Priority and privilege in scientific discovery

Hannah Rubin and Mike D. Schneider

Abstract

The priority rule in science has been interpreted as a behavior regulator for the scientific community, which benefits society by adequately structuring the distribution of intellectual labor across pre-existing research programs. Further, it has been lauded as part of society's "grand reward scheme" because it fairly rewards people for the benefits they produce. But considerations about how news of scientific developments spreads throughout a scientific community at large suggest that the priority rule is something else entirely, which can disadvantage historically underrepresented or otherwise marginalized social groups.

1 Introduction

In scientific practice, discoveries of sufficient impact appear to generate certain quantities of prestige, which are bestowed upon particular scientists by the scientific community at large. The priority rule is a broad descriptive norm concerning the allocation of that prestige. The inference toward such a norm comes via an observed phenomenon in the history of science whereby, in situations of *multiple discovery*, disputes about who deserves the prestige that comes associated with a discovery are often fought by way of assertions about who was first to make it.

While the phenomenon of multiple discovery in science was recognized by Ogburn and Thomas [1922], who mention in a footnote the common appearance of disputes over priority of discovery in such cases, the investigation of these so-called "priority disputes" was taken up by Merton [1957]. Merton sought to explain, by appeal to the institutions and norms of science, how otherwise dignified and reserved scientists would, when involved in instances of multiple discovery, often fight tooth and nail against the possibility that anyone but them bore responsibility for the discovery.

Merton makes sense of this state of affairs in terms of the perception amongst scientists of intellectual property rights, for which it is the case that the first person responsible for the invention of a new bit of intellectual property ought to enjoy the profit that comes from it. This feature of the account— that the prestige that is generated by a particularly impactful discovery is taken by scientists to properly belong to whoever is truly first to produce it— is what has come to be known as the "priority rule" in science.

Strevens [2003], following a lead from Kitcher [1990], famously formalizes the priority rule in science as a particular manifestation of a "grand reward scheme" (p. 76) that exists in our society, which allocates prestige associated with new discoveries in such a way that incentivizes scientists to optimally distribute themselves across a variety of programs of research. In other words, Strevens offers Merton's priority rule as the basis for a social mechanism concerning the disbursal of prestige in science, and the details of how that social mechanism works are argued to play an epistemic role. He argues that such an epistemic role provides extrinsic justification for the continued existence of the rule. He argues, moreover, that this mechanism is fair, in virtue of rewarding people based on the benefit they confer (since only the first to make the discovery produces a benefit to society, it is regarded as fair that only the first discoverer is rewarded).

However, in Strevens's account, it is assumed that whoever makes the discovery first will receive the prestige from the scientific community at large (or, at least, scientists believe this to be so). This is peculiar, as the history of science is punctuated by the (perceived) failure of the community to adequately implement a social mechanism that does this. Merton's original evidential impetus for talking about priority was that it is priority *disputes* which are commonplace in the history of science and which are fought so animatedly, but these disputes are idealized away in Strevens's formalization of the priority rule. In light of the substantial gap between this idealization of scientific practice (which has become common in philosophy of science) and Merton's original evidence for the priority rule, our primary goal here is to offer a different perspective on the priority rule, regarded as a social mechanism for prestige disbursal, than that which Strevens provides. Our essential point is that individuals engaged in assigning credit for a discovery in accordance with their beliefs about priority are what, altogether, give rise to the community disbursing a quantity of prestige for that discovery. Ignoring the gap between credit attribution and the awarding of prestige misses details that are important to any study of the institution known as the "credit economy" in science.

Having argued for our alternative conception of the priority rule (section 2), our secondary goal here is to evaluate some important consequences of the way scientific communities disburse prestige (section 3). We will do so by means of a simple model, where what we find is that inequities in the underlying social network of the scientists can allow prestige to accumulate in the hands of those historically well-positioned within the scientific community, and meanwhile historically underrepresented or otherwise marginalized social groups can suffer in the context of multiple discovery. Notably, this disadvantage arises due to facts about the social structure of scientific communities, rather than due to any differences in skill or achievement, or any bias against the minority population. These sorts of phenomena are impossible to discover in current credit economy models of science, and, as we will argue, can affect conclusions regarding the ultimate outcomes of different credit incentives.

2 From priority disputes to the priority rule and back

The operational notion of profit in priority disputes is not in terms of immediate payment for services rendered. What is at stake is the accumulation of prestige, or what we might think of as ‘wide-scale credit’, in the eyes of the community at large. Merton [1957] lists several examples of how the community confers prestige to individuals, including eponymy (p. 643), honorifics such as the Nobel Prize, introduction into “honorary academies” like the Royal Society and the French Academy of Sciences (p. 644), and posthumous recognition by historians (p. 645). To this list of examples, Strevens [2003] adds “reputation, a sizable office, the rapt attention of graduate students and the like” (p. 57). Based on these examples then, prestige appears to be a retrospective quantity, conferred on individuals when the community as a whole has come to associate them with the corresponding discoveries.

