How Research Goes Astray: Paths and Equilibria

Richard Startz*

April 2019

Abstract

Some research disciplines persist for long periods despite contributing very little to our understanding of the world. In the presence of imperfect information about the quality of a given item of research and the existence of informational externalities, multiple equilibria can exist in which path dependence leads to disciplines going astray because researchers are incentivized to invest in low value lines of research. Research fields which have high social value are more likely to thrive, as are fields in which results are subject to evidentiary testing for their contribution to knowledge. When the allocation of resources responds strongly to research productivity, disciplines which focus on an agenda of low productivity may gradually disappear. Survivorship bias can lead to a distribution across disciplines in which weak disciplines have more talented researchers than do stronger disciplines.

JEL: B00, O3

Keywords: Research, multiple equilibria, science

* Department of Economics, University of California, Santa Barbara, startz@ucsb.edu. Advice from Shelly Lundberg and Meredith Startz is appreciated. The most recent version of this paper can be found at http://www.econ.ucsb.edu/~startz/Epistemology.pdf
I. Introduction

It is a common belief among academics—a belief certainly shared by many economists—that some disciplines contribute very little to the commonweal. Work in these disciplines may be terribly clever but the product is not of much social value, other than perhaps as a source of entertainment and employment for their practitioners. In a world of complete information and well-functioning markets, inferior products should not long survive. It seems, however, that largely useless branches of research do survive for a very long time.

I present here a model of imperfect information about the value of research, in which informational externalities lead to path-dependent multiple equilibria such that the accidents of history can lead a discipline to be dominated by an inherently valuable research agenda, or unfortunately, by an agenda of much less utility. A high inherent value of a line of research certainly advantages growth of that line, but other factors may outweigh that advantage. The advantage to fields with high inherent value is consistent with the informal observation that stronger disciplines do tend to thrive.

I assume that a new piece of research generates a noisy signal of its value that can be assessed only by others within the discipline. In the model below, cumulative growth in a line of research leads to work in that field being more accurately assessed. There will be a higher marginal incentive to invest in a field in which the outcome of investment is more accurately judged and therefore more likely to be rewarded. This positive feedback means that a field where more work has been done—even if through accident of history—is advantaged in attracting more resources. This leads to a yet greater advantage in cumulative knowledge.
I assume that researchers are rewarded according to the perceived quality of the research they produce. They are endowed with limited resources in the form of time and intellectual energy, and invest these scarce resources into producing better research, trading off the returns available in investing in different fields (subdisciplines) as well as the opportunity to consume rather than invest available resources. Agents act atomistically, taking equilibrium as given. I assume there are no cabals, conspiracies, nor other forms of collective action. Research is neither promoted nor suppressed for strategic reasons.

The first source of imperfect information is that the value of a specific piece of research is imperfectly observed. The model is one in which only a signal of value is observed, in a standard setup in which the signal consists of the true value of the research plus a noisy, random error. Agents are compensated for their research according to its perceived value, solved for in a signal extraction problem in which the parameters are consistent with general equilibrium. This leads fields in which signals are judged more accurately to be more productive.

The second source of imperfect information is that ideas in some fields are more “testable” than in other fields. For example, Smolin (2007) writes, “The current crisis in particle physics springs from the fact that the theories that have gone beyond the standard model in the last thirty years fall into two categories. Some were falsifiable, and they were falsified. The rest are untested—either because they make no clean predictions or because the predictions they do make are not testable with current technology.” The extent of testability might be due to the underlying nature of a discipline, or it might be due to disciplinary social norms. The word “testable” is intended to evoke the idea of bringing empirical evidence to bear and the idea of falsifiability. Some care is needed however in that is clear that in economics and in many other
disciplines the greatest ideas are often advances in theory. Such advances may explain previously unexplainable existing scientific observations, or may be valuable for helping understand stylized facts. What is really meant here by “testable” is that experts in the discipline can evaluate the quality of a contribution to improving our understanding of the world—as distinct from evaluating its cleverness as an improvement to a self-referential intellectual endeavor. In the model below, there is a greater reward to an individual for investing in a testable idea than in a non-testable idea. Ceteris paribus, this leads fields that emphasize testability to produce higher value research. Because of path dependence due to knowledge accumulation though, a field with less testability may come to dominate despite being disadvantaged—potentially pulling down the entire discipline. Alternatively, when a discipline is dominated by a successful field this may provide protection for maintaining a weaker field in that discipline.

