Three Ways To Become An Academic Superstar*

Remco Heesen†

March 30, 2015

Abstract

I show that academic superstars arise quite naturally in a scientific community where scientists are motivated only by epistemic goals, in particular a desire for information. This is consistent with a "scientific competence" explanation for superstars on which being a superstar is an indicator of good scientific work. But two alternative explanations complicate this picture. In particular, if luck plays a role in determining the information scientists obtain, being a superstar does not necessarily reflect well on the epistemic virtues of a scientist, even when all scientists are motivated only by epistemic goals.

*Thanks to Kevin Zollman, Teddy Seidenfeld, Katharine Anderson, Tomas Zwinkels, Liam Bright, Aidan Kestigian, Conor Mayo-Wilson, three anonymous referees, and audiences at Ghent University, Oxford University, the California Institute of Technology, and the Munich Center for Mathematical Philosophy for valuable comments and discussion. This work was partially supported by the National Science Foundation under grant SES 1254291.

†Department of Philosophy, Baker Hall 161, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA. Email: rheesen@cmu.edu

1 Introduction

Prestige is divided unequally in science. One measure of the prestige of a scientist is the number of times she is cited. It has long been known that the vast majority of scientists (the "nobodies") receives no more than a handful of citations, while a rare few (the "superstars") get extremely many (Price 1965, Cole 1970).[1]

This raises several questions. Is this inequality the result of non-epistemic forces from outside of science interfering or is it a natural part of a well-functioning science? What, if anything, does having a lot of prestige say about a scientist's merit as a scientist? And how can this information be used by policy makers?

The answers you give to these questions are tied deeply to your view of science, and scientists' motivations in particular.

On one view, traditionally defended by philosophers, scientists are truth-seekers (Kitcher 1993, chapter 1). Each scientist's goal is to obtain knowledge. If scientists are motivated in this way, it seems natural to think that they would pay more attention to better work (an assumption often taken for granted by empirical work, e.g., Cole and Cole 1967, 1968). I will call this the "scientific competence explanation" of superstars. On this explanation, differences in prominence among scientists are justified from an epistemic point of view, and reflect positively on the scientists that benefit from them (the superstars).

On another view, defended mostly by sociologists, scientists' behavior is largely shaped by the social context or other epistemically irrelevant factors. On this view it is primarily these epistemically irrelevant factors which determines which scientists attain prominence; call this the "sociological explanation". On a small scale, this might mean that the scientist with friends in the right places becomes a superstar, whereas the one without does not (Latour

[1]More specifically, the distribution of citations follows a "power law": 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.

and Woolgar 1986), or that racial, gender, or other biases affect scientists' chances of attaining prominence (Fricker 2007). On a larger scale, the social context may require a certain kind of scientific theory (giving prominence to its defenders) or may affect the role of science in society, thus leading to more or fewer prominent scientists (Simonton 1975, Kroeber 1944). Proponents of this view may even think that the existence of superstars is itself evidence that there are epistemically irrelevant factors influencing who gets cited (a point to which I will return). On this view the divide between superstars and nobodies may seem arbitrary: academic superstars are not special in any epistemically or morally relevant way.

In this paper I pursue a systematic comparison of these two explanations for the existence of superstars.[2] I also introduce a third explanation, which turns out to share features of both in an interesting way.

First I will describe a possible mechanism, based on scientists' desire for information, that produces superstars. I show that this mechanism works as advertised in a model of information exchange among scientists (see sections 2 and 3). My goal in these two sections is to prove a mathematically precise version of the following claim.

Claim 1. *If, in choosing whose work to read, scientists are motivated by gathering as much information as possible given their means, then the patterns of interaction that emerge are highly imbalanced (in a way that is consistent with the empirical evidence): some scientists get a lot of attention, while most get very little.*

If this claim is established, it shows how scientists' desire for information can work as a mechanism that leads to the existence of academic superstars.

[2]There are many nuances to be made to this simple picture, as recent work in social epistemology has shown. In particular, Kitcher (1993) and subsequent work has focused on showing how scientists motivated by what I call epistemically irrelevant factors might still form an epistemically successful community. In contrast, one of the upshots of the present work is that a community of truth-seeking scientists may not have all the nice features one would naively expect (see section 6).

It provides some evidence that superstars arise naturally in a well-functioning science. At the very least, this means that the inference from "there exist superstars" to "there must be epistemically irrelevant factors influencing scientists' decisions" is not justified (in the absence of more specific evidence).

The results in section 3 quite naturally cohere with the scientific competence explanation of superstars I outlined above: if competent scientists obtain more valuable information than others, they become superstars. Thus the mathematical results provide an underlying mechanism to the competence explanation. Together they suggest that superstars are better scientists than nobodies. I flesh out this explanation and explore the view on superstars that it entails in section 4.

In contrast, section 5 sets out the sociological explanation in further detail. The sociological explanation does not depend on scientists' desire for information, and so it works quite independently of my mathematical results. I discuss the contrasting view of superstars that follows from this explanation.