A key observation is that in most cases of disputes about who ought to receive the prestige that corresponds to a particular discovery, at least one of the parties involved perceives there to be a great injustice afoot: they are being denied access to newfound prestige because another party involved is, for whatever reason, unrelated to the content of the discovery, in a better position to win the prestige instead. Merton proceeds to explain how such priority disputes emerge as unsurprising artifacts of the institutions and norms that govern scientific practice. And so is born the notion of a ‘priority rule’ in science, as a norm according to which prestige associated with a discovery rightly belongs to whoever is truly first to produce it.

2.1 Priority, prestige, and plural action

Strevens takes Merton’s work as motivation to formalize the priority rule in such a way that he can assess its impact on scientific inquiry, as a mechanism by which scientists disburse prestige. In particular, he models the priority rule as a reward system characterized by two “parts” (p. 56):

> First, rewards to scientists are allocated solely on the basis of actual achievement, rather than, for example, on the basis of effort or talent invested. Second, no discovery of a fact or a procedure but the first counts as an actual achievement.

Strevens then argues that, from the perspective of a central planner, although the priority rule may seem harmful initially (as compared to, e.g., a rule that rewards scientists for hard work and talent), it is actually beneficial for scientific inquiry. In his formal setup, it is by virtue of the priority rule that intellectual labor is efficiently distributed across various research programs with differing odds of success. The basic idea is that scientists balance the odds that a research program will be successful and the odds that, if their research program is successful, they will be the one to make the discovery. Therefore, scientists do not all abandon other lines of research to join the most promising-looking research program. Instead, they distribute themselves among research programs (with more scientists working within more promising-looking programs). Others have since used this same basic modeling framework to discuss, for example, incentives to publish early and frequently [Heesen, 2018] or share intermediate results [Heesen, 2017], disincentives to replicate previous findings [Romero, 2017], and how scientists are motivated by a combination of truth and credit [Zollman, 2018].

It seems odd that in this formalization of the priority rule, there is no possibility of genuine conflict in cases of multiple discovery. After all, it was a litany of such conflicts in Merton's article that gives rise to an inference toward the priority rule in the first place. This gap between Merton's treatment of priority disputes and this account of the priority rule is highlighted by the shift in Strevens's article away from Merton's historical cases of priority disputes and toward scenarios in which different research programs approaching scientific problems are in a winner-takes-all race toward the resolution of those problems. Unlike in Merton's work, "discoveries" are now the resolutions of those problems and prestige is doled out to whoever wins the race.

While such scenarios certainly occur in science, for instance when it comes to highly anticipated results, they are not the scenarios in which Merton's work about priority is most relevant. Where is the disconnect? We argue that it arises (at least in part) from an ambiguity in plural action claims.

Namely, when Strevens asks "... why does the scientific community disburse prestige in accordance with the priority rule rather than... some alternative scheme?" (p. 58), nestled implicitly in the question is a plural action claim:

> The scientific community disburses prestige.

Following the terms provided by Ludwig [2016], this claim can either be read as a distributive action sentence or as a collective action sentence. Moreover (following the form of Ludwig's argument, p. 131), we may understand the ambiguity between these two readings to wholly consist in an ambiguity of scope:

> On the distributive reading, we mean that each of [the scientists within the community] were separately sole agents of [the disbursal of prestige] in a certain way. On the collective reading we mean that for [the disbursal of prestige] each of [the scientists within the community] (and no one else) was an agent of it in a certain way.

As an example, we may read the sentence "Two people built boats" as claiming either that two people independently built boats (distributive reading) or that two people came together to build boats, e.g., with one sawing and the other hammering pieces together (collective reading). Similarly, scientists could separately take part in the community's disbursal of prestige by each independently engaging in a prestige-relevant task— e.g. individually attributing credit (distributive reading), or they could come together as a whole to disburse the prestige as a group (collective reading). Which reading we accept affects how we interpret the priority rule and its effects on the scientific community.

Likewise, how we idealize prestige disbursal in a scientific community constrains how we may read the plural action claim offered implicitly in support of that idealization. Strevens, and others following his way of modeling the credit economy, can only handle the collective reading (except in degenerate cases, where distributive and collective plural action yield the same results). But in the next section, we provide explicit evidence for a distributive reading, which motivates our own discussion of the nature of the priority rule in section 2.3.

In what follows, to keep our discussion of these two readings as clear as possible, we will use the term 'credit' when referring to individual activities of associating a person with a discovery. We will use the term 'prestige' when talking about the consequences of a community-level association of a person with a discovery (whether that association is the result of individual scientists' assigning credit or else the result of some community-level activity).

