Three Ways To Become An Academic Superstar*

Remco Heesen†

March 3, 2014

Abstract

It is well-known that some scientists are more prominent than others. But what makes one scientist more prominent than another? I propose a possible mechanism that produces differences in prominence: scientists' desire for information. In a model of a scientific community exchanging information, I show that this mechanism indeed produces the kind of patterns of prominence that are actually observed. I discuss the implications of this result for three possible explanations of an individual scientist's prominence: an explanation based on scientific merit, an explanation based on epistemically irrelevant factors (e.g., gender bias or charisma), and an explanation based on epistemic luck. Depending on which of these explanations is correct one may draw different conclusions about a scientist based on prominence. I discuss policy recommendations that result from this, including suggestions about when it is appropriate to use measures of prominence (e.g., citation metrics) in giving out grants and awards.

---

*Thanks to Kevin Zollman, Teddy Seidenfeld, Katharine Anderson, Tomas Zwinkels, Liam Bright, Aidan Kestigian, and audiences at Ghent University, Oxford University, and Carnegie Mellon University for valuable comments and discussion.

†Department of Philosophy, Baker Hall 135, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA. Email: rheesen@cmu.edu

1 Introduction

Academic superstars are a familiar phenomenon. These scientists write the papers that everyone reads and talks about, they make media appearances, give presidential addresses, and they win grants and awards. The work of an academic superstar generally attracts more attention than that of the average scientist.

Quantifying attention as the number of citations to their papers, sociologists found an easy way to identify academic superstars.[1] They also noted that superstars are rare: the vast majority of scientists receives no more than a handful of citations, while a rare few get extremely many (Price 1965, Cole 1970). More recent work confirms that the distribution of the number of papers with a given number of citations follows a "power law".[2]

In this paper I ask why superstars exist. Why does some scientists' work receive more attention than others'? What characteristics of individual scientists contribute to their work receiving more or less attention?[3]

I propose a possible mechanism that could be responsible for producing the patterns that are actually observed (in particular, the existence of superstars). The mechanism I propose is scientists' desire for information. In section 2 I set up a model of information exchange among scientists. I state some theorems of this model in sections 3 and 4. My goal in these three sections is to establish something like the following claim.

Claim 1. *If, in choosing whose work to read, scientists are motivated only by gathering as much information as possible given their means, then the*

[1] Kieran Healy recently did this for the field of philosophy, in a blog post which was widely discussed in informal circles in philosophy (see `http://kieranhealy.org/blog/archives/2013/06/18/a-co-citation-network-for-philosophy/`).

[2] This means that the number of papers that gets cited n times is proportional to $n^{-\alpha}$ for some α. Redner (1998) estimates α to be around 3.

[3] The standard explanation in the literature for power law distributions is through preferential attachment models (Barabási and Albert 1999). However, such models do not include characteristics of individual scientists or papers. A more detailed comparison between preferential attachment models and the model of this paper is given in section 2.

patterns of interaction that emerge are highly imbalanced: some scientists get a lot of attention, while most get very little.

Insofar as I succeed in establishing this claim, it shows that scientists' desire for information (which is surely one of scientists' many motivations) can work as a mechanism that leads to the existence of academic superstars.

Next, I compare and contrast three explanations of how an individual scientist becomes an academic superstar (sections 5, 6, and 7). Two of these are familiar from the literature and one is novel. In each case, I discuss how the mathematical results from sections 3 and 4 bear on the explanation, as well as policy recommendations that follow from the explanation.

2 The Model and the Assumptions

The goal of this model is to capture important aspects of the way scientists exchange information. In order to facilitate mathematical analysis, some abstractions and simplifications have to be made. Here I indicate and defend the most important of those.

A set of scientists I is assumed to be given, where each element $i \in I$ represents an individual scientist. This may be an arbitrary set of scientists, but the most natural way to think of it is either as the practitioners of some given scientific discipline (small or large) or as all of science. I is assumed to be finite.

The scientists are interested in learning something about the world. There is a set of worlds Ω that the scientists consider possible, and scientists are interested in distinguishing between these worlds or sets of them.[4] I make

[4]Different scientists may consider different worlds possible, or may be interested in different distinctions. The assumption of a single set of worlds Ω does not rule this out: if Ω_i is the set of worlds scientist i considers possible, define Ω as the Cartesian product of Ω_i for all $i \in I$, but allow a scientist i to distinguish between two possible worlds only if they differ in their i-th index. Everything in this paper is consistent with this way of setting things up.

no assumptions about the cardinality of Ω, which distinctions the scientists are or are not interested in making, or about any probability or plausibility ordering the scientists might have over worlds.

Each scientist learns something about the world through her own research: say, the outcome of her experiments. Suppose there are m experiments one might do. These are modeled as m probability distributions $F_1, \ldots, F_m$ that one might draw from. Let $n(i, j)$ be the number of times scientist i performs experiment j ($1 \leq j \leq m$). Then her research yields, among other things, $n(i, j)$ random variables $X_{j,i,1}, \ldots, X_{j,i,n(i,j)}$, where each of these random variables has probability distribution F_j.

That scientists may learn from experiments is reflected by the fact that the probability distributions of the experiments may depend on the possible world the scientist is in. I make no assumption on the form this dependence takes.

I refer to the collection of experiments performed by a given scientist i as her information set A_i. So

$$A_i = \{X_{j,i,k} \mid 1 \leq j \leq m, 1 \leq k \leq n(i, j)\}.$$

Each scientist publishes her information in a paper. For simplicity[5] I assume that each scientist publishes a single paper, and that this paper contains all the information in her information set. Thus I can refer to each $i \in I$ interchangeably as a scientist or a paper. Call the ordered pair $\mathcal{C} = (I, \{A_i \mid i \in I\})$ a scientific community.

Now the scientists form "connections". The benefit of a connection is to

[5] A slight generalization of my model would have separate sets of scientists and papers, with an information set for each paper. Connections (as defined below) would then go from scientists to papers. The measure of prominence I define below would be defined for each paper, and all my results (in particular theorem 6) would hold under the same assumptions. The measure of prominence of a scientist could then be taken to be the sum of the measures of prominence of her papers, and the results would be essentially the same, but with more complex notation.

obtain the information in the other scientist's information set.[6] Connections may be interpreted as scientists reading each other's papers: i connects to i' is short for i reads i''s paper and thereby learns the realization of each random variable in information set $A_{i'}$.

Importantly, connections are one-sided (or directed). So a scientist may form a connection to another scientist on her own initiative, without needing the other scientist's consent. By doing so she learns the contents of the other scientist's information set, but the other scientist does not learn anything.

These assumptions yield a realistic model of the type of information exchange that occurs when scientists read each other's work. A scientist can read another's paper without prior consent. By reading a paper the scientist obtains some information from the other scientist but no information flows in the reverse direction.