A two-part assumption generates the dynamic externality that leads to multiple equilibria. The first part of the assumption is that in the short run the value of a specific piece of research can only be judged by experts within the discipline, albeit even then only imperfectly. Referees, editors, and grant reviewers are largely drawn from the pool engaged in closely related work, and are assumed to make optimal-albeit-imperfect evaluations of each piece of research conditional on available information. Outsiders are not able to make useful judgments about individual pieces of work or about the merit of fields within a discipline. The second part of the assumption is that the precision with which contributions in a field are judged is increasing in the cumulative value of knowledge in that field. This is closely related to, but not precisely the same as, the explanation given by Kuhn (1996, p. 10.) “‘normal science’ means research firmly
based upon one or more past scientific achievements, achievements that some particular scientific community acknowledges for a time as supplying the foundation for its further practice.”¹ If the precision of judging work in one field is relatively high, then the signal extraction problem puts more weight on the signal about a given piece of work and less on the average value of work in that field. This increases the marginal return to investing in that field. The greater incentive leads to greater investment. Over time, greater investment generates higher cumulative knowledge in the field. Thus there is a dynamic, positive feedback loop that can generate multiple steady-states in a pretty standard way.

Multiple equilibria are not a necessary outcome in the model. If ideas in a field are sufficiently more testable than in an alternative area, the incentives for research in that field can be strong enough that there is only a single steady-state equilibrium in which that field dominates. Similarly if work in a field is sufficiently inherently more valuable, that field becomes dominant.

The third assumption is that outsiders are able to assess the contribution of a discipline as a whole even though they are unable to evaluate individual contributions, and that such assessments have a strong effect on the flow of resources. By “strong” I mean that an increase in resources to a discipline leads to greater marginal rewards for doing research rather than privately consuming the limited available resource. I make the further assumption that outside evaluations for the purpose of redirecting resources occur at quite low frequencies. Where there are multiple equilibria, one equilibrium is more productive overall than the other. If path

¹ Kuhn also differs in describing a more zero/one mechanism evaluation mechanism, writing about the “standards by which the profession determined what should count as an admissible problem or as a legitimate problem-solution.”
dependence leads to the less productive equilibrium, then when the whole discipline is evaluated by outsiders it may face a reduction in overall resources. This leads in turn for research to be a less attractive investment. At the next outside evaluation resources are further reduced, leading to yet a further reduction in research production. In this way, a discipline dominated by a field in which testability is low can go into a death spiral, eventually collapsing entirely.

The idea that a field may go astray does not seem terribly controversial, although the mechanism described here is novel. The notion is a commonplace conclusion of lunchtime conversations among economists with regard to certain areas in the humanities and the social sciences. In the sciences, Smolin (2007) suggests the entire enterprise of string theory is an example of an area gone astray. Indeed, he writes, “At Harvard, the string theory seminar was called the Postmodern Physics seminar. This appellation was not meant ironically.” Even within economics, prominent scientists occasionally suggest that branches of the discipline are going about research the wrong way. Here are a few examples selected from writings of Nobel laureates and Presidents and Distinguished Fellows of the American Economic Association. Leontief (1971) criticized mathematical theorizing as having an inadequate empirical basis, writing “Uncritical enthusiasm for mathematical formulation tends often to conceal the ephemeral substantive content of the argument behind the formidable front of algebraic signs.” Solow (1985) wrote “economic theory learns nothing from economic history, and economic history is as much corrupted as enriched by economic theory.” A number of the most prominent macroeconomists have decried approaches of macroeconomic research quite publicly; see Lucas (1976), Lucas and Sargent (1989), Akerlof (2002), Blanchard (2016), and
Romer (2016).² Criticizing the paths of both mathematical economics and monetary economics Johnson (1974) wrote,

...there are serious dangers that monetary economics will be diverted by the inherent limitations of the elaborate technologies it has developed into concentration on second-rate but solvable problems and into ceremonially impressive but scientifically unproductive solutions of trivial problems or artificially created pseudo-problems...mathematical economics generates its own criteria of problem-evaluation, problem-solving, and style of communication of results-criteria not purely coincidentally entailing an offensive degree of intellectual snobbishness and elitism—which do not necessarily promote the progress of economic science...The main problem is to keep control of our propensities to select problems that serve mainly to provide exercise for our analytical and statistical tools, and to ignore the law of diminishing returns from concentrating effort on deeper work on already recognized and analysed problems.

