stephen. 1991. "Consensus and the reliability of peer-review evaluations." *Behavioral and Brain Sciences*, Vol. 14(1), pp. 140-141.

Cole, Stephen. 1983. "The Hierarchy of the Sciences?" *American Journal of Sociology*, Vol. 89, pp. 111-39.

Cole, Stephen, Jonathan R. Cole and Gary A. Simon. 1981. "Chance and Consensus in Peer Review." *Science*, Vol. 214(4523), pp. 881-886.

Collins, Randall. 1998. The *Sociology of Philosophies: A Global Theory of Intellectual Change*. Cambridge: Harvard University Press.

Collins, Randall. 1993. "Emotional Energy as the Common Denominator of Rational Action." *Rationality and Society*, Vol. 5(2), pp. 203-230.

Denrell, Jerker and James March. 2001. "Adaptation as Information Restriction: The Hot Stove Effect." *Organization Science*, Vol. 12(5), pp. 523-538.

Ellison, Glenn. 2002. "Evolving Standards for Academic Publishing: A q-r Theory." *Journal of Political Economy*, Vol. 110(5), pp. 994-1034.

Espeland, Wendy N. and Stevens, Mitchell L. 1998. "Commensuration as a social process." *Annual Review of Sociology*, Vol. 24, pp. 313–331.

Evans, James A. 2010. "Industry Induces Academic Science to Know Less about More." *American Journal of Sociology*, Vol. 116(2), pp. 389-452.

Firebaugh, Glenn. 2001. “Reply: The *ASR* Review Process.” *American Sociological Review*, Vol. 66, pp. 619–621.

Fischer, Claude S. 2009. “The 2004 GSS Finding of Shrunken Social Networks: An Artifact?” *American Sociological Review*, Vol. 74, pp. 657-669.

Flynn, F. and J.A. Chatman. 2001. “Strong Cultures and Innovation: Oxymoron or Opportunity?” In Cartwright, S., Cooper, C., Earley, C., Chatman, J., Cummings, T., Holden, N., Sparrow, P. and Starbuck, W. et al., (Eds.), *International Handbook of Organizational Culture and Climate*. Sussex: John Wiley & Sons, pp. 263-287.

Frank, Nathaniel. 2014. “Michigan Courtroom Deals Fake Anti-Gay Research Strategy A Stinging Blow.” *Slate*, March 21.

Frank, Robert H. and Philip P. Cook. 1995. *The Winner-take-all Society*. New York: Free Press.

Friberg, Richard. 2014. “The Wish List of an Editor: Some Reflections on Editing the *Scandinavian Journal of Economics*.” Pp. 45-52 in 44 in Szenberg, Michael and Lall Ramrattan (Eds.), *Secrets of Economics Editors*. Cambridge: MIT Press.

Frickel, Scott and Neil Gross. 2005. “A General Theory of Scientific/Intellectual Movements.” *American Sociological Review*, Vol. 70(2), pp. 204-232.

Gans, Joshua S. and George B. Shepherd. 1994. “How Are the Mighty Fallen: Rejected Classic Articles by Leading Economists.” *Journal of Economic Perspectives*, Vol. 8(1), pp. 165-179.

Gates, Gary J. et al. 2012. “Letter to the editors and advisory editors of Social Science Research.” *Social Science Research*, Vol. 41, pp. 1350-1351.

Gieryn, Thomas F. 1983. “Boundary-work and the Demarcation of Science.” *American Sociological Review*, 48:781-795.

Granovetter, Mark and Roland Soong. 1986. “Threshold Models of Interpersonal Effects in Consumer Demand.” *Journal of Economic Behavior and Organization*, Vol. 7, pp. 83-99.

Guetzkow, Joshua, Michèle Lamont and Grégoire Mallard. 2004. “What is Originality in the Humanities and the Social Sciences?” *American Sociological Review*, Vol. 69(2), pp. 190-212.

Hackett, Edward J. 2005. “Essential Tensions: Identity, Control and Risk in Research.” *Social Studies of Science*, Vol. 35(5), pp. 787-826.

