Beetles: Biased Promotions and Persistence of False Belief

George Akerlof, Pascal Michaillat*

November 18, 2017

This paper develops a theory of promotion based on evaluations by the already promoted. The already-promoted show favoritism toward candidates with similar beliefs, just as beetles are more prone to eat the eggs of other species. With such egg-eating bias, false beliefs may not be eliminated by the promotion system. The main application is to scientific revolutions: when tenured scientists show favoritism toward tenure candidates with similar beliefs, science may not converge to the true paradigm. We extend the statistical concept of power to science: the power of the tenure test is the probability (absent any bias) of denying tenure to a scientist who adheres to the false paradigm, just as the power of any statistical test is the probability of rejecting a false null hypothesis. The power of the tenure test depends on the norms regarding the appropriate criteria to use in promotion and the empirical evidence available to apply these criteria. Economics and other social sciences are particularly at risk of capture by false paradigms because they have low power. Another application is to hierarchical organizations.

*Akerlof: University of California–Berkeley. Michaillat: Brown University. We thank Robert Akerlof, John Friedman, Mitchell Hoffman, Ronny Razin, Jesse Shapiro, and many seminar participants for valuable suggestions. This work was supported by the Canadian Institute for Advanced Research.

1. Introduction

In his classic essay on economic methodology, Friedman (1953) saw the difference between social and physical sciences as one of degree: largely because social sciences lacked the experimental evidence typically available in the physical sciences. But he did not view this characteristic as preventing false hypotheses from being "weeded-out": only that the weeding would be slow (p. 11). In this paper we develop a model of science and revisit this argument. When scientists are unbiased, since the social sciences have lower statistical power than the physical sciences, the social sciences will indeed converge to the truth more slowly. But there is also homophily, of which Friedman was aware (p. 30). And with homophilic bias in the tenure process, the social sciences, with their lower statistical power, may never converge to the truth. Thus limited statistical power in the social sciences may do more than just slow down scientific progress: it may generate different progress dynamics, according to which superior paradigms may be dominated by inferior ones.

To describe the interaction between statistical power and bias in science, we build a model inspired by the behavior of flour beetles. In the 1950s and 1960s, entomologists at the University of Chicago conducted experiments with flour beetles. They placed populations of two different species of beetle into jars of flour, with no effective constraint on either food or space (the two species are depicted in Figure 1). It was expected that the more biologically "fit" species would dominate; but that is not what happened. Instead, in short order, in each of many independent experiments, one, or the other, type of beetle vanished. On close inspection, the reason was determined: it was found that beetles not only eat their own species' eggs, they are also yet more prone to eat the eggs of other species. As a result of this egg-eating bias, as one species increases relative to the other, it is more likely to increase yet further.[1]

Our paper builds a mathematical, promotion-chain model inspired by the behavior of the beetles in the jars. This model captures the population dynamics of several human situations, since in several respects, human behavior parallels the behavior of the beetles. A beetle encountering an egg in the flour jar makes the implicit decision whether to eat the egg or to let it become an adult beetle like herself. Similarly, in many human institutions, candidates for promotion to a higher rank (like the egg) are evaluated by current holders of that rank (like the beetle). Also, the own-species preference of the beetles corresponds to ingroup favoritism/outgroup bias, which is commonly observed in human promotions.

Our main application is to the evolution of beliefs at the time of a "scientific revolution." Such

[1]The first such experiment is described in Park (1954) and analyzed in Neyman, Park, and Scott (1956). Egg-eating bias is documented in Park et al. (1965) and Park, Mertz, and Nathanson (1968). See Costantino and Desharnais (1991) for an overview of the research on competition between species of flour beetles.

A. Confused flour beetle (*Tribolium Confusum*) B. Red flour beetle (*Tribolium Castaneum*)

Figure 1. The Flour Beetles Used in the Chicago Laboratory Experiments

Sources: Panel A: Sarefo, multi-license with GFDL (http://www.gnu.org/copyleft/fdl.html) and Creative Commons CC-BY-SA-3.0 (http://creativecommons.org/licenses/by-sa/3.0/). Panel B: Eric Day, Virginia Tech.

a revolution occurs, according to Kuhn (1996), after the appearance of a superior paradigm. He says that such paradigms tend to appear after accumulations of anomalous questions regarding an old paradigm commonly held by the scientists in some area. We view one species of scientist as those who adhere to the better paradigm; and another species of scientist as those who adhere to the worse paradigm. The promotion-chain model describes the population dynamics of those adhering respectively to the better, or to the worse, paradigm. Similar to the beetles, for whom an egg may turn into an adult, the dynamics of the population of scientists depends critically on whether advisees trained by established scientists become established scientists themselves. And, similar to the beetles' bias toward eating the eggs of the other species, scientists have biases in favor of those of their own paradigm, and against those of the alternative paradigm. In our model, these preferences manifest themselves in decisions regarding the grant of tenure.

Our model of population dynamics characterizes the situations under which a better paradigm will replace a worse one. With the beetles, the more biologically fit species does not always prevail. Similarly, in science, because of the ingroup favoritism/outgroup bias, the "truer," scientifically-more-"fit" paradigm does not necessarily prevail. Convergence among scientists in belief toward the better paradigm—or toward the worse paradigm—depends critically on the difference between the probability of denying tenure to a worse-paradigm candidate and the probability of denying tenure to a better-paradigm candidate.

Concepts from statistics yield an intuition regarding the nature of this difference. We can view the grant of tenure as an indirect test of the validity of the paradigm followed by the tenure candidate. If we take as the null hypothesis that the candidate believes in the more correct paradigm, then, in the language of statistics, the probability of rejection of a worse-paradigm candidate is the statistical power of the test; and, correspondingly, the probability of rejection of a better-paradigm candidate is the statistical significance of the test. The difference between the power of the test and its significance is the Youden index, introduced by Youden (1950).

In our baseline, no-bias model, the convergence of belief to one paradigm or another then depends upon the sign of the Youden index. With ingroup favoritism/outgroup bias, and with other deviations from the baseline model, convergence still depends upon the Youden index. In these cases, however, the condition on the Youden index needs to be appropriately adjusted to take into account deviations from the baseline model.

With ingroup favoritism/outgroup bias, the gap between the Youden index and the bias determines when convergence in belief among scientists is toward the worse paradigm, or on the contrary, toward the better paradigm. For a given composition of the population of scientists, science is more likely to converge to the truth when the Youden index is larger and the bias is smaller. This means that sciences with low power and thus low Youden index are more prone to capture by false paradigms. It also means that science can start converging to the truth if power suddenly increases.

Our analysis of what causes convergence to truer paradigms also leads us to the answer to another question, which Kuhn poses at the end of *The Structure of Scientific Revolutions*. Why, for centuries, has modern science been so continuously successful? There we shall see the importance of the norm for what it means to be a scientist: that a scientist's beliefs should accord with the outcome of high-power tests for the science. Such a norm then gives reason why promotion to membership in the fellowship of established scientists should, likewise, make use of such high-power tests, insofar as they are available. Its emphasis on reliance on high-power tests, and the evidence that comes from them, seems to be one of the special features of science.

An historical example illustrates the use of the concept of high-power tests. Galileo's new telescopes, with their increased powers of observation, played a critical role in the adoption of the heliocentric/rotating-earth cosmology of Copernicus. According to Kuhn (1957), prior to Galileo, when an observation had not quite fit with the Ptolemaic system, it could be easily explained away by adding epicycles. With the power of Galileo's improved telescopes such explanations became much more difficult to countenance, and adherence to the Copernican system took off.

Our first application pictures only one promotion: from advisee to tenured scientist. The paper also has a second application. It considers promotions up an organizational ladder, with promotions to a higher rung drawn from those in the rung just below, and also with the candidates judged by those already in the next-higher rung. Egg-eating bias can result, as the number of rungs in the ladder becomes large, in the capture of the top levels by those with inferior beliefs. And, again, this possibility depends critically on the value of a Youden index.

2. Related Literature

In our model of science, tenured scientists tend to favor the untenured scientists who adhere to their view of the world; they also tend to discriminate against those who support another view. This type of bias, similar to the bias of beetles eating each other's eggs, has been extensively documented by sociologists—who call it "homophily"—and by social psychologists—who call it "intergroup bias." Sociologists have observed that in many contexts, people tend to connect with and favor others who are similar.[2] While sociologists find that people exhibit homophily based on demographic or psychological characteristics, social psychologists have found that even minimal divisions can create strong biases. In the Robbers Cave experiment, intense rivalries developed amongst two groups of 11-year-old boys who had been separated into different cabins in an Oklahoma state park (Sherif et al. 1961). Later Tajfel and his followers showed that even minimal divisions (for example, telling boys whether they had a preference for Klee or Kandinsky) would produce such favoritism (Tajfel and Turner 1979, 1986).[3]

There is also direct evidence that this type of bias operates in the scientific world. According to Lamont (2009), in academia, people favor others from the same school of thought, just as we assume in our model. She suggests that such homophily is present in academic evaluations at every level: admittance to graduate school, performance in coursework, evaluation of thesis, first job, acceptance of papers, award of grants, invitations to conferences, tenure evaluations, mentoring. Studies also provide evidence of such bias in scientific peer review (for example, Mahoney 1977; Travis and Collins 1991).

Homophilic bias is widespread in hierarchical organizations, as we assume in our model of organizations. In an ethnographic study of a large US corporation, Kanter (1993) has found that "managers tend to carefully guard power and privilege for those who fit in, for those they see as 'their kind' " (p. 48); "excellence . . . was not always the selection criterion" but "predictability and trustworthiness by virtue of membership in the right group . . . were likely to be the factors in the choice of the key managers" (p. 51); "managers reproduce themselves in kind" (p. 63). There is also homophilic bias in recruiting, along several dimensions: productivity (Burks et al. 2015), culture (Rivera 2012), and ethnicity (Bertrand and Mullainathan 2004; Oreopoulos 2011).

In addition, our paper contributes to three literatures in economics. In its application to science, this paper contributes to an emerging literature on the theoretical underpinnings of

[2]For a survey of the research on homophily, see McPherson, Smith-Lovin, and Cook (2001).

[3]For a survey of the research on intergroup bias, see Haslam (2004) and Dovidio and Gaertner (2010).

scientific progress. The closest paper to our own is Brock and Durlauf (1999). They show that convergence of belief in science occurs with some generality, as long as the scientists in the respective field have a tendency to conform to each other. In contrast, this paper models conformity more specifically: being expressed in biases in promotions to tenure. In our model, such promotions are vetted by the already tenured, who favor candidates with the same beliefs as themselves, and who also, symmetrically, disfavor candidates with different beliefs. Thus in the two cases conformity takes different forms—Brock and Durlauf: I will change my beliefs toward yours; here: if you don't share my beliefs, I will reduce my chances of awarding you tenure. Since desire for conformity, albeit in quite different form, is the core of both models, it is not surprising that the two models produce similar outcomes—most notably, convergence of belief.

With its focus on the suppression of ideas and scientific progress, our paper is also similar in concept to Bénabou, Ticchi, and Vindigni (2015), which explains the tensions between scientific progress and religion; opposition to science comes from religion, and is arbitrated by the state. The two papers are quite different, however. We focus on specific biases in the system of advancement of scientific careers, whereas Bénabou, Ticchi, and Vindigni focus on a bias imposed by the State in the presence of a powerful Church.

Furthermore, our paper is complementary to another paper on the history of science: Bramoullé and Saint-Paul (2010). In their model of scientific progress, periods of normal science alternate with periods of scientific revolution. Scientists decide, based on their respective incentives, to continue working with the old paradigm (normal science) or to create a new one (revolutionary science). In contrast, our paper focuses on competition between two given paradigms (a new one and an old) during a period of revolutionary science. Scientists adhering to the two different paradigms are variously promoted to established positions, and the allegiance of these established scientists constitutes the state of scientific knowledge.

The paper also develops a model of shared beliefs in hierarchical organizations. It is particularly related to Van den Steen (2010), who explores the reasons for homophily in recruiting and promotion; to Montgomery (1991), who describes the implications of homophilic recruiting for the labor-market outcomes of workers with different abilities and social networks; and to Besley and Ghatak (2005), who study the design of incentives in organizations in which workers and managers have homophilic preferences. While these papers focus on single homophilic promotions, in contrast, our paper studies the consequences of chains of such promotions. Our paper is also related to Bernhardt, Hughson, and Kutsoati (2006). They consider a model in which workers overinvest in skills that increase their chances of being hired and then promoted, and analyze how promotion chains can exacerbate the investment distortions. Like us, they find that in steady state the distribution of managerial expertise may be suboptimal; the novelty of our paper is to study how statistical power in promotion interacts with homophilic bias.

More generally, our paper contributes to the literature on the emergence of dysfunctional beliefs in social groups.[4] In contrast to that literature, which focuses on social learning, our paper introduces chains of promotions in shaping social knowledge. In our paper, people do not learn: those with certain beliefs are promoted to established positions, and the accumulation of the beliefs of these established experts constitute social knowledge.

3. Model of Science

We present a model describing the evolution of a population of scientists whose beliefs are split between two paradigms: Better and Worse. The Better paradigm gives a better description of the world. Scientific knowledge makes progress when a larger fraction of established scientists believe in the Better paradigm. Established scientists are those who have been granted tenure.

The model is defined by a list of assumptions. Despite the length of the list, the model should be easy to understand: since it is fashioned after the standard system in academia regarding grant of tenure, which is likely to be familiar to most readers. The model is basic, but it delivers the main results and conveys intuition. Section 5.3 will extend the basic model in several directions.

3.1. Two Paradigms

There are two distinct paradigms: Better and Worse. The Better paradigm gives a more correct description of the world; the Worse paradigm gives a less correct description. Each scientist adheres either to the Better, or to the Worse, paradigm. A scientist who adheres to a given paradigm performs empirical and theoretical investigations articulated around this paradigm.

At time t, $B(t)$ tenured scientists believe in the Better paradigm; $W(t)$ tenured scientists believe in the Worse paradigm. The fraction of tenured scientists who believe in the Better paradigm is $\sigma(t)$:

$$\sigma(t) = \frac{B(t)}{B(t) + W(t)}.$$

Knowledge is embodied by established scientists, so the strength of a paradigm is measured by the fraction of its adherents among tenured scientists. Since the Better paradigm offers a

[4]A central finding of this literature is that when people use Bayesian logic to infer from the previous actions of others, informational cascades occur, in which groups herd on wrong beliefs (see Banerjee 1992; Bikhchandani, Hirshleifer, and Welch 1992). Such cascades may be even more likely when people are slightly naive in their inference (Eyster and Rabin 2010). False beliefs also emerge in other networks, with various learning mechanisms (for a recent example, see Bloch, Demange, and Kranton 2016).

superior description of the world, knowledge makes progress when $\sigma(t)$ becomes closer to 1.

3.2. Beliefs of Tenured Scientists and Advisees

Tenured scientists train advisees at rate λ. Advisees adhere to the same paradigm as their advisor. During their entire career, scientists adhere to the same paradigm: there is no defection.

Of course, when the Better paradigm is invented, a few tenured scientists defect from the Worse paradigm and spontaneously adhere to the Better paradigm, maybe because they were dissatisfied with the Worse paradigm and are convinced by the Better one. The number of scientists who are converted early is $B(0)$. We do not model the stage of early adoption—we take $B(0)$ and $W(0)$ as given. Our focus is on the systematic competition between Better and Worse paradigms through the tenure system once early adopters start teaching students about the Better paradigm.

Once an advisee is trained, she becomes an untenured scientist and produces research articulated around her paradigm. Then she is brought up for tenure. If she receives tenure, she continues doing research, advises students, and retires at rate δ. If she does not receive tenure, she quits academia.

3.3. The Quality of Research

The research of scientists is based on the paradigm in which they believe. A scientist uses her paradigm as a theoretical framework to guide empirical measurement, to explain empirical observations, to make theoretical predictions, and to further articulate the paradigm.

The quality of a scientist's research is partially determined by the implications of her paradigm. The Better paradigm offers a more correct description of the world: it generates more fruitful empirical investigations, explains more observations, makes more accurate predictions, and can more easily be adjusted to resolve empirical anomalies. As a result, on average, the research of Better scientists is of higher quality than that of Worse scientists. However, there is uncertainty in research quality because no paradigm perfectly describes the real world; depending on the phenomena studied, the paradigm will be more or less successful and the research of higher or lower quality.