It should be said that nothing in what follows attempts to provide a general theory of the production of knowledge. Obviously, sometimes a vein of research is simply worked out or a new discovery invalidates an existing research agenda. Rather, assumptions are made precisely to lead to an explanation of the long-term survival of fields of low merit and how this can lead to the gradual collapse of disciplines. The phenomenon to be explained is that a field of low merit can dominate a discipline despite there being a high merit alternative. Throughout, specific functional forms are assumed when convenient for analysis. For example, I assume certain variables obey joint Gaussian distributions. I also make assumptions that are sharper than precisely what we see in the world. For example, I assume that only researchers within a discipline are called upon to evaluate the quality of a specific piece of research where in

---

² This is not to say that the articles cited here suggest the mechanism discussed in the present paper. Smolin (2007) and Romer (2016) attribute over concentration in the wrong field to excessive respect for the personal authority of prominent scientists, which is closer to the mechanism suggested in Akerlof and Michaillat (2018).
practice boundaries between areas of research are somewhat porous. Nor do the results here indicate there are no cabals or conspiracies in the dissemination and evaluation of research—they simply suggest that one can explain aspects of the evolution of disciplines without their presence.

In the next two sections I present and analyze a formal model to explicate these ideas. While I present a specific form of imperfect information and dynamic informational externality that I think is realistic, other mechanisms might well give similar results. Board et. al. (2017) offer a model in which firms with better managers recruit better talent and the better talent evolves into future better management. One could model selection into disciplines similarly, keeping the notion that disciplines with more able senior researchers recruit more able juniors. One element that is quite different from the model below is that in my model the talent pool in strong and weak disciplines is equal; failing disciplines have talented researchers working on the wrong problems.

Akerlof and Michaillat (2018) offer a model in which weak paradigms can last a long time, to be replaced only when powerful evidence becomes available. The persistence of a weak paradigm is due to homophily. As the authors write “Tenured scientists display homophily: they favor tenure candidates belonging to their paradigm.” As below, an endogenous suboptimal equilibrium can develop that is stable with respect to small amounts of evidence against the weaker paradigm. The assumption of homophily differs from the assumption in my model in that I assume that researchers have no operative self-interest in promoting one area of research over the other, even though the social equilibrium matters to the researchers. One suspects that in addition to exogenous homophily that there have been instances in some
disciplines in which endogenous cartels and cabals appear which for a period of time are able to act out of self-interest to promote their own work above the work of others. But that is not the story offered here.

Brock and Durlauf (1999) offer a model of (potentially) multiple steady-states in the choice between two theories that is also due to social factors. The Brock and Durlauf mechanism is quite different from the mechanism below, focusing on the role of social conformity in the adoption of competing theories.

The model below does not consider herding, offering a somewhat different self-reinforcing informational externality. One can imagine a model in which early discovery in one field makes that field appear to be fertile for further investigation leading to herding. Strevens (2013) talks through a related model in which herding can lead to suboptimal allocation of effort between disciplines.

In the next section, I provide a formal model of research investment, knowledge accumulation, and resource allocation.

II. A formal model

This is a model in which research agendas can stray from the social optimum. In order to have a place for agendas to go astray to, I begin with the idea that a researcher in a discipline allocates resources between two fields within that discipline. Call the two fields $Y$ and $T$. $Y$ is socially preferable to $T$ because it has a higher value in contributing to utility of the general population, although there are marginal tradeoffs in production between the two areas. Going “astray” means producing more $T$ than $Y$, even though the reverse is socially preferable.
Sometimes I will refer to whichever field has greater output as “dominant.” Researchers respond to incentives. If private incentives and social margins are equal, then equilibria will be Pareto optimal of course. The model below explores a dynamic, informational externality causing private and social margins to deviate.

In the formal model functions and parameters are chosen for ease of analysis rather than any attempt at generality. In particular, when we get to the solution the specific functional forms chosen allow for a reduction in the state space to two variables. More general choices would lead to similar possible outcomes, but would require a four variable state space. Throughout, I assume a large set of agents with unit measure so there is little need to distinguish individual and aggregate variables.

