A theory of citations

Wojciech Olszewski[*]

October 2018

Abstract

We propose a model in which researchers maximize the number of times they are cited in later papers. The equilibrium is inefficient, because researchers distort their effort toward writing on popular topics. This inefficiency is affected by various factors, policies and customs. We explored the effect of a variety of such factors. In particular, we argue that the inefficiency is likely to be higher in disciplines (areas of research) in which talent is uniform across topics rather than more topic specific. We also determine conditions under which assigning a higher weight to citations in papers published in higher-ranked journals enhances efficiency.

1 Introduction

Citations have always played an important role in making assessments in academia. Papers with an extremely high number of citations have always been admired in some ways, and researchers have used citation frequency in their evaluations of other researchers. In economics, there seems to have been a recent surge of interest in using citations to assess researchers, academic journals, and various research institutions. Academic journals have increased their emphasis on citation indices. References to citations have become more common in peer evaluations. This trend may be related to researchers becoming more specialized, and finding it more difficult to evaluate research which does not belong to their research area, or it may simply be related to a need to search for more objective evaluation measures. In addition, Google Scholar delivered an easily available, and quickly updatable citation count.

Of course, the use of citations for assessing researchers provides them with strategic incentives for seeking projects that generate citations. The existing research on citations does not explore such strategic incentives, but is focused instead on measurement issues. Various citation indexes have been proposed as measures of

[*] Department of Economics, Northwestern University, Evanston, IL 60208 (e-mail: wo@northwestern.edu). I thank Simon Board, Aisleen Bohren, Natalia Fabra, Matt Jackson, Phil Reny, Joel Sobel, and the audiences at the Baptist University of Hong Kong, the City University of Hong Kong, Hong Kong University, the National University of Seoul, the National University of Singapore, Northwestern University, the State University of New York at Stony Brook, and the Canadian Economic Theory Conference in Toronto for very helpful comments, suggestions, and discussion.

scientific influence, among which the h-index (Hirsch, 2005) is probably the index most commonly referred to. In a recent article, Perry and Reny (2016) propose their own index, and provide a summary of the literature on citation indexes. Other papers address the question of whether citation indexes can provide a "compelling" assessment of scientists and academic departments (see Ellison, 2010 and 2012), or how to use indirect citations to obtain a more accurate assessment (see Chung et al., 2017).

We are broadly interested in how the importance of citations in evaluating researchers affects research. In the present model, we explore the following basic trade-off in the context of citations: Researchers choose topics on which they write their papers. Researchers face individual incentives coming from a higher number of future citations, because this enables them to compare favorably to other researchers. This makes their incentives misaligned with what is socially optimal. So, they are willing to sacrifice some social value to write papers that they expect to be cited more frequently. This creates a spillover effect, and more researchers with misaligned incentives sacrifice even more social value to write papers on popular topics. The resulting outcome is socially inefficient. The size of this inefficiency depends on various factors, such as the relative strength of the incentives coming from citations compared to the incentives coming from social value, the distribution of researchers' talent, or how prone a field is to trends. We provide comparative statics with respect to such factors.

All our results follow from the simple trade-off between enduring quality (which we view as a synonym for social value) and strategic citation-seeking behavior caused by misaligned incentives. We believe that this trade-off should be present in any model of citation-driven research. We elaborate on this point in Section 3.1, after we present our model. However, a plethora of other motives—such as the desire to pioneer research in unexplored areas, or motives not related directly to citation counts—may be important for answering particular questions. In that light, many of our results are best viewed as thought-provoking, rather than as intended to offer definite answers to the questions they address. Despite this caveat, our basic model provides numerous insights regarding practical issues.

One can, for example, wonder whether assigning a higher weight to citations in papers published in higher-ranked journals would enhance efficiency. Our model suggests that it would do so in fields in which higher-quality researchers execute ideas which they find more socially valuable, and lower-quality researchers "shop" for projects that will generate citations. The reason is that the spillover effect is weaker when citations are weighted, because the average quality of researchers seeking citations is lower than the average quality of all researchers. In addition, the researchers who shop for citations publish (on average) in lower-ranked journals.

We also show that, perhaps surprisingly, higher-quality papers are not necessarily cited more frequently than others, and we argue that disciplines (areas of research) in which talent is uniform across topics rather than more topic specific tend to be less efficient.

Another set of potentially interesting questions concerns research dynamics. For example, one may wonder how the flow of papers on various topics would respond to a revival of interest in certain areas. We show that researchers motivated by citations switch to writing on topics in which breakthroughs are likely to occur (e.g., due to new techniques or technology, or the anticipated availability of new data sets), and they do so as soon as they anticipate such breakthroughs. Paradoxically, the inflow of papers on such topics at the time a breakthrough *begins* to be anticipated may exceed the inflow of papers on such topics at the time the breakthrough actually occurs. Our model suggests that this paradox is likely to be observed when researchers need not sacrifice much quality to write papers that they expect to be cited more frequently. Intuitively, the anticipated inflow of papers on some topic in some period T provides incentives to researchers living in period $T-1$ for writing on this topic. And if there are many researchers affected by these incentives, researchers living in period $T-2$ are provided even stronger incentives for writing on this topic. As a result, stronger incentives are transmitted to earlier periods.

Our model also supports the view that assigning lower weights to citations in papers that appear shortly after a cited paper, and higher weights to citations in papers that appear a longer time after the given paper.

Finally, we emphasize that the researchers in our model seek a higher number of future citations, rather than maximizing the value of some more-involved citation indices. Replacing simple citation counts with such indices is likely to reduce the inefficiency described in our paper, and our results support a number of conjectures along these lines. However, such indices seem less useful, and less used, for many practical decisions regarding researchers.[1]

The rest of the paper is organized as follows. We first present our basic model in Section 2, and discuss its equilibria in Section 3. (Some of this discussion is relegated to the Appendix.) This is followed by the results on comparative statics. The most substantive part of our analysis is contained in Sections 5 and 6, in which we apply our model to answer some specific questions of practical importance. We conclude, discuss some extensions, and suggest some additional, potentially interesting questions in Section 7.

2 Basic Model

A random i.i.d. number of researchers live (are active) in each period. This random number is allowed to have an infinite expected value. A researcher may conduct research (which, in our model, corresponds to writing one paper) on one of two topics: topic X or topic Y. For example, if the population represents macroeconomists, then each one may work on modeling the life-time income process, or study issues related to public debt (among a variety of other possible topics, which for simplicity are excluded from the basic model). Alternatively, if the population represents microtheorists, each of them may work on either mechanism design or repeated games (again, among a variety of other topics).

We assume that the two topics cannot be told apart by the market. That is, the market cannot divide

[1] For example, Eigenfactor.org offers some advanced methods, but their scores are calculated based on the citations received over a five-year period.

the papers for those on X and those on Y, and compare the citations of researchers writing on X only with others writing on X, and separately compare the citations across the researchers writing on Y. In practice, the market can tell topics apart to some extent. However, a very fine classification of topics would be probably impractical, or even impossible.[2]

Each researcher is characterized by a 3-dimensional type. First, she may be *partisan* or *strategic*. The researcher is of the former type with probability $1-p$ and of the latter type with probability p. In addition, she is characterized by the expected social value of her papers $(q_X, q_Y) \in [0, \infty)^2$. That is, if the researcher writes a paper on topic $i = X, Y$, this paper's social value is expected to be q_i. For simplicity, we will refer to q_i as social value, omitting the adjective "expected." In particular, a researcher may have an advantage in writing on topic X ($q_X > q_Y$), or an advantage in writing on topic Y ($q_X < q_Y$). Her type is each researcher's private information. We will from time to time call (q_X, q_Y) the researcher's type, if it follows from the context whether we mean a strategic or partisan researcher. The distribution of vectors (q_X, q_Y) has a Lebesgue-measurable density f, which is commonly known among the researchers.

We assume that a paper on topic $i = X, Y$ by a researcher of type (q_X, q_Y) would (in expectation) be cited q_i times in each later period if all researchers were writing on this topic, but would not be cited at all if all subsequent researchers were writing on the other topic. And proportionally, if a fraction M_X of researchers in a later period writes on X, and a fraction M_Y of researchers in a later period writes on Y, then the paper is expected to be cited $M_X q_X$ and $M_Y q_Y$ times, respectively.

This assumption means that if all researchers were choosing to write papers with a higher social value, citations would the correct, and precise, measure of social value. Notice that the objective of citation indices is usually to measure *impact*. So, in other words, the assumption says that in the absence of strategic considerations, impact and social value would coincide. Of course, one can consider other measures of social value which may not be aligned in this way with citations. Such measures would, however, introduce some exogenous inefficiency to our model, since pursuing social goals would not imply maximizing the number citations, even in the absence of any strategic considerations. In contrast, our basic model will exhibit only endogenous inefficiency coming from strategic considerations.

A partisan researcher is assumed to write on the topic with a higher q, that is, on topic X if $q_X > q_Y$ and on topic Y if $q_X < q_Y$. The choice of topic when $q_X = q_Y$ will be irrelevant. A strategic researcher is assumed to choose the topic which generates a higher expected payoff. Her payoff is given by

$$(1-\delta)\sum_{n=1}^{\infty} \delta^n \tau_n, \tag{1}$$

where δ stands for the common discount rate, and τ_n is the number of times the researcher will be cited n period after she writes her paper. That is, a strategic researcher chooses a topic for her paper based on her

[2] How, for example, would one classify a paper looking for an optimal mechanism in a setting in which agents have some behavioral biases? Would one compare it with other papers on the former or the latter topic? Or would this be a separate class of papers? However, by splitting topics in this manner, one could easily end up with few papers on most topics.

beliefs regarding the fractions of researchers who will be working on each topic in later periods. This payoff function should be interpreted as follows: Researchers are interested in being cited, because a higher number of citations enables them to compare favorably to others, and acquire a larger share of the benefits that are to be shared across all researchers (e.g., salaries, prizes and awards, or prestigious positions in various institutions and organizations).

Notice that in the present model by maximizing the absolute number of citations, researchers also maximize their expected ranking with respect to the number of citations within their cohort (that is, among the researchers living at the same time).

3 Equilibria

3.1 First-best outcome

The social optimum is attained when each researcher writes on the topic that gives her an advantage. This is topic X if $q_X > q_Y$, and topic Y if $q_X < q_Y$; this social optimum would be attained if all researchers were partisans. To see why strategic researchers may have distorted incentives, consider a researcher maximizing (1) who happens to live in a society in which all other researchers are partisan. Which topic would such a researcher choose? Denote by μ_X and μ_Y the measures of sets $\{(q_X, q_Y) \in [0,\infty)^2 : q_X > q_Y\}$ and $\{(q_X, q_Y) \in [0,\infty)^2 : q_X < q_Y\}$, respectively. That is,

$$\mu_X = \int_0^\infty \left(\int_{q_Y}^\infty f(q_X, q_Y) dq_X \right) dq_Y,$$

and

$$\mu_Y = \int_0^\infty \left(\int_{q_X}^\infty f(q_X, q_Y) dq_Y \right) dq_X.$$

If $\mu_X = \mu_Y$, then the strategic researcher would make the same choice as a partisan researcher of the same type (q_X, q_Y). But if $\mu_X > \mu_Y$, then the payoff of a strategic researcher of type (q_X, q_Y) from writing on topic i is $\delta q_i \mu_i$; this means that such a researcher would write on topic Y only when

$$\frac{q_Y}{q_X} > \frac{\mu_X}{\mu_Y} > 1.$$

That is, strategic researchers have incentives distorted toward writing on X. Symmetrically, strategic researchers have incentives distorted toward writing on Y when $\mu_X < \mu_Y$. Throughout the paper, we assume that $\mu_X > \mu_Y$. It will be convenient to normalize μ_X and μ_Y to the values such that $\mu_X + \mu_Y = 1$.

As can now be anticipated, the inefficiency in the equilibria of our model will come from strategic researchers with $q_X < q_Y$ choosing to write on topic X. One might argue that such fads can be socially desirable. Indeed, it may be socially efficient when larger groups of researchers focus on studying a specific topic, e.g., because the topic is of particular practical importance, or doing research that is of interest for larger groups of peers may be much more exciting for researchers. Our interpretation is that this kind of factors are included into the social values, captured by q_X and q_Y, and their distribution f. And the fact that some strategic researchers with $q_X < q_Y$ choose to write on topic X is a result of their individual incentives for "consuming a larger share of the pie" that must be shared among all researchers. Since such individual incentives are always present in practice, we believe that the trade-off which we focus on should be present in any model of citation-driven research.

3.2 Description of equilibria

In the previous subsection, we informally argued that if $\mu_X \neq \mu_Y$, then the efficient behavior is not an equilibrium. The game has, however, an inefficient equilibrium and, for some primitives of the model, even multiple equilibria. We must first formally introduce our equilibrium concept. A strategy prescribes a decision regarding the choice of topic for each type of each researcher, contingent on the past choices of other researchers. This implicitly includes the calendar time of making the decision. Strategies are assumed to be Lebesgue measurable, i.e., the set of types choosing each topic is Lebesgue measurable.