The research process brings additional uncertainty to the quality of a scientist's research: empirical observations are difficult to obtain and subject to measurement error; theoretical explanations and predictions are hard to formulate; and scientists vary in skill, effort, and imagination. Hence, research quality is noisy and only partially determined by the underlying paradigm.

3.4. The Tenure Process

The evaluator of a candidate for tenure is randomly chosen from the pool of tenured scientists. Thus the untenured scientist is evaluated by a scientist with belief in the Better paradigm with probability $\sigma(t)$ and by a scientist with belief in the Worse paradigm with probability $1 - \sigma(t)$.

Let's first consider the case with no egg-eating bias in the grant of tenure: that is, tenured scientists are neither more nor less prone to grant tenure to candidates whose beliefs are in agreement with their own, than to candidates with the opposite beliefs. In that case, tenure is entirely determined by the quality of research: all candidates whose research quality is above a certain threshold are granted tenure; all others are denied.

We denote as α the probability of denying tenure to those who believe in the Better paradigm, and as β the probability of granting tenure to those who believe in the Worse paradigm. We have seen that research quality is noisy: it is affected by luck, the competence of the scientist, and the accuracy of the paradigm. Hence, not all Better candidates are granted tenure and not all Worse candidates are denied. Nevertheless, since the Better paradigm is more correct, Better candidates tend to produce better research than Worse candidates. Better candidates are therefore more likely to receive tenure than Worse candidates: $1 - \alpha \geq \beta$.

3.5. Egg-Eating Bias

To the concepts of α and β we now add the egg-eating bias, $\epsilon > 0$. As the untenured scientist believes in the Better or Worse paradigm, she accordingly uses that paradigm as the basis for her empirical or theoretical investigations. And, then, the agreement—or disagreement—of belief between the untenured scientist and her tenured evaluator affects the tenure decision. A Better evaluator has increased probability ϵ (relative to the unbiased test) of giving tenure to an advisee who believes in the Better paradigm; and she has decreased probability ϵ of giving tenure to an advisee who believes in the Worse paradigm. Symmetrically, a Worse evaluator has decreased probability ϵ (relative to the unbiased test) of giving tenure to an advisee who believes in the Better paradigm; and she has increased probability ϵ of giving tenure to an advisee who believes in the Worse paradigm.

Formally, when tenured scientists are biased, the tenure probabilities are as follows. A Better evaluator denies tenure to a Better scientist with reduced probability $\alpha - \epsilon$. On the other hand, a Worse evaluator denies tenure to a Better scientist with increased probability $\alpha + \epsilon$. A Better evaluator grants tenure to a Worse scientist with reduced probability $\beta - \epsilon$. Finally, a Worse evaluator grants tenure to a Worse scientist with increased probability $\beta + \epsilon$. Thus, scientists are biased in favor of applicants who support their view of the world and against applicants who support another view of the world.[5]

3.6. Faithful Representation of Kuhn

Our model gives a faithful representation of what Kuhn (1996) calls "revolutionary science"—in contrast to "normal science." Most of the time, scientists engage in normal science. Normal science is the determination of important facts; match of the existing paradigm with these facts; and more detailed articulation of the paradigm (p. 34). During periods of normal science, scientists work within the framework of an accepted paradigm, which is "revealed in its textbooks, lectures, and laboratory exercises" (p. 43), and they aim to improve the paradigm and its fit with nature. Our model focuses, instead, on periods of "revolutionary science": when two paradigms compete. According to Kuhn, such phases of scientific progress arise in response to discovery of anomalies inconsistent with the old paradigm (p. 66, p. 84). In this phase of science, the decision to reject one paradigm is also the decision to accept another (p. 77).

In our model, scientific knowledge is embodied by established scientists, with the strength of a paradigm indexed by the fraction of scientists adhering to it. This representation is also faithful to Kuhn, who says that a paradigm becomes prevalent only after its acceptance by the scientific community. Scientific revolutions are battles of old versus new paradigms for the allegiance of that community (p. 94, p. 145)—in agreement with the central role of the variable σ in our model.

In Kuhn's model, the new followers of a paradigm are not converts from the old one, but are freshly-minted scientists who adhere to the new. Following Kuhn, in periods of normal science, prior to the intrusion of the new paradigm, all scientists will have the same interpretation of the basic science, which, in the 18th and 19th centuries, was represented by the scientific "classics," and, in more modern times, is represented by the textbooks (pp. 19–20).

In our model, adherence to the Better Paradigm comes from two sources. A presumed small number of tenured scientists immediately convert to the Better paradigm at time 0, perhaps because of individual conviction of its superiority to the Worse. And then over time, advisees whose advisors believe in the Better paradigm join their ranks.[6]

Such a representation corresponds to Kuhn's views regarding the origins of Better-paradigm adherents. Thus he approvingly quotes Planck: "a new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it" (p. 151).[7]

[5] The bias ϵ is bounded such that the four probabilities $\alpha + \epsilon$, $\alpha - \epsilon$, $\beta + \epsilon$, and $\beta - \epsilon$ remain in $(0, 1)$.

[6] In a slight extension of our model, advisees who defect from the Worse paradigm are an additional source of adherents to the Better paradigm.

[7] Similarly, Kuhn also quotes Darwin: "Although I am fully convinced of the truth of the views given in this

Kuhn also emphasizes "resistance" against new paradigms by the adherents of the old paradigm. Such resistance is represented by the parameter ϵ in our model. Those biases are, of course, a form of ingroup favoritism/outgroup bias, which has been documented in many different guises by psychologists and sociologists.

In sum, our model—parsimoniously, although perhaps a bit coarsely—captures Kuhn's description of "scientific revolutions."

4. Analogy with Statistics and Youden Index

In practice nobody knows whether the Better paradigm is indeed more correct than the Worse. Thus the evaluation of a tenure case can be interpreted as a statistical test, in the tradition of Neyman and Pearson (1933), in which the null hypothesis is that the tenure candidate believes in the more correct paradigm, and the alternative hypothesis is that the tenure candidate believes in the less correct paradigm. Furthermore, granting tenure can be interpreted as failing to reject the null hypothesis, while denying tenure can be interpreted as rejecting the null hypothesis. Using this analogy with statistics, we introduce three concepts:

DEFINITION 1. *The* statistical significance *of the tenure test is the probability of denying tenure to a Better scientist when there is no egg-eating bias:* α. *The* statistical power *of the tenure test is the probability of denying tenure to a Worse scientist when there is no egg-eating bias:* $1 - \beta$. *The* Youden index *of the tenure test is* $J = 1 - \alpha - \beta$.

The tenure test assesses many attributes of the tenure candidate beside her belief in the better or worse paradigm. Nevertheless, in the absence of egg-eating bias, α can be considered as the probability of Type I error, and β as the probability of Type II error. The analogy with statistics is exact: α is the probability of rejecting the null even though the null is correct, and β is the probability of accepting the null even though the null is incorrect. For this reason, following standard terminology in statistics, we refer to α as the statistical significance of the tenure test and, correspondingly, to $1 - \beta$ as the statistical power of the tenure test. Moreover, in statistics, the Youden index is the statistical power of the test minus its statistical significance.[8] Similarly, we denote as a Youden index the difference between power and significance of the tenure test.

volume . . . , I by no means expect to convince experienced naturalists whose minds are stocked with a multitude of facts all viewed, during a long course of years, from a point of view directly opposite to mine. A few naturalists, endowed with much flexibility of mind, and who have already begun to doubt the immutability of species, may be influenced by this volume; but I look with confidence to the future, to young and rising naturalists, who will be able to view both sides of the question with impartiality" (p. 151).

[8]The Youden index is used to evaluate binary classifiers (statistical methods to classify observations in two distinct categories, such as "healthy" or "sick"). It is popular in medicine (Armitage, Berry, and Matthews 2002, p. 693) and in machine learning (Murphy 2012, p. 183).

The Youden index measures the ability of tenure tests to distinguish between better and worse paradigms. It will play a critical role in the analysis. When no scientist believing in the Worse paradigm is given tenure, $\beta = 0$ and $J = 1 - \alpha$. At the other extreme, all scientists receive tenure with the same probability, so $\beta = 1 - \alpha$ and $J = 0$; then, the tenure evaluation does not distinguish between Better or Worse paradigms. Between these two extremes, a Better scientist coming up for tenure is more likely to get it than a Worse scientist, but Worse scientists have some chance of receiving tenure: $0 < \beta < 1 - \alpha$ and $0 < J < 1 - \alpha$.

When scientists are not biased, α and β are the probabilities of type I and type II errors. When scientists are biased, the tenure probability for a young scientist depends on the identity of the tenured scientist who conducts the tenure evaluation: so that the actual probabilities of type I and type II errors depend on σ, the fraction of tenured scientists who believe in the Better paradigm. With probability σ an untenured Better scientist is evaluated by a Better evaluator, who denies her tenure with probability $\alpha - \epsilon$; with probability $1 - \sigma$ she is evaluated by a Worse evaluator, who denies her tenure with probability $\alpha + \epsilon$. Hence, with egg-eating bias, the actual probability of type I error—the probability of denying tenure to Better scientists—is given by

$$\text{(1)} \qquad \alpha^{adj}(\sigma) = \sigma(\alpha - \epsilon) + (1 - \sigma)(\alpha + \epsilon) = \alpha + (1 - 2\sigma)\epsilon.$$

Similarly, with probability σ an untenured Worse scientist is evaluated by a Better evaluator, who grants her tenure with probability $\beta - \epsilon$; with probability $1 - \sigma$ she is evaluated by a Worse evaluator, who grants her tenure with probability $\beta + \epsilon$. Hence, with egg-eating bias, the actual probability of type II error—the probability of granting tenure to Worse scientists—is given by

$$\text{(2)} \qquad \beta^{adj}(\sigma) = \sigma(\beta - \epsilon) + (1 - \sigma)(\beta + \epsilon) = \beta + (1 - 2\sigma)\epsilon.$$

With egg-eating bias, the probabilities of type I and type II errors are adjusted to capture the social forces at play in the tenure process, and the Youden index must be adjusted accordingly:

DEFINITION 2. *The* adjusted Youden index *of the tenure test is*

$$\text{(3)} \qquad J^{adj}(\sigma) = 1 - \alpha^{adj}(\sigma) - \beta^{adj}(\sigma) = J + 2(2\sigma - 1)\epsilon.$$

Without bias, $\alpha^{adj}(\sigma) = \alpha$, $\beta^{adj}(\sigma) = \beta$, and $J^{adj}(\sigma) = J$. But with bias, $\alpha^{adj}(\sigma)$, $\beta^{adj}(\sigma)$, and $J^{adj}(\sigma)$ depend on the composition of the population of tenured scientists, measured by σ.

The adjusted Youden index is one minus the probability of type I error minus the probability of type II error. A high index means that there are few type I and type II errors: Better scientists get tenure with high probability; Worse scientists get it with low probability. Conversely, a low index means that there are many type I and type II errors.

The adjusted Youden index is linearly increasing in σ, from $J^{adj}(0) = J - 2\epsilon$ to $J^{adj}(1) = J + 2\epsilon$. The adjusted Youden index is minimized when $\sigma = 0$, because then all tenured scientists believe in the Worse paradigm, and these scientists are biased in favor of young scientists who also believe in it and are biased against those who believe in the Better paradigm. On the contrary, the adjusted Youden index is maximized when $\sigma = 1$, because then all tenured scientists believe in the Better paradigm, and these scientists are biased in favor of young scientists who also believe in it and are biased against those who believe in the Worse paradigm.

If $\epsilon \leq J/2$, the adjusted Youden index is positive for all $\sigma \in (0, 1)$. On the other hand, if $\epsilon > J/2$, the index is negative for $\sigma < \sigma^*$, zero at $\sigma = \sigma^*$, and positive for $\sigma > \sigma^*$, where the threshold σ^* is defined by

$$\sigma^* = \frac{1}{2}\left(1 - \frac{J}{2\epsilon}\right). \tag{4}$$

This threshold is a critical element in the analysis below.

5. Analysis of the Model of Science

We analyze our model of science to determine under which conditions the Better paradigm—which provides a better description of the world—eventually prevails, and conversely under which conditions the Worse paradigm prevails. The beliefs of the population of scientists depend critically on two parameters: the Youden index J and egg-eating bias ϵ.

5.1. Science Without Egg-Eating Bias

We begin by analyzing the model in the absence of egg-eating bias ($\epsilon = 0$). This analysis provides a useful point of reference. The evolution toward Better paradigm or Worse paradigm depends critically on the outcome of a basic horse race: regarding which species of scientists—Better or Worse—is increasing at the faster rate.

The role of this horse race is gleaned from the definition of σ as the fraction of Better scientists in the total population: $\sigma(t) = B(t)/[B(t) + W(t)]$. Differentiating this definition with respect to t, and a bit of algebraic juggling, yields

$$\dot{\sigma}(t) = \sigma(t)(1 - \sigma(t))\left(g^B - g^W\right), \tag{5}$$

where $g^B \equiv \dot{B}(t)/B(t)$ is the growth rate of tenured scientists with Better beliefs, and $g^W \equiv$

$\dot{W}(t)/W(t)$ is the growth rate of tenured scientists with Worse beliefs. The behavior of $\dot{\sigma}$, in turn, can be easily calculated, since g^B and g^W are easily inferred from the description of the model.

The growth of tenured Better scientists is the difference between the rate of advancement into tenure of Better scientists and their rate of death. The rate of advancement into tenure of Better scientists is the product of two terms: the first term being the fraction of Better advisees surviving into tenure, $(1-\alpha)$, where α is the statistical significance of the tenure test; the second term being the spawning rate of Better advisors, λ. The death rate of tenured Better scientists is δ. As a result, the growth rate of tenured Better scientists is

$$g^B = (1-\alpha)\lambda - \delta.$$

Similarly, the growth rate of tenured Worse scientists can be calculated as

$$g^W = \beta\lambda - \delta.$$

The first term is the product of the fraction of Worse advisees surviving into tenure, β, and the spawning rate of Worse advisors, λ. (Recall that $1-\beta$ is the statistical power of the tenure test, so that a fraction β of believers in the Worse paradigm will be wrongly granted tenure.) The second term is the death rate, δ, of Worse scientists with tenure.

Using (5) and the expressions for g^W and g^B yields

$$\dot{\sigma}(t) = \lambda\sigma(t)(1-\sigma(t))J. \tag{6}$$

The term $J = 1 - \alpha - \beta$ is the Youden index of the tenure test. The index determines the differential rate of increase of Better relative to Worse scientists in our model, because it governs the difference between the rate at which Better scientists are granted tenure $(1-\alpha)$ and the rate at which Worse scientists are granted tenure (β).

The differential equation (6) is a well-known logistic equation.[9] The equation immediately leads to the following proposition:

PROPOSITION 1. *Without egg-eating bias ($\epsilon = 0$) there are two possible regimes. When the Youden index is zero ($J = 0$), the composition of the population of scientists is constant over time ($\sigma(t) = \sigma(0)$ for all t). When the Youden index is positive ($J > 0$), the Better paradigm*

[9]The logistic differential equation is widely used. It was introduced by Verhulst (1845) in the 19th century to describe population growth. Verhulst's work was later popularized by Lotka (1925). The logistic differential equation was rediscovered twice in the 20th century: first by McKendrick and Pai (1912) in their study of the growth of micro-organisms, and then again by Pearl and Reed (1920) in their study of the US population growth.

eventually prevails ($\lim_{t\to\infty} \sigma(t) = 1$).

The formal proof is in Appendix A, but the intuition is simple: if $J = 0$, $\dot{\sigma} = 0$ so σ remains constant; if $J > 0$, on the other hand, $\dot{\sigma} > 0$ whenever $\sigma \in (0, 1)$, so σ grows until it reaches 1.

The proposition occasions three remarks. First, there can only be scientific progress—that is, convergence toward the Better paradigm—if the Youden index of the tenure test is positive. That is, convergence toward true belief depends not just upon the test's statistical power, which is $1-\beta$; it also depends on its statistical significance, α. Furthermore, the speed of convergence toward true belief is faster when the Youden index is larger.