*Production of research*

I begin with the production problem. At each point in time, each agent receives an idiosyncratic draw of a research idea of value $Y^e$ and a research idea of value $T^e$ respectively, where $Y^e$ and $T^e$ are flows values. Researchers choose to augment their endowments by an additional $Y$ and $T$ by exerting effort, so total production is $\{Y^e + Y, T^e + T\}$. Researchers have a limited flow capacity of intellectual endeavor, $K$, which can be deployed creating $Y$ or $T$ or which can be employed in forms of intellectual recreation, $C$. The constraint, exhibiting constant returns to scale, is

$$K = [Y^\omega + T^\omega]^{1/\omega} + C, \omega > 1$$

(1)

$\bar{Y} = \bar{Y}^e + \bar{Y}$ is the aggregate flow of new research averaged over all researchers and $\ddot{Y}_t$ is the aggregate stock. (And analogously for $T$. In the interest of brevity, some of the $T$ formulas
are not displayed where they are identical to the $Y$ formulas except for subscript $T$ rather than $Y$.) Assume that the stock of research depreciates at the rate $\rho$, common to both fields. Then

$$\tilde{Y}_t = (1 - \rho)\tilde{Y}_t dt + \tilde{Y}_t$$

and the steady-state, $Y^*$, occurs at $Y^* = \tilde{Y}^*/\rho$.

**Signaling and information**

Turn now to the role of imperfect information. Experts within a discipline evaluate the contribution of a new piece of research, but do so subject to error. Each new piece of research generates a signal of its quality, but only some of these signals allow for a test of the social value of the research. As an example of the latter issue, ideas in chemistry are subject to test and, by and large, ideas in alchemy historically were not. The suggestion is that as the discipline moved toward modern, evidence-based chemistry, the value of a contribution in chemistry became possible for experts to evaluate, while for work from alchemists such value was (historically) impossible to verify or refute.

For modelling purposes, a signal not subject to evaluation by evidence is treated as effectively not being a signal at all. With probabilities $\{\lambda_Y, \lambda_T\}$ a testable signal-with-error, $\{S_Y, S_T\}$, is observed by experts. If no testable signal if observed, then the evaluation of the project is simply the unconditional mean, $E(Y^e + Y)$. If a signal $S_Y$ is observed, then experts evaluate the research at $E(Y^e + Y | S_Y)$. The noisy signal of research units is

$$S_Y = (Y^e + Y) + \varepsilon_Y$$

where endowments $Y^e$ and $T^e$ are assumed to be independent between the two fields as are the error terms $\{\varepsilon_Y, \varepsilon_T\}$. 
Assume for analytic convenience that across the population of agents—where individuals are denoted by subscript \( i \) as needed for clarity—\( Y_i^e \) and \( \varepsilon_{Y,i} \) are joint normal and uncorrelated with \( Y_i^e \sim N(0, h_Y^{-1}) \) and \( \varepsilon_{Y,i} \sim N(0, h_{\varepsilon_Y}^{-1}) \), where \( h \) is the precision (inverse variance). The zero mean assumption is inessential but simplifies the analysis. I prove below that \( Y^e + Y \) is also distributed jointly normal with \( \varepsilon_Y \). It follows from standard results that the unconditional and conditional expectations are:

\[
\bar{Y} = E(Y_i^e + Y_i) = E(Y)
\]

\[
E(Y_i^e + Y_i | S_{i,Y}) = E(Y_i^e + Y_i) + \beta_Y (Y_i^e + Y + \varepsilon_{i,Y} - E(Y_i^e + Y + \varepsilon_{i,Y}))
\]

where

\[
\beta_Y = \frac{h_{\varepsilon_Y}}{h_Y + h_{\varepsilon_Y}}, \quad \beta_T = \frac{h_{\varepsilon_T}}{h_T + h_{\varepsilon_T}}
\]

In equilibrium, the marginal impact of the signal, \( \beta_Y \), is common to all agents and noise errors average out. In order to determine individual incentives and general equilibrium, all that is needed going forward is the relation for the typical agent,

\[
E(Y^e + Y | S_Y) = (1 - \beta_Y)\bar{Y} + \beta_Y Y
\]

I turn now to the determination of \( h_{\varepsilon} \), which is the key to the development of a “Dutch disease” in which allocation of resources depends on cumulative production (Krugman 1987). The underlying notion is that expertise in judging work in a field develops as more researchers have worked in that field and as more is known about the field and that as a result error precision is improved (error variance is reduced) as cumulative knowledge in a field increases. Kuhn (1996, p. 36) writes, “To scientists, at least, the results gained in normal research are significant because they add to the scope and precision with which the paradigm can be
applied.” As shown below, in general equilibrium this means that more work in a field improves the ability to read a signal in that field which leads to a greater incentive for further work in that field. In this way, path dependence plays an important role in determining which field becomes dominant within a discipline. The ability to evaluate research in the two fields is modeled as rival, so that increased knowledge in one field compared to the other increases the precision of signal evaluation in the first field at the expense of the second. Assuming precision is rival is not essential, but does allow for reducing the state space to two ratios. For convenience I assume the specific functional forms, where $Q$ is an unimportant constant:

$$h_{\varepsilon Y} = \left(\frac{Y}{T}\right)^{\gamma} Q, h_{\varepsilon T} = \left(\frac{T}{Y}\right)^{\gamma} Q, \gamma > 0$$