2.2 Evidence for the distributive reading

The scientific community, though in some respects quite hierarchical, is not centrally governed. The awarding of prestige, though associated with individual accomplishments, often happens at no particular moment in time. That individuals come to be awarded prestige at no particular time by no particular decision made on behalf of the community strongly suggests that the disbursal of that prestige is not a single act of which each scientist was an agent. Instead, it seems more plausible to view the disbursal of that prestige as following from the members of the community separately taking actions (of an identical kind to each other) that need not individually resemble the awarding of prestige. That is, individual scientists assign credit upon learning about a discovery in relation to a name, and along the way prestige is conferred by the scientific community. (No one gains the respect of a community because one person gives them credit for a discovery, but if the members of a community all generally associate a particular scientist with a discovery, that scientist will receive the associated reputational benefits, etc.)

To support this alternative view, we will offer historical examples drawn from evolutionary biology, (computational) social science, and fundamental physics. Each of the examples below is a case where the scientific community as a whole "got it wrong". However, to be clear, that the scientific community "got it wrong" in these cases is not itself what constitutes evidence for the distributive reading we have advocated.[1] What we want to emphasize is one feature common to all three cases: *individuals attributing credit as they go about their ordinary affairs are what, in aggregate, determines to whom prestige comes to be awarded.*

[1]The community can also be wrong about who to award prestige under a collective reading. However, the collective reading cannot capture what went wrong in the cases below, where each individual agent assigning credit on account of personal commitments concerning priority results in a conflict between the recognition of the priority of one person and the prestige being bestowed on another.

In evolutionary biology, the distinction between proximate and ultimate causes in biology is attributed to Ernst Mayr. As Laland et al. [2011] point out, this distinction was made well before Mayr wrote about it in the 1960s (they cite an article by J. Baker written in the 1930s). Nonetheless, Mayr's article is what ostensibly led to the distinction's widespread acceptance in evolutionary theory. Moreover, even as Laland et al. [2011] flag the trivia that Mayr was not the first to make the distinction, they consistently refer to it as "Mayr's distinction". This is illuminating because, as an evidently conscious decision of the authors in that article, it demands explanation. One prudent explanation is that of historical usage: since people started citing Mayr when talking about the distinction, it subsequently became known as his distinction. Hence, even when the pre-history of Mayr's work is acknowledged, Mayr's legacy continues to enjoy the prestige.

In the social sciences, Thomas C. Schelling undoubtedly enjoys the prestige that surrounds the basic checkerboard model of segregation, which is equivalently called the "Schelling model" (an instance of eponymy), and is taught in any introductory course on the subject of agent-based computational models in the social sciences. As is now recognized, James M. Sakoda beat Schelling to the public invention of computational models of segregation: the latter's model amounts to a special case of one of the former's, which had its origins in his dissertation work twenty years earlier and was printed in the previous issue of the same journal in which the latter's model appeared. But as Hegselmann [2017, p. 5-6] argues, "No crime happened, no conspiracy was involved... as to the main actors, nobody did anything wrong."

Instead, one deciding factor it seems was Schelling's subsequent decision to write a book, developing many of the ideas in his paper, accessible for much broader audiences than just those computer scientists who happen to additionally be interested in modeling social dynamics. Those broader audiences were encouraged to try out small, table-top examples of the checkerboard models under scrutiny, whereupon: "They all had experienced how surprisingly fast, right before their eyes, certain unexpected, dramatic macro structures evolved, generated by fairly innocent looking micro-motives– an eye-opening phenomenon par excellence" [Hegselmann, 2017, p. 87]. This got those broader audiences talking about the demonstrative power of these simple computational models, and they were talking about it in the context of Schelling's work. Whatever reasons we give for why Schelling enjoys the prestige, an indispensable part of the story is that he enjoys it because individuals separately began to associate him with the discovery, rather than because the scientific community wholesale decided that it was he who was responsible for the discovery.

A final case worth mentioning is one in fundamental physics, particularly in the history of quantum mechanics. In 1932, von Neumann published an alleged "no-go" proof of the viability of hidden-variables underpinning quantum mechanical behavior in an ultimately deterministic theory. Grete Hermann evidently discovered a flaw in the scope of the proof in 1935, yet this discovery was "not widely known at the time, and her criticism had no impact whatsoever" [Seevinck, 2016, p. 107]. In 1964, John Bell happened on the same such discovery, in the aftermath of the development of Bohmian mechanics (whose success as a deterministic, hidden-variables alternative to quantum mechanics clearly stood as proof of the alleged impossible). Bell enjoyed the benefit of offering Bohmian mechanics as a demonstration of his point (whereas in 1935, Hermann could only conclude that the possibility of such a hidden-variables theory was left open). There are many reasons why Hermann's contributions at the time were overlooked. Our point here is that following this neglect, when most scientists were for the first time ready to credit *someone* for identifying the ostensible flaw in von Neumann's proof (that is, following the development of Bohmian mechanics), Bell's independent study of the subject led him to be the recipient of that credit, and, eventually following, prestige.