Finally, I present a third explanation on which individuals become superstars based on (epistemic) luck (section 6). Here luck plays a role in determining which scientists get useful information. If it is assumed that scientists have a desire for information, the mathematical results of section 3 apply to this third explanation just as much as they apply to the first. But here there is no reason to assume that the resulting superstars are better scientists than the nobodies.

So this third explanation is of particular interest because it shows that superstars need not be the most epistemically virtuous members of the scientific community, even under the (strong) assumption that scientists' motivations are epistemically pure. I explore the implications of this explanation, in particular for funding agencies. A short conclusion wraps up the paper.

2 A Model of Information Exchange

The goal of this model is to capture important aspects of the way scientists exchange information "in the short run", that is, in the context of one research project or paper. This section describes the model, with a focus on indicating and defending some of the more important modeling choices that have been made.

Consider a scientific community, consisting of some number of scientists. Call the set of scientists I, with each element $i \in I$ representing an individual scientist. I is most naturally thought of as either the practitioners of some given scientific discipline (small or large) or all of science. I is assumed to be finite.

The scientists are interested in learning about the world. This is represented by assuming that there are multiple ways the world could be (collected in a set Ω), and that scientists are interested in distinguishing between some of these (i.e., figuring out which world, or subset of worlds, is or contains the actual world).[3] I make 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 ways the world could be.

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. Each scientist's research involves doing each of these experiments some (possibly zero) number of times. Write $n(i, j)$ to denote the number of times scientist i performs experiment j.

The set of results of the experiments of a given scientist i is called sci-

[3]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.

entist i's information set A_i. Experiments are modeled as random variables, with different experiments corresponding to different probability distributions, so an information set is a set of random variables. Information set A_i thus contains $n(i, j)$ random variables for each experiment j.

Each scientist publishes her information in a paper. For simplicity, I assume that each scientist publishes a single paper, and that this paper contains all the information in her information set.[4]

This leads to a second way in which scientists can learn about the world: they can read each other's papers. Reading a paper means learning the information in the information set of the scientist who wrote the paper.[5] In the context of the model, reading another scientist's paper is an action that a scientist can take. Because of the graphical interpretation I will later give to this action, I refer to it as "forming a connection". So when I say "scientist i connects to scientist i'" this just means that scientist i reads scientist i''s paper.

The type of information exchange that occurs when scientists read each other's work has some features that differentiate it from other ways scientists may exchange information. First, a scientist can read another's paper without prior consent by the other scientist. Second, the scientist who wrote

[4]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 7) 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.

[5]Scientists learn each other's experimental results (or 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).

the paper does not learn anything when someone reads her paper. Zollman (2013, section 2) calls this “one way” information transmission.

In contrast, consider a case where two scientists meet at a conference. Here both of these features are reversed: one scientist can only learn from the other if the other is willing to share her information, and information can potentially flow in both directions if both are willing to share (“two way” transmission in Zollman’s terminology).

A major difference with the information transmission models considered by Zollman (e.g., Jackson and Wolinsky 1996, Bala and Goyal 2000) is that there is no transfer of information in this model. That is, if scientist i has read scientist i'’s paper, and then scientist i'' reads scientist i’s paper, scientist i'' does not thereby learn the information in the information set of scientist i'.

This assumption is made for two reasons. First, scientist i’s paper presumably focuses on reporting scientist i’s experimental results, not those she learned from others. Second, even if some transfer of information happened, and scientist i'' learned something interesting about scientist i'’s work this way, one might expect her to then cite scientist i' as well, thus (for the purposes of this model) forming an independent connection with scientist i'.

Because this model considers only one way information transmission, without information transfer, forming a connection is an action by a scientist that affects only herself. This means that pairwise stability (Jackson and Wolinsky 1996) and other notions of equilibrium are not an issue.

Another difference is that the scientists in most information transmission models have very few individual characteristics other than those given by their place in the network (such as their number of neighbors). For example, in the models of Jackson and Wolinsky (1996) and Bala and Goyal (2000), each scientist has one “unit” of information, different from all other scientists’ units, and equally valuable.

One of very few models to incorporate more diversity among its scientists is found in Anderson (2011). She focuses on the formation of collaboration networks rather than citation networks, and her “skills” (comparable to information in my model) are binary and deterministic. By having information (in the form of a set of random variables) associated with each scientist, my model allows for even more varied communities of scientists.

Another strand of recent work on epistemic networks has compared the performance of different network structures on various epistemic desiderata (Zollman 2010, Grim et al. 2013). In this work the network structures being compared are fixed in advance by the modeler. Such work is thus complimentary to the type of model considered here, which focuses on the formation of the network.

The results obtained from this model are described in the next section. The remainder of this section considers some ways in which this model is unrealistic, and the conclusions to be drawn from that.

An important assumption 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 what number of each of the m types of experiments.