(3)

Greater accuracy is driven by increased precision of $\varepsilon$, $h_{\varepsilon Y}$ and $h_{\varepsilon T}$; if the precision of $\varepsilon$ is infinite then for a testable research quality is perfectly observable. In other words, from equation (2) we see that $h_\varepsilon \to \infty \Rightarrow \beta \to 1$ (private and social incentives will turn out to match) and $h_\varepsilon \to 0 \Rightarrow \beta \to 0$ (there is no marginal effect of greater investment on the signal).

**Demand side and the problem of the researcher**

Turn now to the demand side of the economy. First, assume that in the broader economy the two fields have exogenously given relative unit values $p_Y$ and $p_T$. (So $Y$ being socially preferred to $T$ just means $p_Y > p_T$.) Second, people outside the discipline view a marginal unit from either field as being worth $w$. In other words, the total social value of a unit of research is $p_Yw$ or $p_Tw$ respectively. Note that the outside valuation, $w$, does not affect the relative value of the fields, consistent with the assumption that outsiders assess the discipline as a whole but
cannot make fine judgements between fields. For the moment, \( w \) is exogenous. Later, I allow \( w \) to change dynamically while still assuming that researchers take \( w \) to be exogenous with respect to their own choices.

While the social value of a unit of research is \( pw \), the private return to the researcher needs to be determined. Call the private rewards to a unit \( \pi_Y \) and \( \pi_T \), where the relation between \( pw \) and \( \pi \) needs to be determined in general equilibrium. The compensation to the researcher depends on the evaluation of the research, as described above. Compensation equals the private return to a unit of research times the evaluation of the number of quality units produced. The latter further depends on whether the research is testable—probability \( \lambda \), in which case evaluation is conditional on the signal produced, or not testable—probability \( 1 - \lambda \), in which case the signal is irrelevant and the evaluation is unconditional. Thus compensation is

\[
p_Yw[(1 - \lambda_Y)\bar{Y} + \lambda_Y[(1 - \beta_Y)\bar{Y} + \beta_YY]]
\]

and marginal compensation for produced research is

\[
\pi_Y = p_Yw\lambda_Y\beta_Y \quad \text{and} \quad \pi_T = p_Tw\lambda_T\beta_T
\]  

(4)

We can now solve the problem for each researcher. Researchers receive compensation \( \pi_Y \) and \( \pi_T \) for each quality unit of \( Y \) and \( T \) respectively. They receive commensurable utility \( C^\mu, 0 < \mu < 1 \), from flow consumption of the good \( C \). Each researcher solves the Lagrangian

\[
\max_{\dot{Y},\dot{T},L} L = \pi_YY + \pi_TT + C^\mu - \delta \left( K - [Y^\omega + T^\omega]^{1/\omega} - C \right),
\]

the interior solution to which is given by

\[
\frac{\pi_Y}{\pi_T} = \left( \frac{Y}{T} \right)^{\omega - 1}
\]  

(5)
\[
C = \min \left\{ \left( \frac{\pi_Y}{\mu} \left[ 1 + \left( \frac{\pi_Y}{\pi_T} \right)^{-\frac{\omega}{\omega-1}} \right]^{1-\frac{1}{\omega}} \right)^{\frac{1}{\mu-1}}, K \right\}
\]  

(6)

\[
Y = \left[ K - \left( \frac{\pi_Y}{\mu} \left[ 1 + \left( \frac{\pi_Y}{\pi_T} \right)^{-\frac{\omega}{\omega-1}} \right]^{1-\frac{1}{\omega}} \right) \left( 1 + \left( \frac{\pi_Y}{\pi_T} \right)^{\frac{\omega}{\omega-1}} \right)^{-\frac{1}{\omega}} \right]
\]  

(7)

Most of the analysis below is done in terms of \( Y/T \); if desired, a partial equilibrium solution for \( T \) can be found by combining equations (5) and (7). Note that equation (5) gives one of the two relations we need between the two state variables. Also note that the ratio of production in the two fields does not depend on the level of intellectual resources, \( K \), available to a typical researcher in the discipline. Therefore, \( K \) does not determine which field is dominant. It is not the case (in this model) that the level of intellectual ability in a discipline determines whether the socially preferable field is dominant.