If a strategy is independent of the past choices of other researchers, then we call the strategy *history-independent*. History-independent strategies may, however, depend on calendar time. If, in addition, the strategy is independent of calendar time, then we call the strategy *stationary*.

In equilibrium, the prescribed strategies must give each type of strategic researcher an expected payoff which weakly exceeds the expected payoff from choosing the other topic, given that other researchers make the prescribed decisions. If an equilibrium strategy profile is history-independent or stationary, then we call the equilibrium history-independent or stationary, respectively.

We restrict our attention to studying history-independent equilibria. The reason for this is that in the present model past choices of other players have no direct payoff implications for the players in a continuation game. So, their past choices could serve only as some kind of sunspots. Alternatively by conditioning on the past play, players living in future periods would be providing incentives to players living in earlier periods, despite the fact that the actions of players living in earlier periods have no direct payoff consequences for players living in future periods. History-dependent equilibria may or may not exist, depending on the parameters. We informally describe a history-dependent equilibrium in the Appendix. However, we think that history-dependent equilibria are less reasonable, and thus disregard them in the main text. Equilibria in history-dependent strategies would seem more reasonable in a richer model (not studied in the present paper) in which players' research on a topic depends on previous research on the topic.

We begin the analysis with two examples.

Example 1. In this example, we explore a special case of our model in which only one researcher lives in each period. Suppose that $p = 1/2$, i.e., the chance that each researcher is strategic or partisan is fifty-fifty; and suppose that the density f is equal to $4/3$ on $\{(q_X, q_Y) \in [0,1]^2 : q_X > q_Y\}$, and is equal to $2/3$ on $\{(q_X, q_Y) \in [0,1]^2 : q_X < q_Y\}$. Then, q_i, $i = X, Y$, represents the probability that a researcher who writes on topic i will be cited by a later researcher who writes on the same topic. (Recall that no researcher will ever be cited by later researchers writing on the other topic.)

We will explore only stationary equilibria. Denote by M_X an equilibrium probability that a researcher writes on topic X, and by $M_Y = 1 - M_X$ the equilibrium probability that a researcher writes on topic Y. Due to the presence of partisan researchers, both M_X and M_Y are positive. The types (q_X, q_Y) such that $\delta q_X M_X = \delta q_Y M_Y$, or equivalently, such that

$$q_Y = \frac{M_X}{M_Y} q_X$$

are indifferent between writing on X and writing on Y. The segment of points (q_X, q_Y) that satisfy this equation divides the square $[0,1]^2$ into two parts. All strategic researchers with type (q_X, q_Y) to the right of this segment write on topic X, and all strategic researchers with (q_X, q_Y) to the left write on topic Y.

Case 1 ($M_X/M_Y \leq 1$). In this case,

$$M_X = \frac{1}{2}\frac{4}{3}\frac{1}{2}\frac{M_X}{M_Y} + \frac{1}{2}\frac{4}{3}\frac{1}{2}.$$

The first component of the right-hand side represents the probability that a researcher is strategic and writes on topic X, and the second component represents the probability that a researcher is partisan and writes on topic X.

Since $M_Y = 1 - M_X$, this equation says that

$$M_X = \frac{1}{3}\frac{M_X}{1 - M_X} + \frac{1}{3}.$$

It is easy to verify that this quadratic equation has no solution. Therefore, there is no equilibrium in this case.

Case 2 ($M_X/M_Y > 1$). In this case,

$$M_Y = \frac{1}{2}\frac{2}{3}\frac{1}{2}\frac{M_Y}{M_X} + \frac{1}{2}\frac{2}{3}\frac{1}{2}.$$

Since $M_X = 1 - M_Y$, this equation says that

$$M_Y = \frac{1}{6}\frac{M_Y}{1 - M_Y} + \frac{1}{6}.$$

It is easy to verify that this quadratic equation has this unique solution:

$$M_Y = \frac{3 - \sqrt[2]{3}}{6}.$$

Therefore, there is a unique stationary equilibrium, up to the choices of types such that $q_X M_X = q_Y M_Y$, which have measure zero. This equilibrium is depicted in Figure 1.

In the equilibrium, the types below the line with slope $M_X/M_Y = 2 + \sqrt[2]{3}$ write on topic X, and types above that line write on topic Y. The area between this line and the line with slope 1 represents inefficiency, or more precisely, the types making inefficient decisions. These types have an advantage in writing on Y but for strategic reasons write on X.

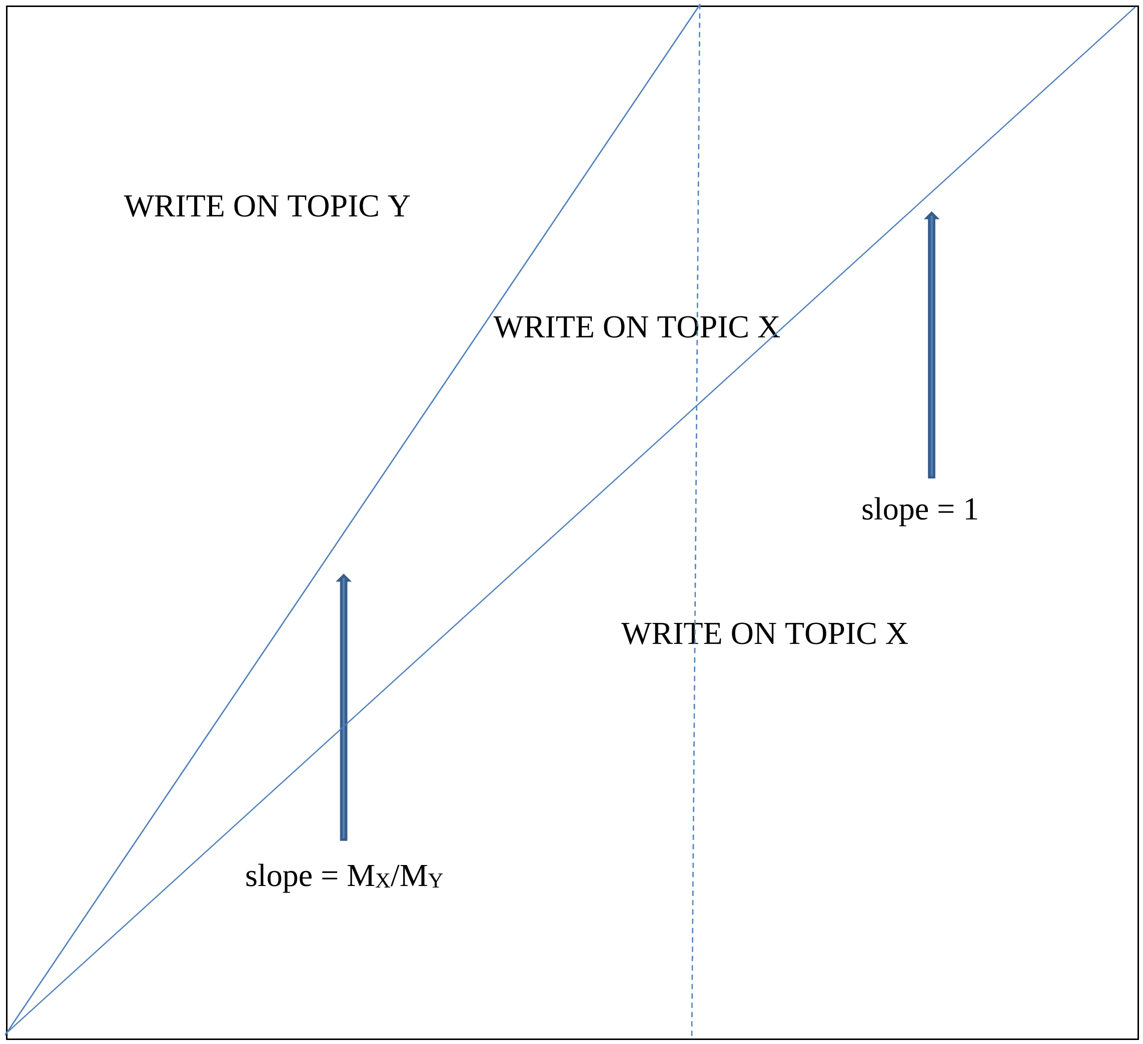

Figure 1. The unique stationary equilibrium strategies for the uniformly distributed types

Example 2. Suppose again that $p = 1/2$, i.e., half of the researchers are partisan, while the other half are strategic. Suppose also that the density f is equal to e^{-x-y} multiplied by 4/3 on $\{(q_X, q_Y) \in [0,1]^2 : q_X > q_Y\}$, and multiplied by 2/3 on $\{(q_X, q_Y) \in [0,1]^2 : q_X < q_Y\}$. Thus, $\mu_X = 2/3$ and $\mu_Y = 1/3$.

We will again explore only stationary equilibria. As in Example 1, denote by M_X an equilibrium fraction of researchers who write on topic X, and denote by M_Y the remaining fraction of researchers, who write on topic Y. The indifferent types (q_X, q_Y) are such that $\delta q_X M_X = \delta q_Y M_Y$, or, equivalently, such that

$$q_Y = \frac{M_X}{M_Y} q_X.$$

The line of such points (q_X, q_Y) divides the quadrant $[0, \infty)^2$ into two parts. All strategic researchers with (q_X, q_Y) below this segment write on topic X, and all strategic researchers with (q_X, q_Y) above it write on topic Y.

Consider the following two cases:

Case 1 ($M_X/M_Y > 1$). In this case, we have that

$$M_Y = \frac{1}{2}\frac{2}{3}\int_0^\infty \left(\int_0^{y\frac{M_Y}{M_X}} e^{-x} dx\right) e^{-y} dy + \frac{1}{2}\frac{2}{3}\frac{1}{2}.$$

The first component of the right-hand side represents the probability that a researcher is strategic and writes on topic Y, and the second component represents the probability that a researcher is partisan and writes on topic Y.

Integrating, we obtain the following formula:

$$\int_0^\infty \left(\int_0^{y\frac{M_Y}{M_X}} e^{-x} dx\right) e^{-y} dy = M_Y,$$

so that

$$M_Y = \frac{1}{3} M_Y + \frac{1}{6} = \frac{1}{4},$$

which is consistent with $M_X/M_Y > 1$.

Case 2 ($M_X/M_Y \leq 1$). In this case,

$$M_X = \frac{1}{2}\frac{4}{3}\int_0^\infty \left(\int_0^{x\frac{M_X}{M_Y}} e^{-y} dy\right) e^{-x} dx + \frac{1}{2}\frac{4}{3}\frac{1}{2},$$

which yields

$$M_X = \frac{2}{3} M_X + \frac{1}{3} = 1,$$

which contradicts the assumption that $M_X/M_Y \leq 1$.

Therefore, the model has a unique stationary equilibrium (up to a type set of measure zero), in which 75 percent of researchers work on topic X, and the remaining 25 percent work on topic Y. In contrast, in the efficient outcome, only 2/3 of researchers would work on topic X, and the remaining 1/3 would work on topic

Y. The inefficient decisions are made by strategic researchers with types (q_X, q_Y) between lines $q_Y = q_X$ and $q_Y = 3q_X$.

In order to characterize equilibria in the general case, it will be convenient to first introduce some notation. Given a history-independent strategy profile, let M_i^t be the expected fraction of researchers who choose topic $i = X, Y$ in period t. This number is independent of the history of play, because we are studying history-independent strategies. When strategies are also stationary, then $M_i^t = M_i$ is constant-over-time. Given any (history-independent, but not necessarily stationary) strategies of researchers living in periods $t, t+1, ...$, let

$$\mathbf{M}_i^t = (1-\delta)\sum_{n=t}^{\infty} \delta^{n-t} M_i^n.$$

This value is often called in the literature the "occupation measure" - in this case - of topic i from time t on, and it represents the weighted average of M_i^n over periods $n = t, t+1, ...$ If the strategies are stationary, then $\mathbf{M}_i^t = M_i$. Finally, let

$$g_X(M_X) := p \Pr\{(q_X, q_Y) : q_X M_X > q_Y(1 - M_X)\} + (1-p)\mu_X$$

and

$$g_Y(M_Y) := p \Pr\{(q_X, q_Y) : q_Y M_Y > q_X(1 - M_Y)\} + (1-p)\mu_Y.$$

Notice that functions g_X and g_Y are fully determined by the distribution of researchers' types. The value of $g_i(M_i)$, $i = X, Y$, represents the expected fraction of researchers writing on topic i in the current period if, on average, this fraction in future periods is expected to be M_i.

The settings from Examples 1 and 2 had unique stationary equilibria. Below, we provide a condition under which the model has a unique (history-independent) equilibrium, and this equilibrium is stationary. We will assume this condition through the main text. In the Appendix, we will characterize equilibria in the general case when the condition may not satisfied.