Another assumption I make is that the process of forming connections happens relatively quickly compared to the process of doing experiments. More precisely, in this model scientists do not perform new experiments while they are forming connections. What experiments each scientist has done is assumed to be fixed background information when they make decisions about whom to connect to. Moreover, individual scientists know this background information: they are aware of which scientists have performed which experiments.[7]

I argue that these assumptions are close enough to the truth for present purposes. Relative to the time and cost involved in designing and running an experiment, reading (and subsequently citing) someone else's paper is a very

---

[6]Scientists learn each other's evidence, not each other's conclusions as expressed, say, in a posterior probability. In this sense my model differs from that of Aumann (1976). One reason for doing it this way is to make sure substantial information is exchanged. If scientists only learn each other's posterior on some set of possible worlds, Aumann's result guarantees that repeated exchange of posteriors will make them equal, but this does not necessarily mean that anyone has learned anything (see Geanakoplos and Polemarchakis 1982, proposition 3).

[7]In other words, every scientist knows the numbers $n(i, j)$ for each i and j.

short-term activity. The assumption that scientists know whom to connect to to get certain information is justified by the further observation that the time required to search for papers on a certain subject (perhaps looking at some titles and abstracts) is itself negligible compared to the time required to actually read papers and obtain the information in them. Additionally, in relatively small scientific communities this assumption may be justified because everyone knows what everyone else is working on through informal channels.

The next question is which connections are actually formed. This obviously depends on the goals of the individual scientists. I do not assume that scientists are fully rational. Instead I make some specific, much weaker, assumptions about their behavior that amount to a kind of bounded rationality. I state and defend these assumptions in section 3.

Individual scientists are allowed to choose their connections sequentially. That is, their decision which scientist to connect to next may depend on the information learned through previous connections.

I think these bounded rationality assumptions and the sequential decisions assumption allow enough flexibility to capture the ways scientists might act on their desire for information. However, I will also defend these assumptions from a Bayesian perspective by proving that fully rational Bayesian scientists may satisfy them (see theorems 11 and 12 in section 4).

Consider the graph or network formed by viewing each scientist as a node, and drawing an arrow (called an arc or directed edge in graph theory) from node i to node i' whenever scientist i forms a connection with scientist i'. More formally, define the network $G = (I, \{(i, i') \in I^2 \mid i \text{ connects to } i'\})$.

In order to study the prominence of individual scientists in G, I need a measure of prominence. A natural measure suggests itself: the number of scientists that read the individual's work. In the network, the number of scientists who read i's work is simply the number of arrows ending at i. In graph-theoretical terms, this is the in-degree of node i. So the in-degree can be used as a measure of the prominence of scientists in the community. This idea is illustrated in figure 1.

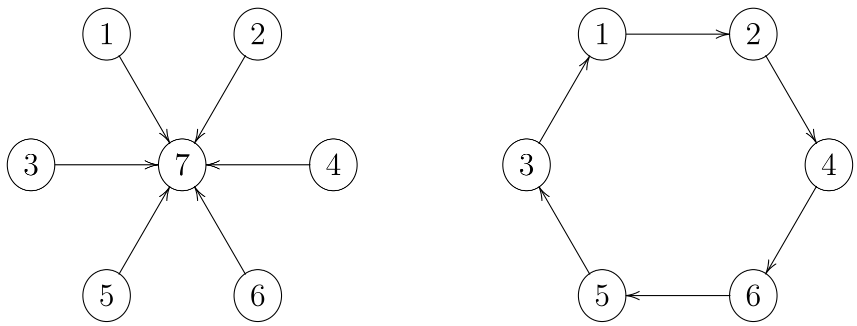

Figure 1: Two networks for small scientific communities. On the left, scientist 7 is prominent because she has an in-degree of 6 while the other scientists in her community have an in-degree of 0. On the right, all scientists are equally prominent, having an in-degree of 1.

If a scientist learns something from another scientist, she will usually acknowledge this fact in future work by citing the paper she read. In general, one may expect the papers that a given scientist cites to be highly correlated with the papers she read. So the measure of prominence based on in-degree I have just defined should in practice match up closely with citation metrics.

If, as I have suggested, some scientists get many citations and some very few, then one should expect large differences in in-degree among scientists. I will prove that this is indeed the case in my model (assuming the community is large enough). This is how I substantiate claim 1.

It may be remarked that in the literature there is already a standard explanation for this pattern of large differences in in-degree. This explanation is fleshed out in so-called preferential attachment models. In a preferential attachment model new nodes (papers) form links (citations) to older nodes proportional to the number of links that older node already has. So any paper that already has twice as many citations as some other paper is also twice as likely to be cited by future papers.

It can be shown that this generates a power law distribution of citations with an exponent equal to three (Barabási and Albert 1999). This is very close to what is observed in real citation data (Redner 1998). While this is interesting and illuminating as far as it goes, it is not fully satisfying for at least two reasons.

First, it gives no insight in why the difference between the two papers appeared in the first place (the model needs to start with some citations already in place for the probabilities of new connections to be well-defined). By not including any characteristics that distinguish papers from one another, it offers nothing to someone who wants to predict in advance which of two forthcoming papers will be more highly cited.

Second, it says nothing about what motivates scientists in choosing to cite one paper rather than another. Preferential attachment models merely stipulate that scientists cite older work proportional to its number of existing citations, without giving any decision theory explaining what makes this behavior rational (or irrational).

For these two reasons preferential attachment models are a non-starter if one is interested in determining whether it is good or bad for science that citations follow a power law. It should be clear from the above that the model in this paper avoids these issues.

My model is obviously unrealistic in portraying science as consisting of one round of experiments and one round of connections. In reality, scientists may form many connections over time, interspersed with experiments. The model can be viewed as looking only at a small period of time in a scientist's career, say, the time connected with a single research project: doing some experiments and exchanging results with epistemic peers.

In terms of the dynamic model of epistemic inquiry by Kelp and Douven (2012), my model may be viewed as zooming in on one deliberative round and one disclosive round. Afterward the scientists take what they have learned as prior information into the next set of rounds.

In conclusion, I do not claim to have given a definitive model of information exchange in science. My goal is merely to describe some aspect of it, in particular the idea that each scientist has access to different information (scientists are heterogeneous in this respect) and that this plays a role in motivating other scientists to read or cite them. The model ignores certain complicating factors, e.g., how sequential choices to form connections may interact with the choice of experiment (the latter being taken as exogenous).

In general, if the model seems too simple to be realistic I would argue that this is a virtue. Similar results should be expected in any (more realistic) model that includes my model (or something close to it) as a special case. Specific arguments would be needed to show that making the model more realistic would undo my results.

3 Superstars in the Model

Say that two scientists i and i' have the same information set ($A_i = A_{i'}$) if they have done each experiment the same number of times: $n(i,j) = n(i',j)$ for all $j \in \{1,\ldots,m\}$.[8] I can then ask what proportion of scientists in the community has the same information set as some given scientist.