Second, when the fraction of Better scientists converges to 1, several things can happen to the populations of Better and Worse scientists. If $(1-\alpha)\lambda > \delta > \beta\lambda$, the number of Worse scientists converges to 0, whereas the number of Better scientists converges to ∞. If $(1-\alpha)\lambda > \delta = \beta\lambda$, the number of Worse scientists is constant, whereas the number of Better scientists converges to ∞. If $(1-\alpha)\lambda = \delta > \beta\lambda$, the number of Better scientists is constant, whereas the number of Worse scientists converges to 0. If $(1-\alpha)\lambda > \beta\lambda > \delta$, the size of both groups goes to ∞, but the fraction of Better scientists in the population goes to 1. Finally, if $\delta > (1-\alpha)\lambda > \beta\lambda$, the size of both groups goes to 0, but the fraction of Better scientists in the population goes to 1.

Third, equation (6) points to another possibility than evolution to true belief, even with a presumably unbiased process of evaluation, such as that described by the grant of tenure. Such a process will lead to false belief if it uses a wrong-minded test: if the Youden index is negative, σ will not converge to 1; instead, it will converge to 0.

5.2. Science With Egg-Eating Bias

We now see how the properties of the model are altered by the presence of egg-eating bias. Now, because the Better and the Worse scientists with tenure are the evaluators; because they evaluate untenured scientists differently; and because the evaluators are chosen randomly from the population of tenured scientists, the growth rates of tenured Better scientists and of tenured Worse scientists depend upon the fraction σ of Better scientists in the population of tenured scientists. We therefore denote these growth rates as $g^B(\sigma)$ and $g^W(\sigma)$. Again (5) will hold so that

$$\dot{\sigma}(t) = \sigma(t)(1-\sigma(t))\left[g^B(\sigma(t)) - g^W(\sigma(t))\right]. \tag{7}$$

With egg-eating bias the growth rates g^B and g^W depend on σ, which was not the case without egg-eating bias. It remains to obtain formulas for $g^B(\sigma)$ and for $g^W(\sigma)$. We can easily calculate $g^B(\sigma)$ as follows:

$$g^B(\sigma) = \left(1 - \alpha^{adj}(\sigma)\right)\lambda - \delta.$$

The first term reflects that the Better scientists are training advisees at rate λ; and with probability $1 - \alpha^{adj}(\sigma)$, where $\alpha^{adj}(\sigma)$ is given by (1), such advisees are granted tenure. The second term reflects that tenured scientists are retiring at rate δ. Similarly we can calculate $g^W(\sigma)$:

$$g^W(\sigma) = \beta^{adj}(\sigma)\lambda - \delta.$$

The first term reflects that the Worse scientists are training advisees at rate λ; and with probability $\beta^{adj}(\sigma)$, where $\beta^{adj}(\sigma)$ is given by (2), such advisees are granted tenure. The second term reflects that tenured scientists are retiring at rate δ.

These equations give us

$$g^B(\sigma) - g^W(\sigma) = \lambda J^{adj}(\sigma),$$

where $J^{adj}(\sigma)$ is the adjusted Youden index, given by (3). Accordingly, (7) can be written as

$$\dot{\sigma}(t) = \lambda\sigma(t)(1 - \sigma(t))J^{adj}(\sigma(t)). \tag{8}$$

In the same way that the Youden index J determined the evolution of the share of Better tenured scientists when egg-eating bias is absent, the adjusted Youden index $J^{adj}(\sigma)$ determines the evolution of the share of Better tenured scientists when egg-eating bias is present. The adjusted Youden index plays this role because it governs the difference between the rates at which Better scientists and Worse scientists are granted tenure: $1 - \alpha^{adj}(\sigma)$ minus $\beta^{adj}(\sigma)$.

Unlike the unadjusted Youden index J, the adjusted Youden index $J^{adj}(\sigma)$ is not constant: it is linearly increasing in σ. Thus, egg-eating bias makes the dynamics of the model more complex: the differential equation governing the dynamics of the population of scientists (equation (8)) is not a regular logistic equation like equation (6); instead, it is a logistic equation with threshold. The threshold is the value σ^*, defined by (4), at which the adjusted Youden index is 0. When the threshold is between 0 and 1, the dynamics of the model are fundamentally altered:

PROPOSITION 2. *With egg-eating bias ($\epsilon > 0$), there are two possible regimes, depending on the amount of bias relative to the Youden index (J). When $\epsilon \leq J/2$, the Better paradigm eventually prevails ($\lim_{t\to\infty}\sigma(t) = 1$), irrespective of the initial fraction of Better tenured scientists ($\sigma(0)$). But when $\epsilon > J/2$, the eventual outcome is determined by the initial value of the fraction of Better tenured scientists relative to the threshold σ^* given by (4). If $\sigma(0) > \sigma^*$, the Better paradigm prevails ($\lim_{t\to\infty}\sigma(t) = 1$); but if $\sigma(0) < \sigma^*$, the Worse paradigm prevails ($\lim_{t\to\infty}\sigma(t) = 0$).*

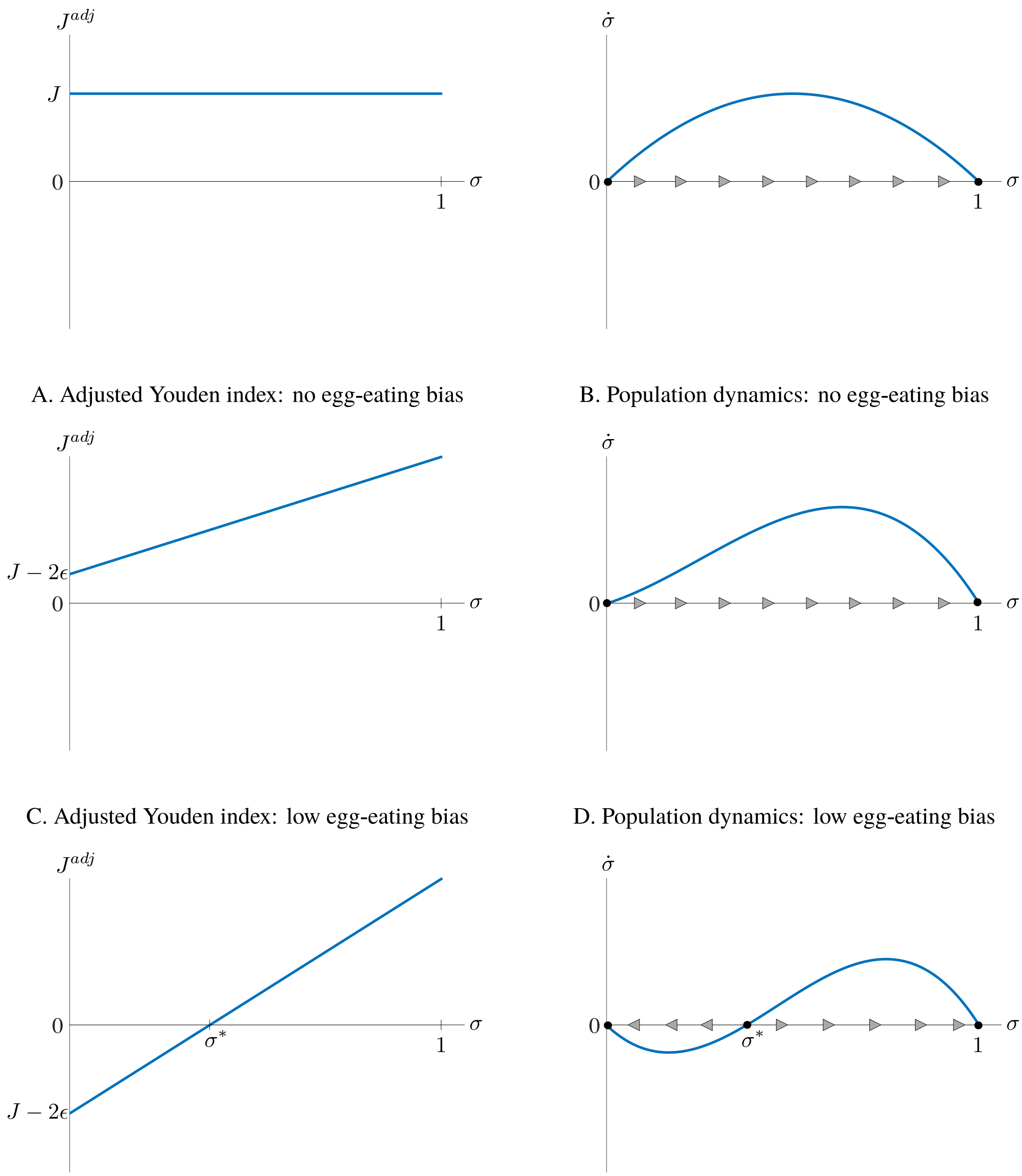
The formal proof is given in Appendix A, but the logic is simple. If there was no egg-eating bias (top row in Figure 2), the adjusted Youden index J^{adj} would always be positive and equal to the regular Youden index J. As a consequence, $\dot{\sigma}$ would be positive for all values of σ between 0 and 1, and σ would always grow until it reaches 1. This is the case covered by Proposition 1. In this proposition, we consider the case with egg-eating bias. When $0 < \epsilon \leq J/2$ (middle row in Figure 2), the adjusted Youden index is positive for all values of σ between 0 and 1, so $\dot{\sigma}$ is also positive for these values, and σ grows until it reaches 1. On the other hand, when $\epsilon > J/2$ (bottom row in Figure 2), the sign of the adjusted Youden index depends on whether σ is above or below σ^*. For all values of σ greater than σ^*, $\dot{\sigma}$ is positive, so σ converges to 1; for all values of σ less than σ^*, $\dot{\sigma}$ is negative, so σ converges to 0.

Proposition 2 occasions five remarks. First, even if the Youden index of the tenure test J is positive, if the bias of scientists is sufficiently large, scientific opinion may gravitate toward the Worse paradigm. This is true even though the Worse paradigm describes the world less correctly than the Better paradigm, and Worse scientists do not have a larger bias than Better scientists. This is because, once there are sufficiently many tenured Worse scientists, and given that they are sufficiently biased against Better scientists and in favor of Worse scientists, the tenure probability of Better scientists falls below that of Worse scientists (formally: the adjusted Youden index becomes negative). Thus the number of tenured Worse scientists grows faster than the number of tenured Better scientists. This shows one effect of biased eating of each other's eggs. It follows from (4) that an increase in egg-eating bias also raises the threshold σ^* below which science converges to falsehood.

Second, the proposition highlights the importance of the Youden index J—which is the statistical power minus statistical significance of the tenure test. Indeed, for a given egg-eating bias ϵ, branches of science with low statistical power and thus low Youden index are at greater risk of convergence to falsehood. One, they are more likely to be in a regime where convergence to falsehood is a possibility (because it is more likely that $J < 2\epsilon$ when J is lower). And two, in this regime, the population of scientists is more likely to be in the region where convergence to falsehood occurs (because the threshold σ^* is higher when J is lower).

Third, the model also points to an important determinant of scientific revolutions. We can compute the increase in Youden index required to start a scientific revolution. Assume that the share of Better scientists is σ and is converging to 0. To initiate a scientific revolution, we need the share of Better scientists to start converging to 1. This requires that the Youden index J increases sufficiently to be above the threshold

$$J^* = 2(1 - 2\sigma)\epsilon.$$



A. Adjusted Youden index: no egg-eating bias

B. Population dynamics: no egg-eating bias

C. Adjusted Youden index: low egg-eating bias

D. Population dynamics: low egg-eating bias

E. Adjusted Youden index: high egg-eating bias

F. Population dynamic: high egg-eating bias

Figure 2. Adjusted Youden Index and Scientific Progress

Notes: The panels on the left display the adjusted Youden index as a function of the fraction of Better tenured scientists: $J^{adj}(\sigma) = J + 2(2\sigma - 1)\epsilon$. The panels on the right display the phase lines for the differential equation governing the dynamics of the fraction of Better tenured scientists: $\dot{\sigma} = (1 - \sigma) \cdot \sigma \cdot J^{adj}(\sigma) \cdot \lambda$. Each row considers a different value of the egg-eating bias: $\epsilon = 0$, $\epsilon < J/2$, and $\epsilon > J/2$.

The threshold is computed such that the adjusted Youden index $J^{adj}(\sigma)$ is just 0 at σ. When the regular Youden index J is above the threshold J^*, the adjusted Youden index at σ is positive and science starts converging to the truth.

Fourth, convergence in science is often interpreted as a sign of progress. In the model, this is not necessarily the case: with egg-eating bias, there may be convergence toward Worse belief.

Fifth, convergence in the model may be surprising for another reason: because there may be no statistical or social reason to expect a single paradigm to prevail. Consider a situation in which the Youden index is zero ($J = 0$). Thus there is no statistical reason why either of the two paradigms should dominate. Further, since Better scientists have the same bias as Worse scientists, there is no social reason why either of the two paradigms should dominate. Nevertheless, egg-eating bias interacts with σ, thereby introducing instability into the system. Thus, if the initial fraction of Better scientists is below $\sigma^* = 1/2$, science converges to Worse beliefs; and if the initial fraction of Better scientists is above $\sigma^* = 1/2$, science converges to Better belief.

5.3. Extensions

Here we propose three extensions of the basic model of science and summarize their properties. These extensions enrich the model by providing additional mechanisms that can pull science toward the truth or toward falsehood. All the results are derived in Appendix B.

Heterogeneous Egg-Eating Bias. Given that Better and Worse scientists have different views of the world, it is natural to allow them to have different egg-eating biases. The results can easily be generalized if Better and Worse scientists have different biases.

We assume that a Better evaluator has an increased probability ϵ^B of giving tenure to an advisee who believes in the Better paradigm; and she has a decreased probability ϵ^B of giving tenure to an advisee who believes in the Worse paradigm. Symmetrically, a Worse evaluator has a decreased probability ϵ^W of giving tenure to an advisee who believes in the Better paradigm); and she has an increased probability ϵ^W of giving tenure to an advisee who believes in the Worse paradigm. With distinct egg-eating biases, the qualitative properties of the model remain the same, but a few quantitative properties need to be adjusted.

First of all, increased bias by the Worse scientists (ϵ^W) make it more likely to be in a regime in which convergence to the Worse paradigm is possible. Indeed, when $\epsilon^W \leq J/2$, the Better paradigm eventually prevails, irrespective of the initial fraction of Better tenured scientists, but when $\epsilon^W > J/2$, the Worse paradigm eventually prevails if the initial fraction of tenured Worse scientists is large enough. It is noteworthy that the bias of Better scientists (ϵ^B) has no effect on the possibility of convergence to Worse belief. This result occurs because for σ close to zero there are almost no Better scientists to make this bias effective.

Nevertheless, once science is in a regime in which convergence to Worse belief is possible, ϵ^B does affect gravitation toward Worse belief. With heterogeneous bias, the threshold σ^* that separates initial conditions leading to the Better or the Worse paradigm is given by $\sigma^* = (\epsilon^W - J/2)/(\epsilon^B + \epsilon^W)$. Hence, an increase in ϵ^B unambiguously decreases the threshold σ^*, and thus reduces the likelihood that science gravitates toward Worse belief. Correspondingly, in this regime, an increase in ϵ^W increases σ^* and thus raises the likelihood that science gravitates toward Worse belief.

Finally, we find that to start a scientific revolution, the Youden index J must increase sufficiently to be above the threshold $J^* = 2\left[\epsilon^W - \sigma\left(\epsilon^W + \epsilon^B\right)\right]$. Hence, an increase in the Youden index triggers a revolution for any $\sigma > 0$ only if the bias of Worse scientists is below $1/2$. If the bias of Worse scientists is above $1/2$, even a Youden index of 1 (the maximum value) is not sufficient to trigger a scientific revolution as σ gets close to 0. Thus, a large enough bias from Worse scientists can prevent knowledge from ever converging to the truth; that is, there exist initial conditions such that for any Youden index beliefs converge to the Worse paradigm.

Heterogeneous Productivity. The race towards having a larger number of students, papers, and grants suggests that productivity is an important aspect of knowledge creation. Maybe some paradigms lend themselves to be more productive and produce more students. For instance, maybe more complicated paradigms open the door to many extensions, which favor the productivity of scientists in that paradigm. In this section we extend the model by introducing different productivities across paradigms. We find that differential productivity affects the Youden index and therefore the growth of the scientific field. Indeed, productivity interacts with statistical power and bias in determining whether science converges to the truth.