Assumption 1. Function g_X has a unique fixed point.[3]

Notice that Assumption 1 is a condition on the primitives of our model, since function g_X is expressed in terms of our primitives. Notice also that $g_i(M_i)$ always has a fixed point M_i. Indeed, function g_X is continuous (because the distribution of types (q_X, q_Y) is continuous), and $g_X(0) > 0$ and $g_X(1) < 1$ (because of the presence of partisan researchers).

For some distributions of researchers' types, g_X has multiple fixed points. But for many "regular" distributions of interest - including uniform and exponential ones (as we have seen in Examples 1 and 2) or normal ones truncated to the quadrant $q_X, q_Y \geq 0$ - the function g_X has a unique fixed point. In contrast, the densities f of distributions with multiple fixed points must have multiple "ups and downs." This is our motivation for making Assumption 1 throughout the main text.

[3] Since $M_X + M_Y = 1$, it follows that $g_Y(M_Y) + g_X(M_X) = 1$. So, this is equivalent to assuming that function g_Y has a unique fixed point.

Our characterization of history-independent equilibria from the Appendix shows that M_X must converge over time to a fixed point of g_X, in every history-independent equilibrium. That is, M_X is approximately a fixed point of g_X in the long run in all (history-independent) equilibria. However, if there are multiple fixed points, g_X crosses the diagonal from above to below at for some of them, and crosses the diagonal from below to above at others. This clearly affects some comparative statics.

The following proposition characterizes the equilibria of our basic model under Assumption 1. Its proof is relegated to the Appendix.

Proposition 1 *The model has a unique equilibrium. This equilibrium is stationary. In the equilibrium, M_i is the fixed point of g_i. A researcher chooses topic X when $M_X q_X > M_Y q_Y$, and chooses topic Y when $M_Y q_Y > M_X q_X$. When $M_X q_X = M_Y q_Y$ (an event of probability 0), a researcher is indifferent between the two topics.*

4 Comparative statics

Before we address more substantial questions, we perform some comparative statics. We begin with studying the effects of an increase in the fraction of partisan researchers in the population, and of an increase in the asymmetry across topics. The former increase is modeled by decreasing p. In the latter case, we wish to increase the ratio of μ_X to μ_Y without introducing any changes in the relative density across types such that $q_Y > q_X$, or across types such that $q_Y < q_X$. Therefore, the latter increase will be modeled by multiplying the density f on the set $\{(q_X, q_Y) : q_Y > q_X\}$ by an $\alpha < 1$, and multiplying the density f on the set $\{(q_X, q_Y) : q_Y < q_X\}$ by $\beta > 1$ such that $\alpha\mu_Y + \beta\mu_X = 1$.

Proposition 2 *(i) If $p'' < p'$, then $M''_Y > M'_Y$. (ii) For any $\alpha < 1$, $M^{\alpha}_Y/\mu^{\alpha}_Y < M_Y/\mu_Y$; in particular, $M^{\alpha}_Y < M_Y$.*

Proposition 2 (i) implies that an increase in the fraction of partisan researchers in the population leads to a more efficient outcome. This happens for two reasons. The first is due to the direct effect: the fraction of partisan researchers is higher, and such researchers make efficient decisions. The second reason is due to the strategic effect: a smaller set of strategic researchers whose advantage is in writing on topic Y strategize by writing on topic X instead.

By Proposition 2 (ii), an increase in the asymmetry across topics reduces the set of types of strategic researchers whose advantage is in writing topic Y who do write on topic Y. Proposition 2 (ii) shows even more: the size of this set decreases even as a fraction the set of all researchers with an advantage in topic Y, though the size of this latter set also decreases. This strategic effect reduces efficiency. But the direct effect of shrinking the fraction of researchers with advantage in topic Y in the population enhances efficiency. So, the total effect on aggregate efficiency is ambiguous. For example, the strength of the direct effect depends on the distribution of types who originally strategize (that is, the distribution f contingent on the set $\{(q_X, q_Y) : q_X < q_Y < M_X q_X / M_Y\}$), whereas the strength of the strategic effect is independent of this distribution.

Proof. It follows from equation $g_Y(M_Y) = M_Y$ and Assumption 1 that the graph of g_Y intersects the diagonal from above to below at $M_Y < 1/2$. When p decreases, the graph of g_Y for $M_Y < 1/2$ moves up, because $\Pr\{(q_X, q_Y) : q_Y M_Y > q_X(1 - M_Y)\} < \mu_Y$ for $M_Y < 1/2$. So, the unique fixed point M_Y of g_Y becomes larger.

When the density f on the set $\{(q_X, q_Y) : q_Y > q_X\}$ is multiplied by an $\alpha < 1$, the graph of g_Y moves down for all M_Y. So, the unique fixed point M_Y of g_Y becomes smaller. Moreover, the values of g_Y for all $M_Y \leq 1/2$ are scaled down by α. Since $g_Y(1/2) = \mu_Y$, $\alpha g_Y(1/2) = \mu_Y^\alpha$, and $M_Y = g_Y(M_Y)$, we would have $M_Y^\alpha / \mu_Y^\alpha = M_Y / \mu_Y$ if M_Y^α were equal to $\alpha g_Y(M_Y)$. However, $M_Y^\alpha < M_Y$, so $M_Y^\alpha = \alpha g_Y(M_Y^\alpha) < \alpha g_Y(M_Y)$, which implies that $M_Y^\alpha / \mu_Y^\alpha < M_Y / \mu_Y$. ■

4.1 Uniform talent across topics versus topic-specific talent

An interesting comparative statics question is whether fields in which talent is uniform across topics are more efficient than fields in which talent is rather topic specific. One can also interpret this question in terms of entry costs.

That is, in some fields of research, it takes time until researchers learn enough to write papers; this typically occurs in more mature fields, which have also become more specialized over time. In such areas, the fact that researchers almost invariably write on more specialized topics could be interpreted to mean that their talent is more topic-specific.

We model fields in which talent is less uniform across topics by moving the mass of (q_X, q_Y) away from the diagonal. More precisely, for any pair of random variables (q'_X, q'_Y) and (q''_X, q''_Y) that take values in $[0, \infty)^2$, we say that (q''_X, q''_Y) is obtained from (q'_X, q'_Y) by *moving mass away from the diagonal* if $(q''_X, q''_Y) = (q'_X, q'_Y) + (\varepsilon_X, \varepsilon_Y)$, where $(\varepsilon_X, \varepsilon_Y)$ is a random variable, which depends on the realization of (q'_X, q'_Y), and which has the following properties, which are illustrated in Figure 2:

(i) When the realization of (q'_X, q'_Y) is such that $q'_X > q'_Y$, then ε_X takes only nonnegative values and ε_Y takes only nonpositive values; and

(ii) when the realization of (q'_X, q'_Y) is such that $q'_X < q'_Y$, then ε_X takes only nonpositive values and ε_Y takes only nonnegative values.

Although the distribution of $(\varepsilon_X, \varepsilon_Y)$ depends on the realization of (q'_X, q'_Y), we will omit this relation in our notation.[4]

[4] Actually, as Satoru Takahashi has been pointed out to the author, the following proposition requires a weaker assumption: it is sufficient to assume that $(q''_X, q''_Y) = (q'_X, q'_Y) + (\varepsilon_X, \varepsilon_Y)$ where $(\varepsilon_X, \varepsilon_Y)$ is such that (a) when $q'_X/q'_Y > 1$, then $q''_X/q''_Y > q'_X/q'_Y$; and (b) when $q'_X/q'_Y < 1$, then $q''_X/q''_Y < q'_X/q'_Y$.

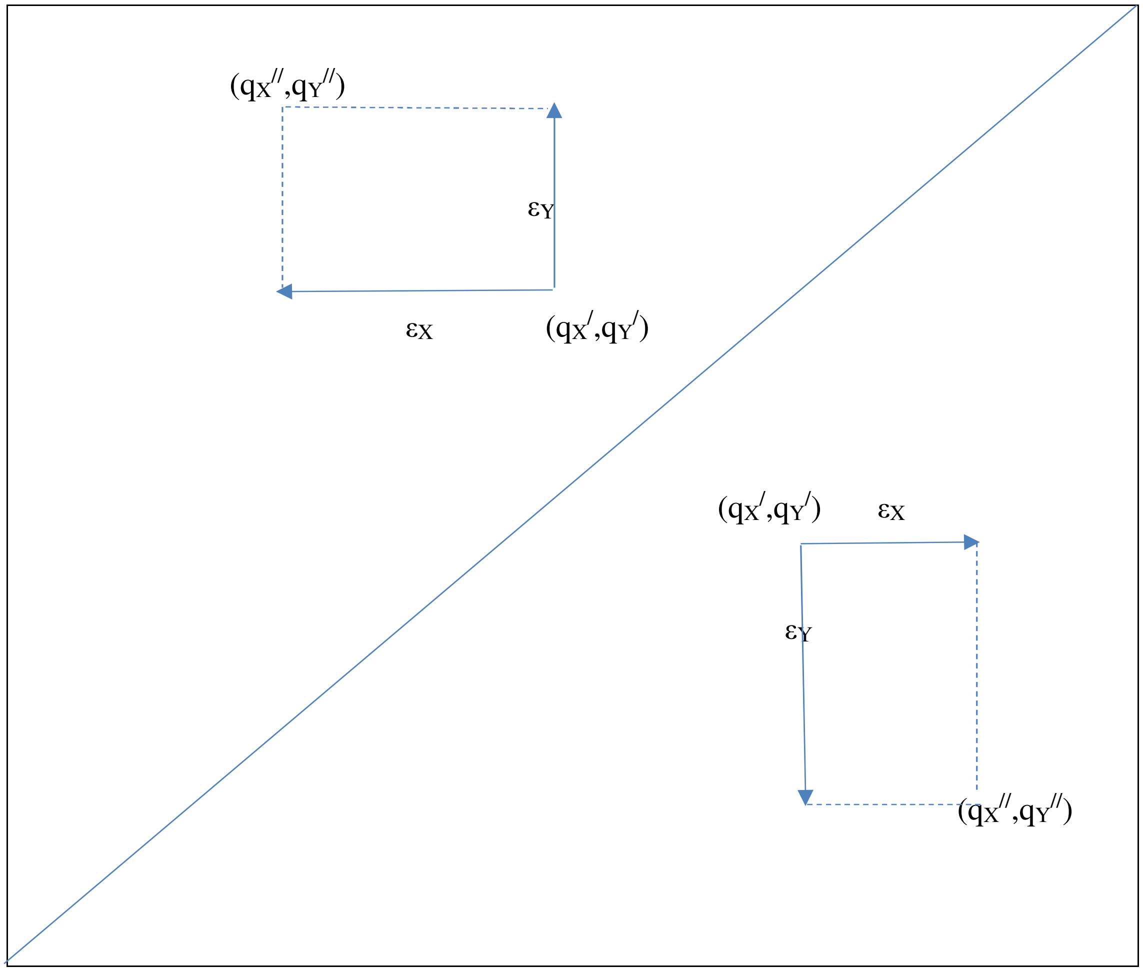

Figure 2. Moving mass away from the diagonal

This notion is a two-dimensional version of the simple mean-preserving spread in Diamond and Stiglitz (1974).[5] Their concept, defined for one-dimensional random variables with cdf's H and G, requires that $G(t) \leq H(t)$ for all $t \geq t^*$ and $H(t) \leq G(t)$ for all $t \leq t^*$ for some t^*. It can be shown that (q''_X, q''_Y) is obtained from (q'_X, q'_Y) by moving mass away from the diagonal if and only if the probability of any set of the form $\{(q''_X, q''_Y) : q''_X > q^*_X \text{ and } q''_Y < q^*_Y\}$, where $q^*_X > q^*_Y$, is higher than the probability of the set $\{(q'_X, q'_Y) : q'_X > q^*_X \text{ and } q'_Y < q^*_Y\}$, and the probability of any set of the form $\{(q''_X, q''_Y) : q''_X < q^*_X$ and $q''_Y > q^*_Y\}$, where $q^*_X < q^*_Y$, is higher than the probability of the set $\{(q'_X, q'_Y) : q'_X < q^*_X \text{ and } q'_Y > q^*_Y\}$.

Proposition 3 *If (q''_X, q''_Y) is obtained from (q'_X, q'_Y) by moving mass away from the diagonal, then $M''_Y > M'_Y$.*

In one interpretation, there is more strategizing when talent is uniform across topics than when talent is topic specific. A larger number of researchers makes inefficient decisions. Indeed, the set of types who write on the topic which is not to their advantage is larger in the former case than in the latter case. And if (q'_X, q'_Y) makes an efficient decision, so does $(q''_X, q''_Y) = (q'_X, q'_Y) + (\varepsilon_X, \varepsilon_Y)$. Of course, this strategizing is inefficient. However, the total effect on aggregate efficiency is ambiguous, because moving mass away from the diagonal has a direct negative effect on the efficiency of individual decisions. More specifically, the efficiency loss coming from the types who in both cases write on topic X but have an advantage in writing on topic Y is larger for (q''_X, q''_Y) than for (q'_X, q'_Y).