My justification for this assumption is as follows. Relative to the time and cost involved in designing and running an experiment, reading (and subsequently citing) someone else's paper is a very 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.

Even if this is accepted, the model unrealistically portrays science as consisting of only one round of experiments and one round of connections. In reality, scientists may form many connections over time, interspersed with experiments. But the model can be viewed as looking only at a small period of time in a scientist's career, say, the time associated 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. Depending on the details of how the content of a scientist's information set is constructed (which I discuss extensively in sections 4 and 6), the effect of considering multiple rounds may be to either amplify or dampen the results I obtain for a single round.

A different kind of objection questions whether there is any need for this model in the first place. It may be remarked that the characteristic pattern of information exchange and citations in the sciences (the so-called power law, with a few superstars and lots of nobodies) already has a standard generating mechanism in the literature. This mechanism is described by so-called preferential attachment models (Barabási and Albert 1999). 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 into why the difference between the two papers appeared in the first place (the preferential attachment model needs to start with some citations already in place for the probabilities of new citations 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, preferential attachment models give a fairly simplistic picture of scientists' motivations in citing a paper. They require that scientists have a preference for papers that are already highly cited (perhaps because those are more easily found) without giving any details for why it is rational for them to do so, or why they occasionally (with low probability) deviate from it.

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 there are superstars, and how being a superstar reflects on an individual scientist. In contrast, section 3 shows that in my model the information content of a paper acts as a predictor of future citations of the paper (theorem 7). The model is compatible with a wide range of motivations that one might ascribe to scientists, and some results are shown for the particular case in which the scientists are modeled as Bayesian expected utility maximizers (theorems 5 and 6). This work is highly suggestive in terms of evaluating superstars, as sections 4, 5, and 6 discuss.

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 (see the next section). The model ignores certain complicating factors.

But 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

How do scientists choose which papers to read (i.e., which connections to form)? One might want to assume that scientists have some form of utility function which they maximize. But this has a number of problems: whether (expected) utility maximization can provide a good model of rationality is controversial; even if it is a good model of rationality scientists may not act rationally so the descriptive power of the model may be poor; and even if those two points are dealt with one would need to argue for the specific form of the utility function, requiring a detailed discussion of scientists' goals.

Here I take a different approach. I state two assumptions, or behavioral rules, that constrain scientists' choices to some extent (although they still leave a lot of freedom). I then show that these assumptions are sufficient for the appearance of superstars in the model.

For those who think that, despite the problems I mentioned, scientists' behavior should be modeled using (Bayesian) expected utility theory, I show that a wide range of utility functions would lead scientists to behave as the assumptions require (see theorems 5 and 6). For those who are impressed by the problems of that approach, I argue that one should expect scientists to behave as the assumptions require even if they are not maximizing some utility function.

So what exactly do the scientists need to decide, and what do they know when making this decision? Each scientist needs to decide which other scientists to connect to, that is, whose papers to read. Before making any decisions, they know what experiments have been performed by each scientist (that is, the values of $n(i, j)$ for all i and j) as well as the results of their own experiments. I do not consider the order in which the different scientists make their decisions: that is, scientists do not know which connections other scientists are forming, or if they do, they ignore this information.

I allow that scientists make their decisions sequentially: after each connection they learn the results of the experiments of the scientist they connected to, and they may use this information in deciding which connection to form next (or whether to stop connecting). This reflects the idea that one may only become interested in reading a paper after reading some other paper. So scientists effectively choose sequential decision procedures, which specify which connections to form as a function of information gained from earlier connections. Because that information takes the form of random variables, the decision procedure itself is also random. A simple example illustrates this phenomenon.

Example 2. Suppose that there are two possible worlds, a and b. Suppose that there is one experiment and each scientist has performed that experiment once. In world a, the experiment outputs either a zero or a one, each with probability 1/2. In world b the experiment always outputs a one. So upon observing a zero a scientist is certain to be in world a.

Consider a scientist who initially thinks she is equally likely to be in either world and uses the decision procedure "form connections until you are at least 99% certain which world you are in". Assume world a is the actual world. Then she forms no connections with probability 1/2 (if her own experiment yields a zero), one connection with probability 1/4 (if she saw a one but the first scientist she connects to saw a zero), and so on. So a decision procedure does not specify which scientists to connect to, but it specifies the probability of connecting to them.

Now I can state the two assumptions I make on scientists' decision procedures. The first assumption says that scientists prefer to get more information rather than less from a connection.

Assumption 3 (Never Consider Subsets)**.** *A scientist does not connect to a second scientist i as long as a third scientist i' is available to connect to and information set $A_{i'}$ contains strictly more information than information set A_i (written $A_i \sqsubset A_{i'}$).*[6]

[6] $A_i \sqsubset A_{i'}$ if and only if $n(i,j) \leq n(i',j)$ for all j, where at least one of the inequalities is strict. This relation among information sets is closely related to the usual set-theoretic relation of inclusion. See Heesen (2014) for more on this.