Since the solution for \( Y \) in equation (7), which is independent of endowments, is conveniently the same for all researchers, \( Y_i^e + Y \) is also normal. The required joint normality of the signal and error is established. This is as in Lundberg and Startz (1983), with a slight variant of the cost function.

**Social valuation of the discipline**

The model to this point determines the behavior of a discipline in isolation. In fact, research generally requires outside resources. I assume that funders outside the discipline are able to get a good sense of the overall value produced by the discipline and fund the discipline as a
whole, but are unable to distinguish the value of individual projects or individual fields. I also assume that assessments are revised only at relatively long intervals. For convenience, assume that \( w \) responds to the steady-state value of a discipline’s production, \( p_Y Y^* + p_T T^* \).

\[
 w = f (p_Y Y^* + p_T T^*), \quad f'(\cdot) > 0 \quad (8)
\]

Remember that \( w \) is the outside value of a unit of research, so that the assumption on the sign of the derivative is not entirely innocuous. The implication is that—over the relevant range—that an increase in the value of research in a discipline is eventually positively rewarded, which creates (see below) an incentive for further improvements.

III. Analysis

Let \( Y \) be the socially preferred field of research by picking parameters such that \( Y \) has two inherent advantages over \( T \). First, \( Y \) is inherently more valuable—by which I mean that \( p_Y \geq p_T \). Second, research in \( Y \) is assumed to be more readily testable than \( T \). This means that an item of research on \( Y \) is more likely to produce a quality signal that can be evaluated, with \( \lambda_Y \geq \lambda_T \). Greater information about a typical item of research on \( Y \) means that private and social incentives are closer in the \( Y \)-field than in the \( T \)-field. Ceteris paribus both these characteristics lead to relatively greater production of \( Y \) than of \( T \). It is important though that \( Y \) is not easier to produce than \( T \). In particular, work done on \( T \) does not require less intellectual capacity than \( Y \). The story here is one in which completely capable scholars end up engaged in not very productive activity.

The choice of functional forms leads to a convenient analysis of steady-states in terms of the ratios \( \beta_Y / \beta_T \) and \( Y / T \). The production equation using equations (5) and (4) is given by
\[
\frac{Y}{T} = \left( \frac{\beta_Y}{\beta_T} \right) \frac{1}{\omega-1} \left( \frac{p_T \lambda_Y}{p_T \lambda_T} \right)^{\frac{1}{\omega-1}}
\]  

(9)

Note that at the steady-state the flow and stock ratios of \(Y\) to \(T\) are the same, so equation (9) also gives the steady-state ratios.

The relations for each field between error precision and the marginal value of a signal in equation (2) together with the judgement accumulation equation (3) gives the signaling relationship

\[
\frac{\beta_Y}{\beta_T} = \frac{h_{\varepsilon_Y}}{h_Y + h_{\varepsilon_Y}} = \frac{h_{\varepsilon_T}}{h_T + h_{\varepsilon_T}}
\]  

(10)

Inserting the judgment accumulation equations (3) gives the signaling relationship

\[
\frac{\beta_Y}{\beta_T} = \frac{\left( \frac{\bar{Y}}{\bar{T}} \right)^y Q}{h_Y + \left( \frac{\bar{Y}}{\bar{T}} \right)^y Q} + \frac{\left( \frac{\bar{Y}}{\bar{T}} \right)^{-y} Q}{h_T + \left( \frac{\bar{Y}}{\bar{T}} \right)^{-y} Q} \right]^{-1}
\]  

(11)

Steady-state equilibria can be found by evaluating equations (9) and (11) at steady-state stock values. Note that both relations are upwards sloping.

Figure 1\(^3\) illustrates an example steady-state solution. The signaling curve shows the relation between \(\beta_Y/\beta_T\) and \(\bar{Y}/\bar{T}\), equation (11), as the solid line. The flow production curve, equation