It is easy to see that the total effect on aggregate efficiency is indeterminate. For example, the total effect is negative when "moving away from the diagonal" is nontrivial only within the region of types who strategize. In such a case, there is no strategic effect, but the direct effect is negative. Conversely, the total effect is positive when moving away from the diagonal only takes mass away from the region of types who strategize to the region of types who behave as partisan researchers. In such a case, both the direct effect and the strategic effect are positive.

Proof. Observe that $\Pr\{(q'_X, q'_Y) : q'_Y M_Y > q'_X(1 - M_Y)\} < \Pr\{(q''_X, q''_Y) : q''_Y M_Y > q''_X(1 - M_Y)\}$ for all M_Y. Indeed, for any given M_Y, when we add $(\varepsilon_X, \varepsilon_Y)$ to (q'_X, q'_Y), (i) we may move some mass from the region of types who choose X to the region of types who choose Y; or (ii) we may move some mass within the region of types who choose X. However, we never move any mass from the region of types who choose Y to the region of types who choose X. This argument is illustrated in Figure 3. So, the graph of g_Y moves up for all M_Y when we replace random variable (q'_X, q'_Y) with random variable (q''_X, q''_Y). This yields $M''_Y > M'_Y$. ■

5 Implications

In this section, we discuss some practical implications which follow from our analysis.

[5] Another version of a simple mean-preserving spread, closer in form to the version used in the present paper is used in Klabjan et al. (2014).

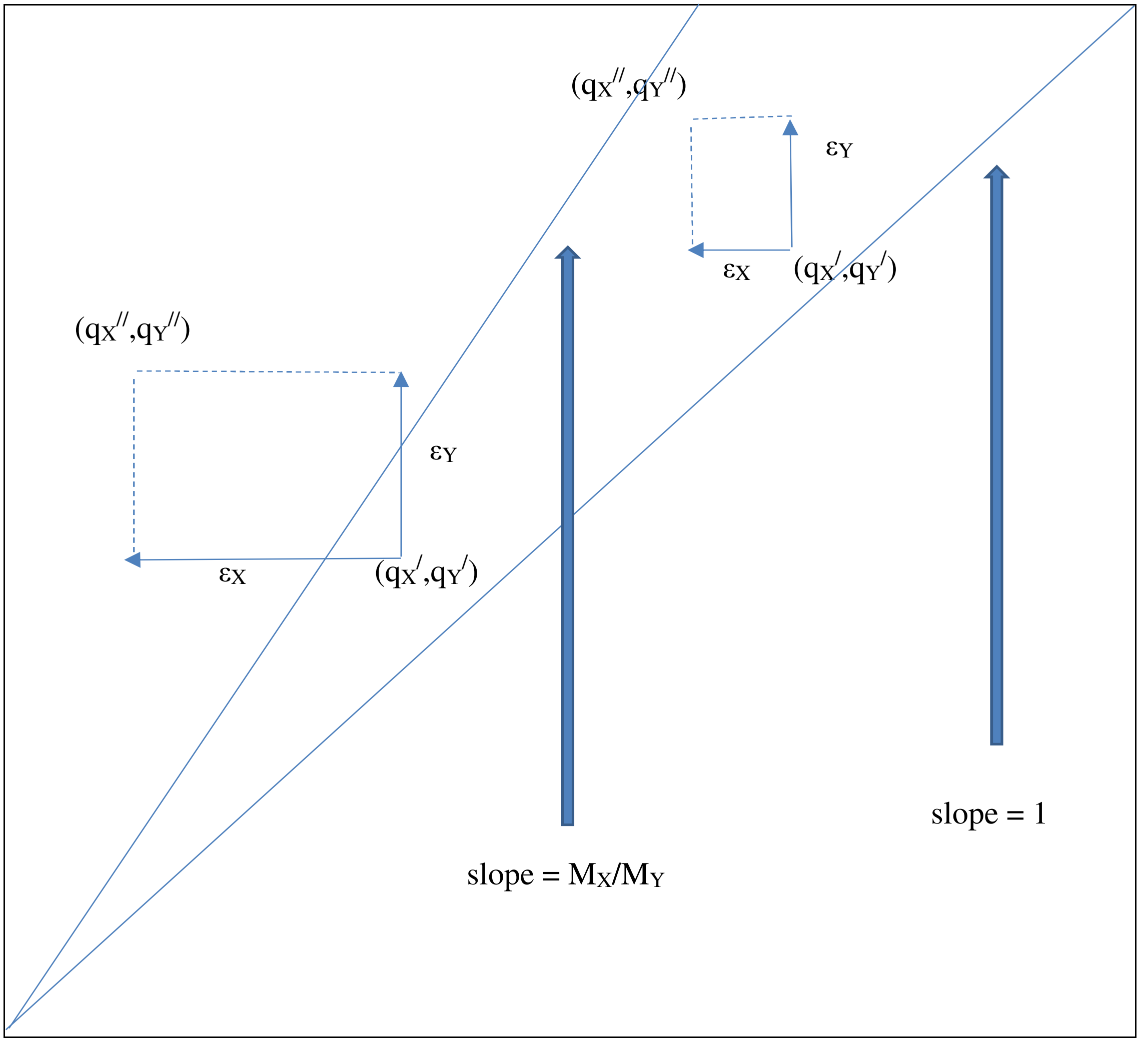

Figure 3. For the lower (q_X',q_Y'), adding $(\varepsilon_X,\varepsilon_Y)$ results in the effect described in (i); and for the upper (q_X',q_Y'), adding $(\varepsilon_X,\varepsilon_Y)$ results in the effect described in (ii).

5.1 Are better papers cited more frequently?

We will consider the papers of higher social value to be "better papers;" these are defined in the model as the papers having a higher q. So by definition, they have an advantage in terms of citations. They are more likely to be cited by subsequent cohorts who write on the same topic. This does not yet mean that they are on average cited more frequently in equilibrium.

First, the fraction of papers of quality q with an advantage in Y may increase in q, which more negatively affects the average number of citations of papers with a higher q, even in the absence of strategic researchers. But even when we assume away this exogenous effect (that is, when we assume that the fraction is constant across all q), there is a strategic effect which may reduce, and even overturn, the direct effect. Specifically, there is an outflow of potential papers of any quality q on topic Y to papers of lower quality on topic X, and an inflow of papers of quality q on topic X from potential papers of higher quality on topic Y. This outflow and this inflow are depicted in Figure 4. They both increase the average number of citations per paper of any quality q, but the strength of the strategic effect may be different for different q's. The way in which the strategic effect varies with q depends on the distribution of types.

To obtain some insight we will assume that the density f is obtained from a density f' that is symmetric across the diagonal, by multiplying f' on the set $\{(q_X, q_Y) : q_Y > q_X\}$ by an $\alpha < 1$, and multiplying f' on the set $\{(q_X, q_Y) : q_Y < q_X\}$ by the β such that $\alpha(1/2) + \beta(1/2) = 1$, that is, $\beta = 2 - \alpha$. This assumption implies that the fraction of papers of quality q with an advantage in Y would be constant across q's in the absence of strategic researchers.

Let $F(y \mid x)$ be the cdf that corresponds to density f, conditional on x. Then, the average number of citations per paper of quality q is given by

$$q \cdot \frac{M_Y \alpha F\left(\frac{M_Y}{M_X} q \mid q\right) + M_X(2-\alpha)F(q \mid q) + M_X \alpha \left[F\left(\frac{M_X}{M_Y} q \mid q\right) - F(q \mid q)\right]}{\alpha F\left(\frac{M_Y}{M_X} q \mid q\right) + (2-\alpha)F(q \mid q) + \alpha \left[F\left(\frac{M_X}{M_Y} q \mid q\right) - F(q \mid q)\right]},$$

where $\alpha F\left(\frac{M_Y}{M_X} q \mid q\right)$ represents the mass of papers of quality q on topic Y, $(2-\alpha)F(q \mid q)$ represents the mass of papers of quality q on topic X written by researchers with an advantage in X, and $\alpha \left[F\left(\frac{M_X}{M_Y} q \mid q\right) - F(q \mid q)\right]$ represents the mass of papers of quality q on topic X written by researchers who would write higher-quality papers on topic Y.

The following condition guarantees that the strategic effect increases by more the number of citations per paper for lower values of q. In other words, the strategic effect works against quality.

Condition 1 *For any $\underline{c} < \overline{c}$, the ratio*

$$\frac{F(\underline{c}q \mid q)}{F(\overline{c}q \mid q)}$$

increases in q.

Condition 1 is implied by the increasing likelihood ratio, and is satisfied by many commonly used distributions, including exponential distributions and normal distributions truncated to the positive quadrant.

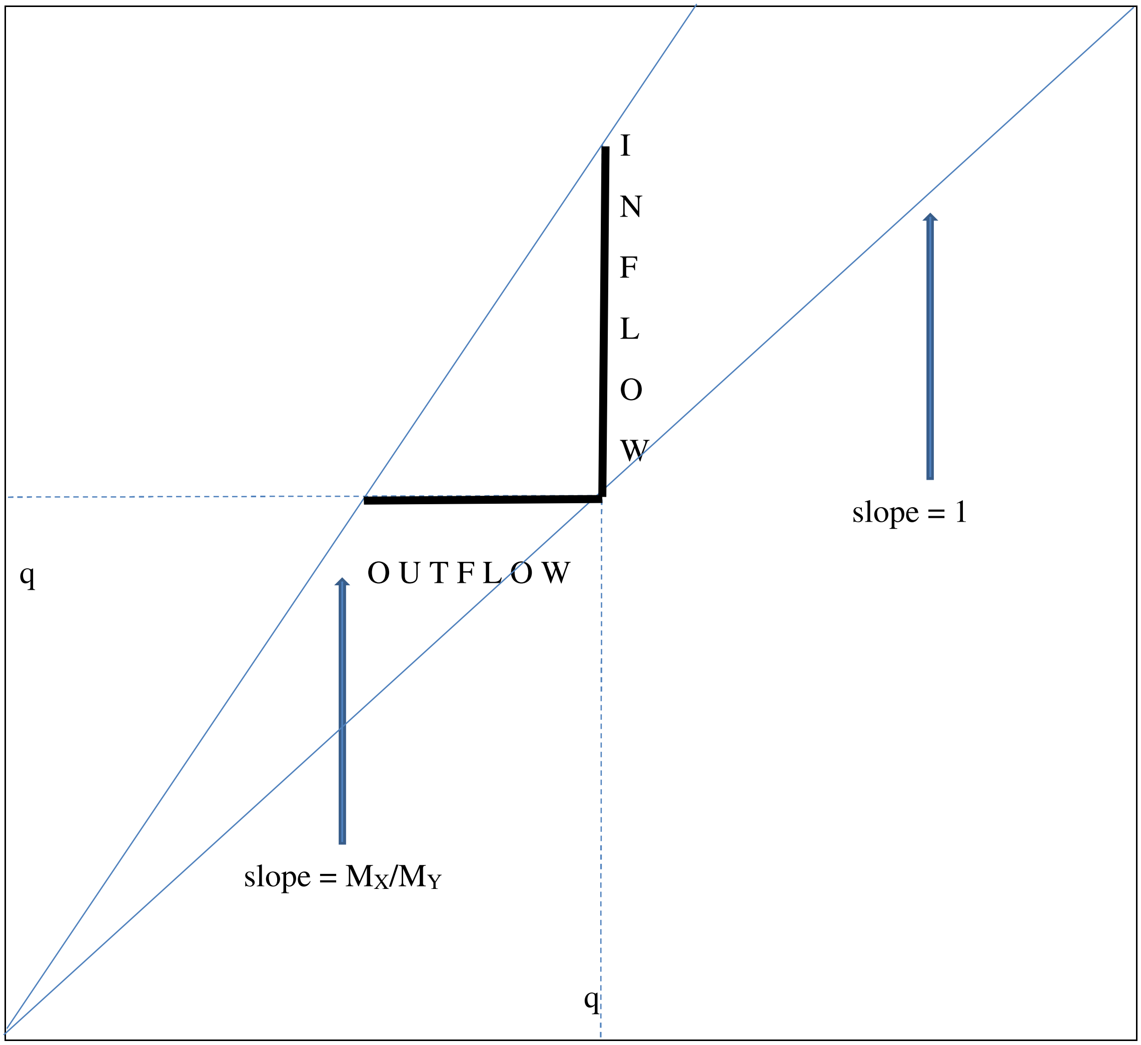

Figure 4. A strategic outflow and inflow of papers of a certain quality

Proposition 4 *Suppose that Condition 1 is satisfied. Then, the fraction of papers on topic X, among all papers of quality q, decreases with q.*

Since a paper of quality q on topic X is cited more frequently than a paper of quality q on topic Y, this result implies that the strategic effect works against quality. For higher values of q, the reduction in the number of citations per paper is greater. Put together with the direct effect that the papers of quality q are cited by subsequent cohorts more frequently, the relation between the average number of citations and the quality of papers is indeterminate. One can verify that the direct effect dominates for exponential distributions and normal distributions, but it is not difficult to construct examples in which higher quality papers have a lower average number of citations.