Productivity determines the rate at which a tenured scientist trains advisees. We assume that tenured scientists believing in the Better paradigm train advisees at rate λ^B, and tenured scientists believing in the Worse paradigm train advisees at rate λ^W. With heterogeneous productivity, the Youden index needs to be redefined to combine three elements: statistical power, statistical significance, and productivity. The Youden index becomes

$$J = 1 - \alpha - \frac{\lambda^W}{\lambda^B}\beta.$$

Thus, an increase in the productivity of the Worse paradigm has the same effect as a reduction in the statistical power of the tenure test, and conversely, an increase in the productivity of the

Better paradigm has the same effect as an increase in statistical power.

Given the new expression for the Youden index, a critical implication of heterogeneous productivity is that even if the statistical power of the tenure test is larger than the statistical significance ($1 - \beta > \alpha$) and there is no egg-eating bias, the Worse paradigm may prevail. This happens when Worse scientists are sufficiently more productive that the Youden index J becomes negative. The formal condition for the Worse paradigm to prevail is $\lambda^W/\lambda^B > (1-\alpha)/\beta$. If Worse scientists are not productive enough, such that the Youden index remains positive, the Better paradigm eventually prevails.

When there is egg-eating bias, the population dynamics of science are also affected in important ways by heterogeneous productivity. First, consider a small egg-eating bias:

$$\epsilon \leq \frac{|J|}{1 + \lambda^W/\lambda^B}.$$

Then if the Youden index J is negative ($\lambda^W/\lambda^B > (1-\alpha)/\beta$), the Worse paradigm eventually prevails, irrespective of initial conditions. But if the Youden index J is positive ($\lambda^W/\lambda^B < (1-\alpha)/\beta$), the Better paradigm eventually prevails, irrespective of initial conditions.

Second, consider a large egg-eating bias:

$$\epsilon > \frac{|J|}{1 + \lambda^W/\lambda^B}.$$

Then initial conditions determine whether science converges to Better or Worse belief. The threshold σ^* that separates initial conditions leading to the Better or the Worse paradigm is given by

$$\sigma^* = \frac{1}{2}\left[1 - \frac{J}{(1 + \lambda^W/\lambda^B)\epsilon}\right].$$

Given that $J = 1 - \alpha - (\lambda^W/\lambda^B)\beta$, the threshold σ^* is increasing in λ^W/λ^B. Hence, an increase in λ^W/λ^B unambiguously increases the threshold σ^*, and thus raises the likelihood that science gravitates toward Worse belief. Correspondingly, a decrease in λ^W/λ^B lowers σ^* and thus reduces the likelihood that science gravitates toward Worse belief.

Overall, productivity has an important effect on the evolution of scientific belief. Beyond statistical power and bias, productivity provides an additional mechanism that can pull science toward the truth or toward falsehood. The paradigm with the highest productivity has a significant edge: in any circumstances, a paradigm whose productivity increases becomes more likely to capture the scientific field.

Defections. In the results presented above, when one paradigm prevails (Better or Worse), everybody believes in it: there are no scientists who believe in the other paradigm. In reality, however, there is always a small group of scientists who work on the alternative, unpopular paradigm. This group of scientists could be composed of scientists attached to the Worse paradigm after a revolution. Or it could be composed of scientists developing and improving the Better paradigm before a revolution. Then, a scientific revolution occurs when the statistical power of the tenure test increases, allowing this group of scientists to expand and eventually dominate the field, until the next revolution overturns them.

A simple way to capture this additional feature is to assume that among the scientists trained by an advisor, a small share defect to the opposite paradigm. When there are advisees who do not follow in the footsteps of their advisors, the results of the model are modified in a natural fashion to become less extreme. The main implication of this assumption is that the steady states in which all scientists believe in the same paradigm disappear. Instead, in any steady state, a positive share of scientists are in the minority and believe in the unpopular paradigm.

We assume that tenured scientists train two types of advisees, Better and Worse. They train advisees of their type at rate λ and advisees of the opposite type at rate μ. We assume that advisors train fewer defecting advisees than regular advisees: $\mu \leq \lambda$.

When the Youden index is zero, there is egg-eating bias, and there is no defection, we have seen what happens: science converges to the Better paradigm if initially Better scientists are more than half, and it converges to the Worse paradigm if initially Better scientists are less than half. With a little bit of defection ($\mu/\lambda \leq \epsilon/(2\beta - \epsilon)$), the results are similar but not as extreme. If initially Better scientists are more than half, science converges to a steady state where the vast majority, but not all, of the scientists believe in the Better paradigm. If initially Worse scientists are more than half, science converges to a steady state where the vast majority, but not all, of the scientists believe in the Worse paradigm. Once defection becomes frequent ($\mu/\lambda > \epsilon/(2\beta - \epsilon)$), the results change. Then, science converges to a state of disagreement in which scientists are equally split between Better and Worse paradigms, irrespective of initial conditions.

We also know what happens when the Youden index is positive but low, there is egg-eating bias, and there is no defection. In that case, science may converge to the Better paradigm or to the Worse paradigm depending on initial conditions. With a little bit of defection ($\mu/\lambda \leq \epsilon/(2\beta - \epsilon)$), the result remains the same except that the two possible steady states do not have only Better scientists or only Worse scientists but they have a majority of Better scientists or a majority of Worse scientists. When defections are frequent ($\mu/\lambda > \epsilon/(2\beta - \epsilon)$), the results change. Then, irrespective of initial conditions, science converges to a unique steady state in which Better scientists are in a majority.

Finally, we know what happens when the Youden index is large, there is egg-eating bias, and there is no defection. In that case, science converges to the Better paradigm irrespective of initial conditions. With defection, the results remain the same except that the steady state does not have only Better scientists, but a majority of them.

6. Statistical Power and Acceptance of Superior Paradigms

This section discusses the main implication of our model of science: the relationship between the statistical power of the tenure test and the adoption or nonadoption of superior paradigms. At the end of *The Structure of Scientific Revolutions*, Kuhn left unanswered a question he deemed important: why has science been continually so successful for a period of centuries? In our model, the share of Better scientists in the community of tenured scientists converges to one if the Youden index of the tenure test—its statistical power minus its statistical significance—outweighs the egg-eating bias.

There are two reasons why the Youden index could be low. First, there may be no high-power scientific tests that discriminate between better and worse paradigms. But, second, even if such tests do exist, such tests may be play little, or no role, in promotions into senior fellowship.

But, returning to Kuhn's question, two special features of science have played an important role in its rapid progress. Historically, the physical sciences have made remarkable discoveries of high-power tests capable of distinguishing between better and worse paradigms. Additionally, norms of science say that tenure and other advancements should be based on high-power tests, if they exist: because, as we will see, being a scientist entails adhering to these tests.

A series of examples will illustrate these points. Two examples of scientific revolutions will demonstrate the near-coincidence of discoveries of high-power tests and adoption of new, better paradigms. We will also see an example of demotion from the fellowship of scientists, which demonstrates the norm of adherence to evidence from high-power tests. But we shall also look at examples in which new, better paradigms have been adopted only after long delays. Three examples from medicine illustrate how failure to use random control trials resulted in continued use of harmful treatments. Another example, from macroeconomics, shows a new paradigm that languished for almost half a century, in the absence of any observations of sufficiently high power to distinguish between the new paradigm and the old.

6.1. High-Power Tests in Two Scientific Revolutions

The scientific revolutions emphasized by Kuhn—by Copernicus, Newton, Lavoisier, Darwin, and Einstein—were all supported, at least in due course, by high-power scientific tests (Kuhn

1996, pp. 148–153).

Consider, for example, adoption of the heliocentric theory of the solar system. The theory had languished for almost 2,000 years before Copernicus; and, even after the publication of *De Revolutionibus*, "Copernicanism made few converts for almost a century" (Kuhn 1996, p. 150). This lack of converts does not mean that Copernicus had no immediate influence on astronomers. Only eight years after its publication, Erasmus Reinhold based important new astronomical tables on the methods of *De Revolutionibus*. But neither Reinhold nor most other contemporary astronomers believed in the moving earth (Kuhn 1957, pp. 186–188). A first step in gaining acceptance of this central Copernican idea came some 50 years after his death, with observations of the elliptical orbits of the planets by Kepler, who Repcheck (2007, p. 188) called "the first true Copernican after [his devoted disciple] Rheticus."[10] But the real breakers-of-the-ice were observations from Galileo's new high-resolution telescopes, since the observed movements of the moons of Jupiter and Saturn were difficult to reconcile with a stationary earth.

The Lavoisier revolution in chemistry gives a second clear example of the role of the power and significance of tests for the science and the lag to better-paradigm adoption. To recall, Lavoisier discovered a superior paradigm about combustion: occurring when flammable materials combined with a component of the air, which he called "oxygène." In contrast, the old paradigm viewed combustion as occurring when flammable materials released their "phlogiston." The difference between the two could be tested through the use of vacuums and precise weights. The high power and significance of these tests resulted in only a short lag (of just a few years) for acceptance of Lavoisier's paradigm.[11]

6.2. Being a Scientist and the Norm of Abiding by High-Power Tests

The Lavoisier revolution in chemistry does not only illustrate the role of high power test in scientific revolution. Kuhn's reaction to Lavoisier's rival, Priestley, illustrates the existence of a scientific norm: that being a scientist entails acceptance of hypotheses that are confirmed by high-power tests. Kuhn calls out Lavoisier's rival, Priestley, for being "unreasonable" and "illogical": because despite the findings, backed by high-power test, Priestley resolutely continued his belief in the old, "phlogiston" paradigm. Kuhn says Priestley "ceased to be a scientist" (p. 159). In our model, the test for tenure regards promotion of the new trainees into the senior fellowship

[10]This view of Kepler as a Copernican is in agreement with Kuhn (1957), except that Kuhn would also include Kepler's teacher, Michael Maestlin. Maestlin, who did not teach heliocentrism to his students, is called a "stealth Copernican" by Repcheck (2007, p. 186).

[11]See Kuhn (1996, p. 147): "Though neither Priestley's nor Lavoisier's theory, for example, agreed precisely with existing observations, few contemporaries hesitated more than a decade in concluding that Lavoisier's theory provided the better fit of the two."

of scientists. Regarding Priestley, Kuhn, in his role as historian of science, has engaged in a rare demotion: because Priestley's judgment did not accord with the results from high-power scientific tests.

6.3. Eschewal of Statistics in the History of Medicine

The history of medicine gives us some examples in which high-power tests were available; but practicing physicians eschewed these tests, rather than embraced them; and they thus played no significant role in promotion to practicing physician. Physicians had a different norm regarding promotion from scientists. Rather than the norm that the candidates' contribution to science should be evaluated with an eye on the results of high-power tests for the science, physicians' criteria for promotion rested on candidates' ability to carry out existing medical practice. Our three examples are the persistence of bloodletting, of radical mastectomies, and of hormone replacement therapy: amid disavowal of statistical methods for determining their efficacy.

As early as the 1830s Pierre-Charles-Alexandre Louis, a practicing physician in Paris, took matched samples of pneumonia patients: one sample, with bloodletting in the first four days of the disease; the other sample, with bloodletting in days five to nine. Louis' results, at the very least, should have called for further testing, since he found a 76 percent higher fatality rate for those with early treatment (Rangachari 1997, p. 281). That difference was difficult to explain if bloodletting was as beneficial to pneumonia patients as it was cracked up to be: Why not, the earlier the better? Publication of Louis' results in English (Louis 1836) were hailed in the *Journal of the American Medical Society* as "one of the most important medical works of the present century," being "the first formal exposition of the results of the only true method of investigation in regard to the therapeutic value of remedial agents" (Bartlett 1836, p. 102). Yet Louis' use of statistical trials to determine effects of bloodletting did not catch on. Neither did the later, much more conclusive findings of John Hughes Bennett, have significant effect on practice.[12] Bennett found no deaths among 105 patients whom he had treated for pneumonia without bloodletting at the Edinburgh Royal Infirmary; in contrast, during a prior period, when bloodletting had been standard treatment, more than one third of the pneumonia patients had died (Thomas 2014, p. 74). Bloodletting did decline greatly over the course of the 19th century, not because of statistical testing, but instead because physicians were buying into the new theories of causation of disease. The 1909 edition of Sir William Osler's textbook on *The Principals and Practice of Medicine* said that "local bloodletting by cupping or leeches is certainly advantageous in robust subjects" (Osler 1909, p. 782); such statements remained in

[12] Bennett is most famous for his discovery of leukemia.

posthumous editions of this influential textbook, as late as its fourteenth edition in 1942 (Thomas 2014, p. 75).

Medical historian John Harley Warner has culled doctors' letters and reports to explain why physicians, especially in the United States, were so averse to statistical methods. They viewed themselves as professionals with the duty of actively treating individual patients; that treatment would depend upon physicians' ability at observation, which was learned through their experience in practice.[13] With this identity, it was considered denial of duty to base judgment in individual cases on statistical samples of unknown patients in different locales and in different circumstances: "[Doctors] were not prepared to accept even in principle the proposition that they should discard existing therapeutic beliefs and practices, validated by both tradition and their own experience on account of somebody else's numbers" (Warner 1986, p. 201).

Two more recent examples, one from the last half of the 20th century, and another from the early part of the 21st, illustrate the resistance of doctors to testing current procedures. According to a survey of physicians treating breast-cancer by surgery, in 1968, 86 percent of surgical treatments for breast cancer were by radical mastectomy (Lerner 2003, p. 132). This procedure, which had been introduced in the United States in 1892 by Johns Hopkins' William Haldane, was highly debilitating to its survivors. Yet matched statistics from the Cleveland Clinic published in 1961 by Crile (1961) had shown that radical mastectomy yielded no improvement in mortality relative to simple mastectomy or lumpectomy, which were much less invasive (Lerner 2003, p. 117). It took ten more years before a significant-size random control trial was begun, against fierce opposition from the cancer-surgeon establishment—opposition that continued even after the trial was in progress (Lerner 2003, p. 138). The breast-cancer surgeons, like Warner's 19th century physicians, based their resistance on their belief in the powers of clinical expertise (Lerner 2003, p. 115). In an extreme expression of that opposition, the editor of the journal of the American Cancer Association said that use of random control trial to decide on procedures for individual patients was playing "scientific Russian roulette" with their lives (Lerner 2003, p. 115). When its results were published in 1981, 20 years after Crile's article, the random control trial bore out his initial findings: no difference in mortality, but great difference in the condition of the survivors (Fisher et al. 1981).

As another example, completion of large-scale tests for hormone-replacement therapy (HRT) occurred yet later. It had been approved in the United States by the Food and Drug Administration

[13]Thus Warner (1986) writes: "Through the mid-nineteenth century professional identity was based on proper behavior and on a medical theory that stressed the principle of specificity, the notion that treatment had to be matched to the idiosyncratic characteristics of individual patients and their environments" (p. 1). And he adds: "Extensive knowledge about the basic sciences was desirable, but not essential to proper professional identity. What was essential was that the physician be able to act, and to do so in accordance with regular values. The common defining body of learning that all regular practitioners shared was knowledge about practice" (p. 13).

in 1942 (Shook 2011, p. 39). The results of the control trials in the United States and Great Britain were only published after 2000. By that time, 44 percent of US post-menopausal women were ever-users (National Health and Nutrition Examination Survey 2003). The trials revealed greatly increased incidence of breast cancer for those taking some forms of HRT.[14]

In contrast to the physical sciences, in medicine there appears to have been no norm for the use of high-power tests, making promotion of believers in the old, inferior paradigm much more likely. Indeed, in medicine, results from high-power tests played no role in promotion to the status of practicing physician. Instead, promotion depended—especially in surgery—on ability to execute current technique. Thus, for example, trainees in breast-cancer surgery were admitted as practicing surgeons themselves, based on their ability to carry out radical mastectomies; so that the test for promotion to elder of the profession had little or no regard for Crile's findings of little difference in mortality—but much difference in patient welfare—between radical and simple mastectomy.