In each of these historical anecdotes, there are myriad reasons one could give for why a particular person enjoys the prestige. The common feature across each of these histories is that the recipient of the prestige associated with a discovery is the individual who first became largely known to be associated with the discovery, in virtue of a plurality of individuals each attributing credit to them. We take this as evidence from the history of science for a distributive reading of the claim "the scientific community disburses prestige": in each of these cases, each of the scientists within the relevant community happened to attach credit for the relevant discovery to some or other scientist. Meanwhile, by virtue of each such scientist acting in this way separately, i.e. as a sole agent, prestige came eventually to be disbursed by the community to the individual that most of the scientists credited.

2.3 Credit attribution contests

Taking a lead from the historical cases just discussed, we offer the following perspective. Scientific developments are not, in general, immediately known by everyone in a scientific community. News of them spreads throughout the community, rather, through an informal social network that spans the community. Insofar as a development proves valuable to the community, the party known to be responsible for that development enjoys a corresponding quantity of prestige. Of course, it need not be the case that *every* individual in the community has assigned credit; the community may just as well bestow prestige when the vast majority has learned about the development (whereupon any stragglers learn about the development as do those outside of the community, as discussed below).

In cases of multiple discovery, just as in any other case, when individuals within the community go looking for news of pertinent developments, they generally come to associate a development with whomever they first learn is responsible for it. The difference is that in the case of multiple discovery, a large majority of the community may all come to associate the development with one of the parties involved. If so, prestige comes to be bestowed upon that party (as it is normally), while any other party that independently produced the same development is neglected.

At risk of being denied the prestige that they believe they are due on the basis of their work, any party that does not enjoy the support of the majority will protest that they ought to receive the prestige that in other circumstances would have been awarded to them, for instance had the social network just happened to have been structured differently. As Merton [1957] notes, those in the minority who believe the wrong person received the prestige will often protest, too. On the hypothesis that the conferring of prestige by the community occurs when a sufficient number of individuals within the community associate a particular person with a discovery, the function of such a protest is obvious: to change individuals' associations as to who typically gets credit for a discovery.

Returning to the priority rule, one can grant that scientists ought to disburse prestige on the basis of actual achievement, and none but the first counts as an actual achievement. Still, a question is left open: how does prestige *get allocated*, when scientists are each acting separately as sole agents assigning credit, consonant with the norm just articulated? Based on what we have said above about the nature of priority disputes, we conclude that prestige associated with a particular scientific development is bestowed wholesale upon a particular party as a consequence of a large majority of the surrounding community having all individually come to associate, as a matter of priority, that development with them. And in instances of multiple discovery, we may imagine 'credit attribution contests' occurring among individuals within the community who independently produce similar developments at the same time.[2]

One might object that a large majority is not enough – surely all or nearly all of the community must agree on a discoverer for that person to get the prestige. But there are reasons we think a large majority would suffice. First, in the instance of a priority dispute between the two competing parties, one might imagine that those who enjoy the support of the large majority are more likely to win the priority dispute: along the lines of the observation by Merton just mentioned, those who enjoy the support of the large majority also enjoy a larger collection of possible defenders, ready to fight against an instance of perceived injustice.

Second, even without a priority dispute, a large majority of scientists coming to associate one person with a discovery might be sufficient for prestige to be bestowed on that individual. For instance, when using reference to a name as short-hand for an idea that one's interlocutor will understand (e.g. using "Schelling's segregation model" to refer to a mathematical model whereby minimal conditions for segregation are demonstrated), the most effective name to choose is the name most well-known in connection to that idea (one's interlocutor is less likely to know of Sakoda's work). One might imagine that there is some tipping point, or threshold, at which it becomes prudent for individuals wishing to communicate in this way to defer to the use of the more well-known name, irrespective of who they individually have associated with it.[3] For exam-

[2]Of course, what is supposed to count as "similar developments" and "at the same time" is highly contextual. As the historical cases suggest, "at the same time" is sometimes better understood as "both occurring before that time at which the content of the discovery becomes a hot topic".

[3]Thanks to the Norms and Networks Cluster at University of Groningen for discussions on

ple, Laland et al. [2011] refer to "Mayr's distinction" while noting that Mayr does not deserve priority for that distinction. Similarly, new members of the community, hoping to signal their understanding of the field, will think to provide the name most people within their community associate with the discovery, lest they be thought ignorant. Finally, it is reasonable to expect that the more well-known name would be used in review articles written by members of the community, to communicate the results of their sub-field to a wider audience (and to the few members of the community who may not have already heard of the discovery), and so both the broader scientific community and, eventually, the public at large and history books would come to recognize the more well-known name.[4]

In the next section, we will argue that this more accurate picture of how priority interacts with prestige allocation in science does not necessarily align with our ideas of fairly rewarding people for benefits they confer (as in society's grand reward scheme, which Strevens suggests does exactly that). This discussion is valuable in its own right, as it highlights one plausible way by which the institution of the credit economy can foster structural disadvantages in socially diverse science. But as we elaborate, it also suggests that what is incentivized by a scientific community's sustained embrace of the priority rule is most likely not the optimal division of cognitive labor that Strevens envisages.