Proof. The fraction of papers on topic X among all papers of quality q is given by

$$\frac{(2-\alpha)F\left(q \mid q\right)+\alpha\left[F\left(\frac{M_X}{M_Y}q \mid q\right)-F\left(q \mid q\right)\right]}{\alpha F\left(\frac{M_Y}{M_X}q \mid q\right)+(2-\alpha)F\left(q \mid q\right)+\alpha\left[F\left(\frac{M_X}{M_Y}q \mid q\right)-F\left(q \mid q\right)\right]},$$

which is equal to

$$\frac{(2-2\alpha)+\alpha\dfrac{F\left(\frac{M_X}{M_Y}q \mid q\right)}{F\left(q \mid q\right)}}{\alpha\dfrac{F\left(\frac{M_Y}{M_X}q \mid q\right)}{F\left(q \mid q\right)}+(2-2\alpha)+\alpha\dfrac{F\left(\frac{M_X}{M_Y}q \mid q\right)}{F\left(q \mid q\right)}}.$$

Denote $F\left(\frac{M_X}{M_Y}q \mid q\right)/F\left(q \mid q\right)$ by $\phi(q)$, and $F\left(\frac{M_Y}{M_X}q \mid q\right)/F\left(q \mid q\right)$ by $\psi(q)$. Since $M_X > M_Y$, by virtue of Condition 1, ϕ decreases in q, and ψ increases in q. Rewriting the formula for the fraction of papers on topic X among all papers of quality q, we obtain

$$\frac{(2-2\alpha)+\alpha\phi(q)}{\alpha\psi(q)+(2-2\alpha)+\alpha\phi(q)}.$$

The derivative of this function is

$$\frac{\alpha^2\phi'(q)-\alpha(\beta-\alpha)\psi'(q)-\alpha^2\phi(q)\psi'(q)}{\left[\alpha\psi(q)+(\beta-\alpha)+\alpha\phi(q)\right]^2}<0.$$

■

It follows from the proof that Condition 1 plays a double role in our result. It implies that (i) the outflow of papers of quality q on topic Y, and (ii) the inflow of papers of quality q on topic X both work against quality, and each affects more positively papers of lower quality.

5.2 Should citations in higher-ranked journals carry higher weight?

In this, and some other subsequent subsections, we will explore the possibility of somewhat modifying the researchers' payoff, by counting citations in a different manner. We interpret this analysis as a comparison of various customs or standards in academia (or in a particular profession). We refrain, however, from a full-fledged mechanism-design exercise, since our model is too simple for this kind of exercise. Our entire analysis is based on some basic trade-off between quality and popularity, and this trade-off may not be crucial, or even present, if the researchers' payoff depends on other performance measures.

We will no longer assume in this subsection that only researchers know their q_X and q_Y. We will assume instead that if a researcher writes a paper on topic i, her q_i becomes public information. The rationale for making this assumption is that once a paper is disseminated, there is no reason to think that the author has much private information concerning the quality of her paper; typically, at least some other researchers are able to estimate the quality, and they can publicize this information, for example, by means of their peer reviews.

We are interested if the society can benefit, that is, enhance efficiency, by increasing the weight assigned to citations in papers with a higher q, that is, the expected quality of the paper citing a given paper. In one interpretation, the question is whether by increasing the weight of citations in papers published in journals of higher reputation will enhance efficiency. Alternatively, would it be more efficient to refer to sources which seem more selective in counting citations (e.g., Web of Science seems more selective than Google Scholar)[6]?

To address this question, suppose that citations are weighted by a ranking of journals in which the papers which cite them are published.[7] More specifically, we assume that once a paper is disseminated, its (expected) quality q becomes public information, and becomes perfectly reflected by the ranking of the journal in which the paper is published.

We will explore only a special case of our model in which only one researcher lives in each period. Let $w : [0, \infty) \to [0, 1]$ be a strictly increasing function, and redefine the payoff function (1) by assuming that a citation in a paper of quality q contributes $w(q)$ to the payoff. That is, the payoff is now given by

$$(1-\delta)\sum_{n=1}^{\infty}\delta^n \chi_n w(q_n),$$

where q_n is the quality of the paper which appears n periods after the researcher disseminated her paper; and $\chi_n = 1$ if the paper from the n periods after is on the same topic as the paper of the researcher, and $\chi_n = 0$ if the paper from the n periods after is on the other topic. Up to now $w(q)$ was equal to 1 for all q.

Proposition 1 generalizes to the new payoffs as follows: Let M_X be the probability of a researcher writing on X, and M_Y be the probability of researcher writing on Y. Let

$$N_Y = \int_0^1 \left(\int\limits_0^y w(y) f(x,y)dx \right) dy - p \int_0^1 \left(\int\limits_{yM_Y/M_X}^{y} w(y) f(x,y)dx \right) dy \tag{2}$$

[6] Of course, our analysis disregards the important fact that Google Scholar provides an easier available, and faster updatable citation count than Web of Science.

[7] One could also explore alternative, endogenous methods of weighting citations, e.g., by the number of citations of the papers which cite them. We conjecture that this would only magnify the inefficiency compared to weighting by the ranking of journals, because this method would assign to papers on more popular topics a higher weight than that assigned to papers of similar quality on less popular topics.

and

$$N_X = \int_0^1 \left(\int\limits_0^x w(x) f(x,y) dy \right) dx + p \int_0^1 \left(\int\limits_{yM_Y/M_X}^{y} w(x) f(x,y) dx \right) dy \tag{3}$$

be the payoff of researchers writing on X and on Y, respectively, in a period in which the types (q_X, q_Y) above the line $q_Y = M_X q_X / M_Y$ write on Y, and the types below this line write on X.

The first term in both formulas represents the expected payoff when all researchers with an advantage in topic i write on topic i; this payoff is decreased for $i = Y$ and increased for $i = X$ by the second term, which represents the flow of researchers with an advantage in topic Y who choose to write on topic X.

Functions g_i, $i = X, Y$, are now defined by

$$g_i^w(N_i) := p \Pr\{(q_X, q_Y) : q_i N_i > q_j N_j\} + (1-p)\mu_i.$$

We also assume that there are unique values M_i such that $M_i = g_i^w(N_i)$ for N_i, $i = X, Y$, as defined by (2) and (3). This assumption guarantees the uniqueness of the (history-independent) equilibrium. In this equilibrium, the types (q_X, q_Y) above the line $q_Y = M_X q_X / M_Y$ write on Y, and the types below this line write on X.

Weighting citations has three effects: The first effect reflects exogenous differences in the distribution of q's across topics. If, for example, there are many types with a high q_Y but only few types with a high q_X, then weighting citations naturally reduces inefficiency, because it increases the payoff to writing on topic Y, and decreases the payoff to writing on topic X. In some extreme cases, it may even reverse the direction of the flow of researchers, resulting in researchers with an advantage in topic X writing on topic Y. We will not attempt to quantify this effect; instead, we will consider a setting in which this effect does not appear, and describe the two remaining, strategic effects.

Specifically, let f' be a density on $[0,1]^2$ that is symmetric across the diagonal, that is, $f'(x,y) = f'(y,x)$, and let $f(x,y) = \alpha f'(x,y)$ for $y > x$ and let $f(x,y) = \beta f'(x,y)$ for $y < x$, where $\alpha(1/2) + \beta(1/2) = 1$.[8] Consider an auxiliary scenario under which citations are weighted by w, but the papers on topic X of the researchers with advantage in Y are assigned the same weight as would be assigned their papers on topic Y. That is, each such researcher still writes a paper of quality q_X, but being cited in that paper is weighted as it were a paper of quality q_Y. Of course, quality q_Y is not publicly observed. One can temporarily interpret the auxiliary scenario by assuming that papers on X by researchers with an advantage in Y are published in journals of a higher quality than they deserve. However, we will interpret the auxiliary scenario in another, more realistic manner in the context of our results.

[8] Our results require making only a weaker assumption that

$$\int_0^1 \left(\int\limits_0^y w(y) f(x,y) dx \right) dy \Big/ \int_0^1 \left(\int\limits_0^y f(x,y) dx \right) dy =$$

$$= \int_0^1 \left(\int\limits_0^x w(x) f(x,y) dy \right) dx \Big/ \int_0^1 \left(\int\limits_0^x f(x,y) dy \right) dx.$$

We will explore the relation between:

- M^o_X/M^o_Y be the equilibrium ratio of the probability of writing on X to the probability of writing on Y when all citations are counted equally;
- M^w_X/M^w_Y be the equilibrium ratio when citations are counted with weight w;
- M^a_X/M^a_Y be the equilibrium ratio under the auxiliary scenario.

The three superscripts o, w, and a refer to 'original,' 'weighted,' and 'auxiliary,' respectively. We decompose the change from M^o_X/M^o_Y to M^w_X/M^w_Y into (i) the change from M^o_X/M^o_Y to M^a_X/M^a_Y, and (ii) the change from M^a_X/M^a_Y to M^w_X/M^w_Y.

The effect of change (i), from the original scenario to the weighted scenario, depends on how the average quality of types with an advantage in Y who write on X (under the original scenario with equal weights) compares to the quality of all types with an advantage in Y. More precisely, we make the following claim:

Claim 1 *(a) If the distribution of q_Y contingent on types with an advantage in Y who write on X first-order stochastically dominates the distribution of q_Y contingent on types with an advantage in Y, then $M^o_X/M^o_Y \leq M^a_X/M^a_Y$.*

(b) Conversely, if the distribution of q_Y contingent on types with an advantage in Y first-order stochastically dominates the distribution of q_Y contingent on types with an advantage in Y who write on X, then $M^a_X/M^a_Y \leq M^o_X/M^o_Y$.

This claim implies that the auxiliary scenario is less efficient than the original scenario in case (a), and is more efficient in case (b).

Proof. We will prove the first part of the claim. The second part can be proved by analogous arguments. The assumption that the distribution of q_Y contingent on types with an advantage in Y who write on X first-order stochastically dominates the distribution of q_Y contingent on types with an advantage in Y means that

$$\int_0^1 \left(\int\limits_{yM_Y/M_X}^{y} w(y)f(x,y)dx \right) dy \Big/ \int_0^1 \left(\int\limits_{yM_Y/M_X}^{y} f(x,y)dx \right) dy >$$

$$> \int_0^1 \left(\int\limits_0^y w(y)f(x,y)dx \right) dy \Big/ \int_0^1 \left(\int\limits_0^y f(x,y)dx \right) dy.$$

By symmetry,

$$\int_0^1 \left(\int\limits_0^y w(y)f(x,y)dx \right) dy = \int_0^1 \left(\int\limits_0^x w(x)f(x,y)dy \right) dx.$$

This implies that

$$\frac{N_Y^a}{N_X^a} = \frac{\int_0^1 \left(\int\limits_0^y w(y)f(x,y)dx\right) dy - p\int_0^1 \left(\int\limits_{yM_Y/M_X}^y w(y)f(x,y)dx\right) dy}{\int_0^1 \left(\int\limits_0^x w(x)f(x,y)dy\right) dx + p\int_0^1 \left(\int\limits_{yM_Y/M_X}^y w(y)f(x,y)dx\right) dy} <$$

because

$$< \frac{M_Y^o}{M_X^o} = \frac{\int_0^1 \left(\int\limits_0^y f(x,y)dx\right) dy - p\int_0^1 \left(\int\limits_{yM_Y/M_X}^y f(x,y)dx\right) dy}{\int_0^1 \left(\int\limits_0^x f(x,y)dy\right) dx + p\int_0^1 \left(\int\limits_{yM_Y/M_X}^y f(x,y)dx\right) dy}.$$

Thus, $g_Y^w(N_Y^a) < g_Y(M_Y^o) = M_Y^o$ and $g_X^w(N_X^a) > g_X(M_X^o) = M_X^o$; this in turn implies that $M_Y^a < M_Y^o$ and $M_X^a > M_X^o$. ■

The effect of change (ii), from the auxiliary scenario to the weighted scenario, is always positive in terms of efficiency.

Claim 2 $M_X^w/M_Y^w \leq M_X^a/M_Y^a$

This is so simply because the quality of papers of researchers with an advantage in Y who write on X is lower compared to what it would be if they were writing on Y.

The intuition for the two effects can be explained as follows. In our model, more researchers have, by assumption, an advantage in X than an advantage in Y. This provides incentives for strategic researchers with an advantage in Y to write on X; as a consequence, the researchers whose advantage in Y over X is not too high write on X. This in turn magnifies the incentives for writing on X, and the researchers with an even higher advantage in Y write on X. This process must stop, however, since the researchers with a sufficiently high advantage in Y will never write on X. This is a consequence of the presence of partisan researchers.