Hackett, Edward H. 1997. “Peer Review in Science and Science Policy.” In *Evaluating Science and Scientists*, Mark S. Frankel and Jane Cave (Eds.). Budapest: Central European University Press.

Hargens, Lowell L. 1988. “Scholarly Consensus and Journal Rejection Rates.” *American Sociological Review*, Vol. 53(1), pp. 139-151.

Harvey, Campbell R. 2014. "Reflections on Editing the Journal of Finance, 2006 to 2012." Pp. 67-82 in Szenberg, Michael and Lall Ramrattan (Eds.), *Secrets of Economics Editors*. Cambridge: MIT Press.

Herndon, Thomas, Michael Ash and Robert Pollin. 2013. "Does High Public Debt Consistently Stifle Economic Growth? A Critique of Reinhart and Rogoff." Political Economy Research Institute - Working Paper Series (322).

Horrobin, David F. 1990. "The Philosophical Basis of Peer Review and the Suppression of Innovation." *Journal of the American Medical Association*, Vol. 263(10), pp. 1438-1441.

Ioannidis, J.P.A. 2005. "Why Most Published Research Findings Are False." *PLoS One Medicine*.

Jasso, Guillermina. 1985. "Marital Coital Frequency and the Passage of Time: Estimating the Separate Effects of Spouses' Ages and Marital Duration, Birth and Marriage Cohorts, and Period Influences." *American Sociological Review*, Vol. 50(2), pp. 224-241.

Kahn, Joan R. and J. Richard Udry. 1986. "Marital Coital Frequency: Unnoticed Outliers and Unspecified Interactions Lead to Erroneous Conclusions." *American Sociological Review*, Vol. 51, No. 5, pp. 734-737.

Kahneman, Daniel and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica*, Vol. XLVII, pp. 263-291.

Kendall, Evan. 2014. "Does more power mean more strikeouts?" Accessed May 15, 2014: http://www.beyondtheboxscore.com/2014/1/12/5299086/home-run-strikeout-rate-correlation

Kennelly, Ivy, Sabine Merz and Judith Lorber. 2001. "What is Gender?" *American Sociological Review*, Vol. 66, pp. 598-605.

Kennedy, Donald. 2002. "To Publish or Not to Publish." *Science*, Vol. 295, pp. 1793.

Knight, Frank H. 1921. *Risk, uncertainty and profit*. Boston and New York: Kelley.

Kraemer, Helena Chmura. 1991. "Do we really want more "reliable" reviewers?" *Behavioral and Brain Sciences*, Vol. 14(1), pp. 152-154.

Kuhn, Thomas. 1977. *The Essential Tension: Selected Studies in Scientific Tradition and Change*. Chicago: University of Chicago Press.

Kuhn, Thomas. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Lamont, Michèle. 2009. *How Professors Think: Inside the Curious World of Academic Judgment*. Cambridge: Harvard University Press.

Lamont, Michèle. 1987. "How to Become a Dominant French Philosopher: The Case of Jacques Derrida." *American Journal of Sociology*, Vol. 97, pp. 347-375.

Langfeldt, Liv. 2001. "The Decision-Making Constraints and Processes of Grant Peer Review, and Their Effects on the Review Outcome." *Social Studies of Science*, Vol. 31(6), pp. 820-841.

Latour, Bruno and Steve Woolgar. 1979. *Laboratory Life: The Construction of Scientific Facts*. Los Angeles: Sage.

Leahey, Erin. 2008. "Overseeing Research Practice: The Case of Data Editing." *Science, Technology and Human Values*, Vol. 33(5), pp. 605-630.

Leahey, Erin. 2004. "The Role of Status in Evaluating Research: The Case of Data Editing." *Social Science Research,* Vol. 33(3), pp. 521-537.

Levinthal, Daniel A. and James G. March. 1993. "The Myopia of Learning." *Strategic Management Journal*, Vol. 14(S2), pp. 95-112.