6.4. The Challenges of a Low Youden Index

It may not be the eschewal of high-power tests that leads to convergence to inferior paradigms. In some fields of investigation, such tests may not exist. This is a major difference between the physical and the social sciences.[15]

A little-known example from the history of economics (intentionally chosen to be far from current debates) illustrates. In the 1880s Uriel Crocker, a prominent Boston lawyer, published an article in the *Quarterly Journal of Economics* regarding the possibility of "an excess of [productive capacity] . . . beyond the amount required to meet all demands that are backed by the ability and the willingness to pay for the things demanded" (Crocker 1887, p. 362). S.M. Macvane, a professor at Harvard, wrote a comment to the article and offered the following conclusion: "Demand for savings is the offer of labor for wages. In order that the supply of capital shall exceed the demand for it, there must be more capital offering for labor than the laborers are willing to receive! [exclamation in the original] The mere statement of the case is sufficient to show its absurdity" (p. 367).

[14]Among the findings of Great Britain's Million Women Health Study: "Use of HRT by women aged 50–64 years in the UK over the past decade has resulted in an estimated 20,000 extra breast cancers, 15,000 associated with oestrogen-progestagen; the extra deaths cannot yet be reliably estimated" (Million Women Study Collaborators 2003, p. 419). Extrapolation of this number to the United States on the basis of population would be conservative, since HRT was more common in the United States than in Britain.

[15]In a metastudy of papers published in the social and behavioral sciences between 1960 and 2011, Smaldino and McElreath (2016) estimate that for a statistical significance of $\alpha = 0.05$, the average statistical power is $1 - \beta = 0.24$ (Figure 1). This leads to a low Youden index of $J = 0.24 - 0.05 = 0.19$. Moreover, despite presumably better methods and richer data, statistical power has not increased over time.

Undaunted, Crocker wrote *The Cause of Hard Times*, published in 1895. But rather than becoming known as precursor to Keynes, Crocker's views led him (and those of the "underconsumptionists" who followed him) nowhere. His distress is expressed in the last chapter: "in closing, it may be well to say that no professional economist has ever publicly recognized the validity of the theories and arguments set forth in this book" (Crocker 1895, p. 103). Among the economists, who "have published attempted refutations" or who "have privately expressed to this author their complete dissent from his views" were luminaries of the profession, including J. Laurence Laughlin, Thorstein Veblen, William Graham Sumner, and Frank Taussig. The Great Depression generated a powerful test of the old paradigm that supply creates its own demand; and, after Keynes' *General Theory*, the economics profession no longer dismissed underconsumptionism as "absurd."

6.5. Relation to "The Methodology of Positive Economics"

Our model and its findings yield a perspective on the classic essay on economic methodology by Friedman (1953). Writing before the birth of experimental economics, Friedman saw the difference between social and physical sciences as one of degree: largely because social sciences lacked the experimental evidence typically available in the physical sciences. But he did not view this characteristic as preventing false hypotheses from being "weeded-out": only that the weeding would be slow (p. 11).

According to our model, when scientists are unbiased, social sciences will indeed converge to the truth, even if their Youden index is low. But there is also homophily, of which Friedman was aware. He cited, as example, differences between sociologists and economists in their interpretation of evidence regarding the hypothesis of self-serving profit maximization.[16] And following our model, the social sciences, with their low Youden index and with homophily, may never converge to the truth. Witness the schisms between economics and sociology (indirectly referenced in Friedman's example, and also suggested by Fourcade, Ollion, and Algan (2015)), and, prior to behavioral economics, between economics and psychology. These groups of academics, with their respective beliefs and paradigms, show the survival of different species. They have evolved differently: in separate academic flour jars.

Finally, we cannot necessarily infer the validity of a paradigm—as Friedman (1953) asserts—

[16]Friedman said that social scientists are more likely to accept hypotheses that are familiar to them: "The background of the scientists is not irrelevant to the judgments they reach. There is never certainty in science, and the weight of evidence for or against a hypothesis can never be assessed completely 'objectively.' The economist will be more tolerant than the sociologist in judging conformity of the implications of the hypothesis [of single-minded pursuit of self-interest by employers] with experience, and he will be persuaded to accept the hypothesis tentatively by fewer instances of 'conformity' " (p. 30).

because of “its continued use and acceptance” (p. 30). On the contrary, in the beetles model there may be convergence into a sink of continued, near universal, false belief.

6.6. Recent Evidence from Citation Dynamics

Biases in the system for acceptance and rejection of academic journal articles (described in Zuckerman and Merton (1971)) could also be represented by our promotion-chain model. Acceptance of papers makes their authors prime candidates to be referees (and also editors) for later submissions—just as grantees of tenure become the judges of later tenure candidates.

But, because of problems of determining the quality of articles, attributions of bias must be accompanied by cautionary tales. A creative paper by Wang, Veugelers, and Stephan (2016) illustrates. They index articles' novelty by the number of previously unrecorded combinations of cited journals in the references. They find that “highly novel” articles in the 2001 Web of Science were published in journals with 18 percent lower Impact Factor than articles deemed by the index to have “no novelty at all” (p. 6). But the authors' inference that this difference indicates bias may be overplayed. They do not have a measure of article quality: so that the 18 percent lower impact factor could well result, instead, from differences in the quality of novel and non-novel articles.

Similar caution should be applied to the findings of Siler, Lee, and Bero (2015). They track the fate of more than a year's submissions to three leading medical journals—*Annals of Internal Medicine*, the *British Medical Journal*, and the *Lancet*. They find that among the 1008 submissions in the sample, 808 were ultimately published (p. 360). But *all* of the 14 most-cited articles had been rejected from one of the three journals; 12 of them, desk-rejected (p. 362). However, lacking a measure of quality, this finding is only suggestive of bias. High citation counts are not necessarily indicative of paradigm shift, or even of “high quality.” Absent a list of the articles and independent evaluation, we cannot know that the rejections and the high citations were not the result of unfounded claims.

In this regard, we do have one such list of rejected articles, from Gans and Shepherd (1994). They asked “leading economists” about their articles that had been rejected for publication. At least one such article on the list, “Expectations and the Neutrality of Money” by Robert Lucas, is commonly seen as a paradigm-shifter.

A recent study, by Azoulay, Fons-Rosen, and Graff Zivin (2015), finesses the problem of article quality by use of a clever instrument. They examine the number of publications in narrowly-defined fields of the life sciences, after the early death of a “star.” Given the noisiness of the dying-star instrument, the results are surprisingly large. Publications of previous collaborators of the deceased star declined by 20.3 percent; in contrast non-collaborators' publications increased by 7.9 percent (p. 16, p. 37). Additionally—yet more indicative of the ingroup favoritism/outgroup bias central to our model—the 7.9 percent increase did not come mainly from previous non-collaborators in the field, but instead from new entrants, whose papers were also highly cited on average (p. 22). In further line with predictions of the egg-eating model, high "intellectual coherence" of a field (indicated by a high percentage of within-field citations) acted as a brake on the new entry, just as in our model, a low value of σ deters better-paradigm adherents. Their data also accords with a prediction of the extended beetles model: new entry was lower if the deceased star had produced many academic progeny.

Another caution regarding the use of citation studies is also relevant. In our model citations may be neither predictive of an article's quality or of its paradigm-shifting potential: since the model predicts that citations to new paradigm areas will be rare—or even non-existent—if the new paradigm is not adopted. Underconsumptionism again illustrates. Prior to Keynes' *General Theory*, Crocker's article in the *Quarterly Journal of Economics* and his book are not cited even once in the Social Science Citation index; to date, their total citation count is only seven.

7. Model of Hierarchical Organizations

This section demonstrates another possible application of egg-eating bias: it adapts the previous model of science to promotion systems in hierarchical organizations. Just as there were two types of beetles in the flour jars, we assume there are two types of employees, which we label "Better" and "Worse". We assume that the attributes of the Better make them as productive, or more productive, for the organization than the Worse. We determine conditions under which the promotion system works well, in the sense that the Better are increasingly prevalent at higher levels of the organization. We also determine conditions under which the promotion system fails.

7.1. Assumptions

We assume that the employees' productivity differs, according to whether they are Better or Worse. We assume that the Better are as "fit," or more "fit," than the Worse. Their level in the organization, or their mixture with others, does not affect their attributes or behavior.

While the productivity of an employee is partially determined by her attributes, the production process within the organization brings additional uncertainty to the productivity of an employee: tasks are difficult to execute and subject to imponderables; the attributes of a worker may not be perfectly adapted to all situations; and workers vary in ability, effort, and resourcefulness. Hence, productivity is noisy; and only partially determined by the employee's underlying attributes.

The hierarchical organization has n levels, indexed by $i = 1, 2, \ldots, n$. In level i, there are $B(i)$ Better employees and $W(i)$ Worse employees. The fraction of Better employees is

$$\sigma(i) = \frac{B(i)}{B(i) + W(i)}.$$

The average productivity at level i increases with $\sigma(i)$.

We define level 0 of the organization as the base population from which the organization recruits. That population has $B(0) > 0$ Better and $W(0) > 0$ Worse, with $B(0)$ and $W(0)$ large relative to the size of the organization.

At any time, new members enter the organization; some existing members are promoted; and others exit. The organization promotes members from level i to level $i + 1$; new recruitment is represented as promotion from level 0 to level 1. Members at level i are brought up for promotion at rate $\lambda(i)$.[17] The sequence $\{\lambda(i)\}_{i=0}^{n-1}$ is exogenous; it determines the relative size of the different levels of the organization. A member who is considered for promotion, but is denied, leaves the organization.[18] Other members also leave, for exogenous reasons, at rate δ.

To achieve its goals, the organization aims to promote its most productive members. A member at level i is evaluated by her supervisor, who is randomly chosen among the members at level $i + 1$. Therefore, the supervisor is a Better with probability $\sigma(i + 1)$ and a Worse with probability $1 - \sigma(i + 1)$.

We have seen that productivity is noisy: it is affected by chance and the competence of the worker. Hence, not all Better candidates are promoted, and not all Worse candidates are denied promotion. Nevertheless, since the Better attributes are more conducive to productivity, Better candidates tend to be more productive for the organization than Worse candidates. Accordingly, without egg-eating bias, supervisors would promote a Better with probability $1 - \alpha$ and a Worse with probability β, with $1 - \alpha \geq \beta$.

As in the model of science, we can interpret the promotion test as a statistical test that the attributes of the candidate are superior. Under this interpretation, α is the probability of Type I error and β is the probability of Type II error when supervisors are unbiased. We refer to α and $1 - \beta$ as the statistical significance and statistical power of the promotion test. Accordingly, the Youden index of the promotion test is $J = 1 - \alpha - \beta$.

With egg-eating bias in evaluating a promotion file, the probabilities of promoting Better and Worse candidates are different from $1 - \alpha$ and β. Supervisors are biased in favor of applicants of

[17] Note that $\lambda(n) = 0$ since nobody can be promoted above the highest level of the organization.

[18] We thus assume that the organization follows an up-or-out promotion system. This system is common in law firms, consulting firms, and academia. It is also practiced in the US military, and codified in the 1980 Defense Officer Personnel Management Act (Rostker et al. 1993).

the same type and against applicants of the opposite type. The bias of a supervisor is measured by the parameter $\epsilon > 0$. A Better supervisor denies promotion to a fellow Better with lowered probability $\alpha - \epsilon$; a Worse supervisor denies promotion to a Better candidate with increased probability $\alpha + \epsilon$; a Better supervisor grants promotion to a Worse candidate with lowered probability $\beta - \epsilon$; and a Worse supervisor grants promotion to a fellow Worse with increased probability $\beta + \epsilon$. Thus, promotion probabilities for employees at level i depend on the fraction of Better employees at level $i + 1$. We denote by $\alpha^{adj}(i)$ the probability that a level-i Better employee is dismissed and by $\beta^{adj}(i)$ the probability that a level-i Worse employee is promoted. Without bias, $\alpha^{adj}(i) = \alpha$ and $\beta^{adj}(i) = \beta$. With egg-eating bias, the probabilities are

$$\alpha^{adj}(i) = \sigma(i+1)(\alpha - \epsilon) + (1 - \sigma(i+1))(\alpha + \epsilon) = \alpha + (1 - 2\sigma(i+1))\epsilon \tag{9}$$

$$\beta^{adj}(i) = \sigma(i+1)(\beta - \epsilon) + (1 - \sigma(i+1))(\beta + \epsilon) = \beta + (1 - 2\sigma(i+1))\epsilon. \tag{10}$$

The probabilities $\alpha^{adj}(i)$ and $\beta^{adj}(i)$ are the actual probabilities of Type I and Type II errors between levels i and $i + 1$. These probabilities are not constant but depend on the composition of each level of the organization. The adjusted Youden index at level i is similar to the adjusted Youden index in our model of science:

$$J^{adj}(i) = 1 - \alpha^{adj}(i) - \beta^{adj}(i) = J + 2(2\sigma(i+1) - 1)\epsilon.$$

7.2. Analysis

We study the steady state of the hierarchical organization. For any level $i + 1 = 1, 2, \ldots, n$, inflows into level $i + 1$ equal outflows from level $i + 1$. Thus the numbers of Better and Worse employees in levels i and $i + 1$ are related by

$$(\delta + \lambda(i+1))\, B(i+1) = \left(1 - \alpha^{adj}(i)\right) \lambda(i) B(i)$$
$$(\delta + \lambda(i+1))\, W(i+1) = \beta^{adj}(i) \lambda(i) W(i).$$

Combining these equations and using $B(i)/W(i) = \sigma(i)/(1 - \sigma(i))$, we obtain

$$\frac{\sigma(i+1)}{1 - \sigma(i+1)} = \frac{1 - \alpha^{adj}(i)}{\beta^{adj}(i)} \cdot \frac{\sigma(i)}{1 - \sigma(i)}, \tag{11}$$

where $\alpha^{adj}(i)$ and $\beta^{adj}(i)$ are given by (9) and (10). This difference equation, together with the boundary condition $\sigma(0) = B(0)/(B(0) + W(0))$, determines the sequence $\{\sigma(i)\}_{i=0}^{n}$ representing the steady-state fraction of Better employees at each level of the organization.

Our models of science and of hierarchy differ in their formalism: with differential equations in the case of science; with difference equations in the case of hierarchy. But the analysis is similar in both cases, because in both models the dynamics of σ are driven by the sign of an adjusted Youden index. We saw it in the case of science, and equation (11) also shows it in the case of organizations. Indeed, whether σ is increasing or decreasing in i depends on whether $\left(1-\alpha^{adj}(i)\right)/\beta^{adj}(i)$ is greater or smaller than 1, which in turn depends on whether $1-\alpha^{adj}(i)$ is larger or smaller than $\beta^{adj}(i)$. Equivalently, the dynamics of σ depend on whether $1-\alpha^{adj}(i)-\beta^{adj}(i) = J^{adj}(i)$ is positive or negative. Thus, as with science, the sign of the adjusted Youden index determines whether or not promotion chains filter out false belief.

Thus the propositions regarding hierarchy are nearly identical to those regarding science:

PROPOSITION 3. *Without egg-eating bias ($\epsilon = 0$) there are two possible regimes. When the Youden index is zero ($J = 0$), each hierarchical level has the same composition, equal to the composition of the population ($\sigma(i) = \sigma(0)$ for all i). In particular, the composition of the top hierarchical level reflects the composition of the population. When the Youden index is positive ($J > 0$), the fraction of Better increases up the promotion ladder ($\{\sigma(i)\}_{i=0}^{n}$ is an increasing sequence), and the fraction of Better at the top hierarchical level converges to 1 as the number of levels becomes infinite ($\lim_{n\to\infty}\sigma(n) = 1$).*

PROPOSITION 4. *With egg-eating bias ($\epsilon > 0$), there are two possible regimes, depending on the amount of bias relative to the Youden index (J). When $\epsilon \leq J/2$, the fraction of Better increases up the promotion ladder ($\{\sigma(i)\}_{i=0}^{n}$ is an increasing sequence), and the fraction of Better at the top hierarchical level converges to 1 as the number of levels becomes infinite ($\lim_{n\to\infty}\sigma(n) = 1$), irrespective of the fraction of Better in the population ($\sigma(0)$). But when $\epsilon > J/2$, the dynamics are determined by the value of the fraction of Better in the population relative to the threshold σ^* given by (4). If $\sigma(0) > \sigma^*$, the fraction of Better increases up the promotion ladder ($\{\sigma(i)\}_{i=0}^{n}$ is an increasing sequence), and the fraction of Better at the top level converges to 1 as the number of levels becomes infinite ($\lim_{n\to\infty}\sigma(n) = 1$). If $\sigma(0) < \sigma^*$, the fraction of Worse increases up the promotion ladder ($\{\sigma(i)\}_{i=0}^{n}$ is a decreasing sequence), and the fraction of Worse at the top level converges to 1 as the number of levels becomes infinite ($\lim_{n\to\infty}\sigma(n) = 0$).*

7.3. Discussion

Because of their greater average productivity, in the absence of bias, Better employees are more likely to be promoted than Worse employees. However, if egg-eating bias is sufficiently strong, the organization can be captured by Worse employees: this means that there is increasing concentration of Worse employees at ever-higher levels of the promotion ladder. And, as the number of levels in the organization grows large, the fraction of Worse employees at the top levels approaches one.