The incentives for writing on X are magnified more for a strictly increasing function w than for a constant w if the quality of researchers with an advantage on Y but writing on X is high compared to the population, and conversely the incentives are magnified less if the quality of researchers with an advantage on Y but writing on X is low compared to the population. This is the first of the two effects, described above in (i). In addition, the incentives for writing on X are magnified less with a strictly increasing function w, because switching topics burns some value in terms of citations compared to a constant w; or in other words, the positive externality imposed on the researchers writing on X by those who switch from Y to X is diminished. This is the second of the two effects, described above in (ii).

This discussion is summarized by the following proposition.

Proposition 5 *(a) If the quality q_Y of researchers with an advantage in Y first-order stochastically dominates the quality of researchers with an advantage in Y who write on X under a constant w, then replacing a constant w with a strictly increasing w enhances efficiency.*

(b) If the quality q_Y of researchers with an advantage in Y is first-order stochastically dominated by the quality of researchers with an advantage in Y who write on X under a constant w, then the effect of replacing a constant w with a strictly increasing w is indeterminate.

In particular, we can conclude that in areas in which researchers of relatively low quality are "shopping" for topics, putting a higher weight on citations in higher-quality journals enhances efficiency. This is so because both described effects work in the same direction. In turn, if researchers of relatively high quality are shopping for topics, then the two effects work in the opposite direction, and the total effect of putting a higher weight on citations in higher-quality papers is ambiguous.

How may we know whether shopping for topics is positively or negatively correlated with quality? In the present model, we may never learn about this correlation. However, both in practice and in many richer models, we often receive independent signals of researchers' quality.

One might say that if q_i is revealed, and payoff schemes other than (1) are allowed, then we can restore the efficient outcome by compensating researchers for their publications instead of rewarding their citations. Indeed, if strategic researchers were rewarded in our model according to the quality of journals in which their papers are published, they would make the same decisions as partisan researchers, and the outcome would be fully efficient.

However, the quality of journals seems a better measure at the aggregate level than at the individual level.[9] More importantly, as we have already argued, our basic model is not designed for studying a variety of issues that arise when the researchers' payoff is a function of the ranking of journals in which their papers are published.

Finally, it should be mentioned that in our analysis we disregard the fact that journals are typically keen to publish papers with a higher expected number of future citations, since their ranking depends on the citations of the papers that they publish. This fact affects our analysis quantitatively, but as far as editorial decisions also take into account quality, it seems not to affect the analysis qualitatively.

[9] More specifically, the possibility of publishing in a journal may strongly reflect the taste and views of the current editors, which makes it easier to publish certain kind of papers even when their social value is lower. Over time, those tastes and views are likely to average out (at least to some extent), which makes relying on citations with higher weights assigned to higher-ranked journals more effective than relying on the ranking of the journal in which a paper is published.

5.3 What does the distribution of citations say about efficiency?

Some fields, even within the same discipline, differ substantially in the distribution of citations across papers.[10] In some areas, the top numbers of citations are fairly small but are attained by a larger number of papers; in other areas, many papers are cited seldom but a few papers are cited very frequently. Therefore, it would be useful if we could draw some conclusion regarding efficiency only from the distribution of citations.

In our model with two topics per field, the distribution of citations is bimodal: the papers on topic X are cited more frequently, than the papers of similar quality on topic Y. Therefore, one can compare only the differences between the most and the least cited papers, or equivalently, the average numbers of citations. (Of course, the numbers must be expressed in percents of all citations if we compare fields that differ in the number of researchers.)

It turns out that, according to our model the least efficient fields are those with moderate average numbers of citations (or differences between the most and the least cited papers). We will illustrate this conclusion by the following example.

Example 3. Consider a modified setting from Example 2. As in Example 2, assume that $p = 1/2$, i.e., a half of researchers are partisan, while the other half are strategic. Suppose also that the density f is equal to e^{-x-y} multiplied by $a \in (0,1)$ on $\{(q_X, q_Y) \in [0,1]^2 : q_X < q_Y\}$, and multiplied by $2-a$ on $\{(q_X, q_Y) \in [0,1]^2 : q_X > q_Y\}$. In Example 2, $a = 2/3$. The average number of citations (and, similarly, the difference between the more and the less cited papers) decreases from 1 to 1/2 when a increases from 0 to 1.

As in Example 2, we compute the unique (history-independent) equilibrium, and compute that the aggregate inefficiency is equal to

$$\frac{2a(a-1)^2}{(2-a)^2}.$$

This value initially increases and then decreases, reaching its maximum at $a = (5 - \sqrt[2]{17})/2$, that is, at a just lower than 1/2.

This modification of the setting from Example 2 corresponds to rising asymmetry across topics (when a decreases from 1 to 0). The intuition is the following: When a is close to 1, we observe little inefficiency. Asymmetry creates a negative strategic effect on efficiency; however, rising asymmetry mitigates the strategic effect, bringing the outcome back to full efficiency when a is close to 0.

Similar conclusions (about inefficiency first increasing and then decreasing) hold true when we modify the setting from Example 2 in another way, by considering $b \in (0,1)$ the density f which is equal to e^{-x-y} multiplied by $(2/3)/(1-b)$ on $\{(q_X, q_Y) \in [0,1]^2 : b < q_X/q_Y < 1\}$, and multiplied by $(4/3)/(1-b)$ on $\{(q_X, q_Y) \in [0,1]^2 : b < q_Y/q_X < 1\}$. This modification corresponds of course to talent becoming more uniform across topics (as b increases from 0 to 1).

[10] For example, Ellison (2013) reports that in his data set from top 50 economics departments, the average number of citations varies from a high of 108 per paper in international trade to a low of 30 per paper in economic history.

The observation from Example 3 generalizes as follows: Suppose that the fractions of partisan and strategic researchers are constant across fields. That is, assume that it is not the case that some fields attract more partisan researchers, while other fields attract more strategic researchers. Suppose instead that fields differ with respect to asymmetry of topics. For example, in some fields research is focused around some central topics, while in other fields researchers work on a wider variety of topics. Alternatively, suppose that in some fields talent is more uniform, while in other fields talent is more topic specific. For example, this may be the result of different levels of maturity of fields, and different time investment required to begin producing valuable research. In both these cases, the lowest efficiency corresponds to intermediate average numbers of citations. As in Example 3, fields at the extremes (with the lowest or the highest average numbers of citations) are almost efficient, while the fields in the "middle" are the least efficient. Of course, for some distributions f it may not be the case that inefficiency monotonically increases until it reaches the maximum, and then monotonically decreases, as it happened in Example 2. There may be several ups and down as we move from one extreme to the other.

5.4 Other potential ways of removing or reducing inefficiency, neglected in the analysis

If we could tell topics apart, the inefficiency could be easily removed by comparing citations divided by the average number of citations on a given topic. However, as we have argued earlier, a very fine classification of topics would most likely be impractical, and perhaps even impossible. So, this will rather not completely remove the inefficiency.

One can also argue that the inefficiency would be removed if instead of counting citations (discounted by the period in which the citation arrived), we would count citations divided by the number of cited papers in the paper in which the citation appears. That is, the inefficiency would be removed if researchers were maximizing the value of an index satisfying Invariance to Reference Intensity, introduced by Palacios-Huerta and Volij (2004) in the context of ranking journals. This argument is indeed correct for the present version of the model, and can be interpreted as an argument for using more-elaborate citation indices when it is possible in practice.

This way of counting citations may also affect incentives in practice if papers on trendy topics tend to have unusually long lists of references. However, determining this correlation is a matter of (future) empirical research, so we are not drawing any (premature) conclusions here.

Nevertheless, we would simply like to point out that this new (just described) way of counting citations seems rare in practice (probably because more-elaborate indices are not as easily available, and quickly updatable as the citation counts provided by Google Scholar), and that slightly modified versions of our model are consistent with weak correlation, which makes the new way of counting citations less effective. For example, suppose that researchers writing on less popular topics cite more papers from other, not closely related areas in order to maintain some kind of standard for not having too short lists of references.

6 Dynamic analysis

Up to now, we have been assuming that the distribution of types is constant over time. We will relax this assumption in the present section. Of course, since there is a huge variety of ways in which the distribution can change over time, we will not be able to derive any general results. However, some important insight can be obtained by studying the dynamic evolution of asymmetry across topics without any changes in the relative density of types which have an advantage in X, or of types which have an advantage in Y. We will assume that the distributions have constant fractions of partisan and strategic researchers, but that the density f over types (q_X, q_Y) changes over time. More specifically, some constant-over-time density f' is multiplied at time t by an α^t on the set $\{(q_X, q_Y) : q_Y > q_X\}$; and on the set $\{(q_X, q_Y) : q_Y < q_X\}$ is multiplied by β^t such that $\alpha^t \mu'_Y + \beta^t \mu'_X = 1$.

6.1 Anticipated inflow of ideas on a topic

We will first study the dynamics of research in response to the news that an inflow of ideas is going to occur at some future date. For example, one may think about the invention of a technology that generates big data sets. In this case, we may not yet know in which way these data sets can be used, but we can anticipate that, sooner or later, researchers will find ways of using these data sets to answer important questions.

Some insight will be obtained by studying the process such that $\mu_Y = \mu_X = 1/2$ and $\alpha^t = 1$ in all periods except period $t = T > 1$, in which $\alpha^T = \alpha < 1$ (and so, $\beta^T = \beta = 2 - \alpha > 1$). Similar insight would be obtained by studying a more general process such that α^t gradually, geometrically increases till time T, and gradually, geometrically decreases from time T on.[11]

Since researchers are forward-looking, the decisions of researchers living in periods $t = T, T+1, ...$ are fully efficient. They choose the topic whose q is higher. So, $M_X^T = \beta$, $M_Y^T = \alpha$, and $M_X^k = M_Y^k = 1/2$ for $k = T+1, T+2, ...$; and $\mathbf{M}_X^T = (1-\delta)\beta(1/2) + \delta(1/2)$, $\mathbf{M}_Y^T = (1-\delta)\alpha(1/2) + \delta(1/2)$, and $\mathbf{M}_X^k = \mathbf{M}_Y^k = 1/2$ for $k = T+1, T+2, ...$ For researchers living in periods $k = 1, ..., T-1$, the following holds:

$$M_X^k = p \Pr\left\{(q_X, q_Y) : q_X \mathbf{M}_X^{k+1} > q_Y \left(1 - \mathbf{M}_X^{k+1}\right)\right\} + (1-p)(1/2), \tag{4}$$

$$M_Y^k = p \Pr\left\{(q_X, q_Y) : q_Y \mathbf{M}_Y^{k+1} > q_X \left(1 - \mathbf{M}_Y^{k+1}\right)\right\} + (1-p)(1/2); \tag{5}$$

and

$$\mathbf{M}_X^k = (1-\delta) M_X^k + \delta \mathbf{M}_X^{k+1},$$

$$\mathbf{M}_Y^k = (1-\delta) M_Y^k + \delta \mathbf{M}_Y^{k+1}.$$

[11] We conjecture, but have not proved formally, that similar insight could also be obtained for stochastic processes such that α^T decreases to α randomly, and then increases to 1, also randomly.

Thus, M_X^k, M_Y^k, $\mathbf{M}_X^k$, and $\mathbf{M}_Y^k$ for $k = 1, ..., T-1$ are determined recursively. An argument analogous to that from the proof of Proposition 1 shows the uniqueness of the (history-independent) equilibrium.

Since $M_X^T = \beta > 1$ and $M_Y^T = \alpha < 1$, it follows from (4) and (5) that $M_X^{T-1} > 1/2$ and $M_Y^{T-1} < 1/2$, and by recursion, it follows that $M_X^k > 1/2$ and $M_Y^k < 1/2$ for $k = 1, ..., T-1$.

In one interpretation, the inflow of new ideas on topic X takes place in period T, but strategic researchers begin switching to topic X from period 1, that is, from the time this inflow is first anticipated. The next proposition shows that under some condition, this strategic effect is so strong that over time we observe a declining flow of papers on topic X. That is, the inflow of papers is the largest at the time researchers begin anticipating the new opportunity, rather than at the time when this opportunity actually becomes available.

Proposition 6 *Suppose that*

$$p \Pr\{(q_X, q_Y) : [(1-\delta)\alpha + \delta]/[(1-\delta)\beta + \delta] q_Y < q_X < q_Y\} > \beta - 1/2. \tag{6}$$

Then, $M_X^k > M_X^{k+1}$ *for all* $k = 1, 2, ..., T-1$*.*

So, the anticipated inflow of ideas on a topic is always preceded by an inflow of papers on this topic. And Proposition 6 shows that the number of papers on the topic written before the inflow of ideas occurs may even exceed the number of papers written at the time the inflow of ideas occurs. The extreme scenario described in Proposition 6 is more likely to take place in fields in which talent is uniform across topics, because Condition (6) requires that a sufficiently large mass of types (q_X, q_Y) be concentrated close to the diagonal, and, more precisely, in the region $\{(q_X, q_Y) : [(1-\delta)\alpha + \delta]/[(1-\delta)\beta + \delta]\ q_Y < q_X < q_Y\}$.