Li, Danielle. 2013. "Expertise vs. Bias in Evaluation: Evidence from the NIH." Accessed December 9, 2013 from the World Wide Web. Accessible: http://www.kellogg.northwestern.edu/~/media/Files/Faculty/Research/Li_NIH_May20_2013.ashx

Lindsey, Duncan. 1988. "Assessing Precision in the Manuscript Review Process: A Little Better than a Dice Roll." *Scientometrics*, Vol. 14(1-2), pp. 75-82.

Lotka, A.J. 1926. "The Frequency Distribution of Scientific Productivity." *Journal of the Washington Academy of Sciences*, 16: 317-23.

Lynn, Freda B. 2012. "The Role of Expertise in the Consumption of Status Signals." Stanford, CA: Stanford Graduate School of Business Colloquium.

March, James. 1994. *A Primer on Decision Making*. New York: Free Press.

March, James and Zur Shapira. 1987. "Managerial Perspectives on Risk and Risk Taking." *Management Science*, Vol. 33(11), pp. 1404-1418.

Mark, Noah. 1998. "Beyond Individual Differences: Social Differentiation from First Principles." *American Sociological Review*, Vol. 63, pp. 309-330.

Mauboussin, Michael J. 2012. *The Success Equation*. Boston: Harvard Business Review Press.

McAfee, R. Preston. 2014. "Edifying Editing." Pp 33-44 in Szenberg, Michael and Lall Ramrattan (Eds.), *Secrets of Economics Editors*. Cambridge: MIT Press.

McCain, Katherine and Laura Salvucci. 2006. "How influential is Brooks' Law? A longitudinal citation context analysis of Frederick Brooks' The Mythical Man-Month." *Journal of Information Science*, Vol. 32(3), 277-295.

McLaughlin, Neil G. 1998. "How to Become a Forgotten Intellectual: Intellectual Movements and the Rise and Fall of Erich Fromm." *Sociological Forum*, Vol. 13(2), pp. 215-246.

McPherson, Miller, Lynn Smith-Lovin and Matthew E. Brashears. 2006. "Social Isolation in America: Changes in Core Discussion Networks over Two Decades." *American Sociological Review*, Vol. 71(3), pp. 353-375.

Merton, Robert K. 1968. "The Matthew Effect in Science." *Science*, Vol. 159, pp. 56-63.

Miller, Eleanor and Carrie Yang Costello. 2001. "The Limits of Biological Determinism." *American Sociological Review*, Vol. 66, pp. 592-598.

Miller, Joann and Robert Perrucci. 2001. Back Stage at *Social Problems*: An Analysis of the Editorial Decision Process, 1993-1996. *Social Problems,* Vol. 48(1), pp. 93-110.

Mone, Mark A. and William McKinley. 1993. "The Uniqueness Value and its Consequences for Organization Studies." *Journal of Management Inquiry*, Vol. 2(3), pp. 284-296.

Mueller, Jennifer S., Shimul Melwani and Jack A. Goncalo. 2011. "The Bias Against Creativity: Why People Desire but Reject Creative Ideas." *Psychological Science*, Vol. 22(12), pp. 1-19.

Peters, D.P. and S.J. Ceci. 1982. "Peer Review Practices of Psychological Journals: The Fate of Published Articles, Submitted Again." *Behavioral and Brain Sciences*, Vol. 5, pp. 187-195.

Pfeffer, Jeffrey. 1993. "Barriers to the Advance of Organizational Science: Paradigm Development as a Dependent Variable." *Academy of Management Review*, Vol. 18(4), pp. 599-620.

Podolny, Joel. 1993. "A Status-Based Model of Market Competition." *American Journal of Sociology*, Vol. 98(4), pp. 829-872.

Poundstone, William. 2006. *Fortune's Formula: The Untold Story of the Scientific Betting System That Beat the Casinos and Wall Street*. New York: Hill and Wang.

Rabin, Matthew and Richard H. Thaler. 2001. "Anomalies: Risk Aversion." *Journal of Economic Perspectives*, Vol. 15(1), pp. 219-232.