Definition 2 (relative frequency of information)**.** A function q describes the relative frequency of information in a scientific community $\mathcal{C} = (I, \{A_i \mid i \in I\})$ (I will abbreviate this as "$\mathcal{C}$ satisfies q") if for all information sets A

$$q(A) = \frac{|\{i \in I \mid A_i = A\}|}{|I|}.$$[9]

[8]Technically the two information sets will not be equal. For example if each scientist has only done experiment 1 once then $A_i = \{X_{1,i,1}\}$ and $A_{i'} = \{X_{1,i',1}\}$. These sets are different because the random variables differ in their second index. It would be more correct to define an equivalence relation on information sets such that two information sets are equivalent (rather than equal) if and only if $n(i,j) = n(i',j)$ for all $1 \leq j \leq m$. Any occurrences of equality of information sets in the remainder of this paper should then be replaced with equivalence. This can be done (see Heesen 2014), but I have chosen to avoid these complications here.

[9]$|S|$ denotes the cardinality of the set S, which is simply the number of elements of S if S is finite (as all sets to which I apply $|\cdot|$ in this paper are).

Let $\delta_i^{\mathcal{C}}$ denote the sequential decision procedure for a scientist i. The procedure tells i whom (if any) of the other scientists in the set I to connect to at any given time. Formally, $\delta_i^{\mathcal{C}}$ is a function whose output is a sequence of members of I (interpreted as the scientists i connects to). It is a function of the information sets of the other scientists: the $n+1$-st element of the sequence may depend on the values of the random variables in the information sets associated with the first n elements of the sequence.

Since decisions to connect may depend on the values of certain random variables, such decisions may themselves be viewed as random. This is the significance of any subsequent use of probabilities and expectations.

Abusing notation, I let $\delta_i^{\mathcal{C}}$ denote both the sequential decision procedure and the number of connections made under that procedure. Since the former is a random variable, so is the latter. Similarly, let $\delta_i^{\mathcal{C}}(A)$ denote the number of connections made to scientists with information set A (i.e., the number of i' such that i connects to i' and $A_{i'} = A$). This is again a random variable.

This brings us to the first assumption on the way scientists choose their connections. It says that scientists are unlikely to connect to a large number of other scientists.

Assumption 3 (Uniformly Bounded Connection Probabilities)**.** *For any relative frequency q, information set A, and $\varepsilon > 0$, there exists $N_\varepsilon(A)$ such that for all $n > N_\varepsilon(A)$, for all $\mathcal{C}$ that satisfy q and for all $i \in I$, if $\Pr\left(\delta_i^{\mathcal{C}}(A) \geq 1\right) > 0$ then*

$$n \Pr\left(\delta_i^{\mathcal{C}}(A) \geq n \mid \delta_i^{\mathcal{C}}(A) \geq 1\right) \leq \varepsilon.$$

Note a number of consequences of this definition. First, $q(A) \geq 0$ for all A. Second, $q(A) > 0$ if and only if at least one scientist in I has information set A. Third, q sums to one. Fourth, since only finite sets of scientists I are considered, a function q can meet this definition only if $q(A)$ is a rational number for all A. Fifth, if there exists a set of scientists that satisfies q then there exist arbitrarily large sets of scientists that satisfy q.

When I prove results below for any relative frequency q this should be read as applying to any function q that satisfies this definition for at least one finite set of scientists I.

The bound on the connection probabilities is uniform in the sense that $N_\varepsilon(A)$ is not allowed to depend on i (so the bound is the same for all scientists) or $\mathcal{C}$ (so the bound is the same regardless of how many scientists are available to connect to). If $N_\varepsilon(A)$ were allowed to depend on i and $\mathcal{C}$, assumption 3 would follow simply from assuming $\mathbb{E}\left[\delta_i^{\mathcal{C}}\right] < \infty$.

Assumption 3 has some plausibility as a principle of bounded rationality. It says that there is some number n such that no scientist has more than a negligible probability of reading the work of more than n other scientists. For a defense of this assumption from the perspective of Bayesian rationality I refer to theorem 12.

The second assumption on the way scientists choose says that they prefer to get more information rather than less from a connection.

Assumption 4 (Never Consider Subsets)**.** *A scientist will not connect to a second scientist i whenever a third scientist i' is available to connect to and $A_i \subset A_{i'}$.*[10]

From a bounded rationality perspective, this assumption is plausible if each random variable in an information set provides independently valuable information. Theorem 11 defends this assumption from the perspective of Bayesian rationality.

As indicated in section 2, the in-degree of a node (i.e., a scientist) is a measure of the prominence of that node in the network. The in-degree $d(i)$ of a scientist i is defined by

$$d(i) = \left|\{i' \in I \mid i' \text{ connects to } i\}\right|.$$

Whether i' connects to i depends on $\delta_{i'}^{\mathcal{C}}$. Since $\delta_{i'}^{\mathcal{C}}$ is random, $d(i)$ is random. Its expected value is

$$\mathbb{E}\left[d(i)\right] = \sum_{i' \in I} \Pr\left(i' \text{ connects to } i \mid \delta_{i'}^{\mathcal{C}}\right).$$

[10] $A_i \subset A_{i'}$ means that $n(i,j) \leq n(i',j)$ for all $j \in \{1,\ldots,m\}$ and $n(i,j) < n(i',j)$ for at least one j. A technical remark analogous to footnote 8 applies here.

The only feature that distinguishes among individual scientists is their information set. Scientists with the same information set are, at least as far as the model is concerned, indistinguishable in the eyes of the other scientists. Thus it makes sense to group them together, and consider their average in-degree.

Definition 5 (average in-degree and expected average in-degree)**.** Let $\mathcal{C}$ be a scientific community and let q describe the relative frequency of information in $\mathcal{C}$. The average in-degree of scientists with information set A, denoted $d(A)$, is

$$d(A) = \frac{1}{q(A)\,|I|} \sum_{i \in I : A_i = A} d(i).$$

The expected average in-degree of scientists with information set A, denoted $\mathbb{E}\left[d(A)\right]$, is

$$\mathbb{E}\left[d(A)\right] = \frac{1}{q(A)\,|I|} \sum_{i \in I : A_i = A} \mathbb{E}\left[d(i)\right].$$

With this definition in hand the main result can be stated. It states that if the set of scientists is sufficiently large, the expected prominence of a given scientist increases rapidly (faster than linearly) in the size of her information set.

Theorem 6 (Supermodularity of the expected average in-degree)**.** *For any relative frequency q there exists a number N such that for all communities $\mathcal{C}$ satisfying q and assumptions 3 and 4, if $|I| > N$ then for all information sets A and B with $q(A \cup B) > 0$*

$$\mathbb{E}\left[d(A \cup B)\right] + \mathbb{E}\left[d(A \cap B)\right] \geq \mathbb{E}\left[d(A)\right] + \mathbb{E}\left[d(B)\right].$$

The theorem is important because it shows that the patterns of information exchange in my model reflect the patterns that can be seen in real citation networks (compare this to claim 1). That is, most papers have few citations, while a rare few have a great number of citations (Price 1965, Cole 1970, Redner 1998).