Corporate Culture. Our model of promotion in organizations is related to a literature on corporate culture. This literature views corporate culture as the similar beliefs and values of the stakeholders—especially the employees—of a corporation.[19] In our model, promotions from one rung of an organization to the next results in increased specialization at each higher rung of the ladder. In this way, our model gives a description of a special mechanism for the evolution of similar beliefs and values in organizations—especially at higher rungs.

Also, if the promotions to the next higher rung are made only with considerations regarding the candidate's specific fitness for the job, there is an externality of some potential importance. The role of the promotees in future promotions is typically not factored into the decision regarding advancement from rung i to rung $i + 1$. In a promotion chain, when the promotees are biased, their advancement will affect the composition and quality of future promotions.

One additional effect does not enter our model explicitly but could be easily added. It could also matter a great deal. Promotions at each individual level usually fail to take into account choices between Better and Worse at all higher levels; thus they usually also fail to take into account the overall mix of Better and Worse in the organization as a whole. On the one hand, specialization either toward Better or toward Worse may have the positive externality of making cooperation easier. On the other hand, such specialization can have the negative externality of siloed thinking that is unreceptive to new ideas (Tett 2015). In the presence of this negative externality, the egg-eating bias is particularly problematic, since it leads to overspecialization.

Dysfunction at the Top of Organizations. A vast literature documents dysfunction at the top of organizations: mostly in biographies of individual leaders. Our model gives a view, complementary to this literature, regarding the rise of dysfunctional cream: it's the outcome of an equilibrium.

In a special important case of difference between Worse and Better, the Worse are those lacking normal moral scruples. With egg-eating bias, this personal attribute may play an increasing role in promotion as the unscrupulous rise to the top: if the judges for promotion have similar morals themselves, and have biases in favor of the like-minded. As one extreme application, consider the leaders of Communist states. The Leninist lack of respect for individual rights, and especially for truthfulness in accusations, advantaged the "Worst," who were willing

[19] See Hermalin (2012) for a review of the literature.

to use any method to get ahead (Akerlof and Snower 2016). The unusual cruelty and unusual skill at eating other's eggs of Stalin, Mao, Ceausescu, Hoxha, Honecker, Pol Pot, and many other Communist leaders cannot result from random draws from the populations of their respective countries.

The tug of war between self-promotion and merit that occurred in extreme form under Communism also occurs— much, much damped—in other forms of organization. Farrell (2010) offers one illustration in some detail. He attributes the fall of Merrill Lynch in the 2008 crash to the promotion of Stanley O'Neal as CEO. Farrell says that O'Neal "had not been content to let merit alone determine his success," being helped by a "cabal" of supporters who "advanc[ed] his candidacy [to CEO] in any way possible" (p. 90).[20] O'Neal's ineptitude in finance led him, against contrary advice, to appoint Osman Semerci, a charismatic former rug-dealer from Istanbul, as head of the Fixed Income Department. That was perhaps a yet worse choice of personal salesmanship over competence than the earlier pick of O'Neal by his Board of Directors; in short order, Semerci exposed Merrill to $31 billions of collateralized debt obligations, just before the market went sour in 2007–2008 (p. 18). Merrill Lynch missed bankruptcy only through a last-minute sellout to Bank of America. Our model shows that such ineptitude at the top is an equilibrium occurrence in systems in which eating of each other's eggs is allowed to privilege those with special skills at achieving their own promotion.

Gender Inequality. As another application, the model also adds to other explanations why organizations are so predominantly male and white at the top. There is much evidence of homophilic bias regarding gender (for example, Ibarra 1992; Reskin and McBrier 2000) and ethnicity (for example, Bertrand and Mullainathan 2004; Oreopoulos 2011). Following our model, even with a diverse labor force and even drawing from labor pools of equal competence, organizations may not be diverse at the top. This accords with findings by Kanter (1993): gender polarization is prevalent in high management positions (p. 16). As predicted by our model, there are decreasing fractions of women up the promotion ladder—from first-level management to middle management to top management (p. 17). Homogeneity by managerial ethnicity is also common, and also explained by our model. To give just one example, the executives in the company studied by Kanter were largely Scotch-Irish (p. 54).

[20]In that position, O'Neal won 18th place in a CNBC list of the 20 worst American CEOs of all time. See http://www.cnbc.com/2009/04/30/Portfolios-Worst-American-CEOs-of-All-Time.html.

8. Conclusion

This paper gives a new model regarding the emergence (or non-emergence) of scientific truth. Contrary to the belief that truth will always emerge, the model shows specific conditions in which, on the contrary, inferior paradigms will prevail even when they are in contest with better ones. In particular, if there are powerful scientific tests for inferior versus superior paradigms, and the outcome of those tests play a major role in the determination of admittance into the next generation of scientists, then a superior paradigm will prevail against an inferior paradigm. But, if the scientific tests either lack power, or are little used in determining admittance into the next generation of scientists, then egg-eating bias (in favor of those who think like ourselves) increases the chances of getting stuck in an inferior paradigm. Such bias may not just slow scientific progress; it may bring it to a halt.

Our model also suggests why science has made significant progress: not only because of new scientific tests of high power, but also because of commitment of established scientists to admit into their ranks those whose work respects the findings of such tests.

Furthermore, the principles we find regarding science also apply to promotions in hierarchical institutions with advancements up a promotion ladder. As with science, promotions biased to resemble those in higher rungs, can filter employees with lower productivity up the promotion ladder—with the highest concentration of such workers at the very top.

References

Akerlof, George A., and Dennis J. Snower. 2016. "Bread and Bullets." *Journal of Economic Behavior & Organization* 126: 58–71.

Armitage, Peter, Geoffrey Berry, and J. N. S. Matthews. 2002. *Statistical Methods in Medical Research*. 4th ed. Oxford: Blackwell.

Azoulay, Pierre, Christian Fons-Rosen, and Joshua S. Graff Zivin. 2015. "Does Science Advance one Funeral at a Time?" NBER Working Paper 21788.

Banerjee, Abhijit V. 1992. "A Simple Model of Herd Behavior." *Quarterly Journal of Economics* 107 (3): 797–817.

Bartlett, Elisha. 1836. "Review Article: Researches on the Effects of Blood-Letting in Some Inflammatory Diseases, and on the Influence of Tartarized Antimony and Vesication in Pneumonitis. By P. C.A. Louis. Translated by C.G. Putnam, M.D. With Preface and Appendix, by James Jackson, M.D., Physician of the Massachusetts General Hospital." *American Journal of the Medical Sciences* 18: 102–111.

Bénabou, Roland, Davide Ticchi, and Andrea Vindigni. 2015. "Forbidden Fruits: The Political Economy of Science, Religion, and Growth." NBER Working Paper 21105.

Bernhardt, Dan, Eric Hughson, and Edward Kutsoati. 2006. "The Evolution of Managerial Expertise: How Corporate Culture Can Run Amok." *American Economic Review* 96 (1): 195–221.

Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review* 94 (4): 991–1013.

Besley, Timothy, and Maitreesh Ghatak. 2005. "Competition and Incentives with Motivated Agents." *American Economic Review* 95 (3): 616–636.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch. 1992. "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades." *Journal of Political Economy* 100 (5): 992–1026.

Bloch, Francis, Gabrielle Demange, and Rachel Kranton. 2016. "Rumors and Social Networks." https://sites.duke.edu/rachelkranton/files/2016/12/rumors28-11-2016-1.pdf.

Bramoullé, Yann, and Gilles Saint-Paul. 2010. "Research Cycles." *Journal of Economic Theory* 145 (5): 1890–1920.

Brock, William A., and Steven N. Durlauf. 1999. "A Formal Model of Theory Choice in Science." *Economic Theory* 14 (1): 113–130.

Burks, Stephen V., Bo Cowgill, Mitchell Hoffman, and Michael Housman. 2015. "The Value of Hiring through Employee Referrals." *Quarterly Journal of Economics* 130 (2): 805–839.

Costantino, Robert F., and Robert A. Desharnais. 1991. *Populatio Dynamics and the Tribolium Model: Genetics and Demography*. New York: Springer-Verlag.

Crile, Jr., George. 1961. "Simplified Treatment of Cancer of the Breast: Early Results of a Clinical Study." *Annals of Surgery* 153 (5): 745–758.

Crocker, Uriel H. 1887. "General Overproduction." *Quarterly Journal of Economics* 1 (3): 362–366.

Crocker, Uriel H. 1895. *The Cause of Hard Times*. Boston: Little, Brown.

Dovidio, John F., and Samuel L. Gaertner. 2010. "Intergroup Bias." In *Handbook of Social Psychology*, edited by Susan T. Fiske, Daniel T. Gilbert, and Gardner Lindzey, 1084–1121. New York: Wiley.

Eyster, Erik, and Matthew Rabin. 2010. "Naïve Herding in Rich-Information Settings." *American Economic Journal: Microeconomics* 2 (4): 221–243.

Farrell, Greg. 2010. *Crash of the Titans: Greed, Hubris, the Fall of Merrill Lynch, and the Near-Collapse of Bank of America*. New York: Crown Business.

Fisher, Bernard, Norman Wolmark, Carol Redmond, Melvin Deutsch, Edwin R. Fisher, and Participating NSABP Investigators. 1981. "Findings from NSABP Protocol No. B-04: Comparison of Radical Mastectomy with Alternative Treatments. II. The Clinical and Biologic Significance of Medial-Central Breast Cancers." *Cancer* 48 (8): 1863–1872.

Fourcade, Marion, Etienne Ollion, and Yann Algan. 2015. "The Superiority of Economists." *Journal of Economic Perspectives* 29 (1): 89–114.

Friedman, Milton. 1953. "The Methodology of Positive Economics." In *Essays in Positive Economics*, 3–43. Chicago: University of Chicago Press.

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

Haslam, S. Alexander. 2004. *Psychology in Organizations: The Social Identity Approach*. 2nd ed. London: SAGE.

Hermalin, Benjamin E. 2012. "Leadership and Corporate Culture." In *Handbook of Organizational*

Economics, edited by Robert Gibbons and John Roberts, 432–478. Princeton, NJ: Princeton University Press.

Ibarra, Herminia. 1992. "Homophily and Differential Returns: Sex Differences in Network Structure and Access in an Advertising Firm." *Administrative Science Quarterly* 37 (3): 422–447.

Kanter, Rosabeth M. 1993. *Men and Women of the Corporation*. 2nd ed. New York: Basic Books.

Kuhn, Thomas S. 1957. *The Copernican Revolution*. Cambridge, MA: Harvard University Press.

Kuhn, Thomas S. 1996. *The Structure of Scientific Revolutions*. 3rd ed. Chicago: University of Chicago Press.

Lamont, Michele. 2009. *How Professor Think: Inside the Curious World of Academic Judgment*. Cambridge, MA: Harvard University Press.

Lerner, Barron H. 2003. *The Breast Cancer Wars: Hope, Fear, and the Pursuit of a Cure in Twentieth-Century America*. Oxford: Oxford University Press.

Lotka, Alfred J. 1925. *Elements of Physical Biology*. Baltimore: Williams and Wilkins.

Louis, Pierre-Charles-Alexandre. 1836. *Researches on the Effects of Bloodletting in Some Inflammatory Diseases: and on the Influence of Tartarized Antimony and Vesication in Pneumonitis*. Boston: Hilliard, Gray.

Mahoney, Michael J. 1977. "Publication Prejudices: An Experimental Study of Confirmatory Bias in the Peer Review System." *Cognitive Therapy and Research* 1 (2): 161–175.

McKendrick, A. G., and M. Kesava Pai. 1912. "The Rate of Multiplication of Micro-organisms: A Mathematical Study.." *Proceedings of the Royal Society of Edinburgh* 31: 649–653.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27 (1): 415–444.

Million Women Study Collaborators. 2003. "Breast Cancer and Hormone-Replacement Therapy in the Million Women Study." *The Lancet* 362 (9382): 419–427.

Montgomery, James D. 1991. "Social Networks and Labor-Market Outcomes: Toward an Economic Analysis." *American Economic Review* 81 (5): 1408–1418.

Murphy, Kevin P. 2012. *Machine Learning: A Probabilistic Perspective*. Cambridge, MA: MIT Press.

National Health and Nutrition Examination Survey. 2003. "Use of Hormone Replacement Therapy Among Postmenopausal Women in the United States, 1988–1994." https://www.cdc.gov/nchs/data/nhanes/databriefs/hrtinwomen.pdf.

Neyman, J., and E. S. Pearson. 1933. "On the Problem of the Most Efficient Tests of Statistical Hypotheses." *Philosophical Transactions of the Royal Society of London* 231 (694-706): 289–337.

Neyman, Jerzy, Thomas Park, and Elizabeth L. Scott. 1956. "Struggle for Existence. The Tribolium Model: Biological and Statistical Aspects." In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, vol. 4, 41–79. Berkeley and Los Angeles: University of California Press.

Oreopoulos, Philip. 2011. "Why Do Skilled Immigrants Struggle in the Labor Market? A Field Experiment with Thirteen Thousand Resumes." *American Economic Journal: Economic Policy* 3 (4): 148–71.

Osler, William. 1909. *The Principles and Practice of Medicine: Designed for the Use of Practitioners and Students of Medicine*. 7th ed. New York: Appleton.

Park, Thomas. 1954. "Experimental Studies of Interspecies Competition II. Temperature, Humidity, and

Competition in Two Species of Tribolium." *Physiological Zoology* 27 (3): 177–238.
Park, Thomas, David B. Mertz, Wladyslaw Grodzinski, and Tadeusz Prus. 1965. "Cannibalistic Predation in Populations of Flour Beetles." *Physiological Zoology* 38 (3): 289–321.
Park, Thomas, David B. Mertz, and Michael Nathanson. 1968. "The Cannibalism of Pupae by Adult Flour Beetles." *Physiological Zoology* 41 (2): 228–253.
Pearl, Raymond, and Lowell J. Reed. 1920. "On the Rate of Growth of the Population of the United States Since 1790 and its Mathematical Representation." *Proceedings of the National Academy of Sciences* 6 (6): 275–288.
Rangachari, Patangi K. 1997. "Evidence-Based Medicine: Old French Wine with a New Canadian Label?" *Journal of the Royal Society of Medicine* 90 (5): 280–284.
Repcheck, Jack. 2007. *Copernicus' Secret: How the Scientific Revolution Began*. New York: Simon & Schuster.
Reskin, Barbara F., and Debra Branch McBrier. 2000. "Why Not Ascription? Organizations' Employment of Male and Female Managers." *American Sociological Review* 65 (2): 210–233.
Rivera, Lauren A. 2012. "Hiring as Cultural Matching The Case of Elite Professional Service Firms." *American Sociological Review* 77 (6): 999–1022.
Rostker, Bernard D., Harry J. Thie, James L. Lacy, Jennifer H. Kawata, and Susanna W. Purnell. 1993. *The Defense Officer Personnel Management Act of 1980: A Retrospective Assessment*. Santa Monica, CA: RAND Corporation.
Sherif, Muzafer, O. J. Harvey, B. Jack White, William R. Hood, and Carolyn W. Sherif. 1961. *Intergroup Conflict and Cooperation: The Robbers Cave Experiment*. Norman, OK: University of Oklahoma Book Exchange.
Shook, Lydia L. 2011. "An Update on Hormone Replacement Therapy: Health and Medicine for Women: a Multidisciplinary, Evidence-Based Review of Mid-Life Health Concerns." *Yale Journal of Biology and Medicine* 84 (1): 39–42.
Siler, Kyle, Kirby Lee, and Lisa Bero. 2015. "Measuring the Effectiveness of Scientific Gatekeeping." *Proceedings of the National Academy of Sciences* 112 (2): 360–365.
Smaldino, Paul E., and Richard McElreath. 2016. "The Natural Selection of Bad Science." *Royal Society Open Science* 3 (9): 160384.
Tajfel, Henri, and John C. Turner. 1979. "An Integrative Theory of Intergroup Conflict." In *Psychology of Intergroup Relations*, edited by William G. Austin and Stephen Worchel, 33–47. Monterey, CA: Brooks-Cole.
Tajfel, Henri, and John C. Turner. 1986. "The Social Identity Theory of Intergroup Behaviour." In *Psychology of Intergroup Relations*, edited by Stephen Worchel and William G. Austin, 7–24. Chicago: Nelson-Hall.
Tett, Gillian. 2015. *The Silo Effect: The Peril of Expertise and the Promise of Breaking Down Barriers*. New York: Simon & Schuster.
Thomas, D. P. 2014. "The Demise of Bloodletting." *Journal of the Royal College of Physicians of Edinburgh* 44: 72–77.
Travis, G. D. L., and H. M. Collins. 1991. "New Light on Old Boys: Cognitive and Institutional Particularism in the Peer Review System." *Science, Technology & Human Values* 16 (3): 322–341.
Van den Steen, Eric. 2010. "On the Origin of Shared Beliefs (and Corporate Culture)." *RAND Journal of*

Economics 41 (4): 617–648.