Intuitively, the anticipated inflow of papers on topic X in period T provides incentives to researchers living in period $T-1$ for writing on X. And if there are many researchers affected by these new incentives, researchers living in period $T-2$ are provided even stronger incentives for writing on X. As a result, stronger incentives are transmitted to earlier periods.

Proof. The probability of writing on topic X in period T exceeds 1/2 by $\beta - 1/2$. In period $T-1$, the probability of writing on topic X is

$$p \Pr\{(q_X, q_Y) : q_X[(1-\delta)\beta + \delta] > q_Y[(1-\delta)\alpha + \delta]\} + (1-p)(1/2),$$

because $\mathbf{M}_X^T = (1-\delta)\beta(1/2) + \delta(1/2)$ and $\mathbf{M}_Y^T = (1-\delta)\alpha(1/2) + \delta(1/2)$.

This number exceeds 1/2 by

$$p[\Pr\{(q_X, q_Y) : q_X[(1-\delta)\beta + \delta] > q_Y[(1-\delta)\alpha + \delta]\} - 1/2].$$

Now, the result for $k = T-1$ follows from $1/2 = \Pr\{(q_X, q_Y) : q_X > q_Y\}$. By backward recursion, the result extends to $k = 1, ..., T-2$. ■

Finally, we modelled an inflow of ideas as an uniform increase of density across types (q_X, q_Y) such that $q_X > q_Y$. One could also think of a breakthrough which consists of a number of very good ideas on topic X, that is, the ideas with high q_X's. In such a case, not only is the number of papers in periods preceding the breakthrough is higher, but also their average quality falls below the average quality of papers at the time the breakthrough actually occurs.

6.2 Assigning different weights to citations in different periods

There seem to be good reasons to believe that citations are a better signal of quality (social value) for older than for newer papers. For example, immediate citations may be more a result of current fashion, or the authors' position in the profession, and thus may be less correlated with actual quality. In addition, social value is better assessed only after some time. Our model is in fact based on this view.

In this section, we address the question as to whether it would be more efficient to assign lower weights to citations in papers that appear shortly after a cited paper, and higher weights to citations in papers that appear long time after the given paper.

To make the analysis simple, we will consider only symmetric Markov processes, and will restrict attention to symmetric equilibria. More specifically, let f be a density over types (q_X, q_Y) symmetric across the diagonal. That is, $f(q_X, q_Y) = f(q_Y, q_X)$, in particular, $\mu_X = \mu_Y = 1/2$. For $t = 1, 2, ...$, let $\alpha^t = \alpha < 1$ or $2 - \alpha > 1$; note that $\alpha(1/2) + (2 - \alpha)(1/2) = 1$. If $\alpha^t = \alpha$, then $\alpha^{t+1} = \alpha$ with probability $\theta > 1/2$, and $\alpha^{t+1} = 2-\alpha$ with the complementary probability of $1-\theta$. Symmetrically, if $\alpha^t = 2-\alpha$, then $\alpha^{t+1} = 2-\alpha$ with probability θ, and $\alpha^{t+1} = \alpha$ with probability $1-\theta$. The density f^t in period t on the set $\{(q_X, q_Y) : q_Y > q_X\}$ is equal to f multiplied by an α^t, and on the set $\{(q_X, q_Y) : q_Y < q_X\}$ is equal to f multiplied by $2 - \alpha^t$. That is, if $\alpha^t = \alpha$, there are more researchers with an advantage in topic X, and if $\alpha^t = 2 - \alpha$, there are more researchers with an advantage in topic Y.

In order to conduct our analysis, we must generalize some concepts and results of our basic model to the present dynamic setting. As in Section 3, we restrict attention to equilibria in which the strategies are independent of the past choices of other researchers. However, we assume that researchers learn about the current state of the Markov process, and are allowed to condition their choices on this state. Let $M_i^{t,j}$ be the expected fraction of researchers who choose topic i in period t, when in period t there are more researchers with an advantage in topic j. Given the strategies of researchers living in periods $t, t+1, ...$, let

$$\mathbf{M}_i^{t,X} = (1-\delta)\sum_{n=t}^{\infty} \delta^{n-t} E_t M_i^{n,j_n},$$

where E_t denotes the expected value taken at time t of M_i^{n,j_n}, which depends on the state j_n of the Markov process in period n. That is, $\mathbf{M}_i^{t,X}$ is the occupation measure of topic i starting from time t, when in period t there are more researchers with an advantage in topic X. One can define $\mathbf{M}_i^{t,Y}$ analogously. Obviously, we have that $M_X^{t,j} + M_Y^{t,j} = 1$ and $\mathbf{M}_X^{t,j} + \mathbf{M}_Y^{t,j} = 1$ for $j = X, Y$.

We call strategies *symmetric* (across topics) if, whenever a researcher of type (q_X, q_Y) chooses topic X (topic Y) in state $j = X$, then a researcher of type (q_Y, q_X) chooses topic Y (topic X) in state $j = Y$. In this case, we also have that $M_X^{t,X} = M_Y^{t,Y}$, $M_Y^{t,X} = M_X^{t,Y}$, and $\mathbf{M}_X^{t,X} = \mathbf{M}_Y^{t,Y}$, $\mathbf{M}_Y^{t,X} = \mathbf{M}_Y^{t,X}$. For any stationary strategies, $M_i^{t,j}$ and $\mathbf{M}_i^{t,j}$ are independent of t. If strategies are symmetric and stationary, we denote $\mathbf{M}_X^{t,X} = \mathbf{M}_Y^{t,Y}$ by M^+ and $\mathbf{M}_Y^{t,X} = \mathbf{M}_X^{t,Y}$ by M^-. Of course, $M^+ + M^- = 1$.

Finally, let

$$\begin{aligned} g^+(M^+) &= (1-\delta)[p\Pr\{(q_X,q_Y) : q_X[\theta M^+ + (1-\theta)(1-M^+)] > q_Y[\theta(1-M^+) + (1-\theta)M^+]\} \\ &\quad +(1-p)(1-\alpha/2)] + \delta[\theta M^+ + (1-\theta)(1-M^+)] \end{aligned}$$

and

$$\begin{aligned} g^-(M^-) &= (1-\delta)[p\Pr\{(q_X,q_Y) : q_X[\theta(1-M^-) + (1-\theta)M^-] < q_Y[\theta M^- + (1-\theta)(1-M^-)]\} \\ &\quad +(1-p)(\alpha/2)] + \delta[(1-\theta)(1-M^-) + \theta M^-], \end{aligned}$$

where operator Pr refers to the distribution with density f multiplied by α on the set $\{(q_X,q_Y) : q_Y > q_X\}$, and multiplied by $2-\alpha$ on the set $\{(q_X,q_Y) : q_Y < q_X\}$.

We now make the following assumption (which is similar to Assumption 1 in Section 3):

Assumption 2. Function g^+ has a unique fixed point.

Since $M^+ + M^- = 1$, it follows that $g^+(M^+) + g^-(M^-) = 1$. It would therefore be equivalent to assume that function g^- has a unique fixed point. Notice that both g^+ and g^- always have a fixed point, since each of them is continuous, exceeds 0 at 0 and falls below 1 at 1. The following proposition characterizes the symmetric equilibria under Assumption 2. Its proof is analogous to the proof of Proposition 1.

Proposition 7 *The model has a unique symmetric (history-independent) equilibrium. This equilibrium is stationary. In the equilibrium, M^+ is the fixed point of g^+. A researcher living in a period in which the state of the Markov process is X (is Y) chooses topic X when $M^+q_X > M^+q_Y$ (when $M^-q_X > M^-q_Y$), and chooses topic Y when the opposite inequality holds.*

We can now address our question. We use an increase in the discount factor δ to model the higher weights assigned to citations in papers that appear in more remote periods.[12]

Proposition 8 *The unique fixed point M^+ of g^+ decreases (and thus the unique fixed point M^- of g^- increases) when δ increases. This means that the unique symmetric equilibrium becomes more efficient.*

[12] Our modelling approach is as follows: Let r be a (constant over time) ratio of the weight of a citation made $n+1$ periods after a given paper has appeared to the weight of a citation n periods after. Further, let w be the weight of an immediate citation. Then, the payoff of a strategic researcher is

$$(1-\delta)\sum_{n=1}^{\infty} w\delta^n r^n \tau_n = \frac{w(1-\delta)}{(1-\delta^/)}(1-\delta^/)\sum_{n=1}^{\infty}(\delta^/)^n\tau_n,$$

where $\delta^/ = \delta r$. Normalizing this payoff by $w(1-\delta)/(1-\delta^/)$, which bears no loss of generality as only relative payoffs matter in the present analysis, we obtain that raising the ratio r is equivalent to raising the discount factor $\delta^/$.

Proof. Rewrite equation $M^+ = g^+(M^+)$ as

$$\frac{[1-\delta\theta+\delta(1-\theta)]M^+ - \delta(1-\theta)}{1-\delta} = p\Pr + (1-p)(1-\alpha/2), \tag{7}$$

where $\Pr := \Pr\{(q_X, q_Y) : q_X[\theta M^+ + (1-\theta)(1-M^+)] > q_Y[\theta(1-M^+) + (1-\theta)M^+]\}$, by first subtracting from equation $M^+ = g^+(M^+)$ expression $\delta[\theta M^+ + (1-\theta)(1-M^+)]$, and then dividing the equation so obtained by $(1-\delta)$. Since, by Assumption 2, $M < g^+(M)$ for $M < M^+$ and $M > g^+(M)$ for $M > M^+$, the graph of the LHS of (7) intersects the graph of the constant RHS of (7) at the unique fixed point M^+ from below to above.

Notice further that the RHS of (7) is independent of δ, while the derivative of the LHS of (7) is positive in δ for all M^+. This means that the point M^+ at which the graph of the LHS intersects the graph of the RHS decreases when δ increases. ■

Proposition 8 captures the intuition that by assigning a higher weight to citations in papers which appear in more remote periods, we remove the incentives coming from current trends, i.e., choosing the topic that others currently choose. And this provides incentives for choosing what is likely to have more social value.

One may also be tempted to address the more general mechanism-design question of characterizing the optimal intertemporal pattern of weights. We find this question interesting and important, but obtaining any reasonable insight requires assuming a more concrete payoff function. We view (1) from Section 2 as a "first-order" approximation of researchers payoffs. In practice, researchers are not directly interested in maximizing citations of their papers, but in other goals which can be partially achieved by means of a higher number of citations. In particular, our analysis is rather "local"; changing δ by a little makes sense, but studying deltas of an arbitrary value makes no sense. One should instead assume that the marginal payoff of a citation diminishes with the number of previous citations, even within a period.

We conjecture that in such a richer model there would exist an optimal moment from which we should count citations. The reason is that delaying this moment makes the choice of topics more efficient; however, doing so also reduces the incentives for conducting any research at all, at least in a model in which researchers face some cost of conducting research and have the option of not conducting any research at all.

7 Conclusions and Extensions

In the present paper, we studied a basic model of citation-driven research. Researchers face individual incentives coming from a higher number of future citations, because this enables them to compare favorably to other researchers. Thus, they sacrifice some social value to write papers that they expect to be cited more frequently, and such decisions create social inefficiency. This inefficiency is affected by various factors, policies and adopted "corporate culture" in various areas of research.

We explored the effect of several such factors. In particular, we argued that the inefficiency is likely to be higher in disciplines (areas of research) in which talent is uniform across topics rather than more topic specific. We also determined conditions under which assigning a higher weight to citations in papers published in higher-ranked journals enhances efficiency. These are only some examples of questions addressed in the present paper. Many other important questions have not been addressed in the present paper. For example, some researchers are very generous in citing others, while other researchers prefer citing only papers which, according to the authors, have made a fundamental contribution. One may wonder how these two different attitudes affect efficiency.

We opted for a basic model with only two topics. This model captures the trade-off between pursuing quality and studying trendy topics. We believe that this trade-off is fundamental in all settings with strategic researchers seeking citations. However, since many features of such settings have been omitted, the model has numerous possible extensions. In a version of the model with more than two topics, stationary equilibria may induce interesting configurations, and other history-independent equilibria may exhibit interesting dynamic patterns.

One important extension would have new topics arriving over time, possibly with other topics becoming obsolete. Numerous intriguing questions could be addressed in such an extension. For example, what types of researchers would be most keen on advancing new ideas, which go beyond the existing paradigms? What kind of dynamics would emerge? Would the innovation rates be efficient? Designing such a setting requires a theory of topic arrivals, taking a position on their social value, and specifying the ways in which the distribution of researchers' types evolves with topic arrivals. This more ambitious and more difficult task is left for future papers.

Our theory assumes that papers are rather independent objects. It seems more realistic that new papers build on the existing papers. This is, for example, a standard approach in the recent literature on innovation and patents. So, it would be another important extension, and a promising direction of future research, to incorporate closer relations between various papers.