The theorem is proved in Heesen (2014). The idea behind the proof is as follows. Let A and B be information sets with $q(A \cup B) > 0$. By assumption 4, no scientist will connect to a scientist with information set A or B unless she has already connected to all scientists with information set $A \cup B$ in the community.[11]

If there were infinitely many scientists with information set $A \cup B$, then she would never run out of such scientists. Thus, she would never connect to a scientist with information set A or B, and so $\mathbb{E}[d(A)] = \mathbb{E}[d(B)] = 0$, which makes the result of the theorem trivially true.

It follows that if there are only finitely many scientists with information set $A \cup B$, $\mathbb{E}[d(A)] + \mathbb{E}[d(B)]$ will be small as long as it is unlikely that anyone connects to all of them. Assumption 3 guarantees that this is the case for a sufficiently large community of scientists.

4 Bayesian Scientists

In the previous section I stated an important result (theorem 6), which gives sufficient conditions for the presence of superstars in my model. However, the conditions involved two important assumptions on the way scientists choose whom to connect to.

This section sets up the situation faced by scientists in my model as a Bayesian decision problem. It sheds further light on assumptions 3 and 4 by stating, for each assumption, a set of conditions sufficient to guarantee that the optimal Bayesian sequential decision procedure satisfies it (see DeGroot 2004, chapter 12, for a more comprehensive treatment of Bayesian sequential decision problems and further references).

Recall that Ω is the set of possible worlds considered by the scientists. Let $\mathcal{P} = (\Omega, \mathcal{W}, \xi)$ be a probability space, i.e., $\mathcal{W}$ is a σ-field of subsets of Ω and $\xi : \mathcal{W} \to [0, 1]$ is a probability measure (think of ξ as the scientist's

[11] Assuming $A \subset A \cup B$ and $B \subset A \cup B$. If this does not hold then either A or B is equal to $A \cup B$. In that case the inequality of the theorem holds trivially.

prior). Let $\xi(A_1, \ldots, A_n)$ denote the probability measure obtained by Bayes conditioning ξ with respect to the random variables in the information sets $A_1, \ldots, A_n$.

Recall the definition of information set:

$$A_i = \{X_{j,i,k} \mid 1 \leq j \leq m, 1 \leq k \leq n(i,j)\}.$$

So far the only assumption on the random variables has been that $X_{j,i,k}$ follows some common distribution F_j for all i and k. An important special case is the one where the random variables also satisfy an independence condition.

Definition 7 (simple scientific community)**.** Call the scientific community $\mathcal{C} = (I, \{A_i \mid i \in I\})$ simple relative to a probability space $\mathcal{P} = (\Omega, \mathcal{W}, \xi)$ if for all $W \in \mathcal{W}$ $X_{j,i,k}$ is independent of $X_{j',i',k'}$ given W unless $i = i'$, $j = j'$, and $k = k'$ (conditional independence) and $X_{j,i,k}|W \sim F_j|W$ (conditional identical distributions).

In order to turn the question whom a Bayesian scientist would connect to into a Bayesian decision problem something is needed to evaluate the state the scientist is in after she finishes connecting. For this purpose, I introduce a terminal decision.

Definition 8 (decision problems and procedures)**.** A decision problem $\mathcal{D} = (\mathcal{C}, D)$ is an ordered pair consisting of a scientific community $\mathcal{C}$ and a set D of terminal decisions. A (sequential) decision procedure δ for $\mathcal{D}$ is a function that outputs a sequence $i_1, i_2, \ldots, i_\delta$ of members of I (where, as before, δ is used to denote the length of the sequence, i.e., the number of connections) and a terminal decision $d_\delta \in D$. For any n, i_{n+1} may depend on the information sets $A_{i_1}, \ldots, A_{i_n}$. d_δ may depend on $A_{i_1}, A_{i_2}, \ldots, A_{i_\delta}$. Let $\Delta_{\mathcal{D}}$ denote the set of all decision procedures for $\mathcal{D}$.

It remains to define a way of evaluating decision procedures. This evaluation depends on three things: the loss function (which gives the disutility of a terminal decision in a possible world), the cost of connecting (a disutility associated with each connection)[12] , and the subjective probability of being in a given possible world.

Definition 9 (sequential risk function)**.** Let $\mathcal{P}$ be a probability space and $\mathcal{D}$ a decision problem. Let $\ell(w, d)$ be the loss associated with terminal decision $d \in D$ in world $w \in \Omega$. The risk of an immediate decision is defined (relative to $\mathcal{P}$) as

$$\rho_0(\xi, d) = \int_\Omega \ell(w, d)\, d\xi(w)$$

for all $d \in D$.

Let $c > 0$. Then the risk under loss ℓ and cost of connecting c is defined (relative to $\mathcal{P}$) as

$$\rho(\xi, \delta) = \mathbb{E}\left[\rho_0(\xi(A_{i_1}, A_{i_2}, \ldots), d_\delta) + c\delta\right]$$

for all $\delta \in \Delta_{\mathcal{D}}$.

Chow and Robbins (1963, theorem 1) prove that for any decision problem there is a decision procedure that minimizes the risk. This justifies the following definition.

Definition 10 (optimal procedure)**.** Let $\mathcal{P} = (\Omega, \mathcal{W}, \xi)$ be a probability space and $\mathcal{D} = (\mathcal{C}, D)$ a decision problem. Let ρ be a risk function (for some loss ℓ and cost of connecting c). For any information set A, define $\delta_A^{\mathcal{D}}$ to be an optimal decision procedure after conditioning on A, i.e., a procedure (pick an arbitrary one if there are multiple) that satisfies

[12]This aspect of the model may reflect such real world costs as the opportunity cost of the time spent reading the paper. Alternatively, it may be viewed as a technical device preventing scientists from connecting to every scientist: it is a theorem that Bayesian scientists always prefer more information over less if the information is free (Good 1967). Ruling this out allows me to study whose information the scientists consider to be more valuable.

$$\rho\left(\xi(A),\delta_A^{\mathcal{D}}\right) = \inf_{\delta\in\Delta_{\mathcal{D}}} \rho(\xi(A),\delta).$$

This notation allows me to compare the optimal procedures after a connection has been made. I can now state a theorem to the effect that a close analogy of assumption 4 holds in this Bayesian framework.

Theorem 11. *Let $\mathcal{P} = (\Omega, \mathcal{W}, \xi)$ be a probability space and $\mathcal{D} = (\mathcal{C}, D)$ a decision problem, with $\mathcal{C}$ simple relative to $\mathcal{P}$. Let ρ be a risk function (specified relative to $\mathcal{P}$ for a given loss function ℓ and cost of connecting $c > 0$). Suppose that i's information set contains at least as much information as i''s: $A_i \subseteq A_{i'}$. Then*

$$\mathbb{E}\left[\rho\left(\xi(A_i),\delta_{A_i}^{\mathcal{D}}\right)\right] + c \geq \mathbb{E}\left[\rho\left(\xi(A_{i'}),\delta_{A_{i'}}^{\mathcal{D}}\right)\right] + c.$$

The inequality is strict if and only if there is a set of possible outcomes of $A_{i'}$ with positive probability such that if $X_{j,i,k} = X_{j,i',k}$ for all $1 \leq j \leq m$ and for all $1 \leq k \leq n(i,j)$, $\delta_{A_i}^{\mathcal{D}}$ is not optimal when conditioning on $A_{i'}$:

$$\rho\left(\xi(A_{i'}),\delta_{A_i}^{\mathcal{D}}\right) > \rho\left(\xi(A_{i'}),\delta_{A_{i'}}^{\mathcal{D}}\right).$$

This result says that there always exists an optimal decision procedure that satisfies assumption 4. If all information is relevant (in the sense specified in the theorem) then every optimal decision procedure satisfies assumption 4. A proof of the theorem is in Heesen (2014).