Regnerus, Mark. 2012. "How different are the adult children of parents who have same-sex relationships? Findings from the New Family Structures Study." *Social Science Research*, Vol. 41, pp. 752-770.

Reinhart, Carmen M and Kenneth S. Rogoff. 2010. "Growth in a Time of Debt." *American Economic Review*, Vol. 100(2), pp. 573–78.

Risman, Barbara. 2001. "Calling the Bluff of Value-Free Science." *American Sociological Review*, Vol. 66, pp. 605-611.

Ritchie, Stuart J., Richard Wiseman and Christopher C. French. 2012. "Failing the Future: Three Unsuccessful Attempts to Replicate Bem's 'Retroactive Facilitation of Recall' Effect." *PLoS ONE*, Vol. 7(3), pp. 1-5.

Rosen, Sherwin. 1981. "The Economics of Superstars." *American Economic Review*, Vol. 7(5), pp. 845-858.

Salganik, Matthew J., Peter Sheridan Dodds and Duncan J. Watts. 2006. "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market", *Science*, Vol. 311 no. 5762 pp. 854-856.

Schrodt, Philip. 2013. "Seven Deadly Sins of Quantitative Political Analysis." Accessible: http://eventdata.psu.edu/7DS/7DS.University.chpt.pdf

Shepherd, George B. 1994. *Rejected: Leading Economists Ponder the Publication Process*. Sun Lakes: Thomas Horton & Daughters.

Sherkat, Darren E. 2012. "The Editorial Process and Politicized Scholarship: Monday Morning Editorial Quarterbacking and a Call for Scientific Vigilance." *Social Science Research*, Vol. 41, pp. 1346-1349.

Smith, Richard. 2006. "Peer Review: A Flawed Process at the Heart of Science and Journals." *Journal of the Royal Society of Medicine*, Vol. 99(April), pp. 178-182.

Sociological Science. 2013. "Journal Information" Accessed August 20, 2013 from the World Wide Web: http://www.sociologicalscience.com/journal-information/

Staw, Barry M. 1995. "Why no-one really wants creativity." In *Creative Action in Organizations: Ivory Tower Visions and Real Voices*. Cameron Ford, Dennis Gioia (Eds.), Thousand Oaks: Sage, pp. 161-166.

Starbuck, William H. 2003. "Turning Lemons into Lemonade: Where is the Value in Peer Reviews?" *Journal of Management Inquiry*, Vol. 12(4), pp. 344-351.

Starbuck, William H, Herman Aguinis, Alison M. Konrad and Yehuda Baruch. 2008. "Epilogue: Trade-Offs among Editorial Goals in Complex Publishing Environments." In Y. Baruch, A. Konrad, H. Aguinis & W. H. Starbuck (Eds.), *Opening the Black Box of Editorship*, pp. 250-270. Basingstoke, UK: Palgrave Macmillan.

Stinchcombe, Arthur L. 1994. "Disintegrated Disciplines and the Future of Sociology." *Sociological Forum*, Vol. 9(2), pp. 279-291.

Stinchcombe, Arthur L. and Richard Ofshe. 1969. "On Journal Editing as a Probabilistic Process." *The American Sociologist*, Vol. 4(2), pp. 116-117.

Udry, J. Richard. 2000. "Biological Limits of Gender Construction." *American Sociological Review*, Vol. 65, pp. 443-457.

United States District Court Eastern District of Michigan Southern Division. 2014. *DeBoer vs. Snyder*. Civil Action No. 12-CV-10285.

Wagenmakers, E.J., R. Wetzels, D. Borsboom and H.L. van der Maas. 2011. "Why psychologists must change the way they analyze their data: the case of psi: comment on Bem (2011)." *Journal of Personality and Social Psychology*, Vol. 100(3), pp. 426-432.

Watts, Michael. 2014. "What Editors Do and Want to Tell Us." Pp. 1-32 in Szenberg, Michael and Lall Ramrattan (Eds.), *Secrets of Economics Editors*. Cambridge: MIT Press.