3 Historically underrepresented groups

Here, we will formalize the credit attribution contests just discussed so as to study the influence of network structure on the awarding of prestige in instances of genuine multiple discovery. This model is meant to be descriptive of scientific practices regarding prestige, not prescriptive of how a central planner ought to distribute prestige.[5] As such, we do not compare the efficacy of different *mechanisms for disbursing* prestige; rather, we take prestige to simply be that which is disbursed in the way we have so far described.

3.1 A basic model

We formalize scientific communities as networks of agents, where nodes of the network represent individual scientists and edges, or links, represent regular information channels between them. These links are bidirectional and can be thought of as representing people who talk to each other when working on a new project, or who ask each other if looking to reference a paper on some or other topic.

---

this point. In the next section we provide a particular tipping point of 2/3. While this number is arbitrary, the existence of a tipping point is not arbitrary, and the particular number chosen will not significantly affect results. For instance, similar results have been obtained with 3/5.

[4]Thanks to Jim Joyce for encouraging us to think about the role of review articles.

[5]We use the modifier 'descriptive' only in contrast to 'prescriptive'; we do not mean to imply that our model is descriptively accurate in that it captures all features of the situation at hand.

We model credit attribution contests where information about discoveries spreads throughout these networks. In the model, we start with an instance of multiple discovery. Two scientists each independently make some discovery, and not knowing about the other, they each believe themselves to be the discoverer. In the second time-step, we pick a third node at random. Whoever they are closest to in the network[6] is who they will credit as having priority for the discovery, who they will cite with respect to that idea, etc.[7] Then in the next timestep, another random node is chosen. This fourth scientist gets news of that discovery from whoever is closest to them in the network that already has some belief about who made the discovery. (The idea here is that when one goes looking for a paper on some topic x, for instance so as to cite a discussion of that topic, one might ask their friends if they know about a paper on x, or alternatively, one might hear about some new discovery from a trusted collaborator, or so on.) This fourth scientist then has a belief about who deserves credit for the discovery. This process continues until all scientists have attributed credit. If there is a super-majority of 2/3 in favor of one discoverer over another, we say the former wins the credit attribution contest (and so, gets the prestige associated with that discovery). If not, there is a tie.

Since this information spread occurs over networks of scientists, the structure of the network will likely impact who receives credit for a discovery. At first, we will consider basic Barabási-Albert networks [Barabási and Albert, 1999], though in the next section, we will consider alterations to that model, which include social identity types and network change over time. As explained below, these networks capture several important features of scientific communities.

Barabási-Albert networks are formed in the following way. First, we start with a small number of fully connected nodes (nodes with all possible links between them), m_0. Then, new nodes are added one by one until the network reaches a designated size. Each time a new node enters, it forms a set number of links, m. For the results presented here, $m_0 = m = 4$. New links are formed via preferential attachment. That is, the more links a node already has (i.e., the higher its *degree*), the more likely it is that an entering node will form a link with it. The probability p_i that the new node is connected to node i is:

$$p_i = \frac{d_i}{\Sigma_j d_j} \tag{1}$$

where d_i is the degree of node i, and $\Sigma_j d_j$ is the sum of the degrees of all nodes in the network (not including the new entering node). The higher the degree a node already has, the more likely it is the new node will connect to it.

These networks have a couple of important features. First, in these networks, the 'rich get richer': nodes that already have many links are more likely to get new links. This captures a scenario seen in many real-world networks where the oldest members of the community tend to be the most central and

[6] If there is a tie, a node is chosen at random.

[7] This can either be interpreted as this third node caring about a particular discovery, and searching for it in that moment, or one of the original discoverers discussing their discovery with someone close to them in the network.

well-connected individuals in the network. Moreover, this model takes on a natural interpretation in the context of scientific research communities: as new researchers enter the community (e.g. as graduate students), they often seek out social relationships with the more well established members of the community, the oldest of which are often the most esteemed.

Second, these networks are scale free, meaning their degree distribution follows a power law. In other words, there are many nodes with a few links and a few nodes with a large number of links; there are a few 'hubs' in the network. Many real world networks are (approximately) scale-free, including many types of social networks and collaboration patterns. Among the scale-free networks, Barabási-Albert networks are particularly useful for our purposes. There is evidence that collaboration networks are formed via preferential attachment, similar to the method of preferential attachment used in the formation of Barabási-Albert networks [Newman, 2001, Barabási et al., 2002]. Additionally, since in later sections we will be discussing how networks evolve over time, Barabási-Albert networks are useful because they already stipulate what should happen when new nodes enter the network.