Let $\delta_{A_i}^{\mathcal{D}}(A)$ denote the number of scientists with information set A that procedure $\delta_{A_i}^{\mathcal{D}}$ connects to. This notation is needed to state the final theorem.

Theorem 12. *Let $\mathcal{P} = (\Omega, \mathcal{W}, \xi)$ be a probability space, D a set, and ρ a risk function (specified relative to $\mathcal{P}$ for a given loss function ℓ and cost of connecting $c > 0$). For any relative frequency q, $\varepsilon > 0$, and information set A, there exists $N_\varepsilon(A)$ such that for all $n > N_\varepsilon(A)$, if $\mathcal{C} = (I, \{A_i \mid i \in I\})$ is a scientific community satisfying q with $\mathcal{C}$ simple relative to $\mathcal{P}$, $\mathcal{D} = (\mathcal{C}, D)$*

the associated decision problem, $i \in I$ a scientist, and $\Pr\left(\delta^{\mathcal{D}}_{A_i}(A) \geq 1\right) > 0$, *then*

$$n \Pr\left(\delta^{\mathcal{D}}_{A_i}(A) \geq n \mid \delta^{\mathcal{D}}_{A_i}(A) \geq 1\right) \leq \varepsilon.$$

This shows that a scientific community in which each scientist has the same prior ξ and the same risk function ρ satisfies assumption 3. Note that scientist i gets the information in information set A_i from her own experiments, and updates on that. This is why $\delta^{\mathcal{D}}_{A_i}$ is the correct optimal procedure for her.[13] For a proof of the theorem I refer once again to Heesen (2014).

Theorems 11 and 12 can be read in three ways. The first reading is normative: if the standards of Bayesian rationality apply to scientists in my model, then it follows that they should satisfy the two assumptions of theorem 6. The second reading is descriptive: if actual scientists behave approximately like fully Bayesian agents, then they will (approximately) satisfy the two assumptions. The third reading justifies my earlier claim that the two assumptions represent a form of bounded rationality: the theorems show that the assumptions are necessary requirements of rationality (and it is easy to show that they are not sufficient).

It follows from theorems 11 and 12 that a large enough group of Bayesian scientists with identical priors and risk functions will satisfy the requirements of theorem 6. That is, the expected average in-degrees of these scientists are a supermodular function of their information sets.

Corollary 13. *Let $\mathcal{P} = (\Omega, \mathcal{W}, \xi)$ be a probability space, D a set, and ρ a risk function (specified relative to $\mathcal{P}$ for a given loss function ℓ and cost of connecting $c > 0$). For any relative frequency q there exists a number N*

[13]This is important here, because all scientists are assumed to have the same prior before seeing their own information set. In the setting of theorem 11 only one scientist is considered at a time, so there is no need to do the same thing there: one may assume that her prior incorporates the information from her own information set and any previous connections.

such that for all communities $\mathcal{C} = (I, \{A_i \mid i \in I)$ *satisfying* q *with* $\mathcal{C}$ *simple relative to* $\mathcal{P}$ *the following holds. For any* $i \in I$ *there exists a procedure* $\delta^{\mathcal{D}}_{A_i}$ *that is optimal for scientist* i *facing decision problem* $\mathcal{D} = (\mathcal{C}, D)$ *and such that if every scientist* i *follows procedure* $\delta^{\mathcal{D}}_{A_i}$ *and if* $|I| > N$ *then for all information sets* A *and* B *with* $q(A \cup B) > 0$

$$\mathbb{E}\left[d(A \cup B)\right] + \mathbb{E}\left[d(A \cap B)\right] \geq \mathbb{E}\left[d(A)\right] + \mathbb{E}\left[d(B)\right].$$

This concludes my discussion of the mathematical results of my model. Recall that the main goal of this paper is to investigate potential explanations for differences in prominence among scientists. The key mathematical contributions to this question are theorem 6 and its Bayesian analogue corollary 13.

The former result says that, subject to some assumptions, small differences in the initial information possessed by scientists lead to large differences in the number of connections to those scientists. The latter result shows that a community of Bayesian scientists may satisfy these assumptions. The two results together show that in a wide range of circumstances, academic superstars arise even if scientists are only motivated by their desire for information.

In the next three sections, I discuss in more detail three potential explanations of how an individual scientist might become a superstar. In each case I will say something about how the mathematical results obtained here relate to this explanation.

5 The Scientific Competence Explanation

The first explanation claims that differences in the number of times scientists' papers get read are the result of differences in the quality of their work. The thought is simple: better scientists obtain more information relevant to a given problem in less time, and having more information leads to more people wanting to read their papers.

The idea that high quality papers get cited more is fairly common in the literature. In fact Cole and Cole (1967, 1968) simply identify the two, using citations as a measure of quality in pursuing the question whether quality of publications is important in getting recognition for one's research. Cole and Cole (1971) and Clark (1957, chapter 3) give some support for this identification, while Lindsey (1989) criticizes it. Assuming that reading and citing a paper are highly correlated, it follows immediately that high quality papers get read more. If competent scientists tend to produce high quality papers, then competent scientists can expect their papers to be read more. Rosen (1981) shows that this is true in an economic model where scientists put value on reading papers by scientists with high talent.

One scientist could be more competent than another for any number of reasons, ranging from a higher general intelligence to better training to something as seemingly trivial as having a better eye for accurate measurement readings. Any of these reasons could be sufficient for one scientist ending up with better data than the other.

It is then easy to see how the model considered in this paper fleshes out the details of the explanation. Consider the following toy model. Suppose there is a set P of propositions that scientists want to learn the truth-value of. Assume that a scientist can be either competent or incompetent. Each incompetent scientist learns the truth-value of m propositions in P, while each competent scientist learns the truth-value of n propositions in P, with $m < n$. This is an instance of the model described in section 2.

Since the competent scientists obtain strictly more information than the incompetent ones, it follows from theorem 6 that competent scientists can expect their information to be in higher demand than incompetent scientists. While this is just a toy model, it is easy to see how the model in this paper can also capture situations where the relation between the competence of a scientist and the information she learns is more complicated. Theorem 6 guarantees that more competent scientists can expect their papers to be read more often than less competent scientists in each of these situations, as long as competence is somehow correlated with information.