So if scientist i' has performed more replications of the same experiments than scientist i, one would read scientist i''s paper rather than scientist i's (unless one prefers to just read both; the assumption allows that option as well). For example in the medical sciences, where the number of experiments might refer to the number of patients studied, a higher number of experiments would correspond to more reliable statistical tests, and would as such be preferable. In this way assumption 3 captures (a fairly weak version of) the idea that scientists are truth-seeking by positing a desire for information (cf. claim 1).

In some cases this assumption might seem unrealistic. For example in testing a medicine one might be more interested in very extreme outcomes (severe side effects or even death) that happened in a small sample than in a large sample where things went as expected. But note that it is an assumption of the model that the results of the trials are not known before the paper is read. Thus, before knowing of the extreme outcome, it would make sense to read the paper with a lot of trials before the one with less trials. If the extreme outcome is learned anyway, one would in most cases still want to read the larger study to get an idea of the frequency of the extreme outcome. So I argue that the assumption remains reasonable in most cases.

The second assumption gets its plausibility from the simple observation that there is a finite limit to how many papers a scientist can read, simply because it is humanly impossible to read more. More formally, there exists some number N (say, a million) such that for any scientist the probability that she forms more than N connections is zero.

But rather than making this assumption explicitly, I assume something strictly weaker: that the probability of a very large number of connections is very small (rather than zero).

Assumption 4 (Uniformly Bounded Connection Probabilities)**.** *Let $p_{i,A,n}$ denote the probability that scientist i connects to at least n scientists with* *information set A.*[7] *For every $\varepsilon > 0$, there exists a number N that does not depend on the scientist or the size of the scientific community, such that $n \cdot p_{i,A,n} \leq \varepsilon \cdot p_{i,A,1}$ for all $n > N$.*

Note that the actual statement of the assumption distinguishes the number of connections to a given information set. This is because for technical reasons, I need to distinguish between cases where $p_{i,A,1}$ is zero (i.e., the scientist never connects to any scientists with information set A) and cases where $p_{i,A,1}$ is positive. But for interpreting the assumption this is mostly irrelevant, because the number of connections to scientists with a given information set is always less than the total number of connections.

So the assumption says that for very high numbers of connections, the probability of forming that number of connections is very small, independent of the scientist forming the connections or the size of the scientific community. If, as I suggested earlier, no scientist ever forms more than a million connections, then $p_{i,A,n} = 0$ for all i and A whenever n is greater than a million, and so the assumption would be satisfied.

As I indicated, these assumptions are not only independently plausible, but are also satisfied by Bayesian scientists (who maximize expected utility) under quite general conditions. The most important of these conditions is that there is a fixed cost c for forming a connection. This cost may reflect such real world considerations as the opportunity cost of the time spent reading the paper. Note also that without such a cost the unrealistic and uninteresting result would be that scientists read every paper (Good 1967). For this reason a cost is commonly included in models of this kind (Zollman 2013, section 2).

The relation between my assumptions and Bayesian rationality is expressed in the following two theorems. Proofs of these theorems may be found in Heesen (2014).

[7] A scientist has information set A if her information set contains the same number of realizations of each experiment as A does. See Heesen (2014) for more on what this means formally.

Theorem 5. *If $c > 0$ and if each realization of an experiment is probabilistically independent and has a positive probability of changing the scientist's future choices, then the way a fully Bayesian rational scientist chooses connections satisfies assumption 3.*

Theorem 6. *If $c > 0$ and if each realization of an experiment is probabilistically independent, then a community of fully Bayesian rational scientists with the same prior probabilities over possible worlds and the same utility functions chooses connections in a way that satisfies assumption 4.*

With the two assumptions in place, 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'.[8]

In order to study the prominence of scientists in this network, I need a measure of prominence. A natural idea suggests itself: a scientist is prominent if a large number of scientists read her 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.

If a scientist learns something from another scientist, she usually acknowledges 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 (the superstars) get many citations and some very few, then one should expect large differences in in-degree among scientists. The following theorem says that this is exactly what happens in my model. The theorem relates the average in-degrees of scientists

[8] More formally, the network of interest is $G = (I, \{(i, i') \in I^2 \mid i \text{ connects to } i'\})$, where I is the set of nodes and $\{(i, i') \in I^2 \mid i \text{ connects to } i'\}$ is the set of arcs.