For this basic model, we formed a network of 100 people, then ran 1000 contests on each network to estimate the likelihood of each person getting the prestige for their discovery, then performed 100 replications (i.e., we formed 100 different networks of 100 people, and ran 1000 contests on each). Figure 1 shows the estimated likelihood of winning for each node in the network. That is, of the contests a node was a part of, it shows the percent of contests that node won. In this figure, the lower number a node is, the older it is (nodes 1-4 are all the same age, as the network started with 4 fully connected nodes). As one might expect, older nodes were more likely to win credit attribution contests because they tended to have a higher degree. Intuitively, when there is an instance of multiple discovery, those scientists who are more well connected are more likely to wind up with the prestige associated with their discovery, because the news of their discovery travels faster.

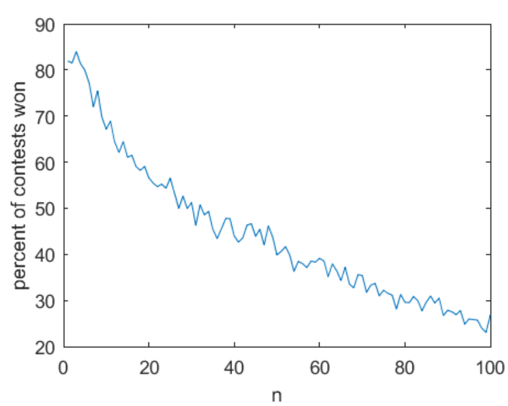


Figure 1: Likelihood of winning for each node in the network

3.2 Types and evolution

This observation points to a possible disadvantage for historically underrepresented groups (HUGs). Since older nodes tend to be members of the historically entrenched group (HEG), the HEG members will tend to be better connected, even when members of a HUG begin to enter the community at an equal rate later on. This means that HEG members tend to receive prestige for their discoveries at a higher rate, even when they make the discovery at the same time as a member of the HUG. Of course, scientific communities also change over time. Older members of the community retire and new scientists enter the community. After a time, if the HUG enters the community at a rate equal to its size in the population, it will eventually achieve proportional representation in the scientific community. This raises the question: will the HEG advantage over the HUG ever go away? (And, if so, how long will it take?)

In order to address these questions, we introduce types into the basic model: HEG members will be type 1 and HUG members type 2. There is nothing intrinsically important about these types; they are not related to scientific competence or likelihood of producing a scientific discovery. They are, however, socially relevant, in that type 1 enters at a higher rate earlier in time. In particular, we used a logistic growth equation to determine how likely it was a new node was type 2 at each point in time. This represents a case where the HUG finds it hard to enter the scientific community at first, but once there is a sufficient number of them it becomes much easier.[8] We will consider results at first where, by the end, the HUG enters at roughly the same rate as the HEG.[9] If we run a similar model as in section 3.1, but assigning types to the nodes as just described, type 1 individuals win credit attribution contests against type 2 individuals about 42.3% of the time, lose about 16.7% of the time, and tie the rest of the time.

These results are already telling, but we also want to consider how social identity impacts network formation and evolution over time. A pervasive feature of real communities comprised of different social identity groups is *homophily*, or the preference for linking to members of one's own social identity group. People are homophilic in a variety of contexts (e.g. when forming friendships [Currarini et al., 2009]), and scientific communities are homophilic as well, especially when it comes to co-authorship [Ferber and Teiman, 1980, McDowell and Smith, 1992, Boschini and Sjögren, 2007, del Carmen and Bing, 2000, West et al., 2013, Wang et al., 2019] and citation patterns [Wardle, 1995, Paris et al., 1998, Ghiasi et al., 2018]. We include homophily in the model by having agents place some weight, H, on their similarity to a node in addition to their degree. We used the following to determine how much the incoming node k values linking with each

[8] Results are shown for $P(type\ 2) = \frac{w}{1+10\times e^{-.05t}}$, where w is the fraction of the larger population the HUG comprises and t is the time-step the node enters. However, none of the results we discuss hang on using this particular equation. We also obtained similar results with an equation where the probability of type 2 increases quickly then asymptotes at .5, $P(type\ 2) = 1 - \frac{1+2e^{-t/20}}{2+e^{-t/20}}$.

[9] That is, we set $w = .5$ in the equation in footnote 8.

of the existing nodes i:

$$v_{ik} = H \times \frac{s_{ik}}{\sum_j s_{jk}} + (1 - H) \times \frac{d_i}{\sum_j d_j} \tag{2}$$

where $s_{ik} = 1$ if nodes i and k are of the same type and 0 otherwise.[10] The probability that the new node links with a particular node, p_{ik}, is then proportional to v_{ik}. The likelihood a node is chosen is thus determined by its value to the new individual, including both homophilic preferences and degree, rather than just its degree.[11]

We incorporate network change over time into the model in the following way. There is a maximum network size of 100 scientists, so as a new scientist enters beyond the first 100, the oldest node is removed from the network (it "retires") along with all its links. When a new node enters, it then forms a set number of links, as in the Barabási-Albert model. Additionally, in order to capture the consequences on the social network structure of a scientific community that trains its younger members, when the network grows beyond 50 nodes, incoming nodes also choose an 'advisor'.[12] For each node the advisor is linked to, the new node has a chance of linking with that node as well (a 50% chance for the results below), in addition to its links formed according to equation 2. This captures a scientific field that grows to a certain size, becomes established, then begins to adopt practices to train new generations of scientists.[13]