\(^3\) Figure 1 is drawn for visual clarity rather than to reproduce the functions above. Using the functions as specified and setting \(h_Y = h_T = 2, Q = 1, y = 0.7, \omega = 1.9, \lambda_Y = \lambda_T = 1, p_Y = p_T = 1, w = 1\). With \(\lambda_T = 1.0\), there are stable steady-states at \([\beta_Y/\beta_T, \bar{Y}/\bar{T}] = [0.142, 0.114]\) and \([\beta_Y/\beta_T, \bar{Y}/\bar{T}] = [7.034, 8.737]\). There is an unstable steady-state at \([\beta_Y/\beta_T, \bar{Y}/\bar{T}] = [0.999, 0.999]\). If the value of \(\lambda_T\) were to decrease to \(\lambda_T = 0.982\), the stable states move to higher \(Y\) values, \([\beta_Y/\beta_T, \bar{Y}/\bar{T}] = [0.214, 0.184]\) and \([\beta_Y/\beta_T, \bar{Y}/\bar{T}] = [9.078, 11.86]\) and...
(9), shows the relation between values of $\beta_Y/\beta_T$ and $Y/T$ as the dashed line for the initial analysis. At points where the signaling curve is above the production curve $\bar{Y}/\bar{T} > \dot{Y}/\dot{T}$ and flow production is insufficient to sustain the ratio on the signaling curve, so the law of motion is downward. Analogously, the law of motion is upward where the solid line is below the dashed line.

The signaling curve intersects with the initial production curve in three places. Since in the steady-steady state $Y/T = \bar{Y}/\bar{T} = Y^*/T^*$, these intersections give steady-state values. The arrows drawn on the production curve give the laws of motion based on the relative positions of the signaling and production curves. The outer two intersections are stable and the interior solution is unstable.

We see that multiple stable equilibria chosen by initial conditions are possible. Note that at the rightmost equilibrium the ratio of $Y$ to $T$ is higher than at the leftmost equilibria. This establishes the first basic claim of the paper: that which field comes to dominate may be determined by accidents of history through path dependence.

---

the unstable steady-state moves to lower values of $Y$, $[\beta_Y/\beta_T, Y^*/T^*] = [0.537,0.512]$. For $\lambda_T = 1.04$ (not shown in Figure 1), there is a single, stable, steady-state at $[\beta_Y/\beta_T, Y^*/T^*] = [11.055,15.081]$. 
A first comparative static result is shown with the dotted production line in Figure 1, drawn for a lower value of $\lambda_T$ relative to $\lambda_Y$. Both stable equilibria move to higher values of $Y$ relative to $T$ and the space in which the model dynamically converges to a high $Y$ value expands, establishing that reduced testability reduces equilibrium production in a field. Further, unlike the symmetric equilibrium using the dashed line, in this example the total value of research is higher at the rightmost equilibrium, where $Y^* > T^*$, than at the leftmost equilibrium.

At yet lower values of $\lambda_T$ the production curve moves up further, eventually reaching a position in which the left crossing with the signaling equation pictured in Figure 1 disappears. At sufficiently low values of $T$ testability the multiple equilibria disappear and there is a single, stable steady-state with relatively high values of $Y$ relative to $T$. 

**Figure 1 Steady-states**
A second comparative static result is that a field which is inherently more valuable, where \( p \) is high, has an increased share at both stable equilibria. This follows immediately from the symmetry of \( p \) and \( \lambda \) in equation (9), in other words, the dotted line can represent either lower \( \lambda_T \) or lower \( p_T \).

Turn now to the longer-run question of how the steady-state mix of subdisciplines affects the overall resources available to the discipline. In other words, what is the result of the interaction of the allocation within the discipline for a given \( w \) and the value of \( w \) determined by productivity of the discipline in equation (8)? The idea here is that social resource allocation moves slowly compared to developments within a discipline. Departments are typically reviewed by their universities only at periods of a number of years. Funding agencies only gradually move funds from one discipline to another. So with hopefully only modest abuse of the language, I use “steady-state” to refer to “long-run” equilibrium within the discipline and now analyze the “longer-run” equilibrium in which \( w \) is determined.

The choice of functional forms helps separate the analysis into the results on steady-states, given above and shown in Figure 1, from the interaction with resourcing. Note that the steady-state subdiscipline ratios depend only on the ratio \( \pi_Y/\pi_T \), and that because \( w \) does not affect this ratio, changes to \( w \) do not affect the steady-state solutions above.

While \( w \) has no effect on \( Y^*/T^* \), \( w \) does affect total production in the expected way. Note in equation (7) that \( w \) enters only once, through the left-most appearance of \( \pi_Y \) and that \( \frac{\partial Y}{\partial w} > 0 \). Since \( T \) rises proportionally with \( Y \), higher \( w \) means higher production in both subdisciplines, as shown in Figure 2.
Figure 2 Longer-run equilibrium and external allocation of resources

In the left panel of Figure 2 the solid line gives the steady-state value of $p_Y Y^* + p_T T^*$ for a set of parameters chosen such that $Y^* > T^*$. To understand the characteristic shape of the steady-state curve, note from equation (6) that as $w \to 0$, $C \to K$ and as $w \to \infty$, $C \to 0$. The former means that at sufficiently small $w$, all resources move away from research into $C$ so research output goes to zero. The latter means that as $w$ increases production asymptotes to a value consistent with all resources going to research.