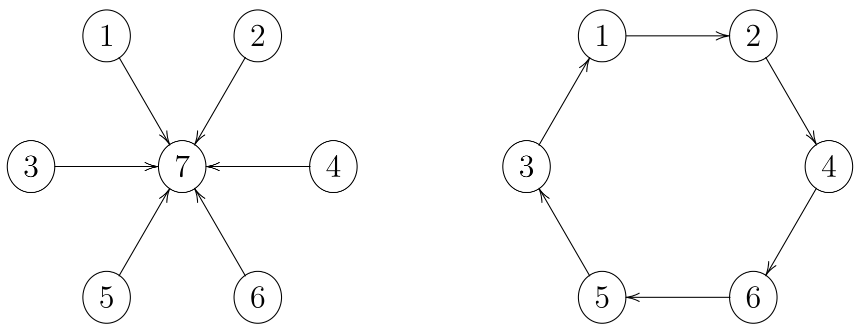

Figure 1: Two networks for small scientific communities. On the left, scientist 7 is prominent because she has an in-degree of 6 while the other scientists in her community have an in-degree of 0. On the right, all scientists are equally prominent, having an in-degree of 1.

(denoted, e.g., $\mathbb{E}[d(A)]$ for the average in-degree of a scientist with information set A).[9]

Theorem 7 (Supermodularity of the average in-degree)**.** *Let I be a set of scientists satisfying assumptions 3 and 4. If I is large enough, then the average in-degree is a supermodular function. That is, for any two information sets A and B such that at least one scientist in I has information set $A \sqcup B$[10],*

$$\mathbb{E}[d(A \sqcup B)] + \mathbb{E}[d(A \sqcap B)] \geq \mathbb{E}[d(A)] + \mathbb{E}[d(B)].$$

Moreover, if neither A nor B contains the same information as $A \sqcup B$[11] and

[9]Recall that a scientist has information set A if her information set contains the same number of realizations of each experiment as A does.

$\mathbb{E}[d(A)]$ denotes an average in two senses. First, it averages over all scientists with information set A. Second, it takes the average (known as the mean in probability theory) over all the possible graphs that may arise due to the probabilistic nature of individual scientists' decisions to form connections.

[10]I write $A_{i''} = A_i \sqcup A_{i'}$ if $n(i'', j) = \max\{n(i, j), n(i', j)\}$ for all j, and $A_{i''} = A_i \sqcap A_{i'}$ if $n(i'', j) = \min\{n(i, j), n(i', j)\}$ for all j. So, loosely speaking, $A \sqcup B$ contains as much information as A and B combined, while $A \sqcap B$ contains only as much information as is shared between A and B. These notions are closely related, but not identical, to the standard set-theoretic notions of union and intersection. See Heesen (2014) for details.

[11]That is, if neither A nor B contains the same number of realizations of each experiment

$\mathbb{E}[d(A \sqcup B)] > 0$ *then the above inequality can be strengthened to*

$$\mathbb{E}[d(A \sqcup B)] > \mathbb{E}[d(A)] + \mathbb{E}[d(B)].$$

What the theorem says is that if the set of scientists is sufficiently large, the average prominence of a given scientist increases rapidly (faster than linearly) in the size of her information set. See Heesen (2014) for a proof.

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. That is, most papers have few citations, while a rare few have a great number of citations (Price 1965, Cole 1970, Redner 1998).

Theorem 7 is how I substantiate claim 1. That is, I interpret the theorem as saying that a community of scientists desiring to gather as much information as possible (assumption 3) given their finite means (assumption 4) will exhibit a pattern of superstars and nobodies similar to the pattern that can be observed in actual science.

In this model, I have assumed only that gathering information is one of scientists' motivations. The model is consistent with many other factors influencing scientists' decisions what to read, including the epistemically irrelevant factors mentioned in section 1 and elaborated upon in section 5. But one special case of the model would be one in which epistemically irrelevant factors do not influence scientists' decisions. By theorem 7, superstars would be present in a community of such "epistemically pure" scientists.

This does not establish whether the superstars of real science arise (primarily) due to epistemic considerations or due to something else. The rest of this paper will take up that question and the implications if one or another answer is true.

---

as $A \sqcup B$.

4 The Scientific Competence Explanation

The previous two sections considered a possible mechanism that produces academic superstars. The crucial component of this mechanism is scientists' desire for information. So in describing this mechanism I relied on scientists being motivated (implicitly or explicitly) by epistemic goals.

How does being a superstar reflect on a scientist? If she is a superstar as a result of other scientists' epistemic goals (her work gets read a lot because of its information content), her status would intuitively seem to be well-deserved. Being a superstar then reflects on her epistemic virtues, i.e., her competence as a scientist.

This section explores that intuition and what follows from it in terms of what an outside observer can conclude from the fact that a scientist is a superstar. It exploits the scientific competence explanation that I outlined in section 1.

The next two sections investigate two contrast cases. The first is based on the alternative approach in which the social context or other epistemically irrelevant factors are the major shaping force. In this case superstars arise due to scientists' non-epistemic goals, independently of the information-based mechanism I have described. Not surprisingly, this leads to a different view on superstars (section 5). The second case challenges the intuition developed above by showing that a scientist could become a superstar as a result of other scientists' epistemic goals (in particular, via the information-based mechanism) without being particularly epistemically virtuous (section 6).