To interpret our results, we define *HEG advantage* as the probability a HEG member wins minus the probability a HUG member wins the credit attribution contest in an instance of multiple discovery, where the two discoverers belong to different social identity groups. In order to discuss how long the HEG advantage will persist when the network evolves, we will talk about HEG advantage over *generations* of the scientific community. A generation is defined as the time it takes to have a complete turnover of scientists. Since in each time-step the oldest scientist retires and a new scientists enters, with 100 scientists a generation is 100 time-steps.

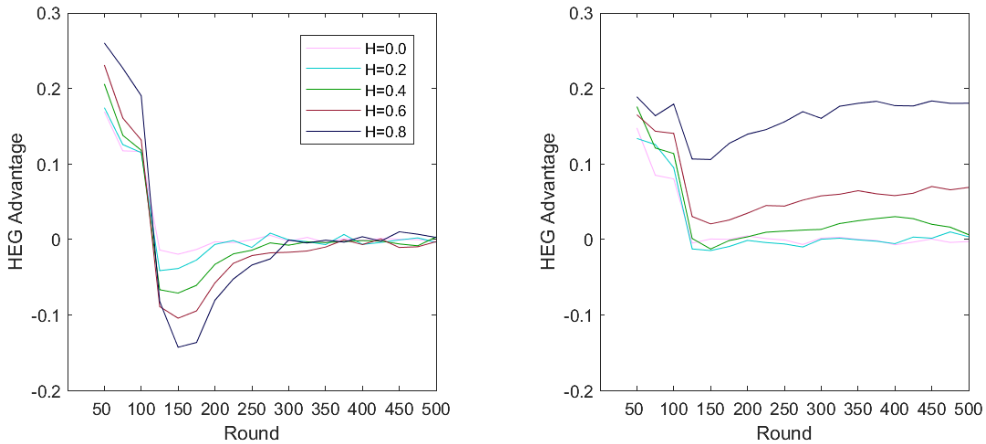
Let us first consider a case where the HUG approaches 50% of the population, where by timestep 100 (after 1 generation) they are entering in equal proportions and after 2 generations they have achieved equal representation. For each level

[10]We incorporate homophily in this way because its influence on the value an incoming node assigns to a new link remains constant as we move node to node, regardless of the degrees of those nodes. Some authors incorporate homophily as a weighting of the degree of a node, and find, similar to our results here, that homophily "makes the rich even richer" [Kim and Altmann, 2017] and that homophily can lead to minority group members occupying less important places in the network [Karimi et al., 2018]. Incorporating homophily in this alternative way does not qualitatively affect the results described below.

[11]One might also think that winning a credit attribution contest would make an existing node more "prestigious" and thus would increase its value in the eyes of incoming nodes. This is not included in the model, but would likely intensify the effects reported here.

[12]The advisor is the first link formed according to equation 2.

[13]Nothing depends on this particular way of doing things. For instance, we obtained similar results when new nodes used a copying mechanism to create new links, modified from Kumar et al. [2000], though homophily had slightly less effect on HEG advantage.



(a) (b)

Figure 2: HEG advantage over time, for different levels of homophily, where (a) the HUG is 50% of the population and (b) the HUG is 10% of the population.

of homophily, we formed 250 networks. For each of these networks, we performed 250 credit attribution contests (where a type 1 and type 2 individual were competing for credit) every 25 rounds to get an idea of how likely it was that each social identity group would get prestige for their discovery, and how these chances changed over time. Figure 2a shows what happens in this case.

We find that the HEG advantage disappears quickly over time. Interestingly, the HUG has an advantage for a short period of time (that is, the HEG advantage goes negative). As homophily increases, this temporary HUG advantage increases. This is likely because, when there are very few members of this group for the initial time period, the one or two that exist serve as focal points for the incoming members. As the HUG starts entering at higher rates, these focal points become highly connected to all the new people such that if a member of their social identity group makes a discovery, they will know about it. Since these focal individuals are so highly connected, they are likely to have at least some connections with new HEG members despite homophily, because people still do care to some extent about forming links with highly connected people. So news of the HUG members' discoveries can spread.

That the HEG advantage disappears quickly in this case looks somewhat promising — if we can get equal representation, eventually no group is disadvantaged.[14] A situation like this is achievable if we are thinking about men and women, but not if we are thinking about minority groups like racial minorities, people with disabilities, etc. In these cases, we can talk about what happens if

[14]It has been argued that in other situations, such as when there is a discriminatory bargaining norm in a community of scientists, merely increasing representation of the minority group will not eliminate inequalities [Schneider et al., 2019].

proportional, rather than equal, representation is achieved.