Verhulst, Pierre-Francois. 1845. "Recherches Mathématiques sur la Loi d'Accroissement de la Population." *Nouveaux Mémoires de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles* 18: 1–41.

Wang, Jian, Reinhilde Veugelers, and Paula Stephan. 2016. "Bias against Novelty in Science: A Cautionary Tale for Users of Bibliometric Indicators." NBER Working Paper 22180.

Warner, John H. 1986. *The Therapeutic Perspective: Medical Practice, Knowledge, and Identity in America, 1820–1885*. Cambridge, MA: Harvard University Press.

Youden, W. J. 1950. "Index for Rating Diagnostic Tests." *Cancer* 3 (1): 32–35.

Zuckerman, Harriet, and Robert K. Merton. 1971. "Patterns of Evaluation in Science: Institutionalisation, Structure and Functions of the Referee System." *Minerva* 9 (1): 66–100.

Appendix A. Proofs

Proof of Proposition 1

When $J = 0$, equation (6) implies $\dot{\sigma}(t) = 0$ so $\sigma(t)$ is constant over time. When $J > 0$, equation (6) can be written $\dot{\sigma}(t) = P(\sigma(t))$ where $P(\sigma) = \sigma(1-\sigma)J\lambda$. The polynomial P has two roots, 0 and 1, and $P(\sigma) > 0$ for all $\sigma \in (0,1)$. Hence the differential equation (6) has two critical points: $\sigma = 0$, which is a source, and $\sigma = 1$, which is a sink. For any initial condition $\sigma(0) \in (0,1)$, $\sigma(t)$ therefore converges toward 1.

Proof of Proposition 2

We define the polynomial P by

$$P(\sigma) = 2\sigma(1-\sigma)\left[2\sigma\epsilon + \frac{j}{2} - \epsilon\right]\lambda.$$

Equation (7) can be written as $\dot{\sigma}(t) = P(\sigma(t))$. The dynamic behavior of $\sigma(t)$ is determined by the properties of P. The properties of the polynomial P depend on J and ϵ. If $\epsilon > 0$, the polynomial P has three roots: 0, 1, and the σ^* given by (4).

The root σ^* has the following properties:

- $\sigma^* < 1$
- $\sigma^* > 0$ if $\epsilon > J/2$
- $\sigma^* = 0$ if $\epsilon = J/2$
- $\sigma^* < 0$ if $\epsilon < J/2$

Since 0, 1, and σ^* are roots of the polynomial P, they are critical points of the differential equation (7). In addition, for any $\sigma \in (0,1)$, if $\sigma > \sigma^*$, then $P(\sigma) > 0$; if $\sigma < \sigma^*$, then $P(\sigma) < 0$. Thus there are three possibilities:

- When $\sigma^* < 0$, 0 is a source while 1 is a sink so $\sigma(t)$ converges to 1 from any $\sigma(0) \in (0,1)$.
- When $\sigma^* \in (0,1)$, 0 is a sink, σ^* is a source, and 1 is a sink. Thus $\sigma(t)$ converges to 1 from any $\sigma(0) \in (\sigma^*, 1)$ and converges to 0 from any $\sigma(0) \in (0, \sigma^*)$.
- When $\sigma^* = 0$, 1 is a sink and 0 is a node so $\sigma(t)$ converges to 1 from any $\sigma(0) \in (0,1)$.

Proposition 2 directly follows from these properties.

Proof of Propositions 3 and 4

We implicitly define a mapping $P : [0,1] \to [0,1]$ by writing equation (11) as $\sigma(i+1) = P(\sigma(i))$.

The mapping $T : x \mapsto x/(1-x)$ is a strictly increasing one-to-one mapping from $(0,1)$ to $(0,+\infty)$. It is invertible and its inverse T^{-1} is a strictly increasing one-to-one mapping from $(0,+\infty)$ to $(0,1)$. If $\epsilon > 0$, the mapping

$$Z : x \mapsto \frac{1-\alpha+(2x-1)\epsilon}{\beta-(2x-1)\epsilon}.$$

is a strictly increasing one-to-one mapping from $(0,1)$ to $((1-\alpha-\epsilon)/(\beta+\epsilon), (1-\alpha+\epsilon)/(\beta-\epsilon))$. We have assumed that $\epsilon < \beta$ to ensure that $\beta-\epsilon > 0$; hence the denominator of $Z(x)$ is strictly positive on $[0,1]$. The mapping P can be expressed as $P = T^{-1} \circ (Z \times T)$. As a consequence, P is a strictly increasing one-to-one mapping from $(0,1)$ to $(0,1)$.

The mapping P has the following properties:

- If $\epsilon = 0$ and $J = 0$, $P(\sigma) = \sigma$ for all $\sigma \in [0,1]$.

- If $\epsilon = 0$ and $J > 0$, the mapping P has two fixed points: 0 and 1. Furthermore, for all $\sigma \in (0,1)$, $P(\sigma) > \sigma$.

- If $\epsilon > 0$, the mapping P has three fixed points: 0, 1, and the σ^* defined by (4). (We have described the properties of σ^* in the proof of Proposition 2.) For any $\sigma \in (0,1)$, $P(\sigma) < \sigma$ if $\sigma < \sigma^*$ and $P(\sigma) > \sigma$ if $\sigma > \sigma^*$.

Propositions 3 and 4 directly follow from these properties.

Appendix B. Extensions of the Model of Science

We formally analyze the three extensions of the model of science discussed in Section 5.3.

Heterogeneous Egg-Eating Biases

With heterogeneous egg-eating bias, a Better evaluator denies tenure to a Better scientist with lowered probability $\alpha - \epsilon^B$; a Worse evaluator denies tenure to a Better scientist with increased probability $\alpha + \epsilon^W$; a Better evaluator grants tenure to a Worse scientist with lowered probability $\beta - \epsilon^B$; and a Worse evaluator grants tenure to Worse scientist with increased probability $\beta + \epsilon^W$. The biases ϵ^B and ϵ^W are bounded such that these four probabilities remain in $(0,1)$. Hence, the actual probability of denying tenure to Better scientists is

$$\alpha^{adj}(\sigma) = \sigma(\alpha - \epsilon^B) + (1 - \sigma)(\alpha + \epsilon^W) = \alpha - \sigma\epsilon^B - (\sigma - 1)\epsilon^W.$$

Similarly, the actual probability of granting tenure to Worse scientists is

$$\beta^{adj}(\sigma) = \sigma(\beta - \epsilon^B) + (1 - \sigma)(\beta + \epsilon^W) = \beta - \sigma\epsilon^B - (\sigma - 1)\epsilon^W.$$

Given these tenure probabilities, the adjusted Youden index is

$$J^{adj}(\sigma) = 1 - \alpha^{adj}(\sigma) - \beta^{adj}(\sigma) = J + 2\sigma\epsilon^B + 2(\sigma - 1)\epsilon^0.$$

The adjusted Youden index is linearly increasing in σ, takes its lowest value of $J - 2\epsilon^0$ at $\sigma = 0$, and takes its highest value of $J + 2\epsilon^B$ at $\sigma = 1$.

If $\epsilon^W \leq J/2$, the adjusted Youden index is positive for all $\sigma \in (0, 1)$. But if $\epsilon^W > J/2$, the index changes sign on $(0, 1)$. We define the threshold

$$\sigma^* = \frac{\epsilon^W - J/2}{\epsilon^B + \epsilon^W}. \tag{A1}$$

If $\epsilon^W > J/2$, then $\sigma^* \in (0, 1)$. Furthermore, the adjusted Youden index is negative for $\sigma < \sigma^*$, zero at $\sigma = \sigma^*$, and positive for $\sigma > \sigma^*$.

As when the biases ϵ^B and ϵ^W are the same, the evolution of $\sigma(t)$ is given the differential equation (8). Since the evolution of $\sigma(t)$ is given by the same differential equation as when the egg-eating biases are the same, and since the adjusted Youden index has the same properties, we immediately obtain a proposition similar to Proposition 2:

PROPOSITION A1. *With egg-eating bias ($\epsilon^B > 0$ and $\epsilon^W > 0$), there are two possible regimes, depending on the amount of bias of Worse scientists (ϵ^W) relative to the Youden index (J). When $\epsilon^W \leq J/2$, the Better paradigm eventually prevails ($\lim_{t\to\infty} \sigma(t) = 1$), irrespective of the initial fraction of Better tenured scientists ($\sigma(0)$). But when $\epsilon^W > J/2$, the eventual outcome is determined by the initial value of the fraction of Better tenured scientists relative to the threshold σ^* given by* (A1). *If $\sigma(0) > \sigma^*$, the Better paradigm prevails ($\lim_{t\to\infty} \sigma(t) = 1$); but if $\sigma(0) < \sigma^*$, the Worse paradigm prevails ($\lim_{t\to\infty} \sigma(t) = 0$).*

Last, we can recompute the increase in Youden index required to start a scientific revolution. Assume that the share of Better scientists is σ and is converging to 0. To initiate a scientific revolution, we need the share of Better scientists to start converging to 1. This requires that the

Youden index J increases sufficiently to be above the threshold

$$J^* = 2\left[\epsilon^W - \sigma\left(\epsilon^W + \epsilon^B\right)\right].$$

When the regular Youden index J is above the threshold J^*, the adjusted Youden index at σ is positive and science starts converging to the truth.

Heterogeneous Productivity

With heterogeneous productivity the growth rates g^B and g^W of tenured Better scientists and tenured Worse scientists are given by

$$g^B(\sigma) = \left(1 - \alpha^{adj}(\sigma)\right)\lambda^B - \delta$$
$$g^W(\sigma) = \beta^{adj}(\sigma)\lambda^W - \delta.$$

These equations give us

$$g^B(\sigma) - g^W(\sigma) = \lambda^B J^{adj}(\sigma),$$

where $J^{adj}(\sigma)$ is a new adjusted Youden index, tailored to the situation with heterogeneous productivity:

$$J^{adj}(\sigma) \equiv 1 - \alpha^{adj}(\sigma) - \frac{\lambda^W}{\lambda^B}\beta^{adj}(\sigma) = J + \left(1 + \frac{\lambda^W}{\lambda^B}\right)(2\sigma - 1)\epsilon.$$

In the definition of the adjusted Youden index, the unadjusted Youden index J is also redefined to account for heterogeneous productivity:

$$J \equiv 1 - \alpha - \frac{\lambda^W}{\lambda^B}\beta.$$

With homogeneous productivity, the Youden index was $J = 1 - \alpha - \beta$. It was necessarily positive because $1 - \alpha > \beta$, as the Better paradigm is superior to the Worse one. With heterogeneous productivity, although $1 - \alpha > \beta$, it is possible that the Youden index $J = 1 - \alpha - (\lambda^W/\lambda^B)\beta$ is negative. In fact this happens when the productivity of Worse scientists is large enough relative to that of Better scientists. Formally, the Youden index J is negative as soon as the ratio λ^W/λ^B is larger than the threshold λ^* given by

$$\lambda^* = \frac{1 - \alpha}{\beta}.$$

Proposition 1 must be adjusted accordingly:

PROPOSITION A2. *Without egg-eating bias ($\epsilon = 0$), there are three possible regimes, depending on the productivity of Worse scientists (λ^W) relative to the productivity of Better scientists (λ^B). When $\lambda^W/\lambda^B > \lambda^*$, the Youden index is negative ($J < 0$), so the Worse paradigm eventually prevails ($\lim_{t\to\infty} \sigma(t) = 0$). When $\lambda^W/\lambda^B = \lambda^*$, the Youden index is zero ($J = 0$), so the composition of the population of scientists is constant over time ($\sigma(t) = \sigma(0)$ for all t). When $\lambda^W/\lambda^B < \lambda^*$, the Youden index is positive ($J > 0$), so the Better paradigm eventually prevails ($\lim_{t\to\infty} \sigma(t) = 1$).*

We now return to the case with bias. The adjusted Youden index is linearly increasing in σ, takes its lowest value of $J-(1+\lambda^W/\lambda^B)\epsilon$ at $\sigma = 0$, and takes its highest value of $J+(1+\lambda^W/\lambda^B)\epsilon$ at $\sigma = 1$.

If $\epsilon \leq |J|/(1 + \lambda^W/\lambda^B)$ and $J > 0$, the adjusted Youden index is positive for all $\sigma \in (0, 1)$. If $\epsilon \leq |J|/(1 + \lambda^W/\lambda^B)$ and $J < 0$, the adjusted Youden index is negative for all $\sigma \in (0, 1)$. But if $\epsilon > |J|/(1 + \lambda^W/\lambda^B)$, the index changes sign on $(0, 1)$. We define the threshold

$$\sigma^* = \frac{1}{2}\left(1 - \frac{J}{(1 + \lambda^W/\lambda^B)\epsilon}\right). \tag{A2}$$

If $\epsilon > |J|/(1 + \lambda^W/\lambda^B)$, then $\sigma^* \in (0, 1)$. Furthermore, the adjusted Youden index is negative for $\sigma < \sigma^*$, zero at $\sigma = \sigma^*$, and positive for $\sigma > \sigma^*$.

Once the new definition of the adjusted Youden index is factored in, the dynamics of $\sigma(t)$ remain given by the differential equation (8). Since the evolution of $\sigma(t)$ is given by the same differential equation as when the productivities are the same, and since the adjusted Youden index has similar properties, we immediately obtain a proposition similar to Proposition 2:

PROPOSITION A3. *With egg-eating bias ($\epsilon > 0$), there are three possible regimes, depending on the amount of bias relative to the Youden index (J). When $\epsilon \leq |J|/(1 + \lambda^W/\lambda^B)$ and $J > 0$, the Better paradigm eventually prevails ($\lim_{t\to\infty} \sigma(t) = 1$), irrespective of the initial fraction of Better tenured scientists ($\sigma(0)$). When $\epsilon \leq |J|/(1 + \lambda^W/\lambda^B)$ and $J < 0$, the Worse paradigm eventually prevails ($\lim_{t\to\infty} \sigma(t) = 0$), irrespective of the initial fraction of Better tenured scientists ($\sigma(0)$). But when $\epsilon > |J|/(1 + \lambda^W/\lambda^B)$, the eventual outcome is determined by the initial value of the fraction of Better tenured scientists relative to the threshold σ^* given by (A2). If $\sigma(0) > \sigma^*$, the Better paradigm prevails ($\lim_{t\to\infty} \sigma(t) = 1$); but if $\sigma(0) < \sigma^*$, the Worse paradigm prevails ($\lim_{t\to\infty} \sigma(t) = 0$).*

Last, we can recompute the increase in Youden index required to start a scientific revolution. Assume that the share of Better scientists is σ and is converging to 0. To initiate a scientific revolution, we need the share of Better scientists to start converging to 1. This requires that the Youden index J increases sufficiently to be above the threshold

$$J^* = \left(1 + \frac{\lambda^W}{\lambda^B}\right)(1 - 2\sigma)\epsilon.$$

When the regular Youden index J is above the threshold J^*, the adjusted Youden index at σ is positive and science starts converging to the truth.