Whitley, Richard. 2000 [1984]. *The Intellectual and Social Organization of the Sciences*. 2nd Ed. Oxford: Clarendon.

Zimmer, Carl. 2011. "It's Science, but Not Necessarily Right." *New York Times*, June 25. Accessible: http://www.nytimes.com/2011/06/26/opinion/sunday/26ideas.html

Zuckerman, Harriet and Robert K. Merton. 1971. "Patterns of evaluation in science: Institutionalisation, structure and functions of the referee system." *Minerva*, Vol. 9(1), pp. 66-100.

Figure 1 – Citation Distributions of Articles in Social Science Journals, 1996-2005

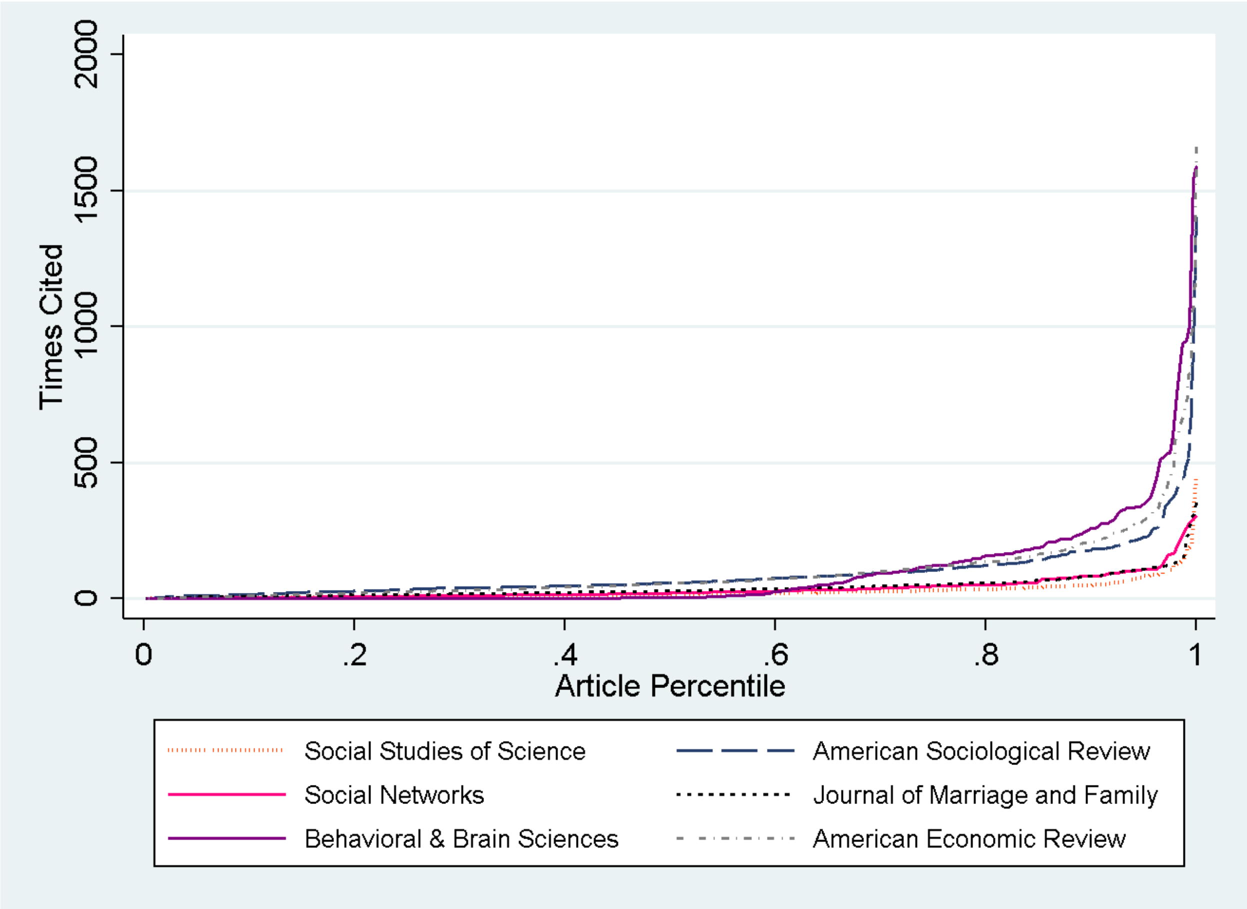

Figure 2 – Hypothetical Distribution of Diminishing Talent in a Winner-Take-All Market