But for now consider the scientific competence explanation: superstars are scientists whose work gets read a lot because their work is of higher quality than that of other scientists. The thought behind combining this explanation with the information-based mechanism 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 common in the literature.[12] 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.

The model of sections 2 and 3 describes how this might work. Consider the following simplified version of that model.[13] 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 $m < n$. Since the competent scientists obtain more information than the incompetent ones, it follows from theorem 7 that competent scientists will be more prominent (in terms of average in-degree) than incompetent scientists.

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.

The latter inference has indeed been made by some philosophers: "I do not need to argue, I think, that a discovery produced by a scientist with a demonstrated record of success [many citations] has more initial credibility than a discovery produced by an unknown" (Strevens 2006, p. 166). Clearly, Strevens thinks that if a scientist has been cited a lot this is evidence in

[12]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.

[13]Here I only consider a very simple way in which more competent scientists may obtain more information. The model of sections 2 and 3 can also capture situations where the relation between the competence of a scientist and the information she learns is more complicated. Theorem 7 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.

favor of the quality of their future work. It appears that Strevens tacitly supports the scientific competence explanation, as this inference is far less straightforward under the explanations I consider in the next two sections.

If scientists choose which papers to read primarily based on the value of the information in those papers, this has some obvious additional benefits (over and above the benefit to the individual scientists of reading more rather than less valuable information). It would result, as theorem 7 shows, in a pattern where the number of citations to a paper correlates strongly with the value of the information therein. This would allow outsiders such as laypeople or funding agencies to identify the best work in a discipline relatively easily by looking at highly cited papers. They could then use citation metrics to rank scientists' competence and, e.g., award grants based on this information.[14]

All seems well so far. But of course the fact that the scientific competence explanation yields a coherent story does not prove that story to be the correct one. In the next two sections I discuss two alternatives.

5 The Sociological Explanation

The sociological explanation provides a very different view on how being a superstar reflects on a scientist. Under the sociological explanation, one or more 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 such terms as being able to get work published, getting citations, or receiving awards, or their work being viewed as "credible"). Some such factors include the scientist's (or her institution's) reputation (the

[14]Note that I am not claiming that this arrangement is in some sense socially optimal. It does not follow from the fact that individual scientists are behaving epistemically optimal in a certain sense (as I assume in this paragraph) that the resulting scientific community is epistemically successful (Mayo-Wilson et al. 2011). But the community-level benefits I do ascribe to it, in particular for policy makers, follow straightforwardly from the work in section 3.

so-called Matthew Effect, see Merton 1968), the reviewers that get assigned to her work (Cole et al. 1981), the scientist's age (Kuhn 1962, Zuckerman and Merton 1972), 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). Additionally, trends within science or in the community at large may lead to certain kinds of scientific work (or science as a whole) to gain or lose prominence (Kroeber 1944). For example, Simonton (1975) investigates the effects of nationalism and political instability on science.

What these factors have in common is that they are presumably epistemically irrelevant: white male scientists at prestigious institutions working on a "hot topic" may get read more, but this in itself 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 its epistemic merit. For example, the scientific work of women is consistently undervalued relative to that of men (Valian 1999). Fricker (2007) and Wylie (2011, p. 168) argue that this constitutes an epistemic injustice (a term coined by Fricker).

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 in fact be rewarding something other than merit. If Fricker and Wylie are right, they may even find themselves perpetuating epistemic injustice.

It is important to emphasize that the model of sections 2 and 3 does not show the sociological explanation to be wrong. Epistemically irrelevant factors certainly exist in real life, and it is plausible that they contribute to the phenomenon that some papers 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 papers would still get read more than others.

Imagine a sociologist of science investigating the phenomenon of superstars. One thing she might do is compare the (epistemic) quality of papers to their citations. Given the extreme skew in the distribution of citations, she would probably find that the differences in quality between papers are much smaller than the differences in the number of times they get cited. As a result, the imaginary sociologist concludes that other factors than epistemic quality are influencing citation patterns.[15]

But she is mistaken. Since the average in-degree is a supermodular function (theorem 7), small differences in quality can account for large differences in prominence in a non-linear fashion. Even differences in prominence that seem disproportionate to the differences in quality are not by themselves evidence of epistemically irrelevant factors being at work. This is one of the upshots of my model.

From the perspective of a relative outsider, in particular one in charge of awarding grants, the sociological explanation suggests conclusions completely opposite to those of the scientific competence explanation. If epistemically irrelevant factors are driving who gets read, citation metrics say little or nothing about the quality of a given paper or scientist.