Defections

We begin by describing the evolution of the population of scientists without egg-eating bias ($\epsilon = 0$). Because there are defections at rate $\mu > 0$, the numbers of tenured scientists who believe in the Better and in the Worse paradigm evolve according to the following differential equations:

$$\begin{aligned}\dot{B}(t) &= -\delta B(t) + (1-\alpha)\left(\lambda B(t) + \mu W(t)\right)\\ \dot{W}(t) &= -\delta W(t) + \beta\left(\lambda W(t) + \mu B(t)\right).\end{aligned}$$

The terms $\mu W(t)$ and $\mu B(t)$ measure the advisees who defect from Worse to Better and from Better to Worse. These equations imply that the growth rates of $B(t)$ and $W(t)$ are given by

$$\begin{aligned}g^B(\sigma) &= -\delta + (1-\alpha)\lambda + (1-\alpha)\mu\frac{1-\sigma}{\sigma}\\ g^W(\sigma) &= -\delta + \beta\lambda + \beta\mu\frac{\sigma}{1-\sigma}.\end{aligned}$$

To obtain these equations we used $O/N = (1-\sigma)/\sigma$. We know that

$$\dot{\sigma}(t) = \sigma(t)(1-\sigma(t))\left(g^B(\sigma(t)) - g^W(\sigma(t))\right).$$

Combining these equations, we obtain the following differential equation:

$$\text{(A3)} \qquad \dot{\sigma}(t) = \sigma(t)(1-\sigma(t))J\lambda + \mu\left[(1-\alpha)(1-\sigma(t))^2 - \beta\sigma(t)^2\right],$$

where $J = 1 - \alpha - \beta$ is the Youden index. The first term on the right-hand side is the same as in the model without defection. The second term is new: it appears because of the defections.

The following propositions describes the dynamics of the population of tenured scientists:

PROPOSITION A4. *With defections ($\mu > 0$), no egg-eating bias ($\epsilon = 0$), and zero Youden* *index ($J = 0$), the tenured scientists are eventually evenly split between the Better and the Worse paradigm ($\lim_{t\to\infty} \sigma(t) = 1/2$).*

PROPOSITION A5. *With defections ($\mu > 0$), no egg-eating bias ($\epsilon = 0$), but positive Youden index ($J > 0$), the Better tenured scientists are eventually in a majority ($\lim_{t\to\infty} \sigma(t) = \sigma^*$ where $\sigma^* > 1/2$).*

Proof. The differential equation (A3) can be written $\dot{\sigma}(t) = P(\sigma(t))$ where

$$P(\sigma) = \sigma(1-\sigma)J\lambda + \mu\left[(1-\alpha)(1-\sigma)^2 - \beta\sigma^2\right].$$

The coefficient on σ^2 is $-J(\lambda - \mu)$. So the polynomial P is of degree 2 with $P(+\infty) = P(-\infty) = -\infty$ if $J > 0$ and $\lambda > \mu$, and it is of degree 1 if $J = 0$ or $\lambda = \mu$. Furthermore,

$$P(0) = \mu(1-\alpha) > 0, \quad P(1/2) = \frac{1}{4}J(\lambda+\mu) \geq 0, \quad \text{and} \quad P(1) = -\beta\mu < 0.$$

We infer that if $J = 0$, P has a unique root $\sigma^* = 1/2$ and $P(\sigma) > 0$ for $\sigma < 1/2$ and $P(\sigma) < 0$ for $\sigma > 1/2$. Thus, 1/2 is the unique critical point of the differential equation (A3) on [0, 1], and this critical point is a sink.

If $J > 0$ but $\lambda = \mu$, P has a unique root

$$\sigma^* = \frac{1-\alpha}{1-\alpha+\beta} \in (1/2, 1)$$

and $P(\sigma) > 0$ for $\sigma < \sigma^*$ and $P(\sigma) < 0$ for $\sigma > \sigma^*$. Thus, $\sigma^* \in (1/2, 1)$ is the unique critical point of the differential equation (A3) on [0, 1], and this critical point is a sink.

Finally, if $J > 0$ and $\lambda > \mu$, P has a unique positive root, σ^*, and this root is in $(1/2, 1)$. We know this because P is of degree 2, the coefficient on σ^2 in P is negative, $P(0)$ is positive, and $P(1/2)$ is positive. The root σ^* is a complicated expression of the parameters:

$$\sigma^* = \frac{1}{2}\left[\frac{\lambda J - 2\mu(1-\alpha)}{(\lambda-\mu)J} + \sqrt{\left[\frac{\lambda J - 2\mu(1-\alpha)}{(\lambda-\mu)J}\right]^2 - \frac{4\mu(1-a)}{(\lambda-\mu)J}}\right]$$

and $P(\sigma) > 0$ for $\sigma < \sigma^*$ and $P(\sigma) < 0$ for $\sigma > \sigma^*$. Hence, $\sigma^* \in (1/2, 1)$ is the unique critical point of the differential equation (A3) on [0, 1], and this critical point is a sink. □

We now describe the evolution of the population of scientists with egg-eating bias ($\epsilon > 0$). The population dynamics become more complex. Using the same logic as above, we find that the fraction σ of tenured scientists who believe in the Better paradigm satisfies

$$\dot{\sigma}(t) = \lambda\sigma(t)(1-\sigma(t))J^{adj}(\sigma(t)) + \mu\left[(1-\alpha^{adj}(\sigma(t)))(1-\sigma(t))^2 - \beta^{adj}(\sigma(t))\sigma(t)^2\right], \tag{A4}$$

where the adjusted Youden index $J^{adj}(\sigma)$ is given by (3), the adjusted statistical significance $\alpha^{adj}(\sigma)$ is given by (1), and the adjusted statistical power $\beta^{adj}(\sigma)$ is given by (2).

We first study the case in which the Youden index is zero:

PROPOSITION A6. *With defections ($\mu > 0$), egg-eating bias ($\epsilon > 0$), and zero Youden index ($J = 0$), there are two possible regimes, depending on the value of defection rate relative to the threshold*

$$\mu^* = \frac{\lambda\epsilon}{2\beta - \epsilon}. \tag{A5}$$

If $\mu \geq \mu^*$, the tenured scientists are eventually evenly split between the Better and the Worse paradigm ($\lim_{t\to\infty}\sigma(t) = 1/2$), irrespective of the initial value of the fraction of Better tenured scientists ($\sigma(0)$). If $\mu < \mu^*$, the eventual outcome is determined by the initial fraction of Better tenured scientists. If $\sigma(0) < 1/2$, Worse tenured scientists are eventually a majority ($\lim_{t\to\infty}\sigma(t) = \sigma^*$ where $\sigma^* < 1/2$). If $\sigma(0) > 1/2$, Better scientists are eventually a majority ($\lim_{t\to\infty}\sigma(t) = 1 - \sigma^*$ where $1 - \sigma^* > 1/2$).

Proof. We define the polynomial P by

$$P(\sigma) = \sigma(1-\sigma)(1-\alpha^{adj}(\sigma) - \beta^{adj}(\sigma))\lambda + \mu\left[(1-\alpha^{adj}(\sigma))(1-\sigma)^2 - \beta^{adj}(\sigma)\sigma^2\right].$$

Equation (A4) can be written as $\dot{\sigma}(t) = P(\sigma(t))$. The behavior of σ is determined by the properties of P. Since $\alpha^{adj}(\sigma)$ and $\beta^{adj}(\sigma)$ are linear in σ when $\epsilon > 0$, P is of degree 3 or less.

We study the case with $J = 0$ and $\epsilon > 0$. Note that in this case

$$1 - \alpha^{adj}(1-\sigma) = 1 - \alpha - \epsilon + 2(1-\sigma)\epsilon = \beta + \epsilon - 2\sigma\epsilon = \beta^{adj}(\sigma).$$

This result implies that $P(1-\sigma) = -P(\sigma)$ and thus that $P(1/2) = 0$. Thus, 1/2 is a root of P. In addition,

$$P(0) = \mu(1-\alpha-\epsilon) > 0, \qquad \text{and} \qquad P(1) = -\mu(\beta-\epsilon) = -\mu(1-\alpha-\epsilon) < 0.$$

From this we infer that either $1/2$ is the unique root of P on $[0,1]$ and $P(\sigma) > 0$ for all $\sigma < 1/2$ and $P(\sigma) < 0$ for all $\sigma < 1/2$ (in that case $P'(1/2) < 0$). Or P has 3 roots on $[0,1]$: 1/2, σ^*, and $1-\sigma^*$. In that case $P(\sigma) > 0$ for $\sigma \in [0, \sigma^*)$, $P(\sigma) < 0$ for $\sigma \in (\sigma^*, 1/2)$, $P(\sigma) > 0$ for $\sigma \in (1/2, 1-\sigma^*)$, and $P(\sigma) < 0$ for $\sigma \in (1-\sigma^*, 1]$ (also in that case $P'(1/2) > 0$). Hence if $P'(1/2) < 0$, then $1/2$ is the unique critical point of equation (A4) on $[0, 1]$, and it is a sink. Or if $P'(1/2) > 0$, then $1/2$ is a source of equation (A4), with two sinks ($\sigma^* \in (0, 1/2)$ and $1-\sigma^* \in (1/2, 1)$) at equal distance on each side.

We compute $P'(1/2)$ to characterize the two different regimes. After some algebra find that

$$P'(1/2) = \epsilon(\lambda + \mu) - 2\mu\beta.$$

Hence, $P'(1/2) > 0$ if and only if $\mu < \mu^*$, where μ^* is given by (A5). Thus, if $\mu > \mu^*$, then 1/2 is a sink. And if $\mu < \mu^*$, then 1/2 is a source. Notice that as ϵ gets larger, 1/2 is more likely to be source.

When 1/2 is a source, there are two sinks on each side of 1/2, at the same distance from 1/2: σ^* and $1-\sigma^*$. It is easy to compute these two other critical points. When 1/2 is a source, P has 3 roots: 1/2, σ^*, and $1-\sigma^*$. It can therefore be written

$$P(\sigma) = -4\epsilon(\lambda - \mu)(\sigma - 1/2)(\sigma - \sigma^*)(\sigma - (1-\sigma^*)).$$

The term of degree 0 in P is $2\epsilon(\lambda-\mu)\sigma^*(1-\sigma^*)$. Using the other expression for P, we see that the term of degree 0 can also be written $\mu(\beta - \epsilon)$. We infer that

$$\sigma^*(1-\sigma^*) = \frac{1}{2} \cdot \frac{\mu}{\lambda - \mu} \cdot \frac{\beta - \epsilon}{\epsilon} \equiv \theta.$$

We verify that for $\mu < \lambda\epsilon/(2\beta - \epsilon)$, it is indeed the case that $\theta \in (0, 1/4)$. We infer that the critical point σ^* is given by

$$\sigma^* = \frac{1 - \sqrt{1-4\theta}}{2}.$$

When $\mu > 0$, then $\sigma^* > 0$ and $1 - \sigma^* < 1$ and the two sinks are interior. □

We now study the case in which the Youden index is positive:

PROPOSITION A7. *With defections ($\mu > 0$), egg-eating bias ($\epsilon > 0$), and positive Youden index ($J > 0$), there are two possible regimes, depending on the value of defection rate relative to the threshold μ^* given by* (A5)*, and depending on the value of the Youden index. With either $\mu \geq \mu^*$, or $\mu < \mu^*$ and J large enough, Better scientists are eventually a majority ($\lim_{t\to\infty} \sigma(t) = \sigma^*$ where $\sigma^* > 1/2$). With $\mu < \mu^*$ and J small enough, the eventual outcome is determined by the initial fraction of Better tenured scientists ($\sigma(0)$). There is some $\sigma^\dagger < 1/2$ such that if $\sigma(0) < \sigma^\dagger$, Worse scientists are eventually a majority ($\lim_{t\to\infty} \sigma(t) = \sigma^\star$ where*

$\sigma^\star < 1/2$*), and if* $\sigma(0) > \sigma^\dagger$*, Better scientists are eventually a majority* ($\lim_{t\to\infty} \sigma(t) = \sigma^*$ *where* $\sigma^* > 1/2$*).*

Proof. We define the polynomial P by

$$P(\sigma) = \sigma(1-\sigma)(1-\alpha^{adj}(\sigma) - \beta^{adj}(\sigma))\lambda + \mu\left[(1-\alpha^{adj}(\sigma))(1-\sigma)^2 - \beta^{adj}(\sigma)\sigma^2\right].$$

Equation (A4) can be written as $\dot{\sigma}(t) = P(\sigma(t))$. Note that $P(0) = \mu(1-\alpha-\epsilon) > 0$ and $P(1) = -\mu(\beta-\epsilon) < 0$.

Next, we define

$$\begin{aligned}
\hat{A}(\sigma) &= 1 - \beta + \epsilon - 2\sigma\epsilon \\
\hat{P}(\sigma) &= \sigma(1-\sigma)(1-\hat{A}(\sigma) - \beta^{adj}(\sigma))\lambda + \mu\left[(1-\hat{A}(\sigma))(1-\sigma)^2 - \beta^{adj}(\sigma)\sigma^2\right].
\end{aligned}$$

Since $\alpha = 1-\beta-J$, $\alpha^{adj}(\sigma) = \hat{A}(\sigma) - J$ and $P(\sigma) = \hat{P}(\sigma) + Q(\sigma)$, where

$$Q(\sigma) = J(1-\sigma)\left[\sigma\lambda + (1-\sigma)\mu\right].$$

For given β, μ, and ϵ, the polynomial $\hat{P}$ is the polynomial studied in Proposition A6. So all the properties of $\hat{P}$ are known.

The difference between the polynomials P and $\hat{P}$ is the polynomial Q. The polynomial Q is simple to analyze because it is only of degree 2 (whereas P and $\hat{P}$ are of degree 3). The polynomial Q has the following properties: $Q = 0$ when $J = 0$, $Q(\sigma) > 0$ on $(0,1)$ when $J > 0$, $Q(1) = 0$, $Q(0) = J\mu$, $Q(\sigma)$ is maximized at $(\lambda - 2\mu)/(2\lambda - 2\mu) \leq 1/2$, and the maximum value of Q on $(0,1)$ is $Q^* = J\lambda^2/[4(\lambda-\mu)]$.

When $\mu \geq \mu^*$, we know that $\hat{P}(\sigma)$ is decreasing on $(0,1)$ and is 0 at 1/2. Furthermore, $Q(\sigma) \geq 0$ for $\sigma \in [0,1/2]$. Since $Q(\sigma) > 0$ and $\hat{P}(\sigma) \geq 0$, then $P(\sigma) > 0$ for $\sigma \in [0,1/2]$. Since $\hat{P}$ and Q are decreasing on $[1/2,1]$ and $P(1/2) > 0$ and $P(1) < 0$, P has a unique root σ^* on (1/2,1). Overall, P has a unique root σ^* on $(0,1)$ and $P(\sigma) > 0$ for all $\sigma < \sigma^*$ and $P(\sigma) < 0$ for all $\sigma > \sigma^*$.

When $\mu < \mu^*$, we know that $\hat{P}(\sigma)$ has three roots on $(0,1)$: 1/2, one root on $(0,1/2)$, and one root on $(1/2,1)$. The value of Q is bounded by $Q^* = J\lambda^2/[4(\lambda-\mu)]$. So for J small enough, we can be sure that there remains two roots of P on $(0,1/2)$. $P(1/2) = (\lambda+\mu)J/4 > 0$ while $P(1) < 0$ so P has at least one root on $(1/2,1)$. Since P cannot have more than 3 roots, P has exactly one root on $(1/2,1)$.

If J is large enough, it is possible to eliminate the two roots on $(0,1/2)$. The root on $(1/2,1)$ always remains. This root is unique for the following reasons. The polynomial $\hat{P}$ has a local maximum at $\hat{\sigma} > 1/2$. The polynomial $\hat{P}$ is decreasing on $(\hat{\sigma}, 1)$. The polynomial Q is also decreasing on $(\hat{\sigma}, 1)$ (in fact it is decreasing on $(1/2, 1)$). Hence the polynomial P is decreasing on $(\hat{\sigma}, 1)$. So there cannot be more than one root on $(\hat{\sigma}, 1)$. Furthermore, there cannot be any root on $(1/2, \hat{\sigma})$ since both Q and $\hat{P}$ are positive. $\square$