If the scientific competence explanation is broadly correct, it yields an easy way of figuring out which scientists are competent and which ones are not: the competent ones are the ones that get cited the most (assuming that being read and being cited are correlated). So under this explanation one can infer in both directions: from competence to many citations and from many citations to competence.

What policy recommendations result from this explanation? If the reason scientists read some scientists more than others is because these scientists are better scientists and thus have more valuable information, it would seem to be best for everyone if this situation was maintained. Individual scientists gain the most by reading those scientists who have the most relevant information, and it also seems to be in the interest of science as a whole that its best work gets the most attention. Policy makers trying to promote the epistemic output of science would do well to stay away from policies that would hinder this process.

All seems well so far (from the perspective of science as a whole). But of course the fact that the scientific competence explanation in combination with theorem 6 makes a coherent story does not prove that story to be the correct one. In the next two sections I discuss two alternatives.

6 The Sociological Explanation

The second explanation I will consider is what I call the "sociological explanation". This explanation claims that certain epistemically irrelevant factors cause some scientists' work to garner more attention than others'. The literature identifies many factors that influence a scientist's prominence (measured in terms of being able to get work published, getting citations, receiving awards, their work being viewed as "credible", etc.). Some such factors include the scientist's (or her institution's) reputation (Merton 1968), the reviewers that get assigned to her work (Cole et al. 1981), age (Kuhn 1962, Zuckerman and Merton 1973), whether the work is available through open access (Greyson et al. 2009), being associated with prestigious scientists (Latour and Woolgar 1986), and prejudice based on gender, race, or academic affiliation (Fricker 2007).

What these factors have in common is that they are presumably epistemically irrelevant: white male scientists at prestigious institutions may get read more, but this is no indication that their work is of higher quality than that of their peers. If these factors are indeed causing some scientists' work to get more attention than others', it would appear that the work of some scientists is getting overvalued (and that of others undervalued) relative to their epistemic merit. For example, Hutchison makes this point about the role of gender in the field of philosophy: "[W]omen in philosophy are particularly susceptible to having their credibility underestimated, and thus being denied authority for the wrong reasons" (Hutchison 2013, p. 111).[14]

From the perspective of, say, funding agencies, this is a serious problem. They try to give grants based on merit, i.e., based on who is likely to make good contributions to science in the future. But if merit is (partially) measured by prominence (e.g., via citation metrics), the agencies will find themselves exacerbating existing biases, rather than rewarding excellence. Because of this and other reasons, if the sociological explanation is true, one should want to reduce or eliminate its effects (and expect overall scientific output to be improved by doing so).

Epistemically irrelevant factors like the ones I just discussed are completely absent in the model of section 2. How attractive a given scientist's work is depends only on the information they have concerning the epistemic problem(s) scientists are facing. I have shown that in this model some scientists get read much more than others. What exactly does this show regarding the sociological explanation?

[14] Healy (see footnote 1) seems to endorse this hypothesis as a way of explaining why women are underrepresented in his dataset (see `http://kieranhealy.org/blog/archives/2013/06/24/citation-networks-in-philosophy-some-followup/`).

It is important to emphasize that my model does not show the sociological explanation to be wrong. Epistemically irrelevant factors certainly exist in real life, and it is entirely possible that they contribute to the phenomenon that some scientists get read more than others.

What the model does show is that the presence of epistemically irrelevant factors is not necessary for the phenomenon of interest to arise. I have shown that even if scientists were (counterfactually) completely blind to epistemically irrelevant factors, some scientists would still get read more than others. Moreover, the degree to which they are read more might seem disproportionate to the difference in quality between them and the scientists whose work gets read less. Thus, I have disproven claims like the following: "The fact that some scientists' work gets read much more than other scientists' work shows (by itself!) that scientists' decisions whom to read are biased by epistemically irrelevant factors".

The above claim would be true if the sociological explanation was the only possible explanation, but I have shown that it is not: it is (mathematically) possible for the phenomenon of interest to arise in the absence of epistemically irrelevant factors, due to scientists having different information.

To summarize: under the sociological explanation, epistemically irrelevant factors cause different scientists' work to get different levels of attention. If this is right, it suggests first that one cannot infer from the number of citations of a paper anything about its (epistemic) quality, and second that policy that encourages scientists to pay (more) equal attention to different scientists' work is a good idea. The model in this paper ignores epistemically irrelevant factors, thereby showing that alternatives to this story exist, but the model certainly does not prove the sociological explanation to be false.

The policy recommendations resulting from the sociological explanation are completely opposite to those resulting from the scientific competence explanation. If epistemically irrelevant factors are driving who gets read, it appears to be a good idea to try to remove those factors or counterbalance their effects. If scientific competence (an epistemically relevant factor) drives who gets read, it would be wholly counterproductive to implement such policies.

It is quite plausible that in reality both epistemically relevant and epistemically irrelevant factors contribute to the differences in prominence among scientists. This complicates the issue even further: all things considered, should policy makers be trying to level out these differences or not? Such attempts could help or harm science, and it is difficult to figure out which.

One thing can be said with relative confidence: policies that remove epistemically irrelevant factors or make scientists less sensitive to them, without directly influencing scientists' decisions (say, about whom to read), should only have positive effects. Such policies should make it more likely that if there are differences among scientists in terms of the amount of attention their work gets, these differences exist for the right reason, i.e., an epistemically relevant one.

However, this conclusion assumes that factors that influence who gets read can be neatly separated into epistemically relevant and epistemically irrelevant ones. The third and final explanation I consider challenges the neatness of this distinction.

7 The Epistemic Luck Explanation

Some scholars have identified dealing with anomalies or unexpected results as a central feature of scientific research (Kuhn 1962, Dunbar and Fugelsang 2005). The third explanation proceeds from the assumption that some amount of luck is involved in getting the kind of unexpected result that leads to an important paper. Under this explanation, it is the lucky rather than the competent scientists who end up with the largest information set and thus get read the most.

Stories involving epistemic luck (or serendipity) are very common in the history of science (Roberts 1989, van Andel 1994). Penicillin's ability to kill bacteria, for example, was discovered when a Petri dish was accidentally left open overnight. Such lucky accidents plausibly have nothing to do with the scientist's competence, or even any specific sociological factor (but see McKinnon 2014 and Merton and Barber 2004, chapter 9, for some discussion of the relation between luck and merit).

A variation on the toy model I considered above shows how this could function as an explanation for the phenomenon of interest. Suppose there is a set P containing n propositions that scientists want to learn the truth-value of. In this simple model, each scientist has a chance α of learning the truth-value of any given proposition, independent of all other propositions and scientists. The lucky scientists who learn the truth-value of all n propositions (which happens with probability α^n) have the highest expected in-degree in this model (this follows from theorem 6; see also Anderson 2011, section 3, for a detailed discussion of this particular model).

This toy model shows that even in the absence of either sociological factors or differences in competence one might get different information sets, which lead to some scientists being read much more often than others. As far as I am aware, the idea that some scientists could be more prominent than others without sociological factors or differences in competence causing this phenomenon has never before been explicitly suggested in the literature. Yet the above shows how this may happen in a mathematically precise sense.