One could also consider relaxing some of our strong assumptions. For example, the distribution of researchers' types and the social values of ideas may in practice be stock-dependent, with early research conducted by a few pioneers having little value by itself, but inspiring lots of later research with more substantial value added, until their original ideas become exhausted and their value becomes marginal. It could also be true that the execution of more socially valuable projects may take more time; this may suggest that it would be socially desirable to pay attention only to a few the most highly cited papers of an individual researcher.

8 References

Chung, K.-S., M.-Y. Liang, and M. Lo (2017): "On the Information Content of Indirect Citations," mimeo.

Diamond, P. A. and J. E. Stiglitz (1974): "Increases in Risk and in Risk Aversion," *Journal of Economic Theory*, **8**, 337-360.

Ellison G. (2012): “Assessing Computer Scientists Using Citation Data,” Massachussetts Institute of Technology and National Bureau of Economic Research, mimeo.

Ellison, G. (2013): “How Does the Market Use Citation Data? The Hirsch Index in Economics,” *American Economic Journal: Applied Economics,* **5**, 63-90.

Hirsch, J. E. (2005): “An Index to Quantify an Individual’s Scientific Research Output,” *Proceedings of the National Academy of Sciences* **102**(46), 16569-16572.

Klabjan, D., W. Olszewski, and A. Wolinsky (2014): “Attributes,” *Games and Economic Behavior* **88**, 190-206.

Palacios-Huerta I. and O. Volij (2004): “The measurement of intellectual influence,” *Econometrica* **72**, 963-977.

Perry M. and P. J. Reny (2016): “How to Count Citations If You Must,” *American Economic Review* **106**(9), 2722–2741.

9 Appendix

9.1 A characterization of history-independent equilibria

The following proposition characterizes the equilibria of our basic model in the general case, without making Assumption 1.

Proposition 9 *In any equilibrium,*

$$\mathbf{M}_i^t = (1-\delta) g_i(\mathbf{M}_i^{t+1}) + \delta \mathbf{M}_i^{t+1}, \tag{8}$$

the sequence $(\mathbf{M}_i^t)_{t=1}^{\infty}$ monotonically increases or decreases, and converges to a fixed point M_i of function g_i.

Proof of Proposition 8: Notice that the payoff of a strategic researcher with type (q_X, q_Y), living in period t, from writing on topic i is $\delta q_i \mathbf{M}_i^{t+1}$. Therefore, the expected fraction of researchers who choose topic $i = X, Y$ in period t is given by $M_i^t = g_i(\mathbf{M}_i^{t+1})$. Thus, by definition, $\mathbf{M}_i^t$ and $\mathbf{M}_i^{t+1}$ must satisfy (8).

It follows immediately from definitions that if $\mathbf{M}_i^{t+1} > \mathbf{M}_i^{t+2}$, then $M_i^t > M_i^{t+1}$ and $\mathbf{M}_i^t > \mathbf{M}_i^{t+1}$; and if $\mathbf{M}_i^{t+1} < \mathbf{M}_i^{t+2}$, then $M_i^t < M_i^{t+1}$ and $\mathbf{M}_i^t < \mathbf{M}_i^{t+1}$. Therefore, sequence $(\mathbf{M}_i^t)_{t=1}^{\infty}$ must be increasing or decreasing. In either way, the sequence must converge to a number between 0 and 1. By definition, so does sequence $(M_i^t)_{t=1}^{\infty}$, and the limit of these sequences must be a fixed point M_i of g_i.

This completes the proof.

Clearly, any sequence $(\mathbf{M}_i^t)_{t=1}^{\infty}$ that satisfies equation (8) determines the following equilibrium: A researcher living in period t chooses topic X when $\mathbf{M}_X^{t+1} q_X > \mathbf{M}_Y^{t+1} q_Y$, and chooses topic Y when $\mathbf{M}_Y^{t+1} q_Y > \mathbf{M}_X^{t+1} q_X$; when $\mathbf{M}_X^{t+1} q_X = \mathbf{M}_Y^{t+1} q_Y$ (which event has probability 0), a researcher living in period t is indifferent between the two topics. In addition, any sequence $(\mathbf{M}_i^t)_{t=1}^{\infty}$ that satisfies equation (8) is monotonic, increasing or decreasing, so it converges to a fixed point of g_i.

Thus, Proposition 8 provides a complete characterization of history-independent equilibria.

In Figure 5, we depict a sequence $(\mathbf{M}_i^t)_{t=1}^{\infty}$ such that the equilibrium determined by this sequence is not stationary. To construct such a sequence, we start from any point $\mathbf{M}_i^1$ higher than some fixed point of $(1-\delta)g_i(M)+\delta M$, and such that the graph of $(1-\delta)g_i(M)+\delta M$ at $\mathbf{M}_i^1$ lies above the diagonal. This point determines $\mathbf{M}_i^2$ by equation (8). At $\mathbf{M}_i^2$ the graph of $(1-\delta)g_i(M)+\delta M$ still lies above the diagonal, and $\mathbf{M}_i^2$ lies between $\mathbf{M}_i^1$ and the fixed point of $(1-\delta)g_i(M)+\delta M$ which is the closest to $\mathbf{M}_i^1$ from the left. This $\mathbf{M}_i^2$ in turn determines $\mathbf{M}_i^3$, and recursively, the remaining elements of the sequence. The sequence converges to the fixed point of $(1-\delta)g_i(M)+\delta M$ (and thus a fixed point of $g_i(M)$) which is the closest to $\mathbf{M}_i^1$ from the left.

In fact, we can construct all equilibria either in the above way, or by starting with any point $\mathbf{M}_i^1$ lower than some fixed point of $(1-\delta)g_i(M)+\delta M$, and such that the graph of $(1-\delta)g_i(M)+\delta M$ at $\mathbf{M}_i^1$ lies below the diagonal.

Finally, we provide the proof of Proposition 1.

Proof of Proposition 1: In a stationary equilibrium, $\mathbf{M}_i^t$ and $\mathbf{M}_i^{t+1}$ are equal to M_i. Equation (8) therefore reduces to $M_i = g_i(M_i)$. By Assumption 1, this implies the uniqueness of the stationary equilibrium. To complete the proof of Proposition 1, we must show that under Assumption 1 the model has no (history-independent) equilibrium other than the stationary equilibrium determined by the fixed points M_X and M_Y.

Suppose that $\mathbf{M}_i^{t+1}$ for some t is higher than the unique fixed point M_i. Since the graph of g_i intersects the diagonal only once from the left to the right, it follows from (8) that $\mathbf{M}_i^t < \mathbf{M}_i^{t+1}$, and therefore sequence $(\mathbf{M}_i^t)_{t=1}^{\infty}$ increases, and cannot converge to M_i. Similarly, if $\mathbf{M}_i^{t+1}$ for some t falls below the unique fixed point M_i, then it must be that $\mathbf{M}_i^t > \mathbf{M}_i^{t+1}$, and therefore sequence $(\mathbf{M}_i^t)_{t=1}^{\infty}$ decreases, and cannot converge to M_i. Thus, $\mathbf{M}_i^{t+1} = M_i$ for all t, and so does $\mathbf{M}_i^1$. As a result, the equilibrium coincides with the unique stationary equilibrium.

9.2 History-dependent equilibria

In this subsection, we informally describe a history-dependent equilibrium. To provide a simple example, assume that the distribution of types f is symmetric across the diagonal; this implies, in particular, that $\mu_X = \mu_Y$. Assume also that only one researcher lives in every period.

The strategies are defined as follows: If all researchers in the past have chosen topic X, then the researcher living in the current period chooses topic X when $q_X/q_Y > \gamma$ for some $\gamma < 1$, and chooses topic Y when $q_X/q_Y < \gamma$. Similarly, if all researchers in the past have chosen topic Y, then the researcher living in the current period chooses topic Y when $q_Y/q_X > \gamma$, and chooses topic X when $q_Y/q_X < \gamma$. If in the past there were researchers writing on X, as well as researchers writing on Y, then the strategic researchers living in the current period choose the topic in which they have an advantage, that is, they behave as partisan

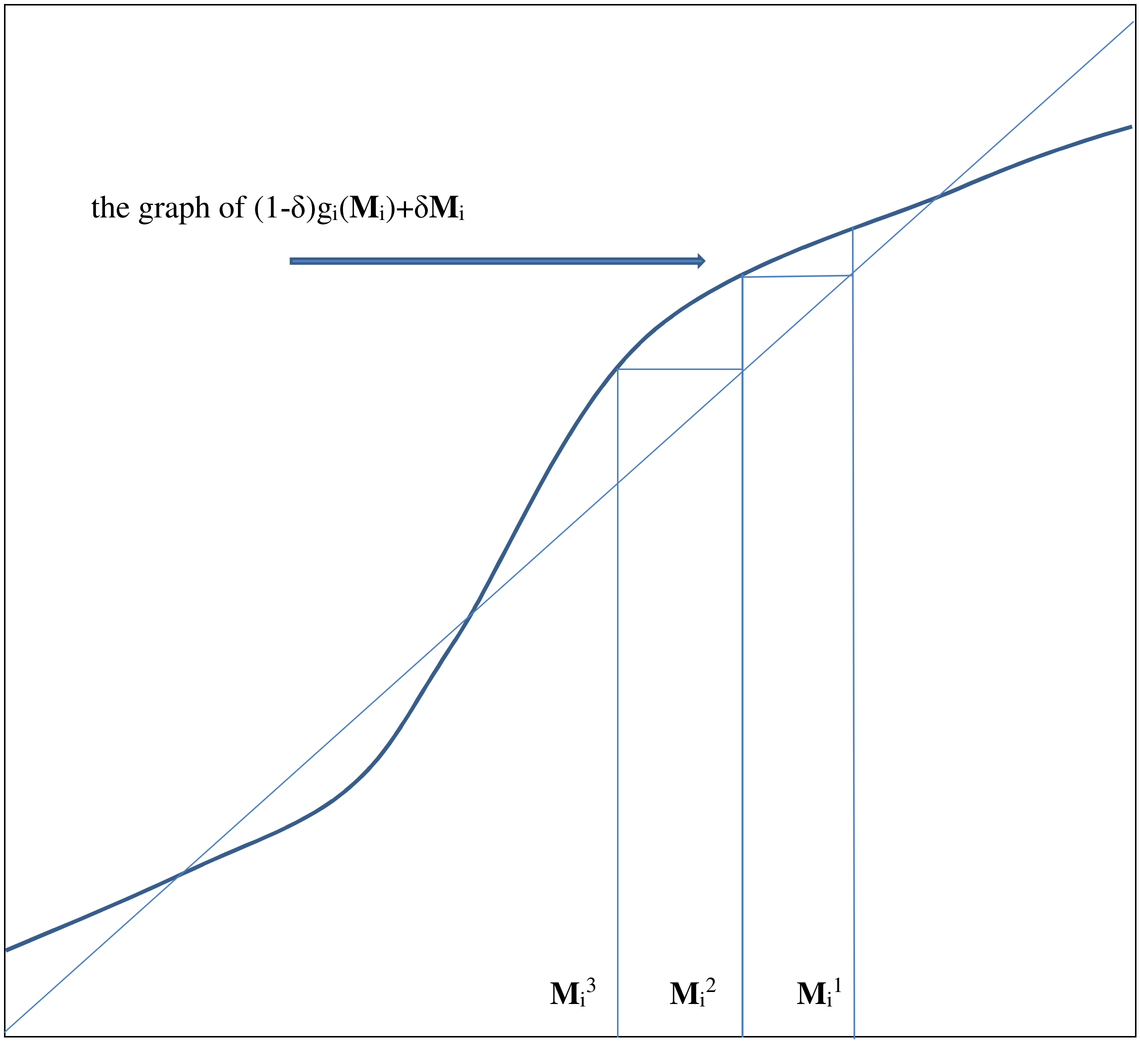
the graph of $(1-\delta)g_i(\mathbf{M}_i)+\delta\mathbf{M}_i$
$\mathbf{M}_i^3$
$\mathbf{M}_i^2$
$\mathbf{M}_i^1$

Figure 5. The construction of a nonconstant sequence $(\mathbf{M}_i^t)_{t=1}^{\infty}$ satisfying the conditions of Proposition 8

researchers. Similarly, the strategic researchers living in the first period behave as partisan researchers.

These strategies are an equilibrium, for some distributions f, since the researchers with a small disadvantage in i, $i = X, Y$, are still willing to choose i, if all past choices were i, since they expect that more than half of researchers will choose i in the future. The distribution f must be such that there is a sufficient mass of types close to the diagonal $q_Y = q_X$ to ensure that the researchers with a small disadvantage in i who write on i provide sufficient incentives for writing on i to such researchers from earlier periods.

In turn, contingent on any mixed past choices, players play a symmetric stationary (history-independent) equilibrium. Obviously, the player living in the first period is indifferent between the two choices.