It is quite plausible that in reality both epistemically relevant and epis-

[15] For example, Medoff finds that "after controlling for author quality, journal quality and article-specific characteristics,. . . an article written by an economist affiliated with Harvard University or the University of Chicago had significantly (statistically and numerically) greater peer recognition [measured in citations] than if it was written by an economist affiliated with a less prestigious university" (Medoff 2006, p. 504). He concludes that "the empirical results found that an institutional Matthew Effect was in operation" (Medoff 2006, p. 504), which is defined as a "disproportionate allocation of peer recognition" (Medoff 2006, p. 503) based on the prestige of the economist's institution. Notably, quality is only allowed to have a linear effect on citations in his model. Thus, this is an example of someone concluding that epistemically irrelevant factors must be at work because the observed differences in citations seem disproportionate (in the sense that a linear model cannot capture them) to the differences in quality.

temically irrelevant factors contribute to the differences in prominence among scientists. This complicates the issue even further: all things considered, what do citation metrics say about the quality of a paper?

If it were possible to reduce the impact of epistemically irrelevant factors or increase the impact of epistemically relevant factors, this would at least make the jobs of those awarding grants easier.[16] Here one can think for instance of programs designed to decrease implicit bias in science. Among other effects (presumably there are benefits from an ethical perspective as well), such programs 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 epistemic reasons.

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.

6 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 "epistemic luck 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). Penicillin's ability to kill bacteria, for ex-

[16]Insofar as this is possible, I think it would be a good idea overall. But this does not straightforwardly follow from the claim that individual scientists' decisions would be motivated more by epistemically relevant factors and less by epistemically irrelevant factors (Mayo-Wilson et al. 2011). See also footnote 14.

ample, 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).

The important point to note here is that, given the results of section 3, relatively minor lucky accidents can have a big impact on prominence. I illustrate this in a simplified version of the model, but the point generalizes.

Suppose there is a set P containing n propositions that scientists want to learn the truth-value of. Suppose 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) write the papers with the highest average in-degree in this model (this follows from theorem 7; see also Anderson 2011, section 3, for a detailed discussion of this particular model).

There is an interesting interplay between epistemically relevant and epistemically irrelevant factors here. On the one hand, the scientists are determining whose work to read based on information, an epistemically relevant factor. But who has the most informative paper is determined by luck, an epistemically irrelevant factor.

The epistemic luck explanation is thus similar to the scientific competence explanation in that it can produce superstars via the mechanism based on scientists' desire for information. Theorem 7 can turn both competent and lucky scientists into superstars. But it is similar to the sociological explanation in that it does not follow that prominent scientists are more competent than others (more on this below).

It is interesting to note that if scientific competence and epistemic luck are brought into play at the same time, the luck factor can easily drown out the competence factor. Once again I use a simplified model to illustrate this point.

Assume 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 paper with high in-degree is written by 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's paper has a high in-degree.

Proposition 8. *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 7 that papers by scientists who learn all n propositions will have the highest in-degree (see also Anderson 2011, theorem 5). Therefore

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

So I can make the proportion of high in-degree papers written by 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 paper gets read a lot this is not good evidence that the author is a good scientist. Thus the inference from many citations to competence (which is valid when the scientific competence explanation is the only correct one) is invalid if epistemic luck is a factor.

Note that repeating this process over multiple papers would give the competent scientists a better chance. Thus one might think that while any individual paper being highly cited could be a result of luck, competent scientists might still end up being more highly cited over the course of a career. But this is not obviously true: proposition 8 could be extended to say that if good scientists are rare enough, the chance that a whole series of well-cited papers (by the same author) are written by a good scientist may be arbitrarily small. Of course the proportion of good scientists would need to be made exceedingly small, perhaps unrealistically so.

But if the claim that luck can drown out competence is taken seriously, it has some interesting implications. Even if scientists are following only "epistemically pure" motivations (e.g., their desire for information) in deciding whom to read, this does not necessarily reflect well on the epistemic virtues of the prominent scientists. This is because the valuable information in the prominent scientists' papers may have arisen either from their competence or from luck. Presumably only the former case reflects well on their virtues as scientists.

This suggests a separation between two questions that might otherwise have been easy to conflate. If one is interested in awarding credit (say, a Nobel 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.[17]

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.

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 either exacerbate or weaken the effects of the differences in information.

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 inferences (e.g., about a scientist's merit) that are straightforward if some particular factor is the cause may be mistaken if another factor is the cause, or if multiple factors are at work.

[17]On most views of luck. See McKinnon (2014) for a possible exception.

7 Conclusion

In this paper I considered a model of information exchange in which academic superstars are seen to arise. The model shows that a mechanism based on scientists' desire for information can be quite effective: small differences in informational value can lead to large differences in prominence.

This suggests a scientific competence explanation for academic superstars. This explanation allows for the use of citation metrics in identifying competent scientists. But two alternative explanations complicate the picture.

The sociological explanation claims that a different mechanism than scientists' desire for information creates superstars. The epistemic luck explanation claims that even when superstars are created as a result of scientists' desire for information, the identification of competence with prominence might not go through.

This is relevant to funding agencies, as it highlights the fact that prominent scientists are not necessarily the most promising. The epistemic luck explanation 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.