It is interesting to note that if the scientific competence and the epistemic luck factor are brought into play at the same time, the luck factor can easily drown out the competence factor.

To see this, consider again the toy model where there are n propositions the scientists want to learn about, and there is a fixed probability of learning any given proposition, independent of the other ones. To reflect the competence factor, assume that there are two types of scientists: average ones, whose probability of learning a proposition is α, and good ones, whose probability of learning a proposition is β ($0 < \alpha < \beta < 1$).

Let p denote the proportion of good scientists (so $1 - p$ is the proportion of average ones). It seems plausible that good scientists are relatively rare:

most scientists are of average quality. Now it turns out that if good scientists are sufficiently rare, the chance that a scientist with high in-degree is a good scientist may be arbitrarily small.

To make this more precise, suppose one draws a scientist at random from the population. Let g denote the proposition that the scientist drawn is a good scientist (so $\neg g$ means drawing an average scientist) and let h denote the proposition that the scientist drawn has a high in-degree.

Proposition 14. *Assume that average scientists learn with probability α and good scientists learn with probability β (where $0 < \alpha < \beta < 1$ and the probability of learning any given proposition is independent of the probability of learning any other proposition). Then for all $\varepsilon > 0$ there exists a proportion of good scientists $p \in (0, 1)$ such that $\Pr(g \mid h) \leq \varepsilon$.*

Proof. Let $\varepsilon > 0$. If $\varepsilon \geq 1$ then $\Pr(g \mid h) \leq \varepsilon$ is true for any p. Otherwise choose

$$p = \frac{\alpha^n \varepsilon}{\beta^n(1-\varepsilon) + \alpha^n \varepsilon}.$$

It follows from theorem 6 that those scientists who learn all n propositions will have the highest in-degree (see also Anderson 2011, theorem 5). Therefore

$$\begin{aligned}
\Pr(h \mid g) &= \beta^n, \\
\Pr(h \mid \neg g) &= \alpha^n, \\
\Pr(g) &= p = \frac{\alpha^n \varepsilon}{\beta^n(1-\varepsilon) + \alpha^n \varepsilon}, \\
\Pr(\neg g) &= 1 - p = \frac{\beta^n(1-\varepsilon)}{\beta^n(1-\varepsilon) + \alpha^n \varepsilon}, \\
\Pr(g \mid h) &= \frac{\Pr(h \mid g)\Pr(g)}{\Pr(h \mid g)\Pr(g) + \Pr(h \mid \neg g)\Pr(\neg g)} \\
&= \frac{\frac{\beta^n \alpha^n \varepsilon}{\beta^n(1-\varepsilon)+\alpha^n\varepsilon}}{\frac{\beta^n \alpha^n \varepsilon + \alpha^n \beta^n (1-\varepsilon)}{\beta^n(1-\varepsilon)+\alpha^n\varepsilon}} = \frac{\alpha^n \beta^n \varepsilon}{\alpha^n \beta^n} = \varepsilon.
\end{aligned}$$

□

So I can make the proportion of high in-degree scientists who are good scientists arbitrarily small by making the overall proportion of good scientists very small. If one thinks that both scientific competence and epistemic luck have a role to play in determining how much valuable data a scientist obtains from her experiments, and if one also thinks that good scientists are quite rare, then if a scientist gets read a lot (and gets cited a lot, if getting cited and getting read are correlated) it is not good evidence that she is a good scientist. Thus the inference from many citations to competence that is valid when the scientific competence explanation is the only correct one is invalid if epistemic luck is a factor.

If the epistemic luck explanation is the correct explanation of the phenomenon that some scientists get read much more than most, what does this mean for policy makers? If the sociological explanation is correct, policy makers should do something, while if the scientific competence explanation is correct, they should do nothing. If the epistemic luck explanation is correct, things are not so clear.

On the one hand, scientists are reading scientists with the most useful information. This seems like a good practice that one might not want to disrupt. On the other hand, the scientists that are being read have the most useful information because they got lucky, not because they are better scientists. The prominence of these scientists says little if anything about the quality of their work. So future work from those scientists is not more likely to be interesting than that of a less prominent scientist.

This suggests a separation between two questions that might otherwise have been easy to conflate. If one is interested in awarding credit (say, a Noble prize) for past contributions to science, it seems reasonable to look primarily at the informational value of the contributions, and not worry about whether this value was primarily the result of exceptional competence or exceptional luck. But if one is interested in who is most likely to make important future contributions (say, when awarding research grants), it would be important to recognize whether past success was due to competence or luck, as presumably competent scientists are more likely than average scientists to produce valuable work in the future, while lucky scientists are not.

If the epistemic luck explanation is largely correct, it makes citation counts specifically and prominence more generally much less useful as a way of separating the wheat from the chaff when decisions concerning future projects need to be made. This would be important to know not just for scientists considering whom to read, collaborate with or hire for new projects, but also for policy makers, funding agencies, future graduate students, and the general public.

The epistemic luck explanation differs from the scientific competence explanation in that it does not support the idea that having a high number of citations can act as a signal indicating the merit of a particular scientist. In this sense the epistemic luck explanation blurs the line between epistemically relevant factors like scientific competence and epistemically irrelevant factors like a scientist's charisma, sex, or race.

As I alluded to in the previous section, it is entirely possible that more than one of the explanations I have discussed is true. Perhaps both scientific competence and epistemic luck contribute to differences in information among scientists, which leads to some scientists' work getting more attention than others', while epistemically irrelevant factors cause biases in scientists' choices of whom to read that either exacerbate or weaken the effects of the differences in information.

If this is correct, it is unclear whether policy addressing this phenomenon will have a positive or negative overall effect (although policy aimed specifically at reducing biases due to epistemically irrelevant factors may still be defensible). One of the lessons from this paper should then be not to jump to the conclusion that just because some scientists are more prominent than others some particular factor must be causing it: there are many factors that could cause this, and policy that would be a good idea if some particular factor is the cause may be counterproductive if another factor is the cause, or if multiple factors are at work.

8 Conclusion

In this conclusion I emphasize two key take-aways from the paper. First, theorem 6 shows that in my model, small differences in the information scientists gain from experiments can lead to large differences in prominence, such that a small group of scientists (the superstars) are widely read, while the vast majority is more or less ignored. This shows that scientists' desire for information can, on its own, produce differences in prominence equal to those observed in actual science. Thus, even when the differences in prominence among scientists may appear to the naked eye to be too large to be due to differences in quality, one should be careful not to conclude too quickly that outside forces or biases have interfered with the scientific process.