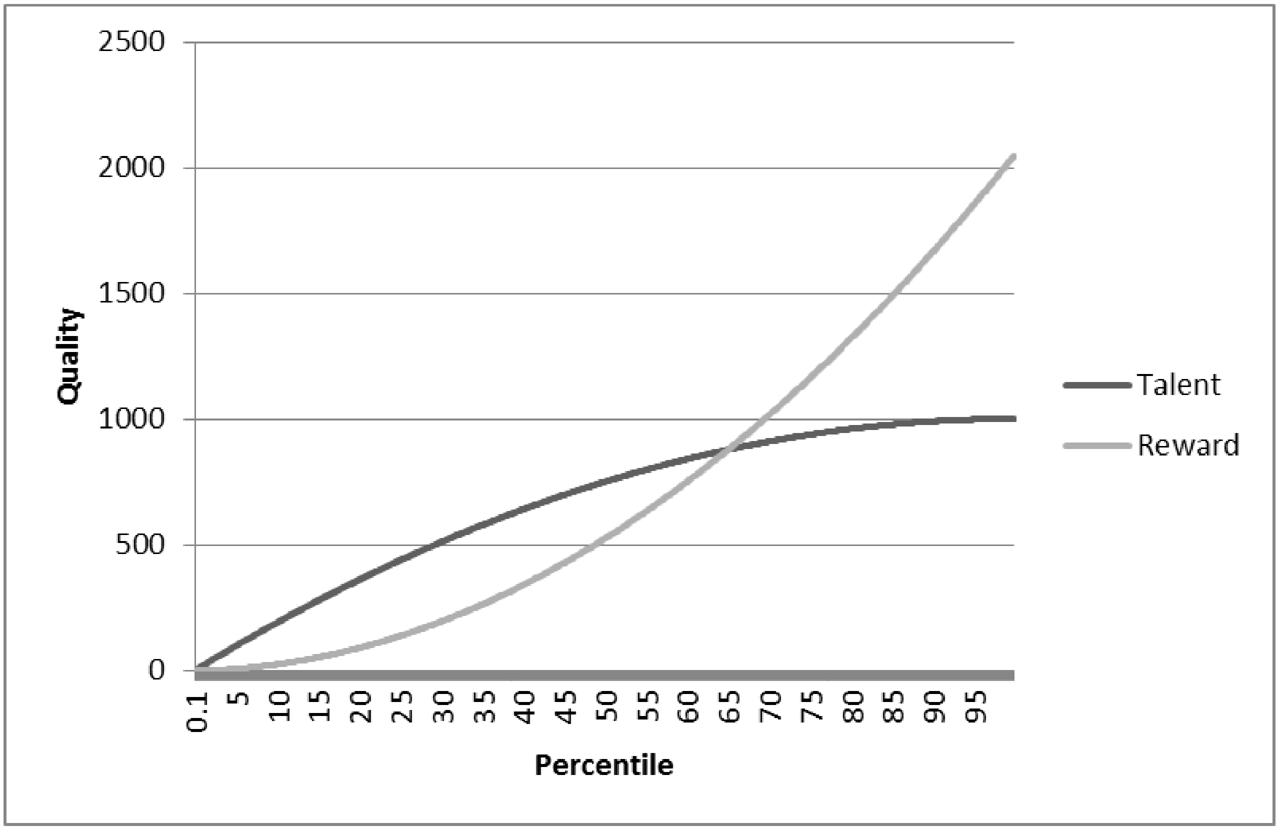

Figure 3 – Payoff Distributions for Ideal Types of Innovations