More generally, care should be taken in drawing conclusions about explanations for academic superstars, or the merits of individual superstars. The existence of an information-based mechanism to create superstars does not warrant the conclusion that actual superstars were created by this mechanism. But, due to the information-based mechanism, one should also not be too quick to conclude that superstars were created by epistemically irrelevant factors. And even evidence that a given superstar was created by the information-based mechanism does not establish whether competence or luck was responsible.

References

Katharine A. Anderson. Collaboration network formation and the demand for problem solvers with heterogenous skills. Manuscript, 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.

Venkatesh Bala and Sanjeev Goyal. A noncooperative model of network formation. *Econometrica*, 68(5):1181–1229, 2000. ISSN 1468-0262. doi: 10.1111/1468-0262.00155. URL http://dx.doi.org/10.1111/1468-0262.00155.

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.

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.

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.

Patrick Grim, Daniel J. Singer, Steven Fisher, Aaron Bramson, William J. Berger, Christopher Reade, Carissa Flocken, and Adam Sales. Scientific networks on data landscapes: Question difficulty, epistemic success, and convergence. *Episteme*, 10:441–464, 12 2013. ISSN 1750-0117. doi: 10.1017/epi.2013.36. URL http://journals.cambridge.org/article_S1742360013000361.

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.

Matthew O. Jackson and Asher Wolinsky. A strategic model of social and economic networks. *Journal of Economic Theory*, 71(1):44–74, 1996. ISSN 0022-0531. doi: http://dx.doi.org/10.1006/jeth.1996.0108. URL http://www.sciencedirect.com/science/article/pii/S0022053196901088.

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.

Philip Kitcher. *The Advancement of Science: Science without Legend, Objectivity without Illusions*. Oxford University Press, Oxford, 1993.

Alfred Louis Kroeber. *Configurations of Culture Growth*. University of California Press, Berkeley, 1944.

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, second edition, 1986.

Conor Mayo-Wilson, Kevin J. S. Zollman, and David Danks. The independence thesis: When individual and social epistemology diverge. *Philosophy of Science*, 78(4):653–677, 2011. ISSN 00318248. URL http://www.jstor.org/stable/10.1086/661777.

Rachel McKinnon. You make your own luck. *Metaphilosophy*, 45(4-5):558–577, 2014. ISSN 1467-9973. doi: 10.1111/meta.12107. URL `http://dx.doi.org/10.1111/meta.12107`.

Marshall H. Medoff. Evidence of a Harvard and Chicago Matthew effect. *Journal of Economic Methodology*, 13(4):485–506, 2006. doi: 10.1080/13501780601049079. URL `http://dx.doi.org/10.1080/13501780601049079`.

Robert K. Merton. The Matthew effect in science. *Science*, 159(3810):56–63, 1968. ISSN 00368075. URL `http://www.jstor.org/stable/1723414`. Reprinted in Merton (1973, chapter 20).

Robert K. Merton. *The Sociology of Science: Theoretical and Empirical Investigations*. The University of Chicago Press, Chicago, 1973.

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.

Dean K. Simonton. Sociocultural context of individual creativity: A transhistorical time-series analysis. *Journal of Personality and Social Psychology*, 32(6):1119–1133, 1975. ISSN 0022-3514. URL `http://dx.doi.org/10.1037/0022-3514.32.6.1119`.

Michael Strevens. The role of the Matthew effect in science. *Studies in History and Philosophy of Science Part A*, 37(2):159–170, 2006. ISSN 0039-3681. doi: http://dx.doi.org/10.1016/j.shpsa.2005.07.009. URL http://www.sciencedirect.com/science/article/pii/S0039368106000252.

Virginia Valian. *Why So Slow? The Advancement of Women*. MIT Press, Cambridge, MA, 1999. ISBN 9780262720311.

Alison Wylie. What knowers know well: Women, work and the academy. In Heidi E. Grasswick, editor, *Feminist Epistemology and Philosophy of Science*, pages 157–179. Springer Netherlands, 2011. ISBN 978-1-4020-6834-8. doi: 10.1007/978-1-4020-6835-5_8. URL http://dx.doi.org/10.1007/978-1-4020-6835-5_8.

Kevin J. S. Zollman. The epistemic benefit of transient diversity. *Erkenntnis*, 72(1):17–35, 2010. ISSN 01650106. URL http://www.jstor.org/stable/20642278.

Kevin J.S. Zollman. Network epistemology: Communication in epistemic communities. *Philosophy Compass*, 8(1):15–27, 2013. ISSN 1747-9991. doi: 10.1111/j.1747-9991.2012.00534.x. URL http://dx.doi.org/10.1111/j.1747-9991.2012.00534.x.

Harriet Zuckerman and Robert K. Merton. Age, aging, and age structure in science. In Matilda White Riley, Marilyn Johnson, and Anne Foner, editors, *Aging and Society*, volume 3, chapter 8, pages 292–356. Russell Sage Foundation, New York, 1972. Reprinted in Merton (1973, chapter 22).