Second, in thinking about what might cause the differences in information that give rise to academic superstars, I have considered the novel proposal that there might be an important role for luck or randomness. A lucky scientist, in the sense of this paper, is one who obtained important scientifically relevant information through luck. This suggests a distinction, relevant to policy makers, that has not previously been emphasized. The distinction is between funding that explicitly rewards past achievement (which a lucky scientist would deserve) and funding that is based on the expectation of future achievement. Since the lucky scientist is not more likely than an average scientist to be successful in the future, it would be wrong to view her past succes as a reason to award her funding of the latter kind. This shows that one should be careful in using measures of past success (like citation metrics) to decide who is likely to do well in the future.

References

Katharine A. Anderson. Collaboration network formation and the demand for problem solvers with heterogenous skills. 2011. URL `http://arxiv.org/pdf/1112.5121v1.pdf`.

Robert J. Aumann. Agreeing to disagree. *The Annals of Statistics*, 4(6): 1236–1239, 1976. ISSN 00905364. URL `http://www.jstor.org/stable/2958591`.

Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999. ISSN 00368075. URL `http://www.jstor.org/stable/2899318`.

Y.S. Chow and Herbert Robbins. On optimal stopping rules. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 2(1):33–49, 1963. ISSN 0044-3719. doi: 10.1007/BF00535296. URL `http://dx.doi.org/10.1007/BF00535296`.

Kenneth E. Clark. *America's Psychologists: A Survey of a Growing Profession*. American Psychological Association, Washington, DC, 1957.

Jonathan Cole and Stephen Cole. Measuring the quality of sociological research: Problems in the use of the "science citation index". *The American Sociologist*, 6(1):23–29, 1971. ISSN 00031232. URL `http://www.jstor.org/stable/27701705`.

Jonathan R. Cole. Patterns of intellectual influence in scientific research. *Sociology of Education*, 43(4):377–403, 1970. ISSN 00380407. URL `http://www.jstor.org/stable/2111839`.

Stephen Cole and Jonathan R. Cole. Scientific output and recognition: A study in the operation of the reward system in science. *American Sociological Review*, 32(3):377–390, 1967. ISSN 00031224. URL `http://www.jstor.org/stable/2091085`.

Stephen Cole and Jonathan R. Cole. Visibility and the structural bases of awareness of scientific research. *American Sociological Review*, 33(3): 397–413, 1968. ISSN 00031224. URL `http://www.jstor.org/stable/2091914`.

Stephen Cole, Jonathan R. Cole, and Gary A. Simon. Chance and consensus in peer review. *Science*, 214(4523):881–886, 1981. ISSN 00368075. URL http://www.jstor.org/stable/1686309.

Morris H. DeGroot. *Optimal Statistical Decisions*. John Wiley & Sons, New Jersey, 2004.

Kevin N. Dunbar and Jonathan A. Fugelsang. Causal thinking in science: How scientists and students interpret the unexpected. In Michael E. Gorman, Ryan D. Tweney, David C. Gooding, and Alexandra P. Kincannon, editors, *Scientific and Technological Thinking*, chapter 3, pages 57–79. Lawrence Erlbaum Associates, Mahwah, 2005.

Miranda Fricker. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press, Oxford, 2007.

John D. Geanakoplos and Heraklis M. Polemarchakis. We can't disagree forever. *Journal of Economic Theory*, 28(1):192 – 200, 1982. ISSN 0022-0531. doi: 10.1016/0022-0531(82)90099-0. URL http://www.sciencedirect.com/science/article/pii/0022053182900990.

I.J. Good. On the principle of total evidence. *The British Journal for the Philosophy of Science*, 17(4):319–321, 1967. ISSN 00070882. URL http://www.jstor.org/stable/686773.

Devon Greyson, Steven Morgan, Gillian Hanley, and Desy Wahyuni. Open access archiving and article citations within health services and policy research. *Journal of the Canadian Health Libraries Association*, 30(2): 51–58, 2009. doi: 10.5596/c09-014. URL http://dx.doi.org/10.5596/c09-014.

Remco Heesen. Interaction networks with imperfect evidence. Technical Report CMU-PHIL-191, Carnegie Mellon University, 2014. URL http://www.hss.cmu.edu/philosophy/techreports/191_Heesen.pdf.

Katrina Hutchison. Sages and cranks: The difficulty of identifying first-rate philosophers. In Katrina Hutchison and Fiona Jenkins, editors, *Women in Philosophy: What Needs to Change?*, chapter 5, pages 103–126. Oxford University Press, Oxford, 2013.

Christoph Kelp and Igor Douven. Sustaining a rational disagreement. In Henk W. de Regt, Stephan Hartmann, and Samir Okasha, editors, *EPSA Philosophy of Science: Amsterdam 2009*, volume 1 of *The European Philosophy of Science Association Proceedings*, pages 101–110. Springer Netherlands, 2012. ISBN 978-94-007-2403-7. doi: 10.1007/978-94-007-2404-4_10. URL http://dx.doi.org/10.1007/978-94-007-2404-4_10.

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

Bruno Latour and Steve Woolgar. *Laboratory Life: The Construction of Scientific Facts*. Princeton University Press, Princeton, New Jersey, 1986.

D. Lindsey. Using citation counts as a measure of quality in science measuring what's measurable rather than what's valid. *Scientometrics*, 15(3-4):189–203, 1989. ISSN 0138-9130. doi: 10.1007/BF02017198. URL http://dx.doi.org/10.1007/BF02017198.

Rachel McKinnon. You make your own luck. In Duncan H. Pritchard and Lee John Whittington, editors, *The Philosophy of Luck*. Routledge, London, 2014. Forthcoming.

Robert K. Merton. The Matthew effect in science. *Science*, 159(3810):56–63, 1968. ISSN 00368075. URL http://www.jstor.org/stable/1723414.

Robert K. Merton and Elinor Barber. *The Travels and Adventures of Serendipity: A Study in Sociological Semantics and the Sociology of Science*. Princeton University Press, Princeton, 2004.

Derek J. de Solla Price. Networks of scientific papers. *Science*, 149(3683): 510–515, 1965. ISSN 00368075. URL `http://www.jstor.org/stable/1716232`.

S. Redner. How popular is your paper? An empirical study of the citation distribution. *The European Physical Journal B - Condensed Matter and Complex Systems*, 4(2):131–134, 1998. ISSN 1434-6028. doi: 10.1007/s100510050359. URL `http://dx.doi.org/10.1007/s100510050359`.

Royston M. Roberts. *Serendipity: Accidental Discoveries in Science*. John Wiley & Sons, New York, 1989.

Sherwin Rosen. The economics of superstars. *The American Economic Review*, 71(5):845–858, 1981. ISSN 00028282. URL `http://www.jstor.org/stable/1803469`.

Pek van Andel. Anatomy of the unsought finding. Serendipity: Origin, history, domains, traditions, appearances, patterns and programmability. *The British Journal for the Philosophy of Science*, 45(2):631–648, 1994. ISSN 00070882. URL `http://www.jstor.org/stable/687687`.

Harriet Zuckerman and Robert K. Merton. Age, aging, and age structure in science. In Norman W. Storer, editor, *The Sociology of Science: Theoretical and Empirical Investigations*, chapter 22, pages 497–559. The University of Chicago Press, Chicago, 1973.