The dotted line depicts $f(p_Y Y^* + p_T T^*)$ as linear in the region of interest. The thought experiment is that in the long-run, production is on the solid line at a historically determined value of $w$. Occasionally, the discipline is re-evaluated and $w$ is reset to a value indicated by the dotted line. Above the dotted line, the value of production is higher than required to sustain
the existing level of \( w \), so the law of motion moves \( w \) to the right (and to the left below the line). As drawn, there are two longer-run equilibria. The circle at the upper right is a stable longer-run equilibrium.

In contrast, the circle at the lower left identifies an unstable equilibria. If the level of disciplinary output is below this level, the corresponding value of \( w \) leads to a yet lower level of output. Over a sufficiently long period the discipline would collapse. This last result is in part a result of the pattern of substitution implied by the functional forms chosen where sufficiently low values of \( w \) private consumption dominates doing research. This last result is also dependent on the suggestion that marginal compensation resources respond strongly to the success of the discipline. For example, if \( w \) were constant the dotted line would be vertical and there would be only a single, stable equilibrium in the Figure.

The dashed line in the left panel of Figure 2 plots steady-state values of discipline research for the case in which \( T^* > Y^* \) for the same underlying parameters values as in the solid line. For any given value of \( w \) the discipline is less productive than if \( Y^* > T^* \). This reiterates the result from Figure 1. However, we now have a further effect. Because total research is less, resources are less than in the case where \( Y \) is dominant. This leads to lower \( w \), lowering the stable equilibrium. Thus we have a further reason why low verifiability hurts the discipline as a whole.

The right panel of Figure 2 reproduces the left panel and adds curves for both \( Y^* > T^* \) and \( T^* > Y^* \) steady-states based on a lower value of \( \lambda_T \). As expected, this increases longer-run research in equilibria where \( Y \) is the dominant subdiscipline. Consider, however, what happens when \( T \) is the dominant discipline given the judiciously drawn position of the resource curve \( f(\cdot) \). There is now no stable longer-run equilibrium. The comparative static reduction in
testability of the $T$ field leads to an eventual increase in resources for the discipline if history has led the $Y$ fields to be dominant. But if the lower testability $T$ field dominates than the entire discipline enters a slow death-spiral.

Turn now to the question of the relation between talent in a discipline and whether the discipline survives. The higher the level of resources, $K$, the higher the solid steady-state curves in Figure 2. So a discipline inhabited by especially smart people expands the attraction basin of the stable equilibrium. Suppose, as pictured in the right panel of Figure 2, that for a given low value of $K'$ a $Y$-dominant discipline would survive but a $T'$-dominant discipline would eventually disappear. For a sufficiently more talented discipline, both kinds of steady-states will be stable. In a world with many disciplines with varying levels of talent, $T'$-dominant disciplines will survive more often when the overall talent level is high. Despite the fact that the level of talent in a discipline has no effect on the allocation of talent between fields, selection by survival may lead to a distribution in which disciplines dominated by a low-merit field have more talented than average researchers.

IV. Conclusion

Casual empiricism among academics suggests that disciplines sometimes go astray in the sense that they become dominated by fields that are not very productive. One reason this can happen is essentially as an accident of history—work in a “fringe” field base has a harder time being rewarded than does mainstream work. A second reason a discipline may go astray is if the dominant field is one in which research is not easily subject to test and verification, either due to the nature of the field or because of social norms which do not emphasize testability. If
the resources available to a discipline are fixed, then a discipline dominated by a weak field may persist in a stable equilibrium. If a resources given to a discipline respond strongly to the productivity of a discipline, then a discipline dominated by a weak field may disappear when it would not have if it had been dominated by a strong field. The last phenomenon opens the possibility that surviving disciplines dominated by fields that do not emphasize testability may have more talented researchers despite being of less social value than more productive fields.

The model here does not include strategic behavior, in particular the values of $\lambda$ are taken to be exogenous. Nonetheless, the obvious implication is that disciplines are more likely to thrive when they can increase the requirement that research contributions are evaluated on evidence as to their addition to social welfare.