As seen in figure 2b, the situation is different when the HUG is a minority group. In this case, even when the HUG achieves proportional representation within two generations, the HEG advantage can be more severe and last longer, especially in the presence of strong homophily. When homophily is low, the HEG advantage disappears around when the HUG achieves proportional representation (with no real period of advantage for the HUG). But as homophily increases, a qualitative change appears: the HEG advantage can persist over time, meaning the HUG may be less likely to get credit for their discoveries indefinitely.

3.3 Research programs

There is also the further question of whether the priority rule can serve the function of promoting an optimal division of cognitive labor, in the way Strevens suggests.[15] In order for the priority rule to provide this benefit, it must be the case that scientists believe they will receive prestige for their discovery (not merely that they ought to), so that they decide which research program to join based on the likelihood of making a discovery. Merton [1957] already provides some evidence that scientists believe recognition is not automatic after discovery, and (to the contrary) is something that is more likely for some than others. He quotes Norbert Weiner explaining:

> I was competitive beyond the run of younger mathematicians, and I knew equally that this was not a very pretty attitude. However, it was not an attitude which I was free to assume or to reject. I was quite aware that I was an out among ins and I would get no shred of recognition that I did not force (p. 649).

If news of a discovery must travel through the network before a credit attribution contest is won and prestige bestowed, scientists likely also choose research programs based on the likelihood they will receive credit for their discoveries within that research program.

There are many ways such considerations could affect scientists' choice of research program. Our purpose in this section is to provide an example of how considerations of network structure and social identity matter to an analysis of the effects of the priority rule on program choice. To this end, we extend the model in section 3.2 to include scientists' decisions regarding which research program to join. We ask: how might the network structure and considerations of social identity pull a network away from the optimal division of labor?

To answer this question, we assume there is an optimal division when the network first reaches its full size, in order to study how quickly the community departs from optimal when scientists choose their research programs based on a combination of likelihood of making a discovery and likelihood of receiving

[15] Thanks to Remco Heesen for this point.

credit for a discovery.[16] In this model, a scientist chooses a research program, chooses an advisor within that program, then forms links with others in the community. While the advisor must be within the research program chosen, the other links formed may be with any scientist in the community.[17]

In the choice of research program, we match Strevens's set up as closely as possible. There are two research programs, each with some probability of success, s_1 and s_2. Both probabilities are functions of the number of people in the program, where it is assumed there are diminishing returns to each new person joining the program. Program 1 is assumed to be the better program, with a higher probability of generating a discovery, i.e. $s_1 > s_2$. The optimal distribution maximizes the overall probability of making a discovery, $s_1 + s_2 - s_1 s_2$.[18] In Strevens's set-up, a scientist chooses a research program based on the likelihood the program makes a discovery and, in the event that both programs make the discovery, the likelihood their research program does so first, w_i. For research program 1, for instance, this would be $s_1 + s_1 s_2 w_1$. Our slight alteration to this set-up is that we interpret w_i as the likelihood a scientist gets credit in the event of multiple discovery, which scientists estimate based on the most recent round of credit attribution contests – how likely was it that a scientist of their type got credit in research program 1 (likewise in research program 2)? This allows us to see whether, when scientists consider the likelihood of receiving credit when choosing a research program, the priority rule will incentivize the optimal division.

Figure 3 shows how far the scientific community is pulled away from the optimal division over successive rounds when the HUG is 10% of the population.[19] We can see that, even though we start with an optimal division at round 100, the scientific community diverges from this division quite rapidly. While homophily makes some difference to these results, there is substantial deviation for all levels of H.[20]

The most plausible explanation is that scientists are choosing research programs based on the likelihood someone of their type will get credit for a discovery

[16] To keep the analysis simple, we also now start advising in the model when the network reaches its full size.

[17] The scientist's choices are made by the same procedure outlined in section 3.2, i.e. their advisor is chosen according to equation 2, their other links are formed by copying their advisor's links and forming their own links, and so on. (Though, as seems realistic, we assume a scientist is twice as likely to form a link with someone in their own research program.)

[18] Our results use the following equations: $s_1 = \frac{n}{n+10}$ and $s_2 = \frac{N-n}{N-n+50}$, where N is the size of the community and n is the number of people in research program 1. With these equations, and a community of size 100, the optimal division is to have 70 people in program 1 and 30 people in program 2.

[19] Distance from optimal is just |*actual # in program 1 - optimal # in program 1*|.

[20] The HEG advantage remains very similar to what is shown in figure 2b. Though, interestingly, if we assume that people are ten times as likely to link with others in their own research program (as opposed to twice as likely, as for the results in figure 3), HEG advantage increases notably, e.g. when H=.8 the HEG advantage is .35 by round 500. This suggests that HEG advantage is not a particularly good predictor of how far the community will be from an optimal distribution. This may be of interest, as HEG advantage could plausibly be measured, while distance from optimal is a theoretical quantity, which is more difficult to directly assess.

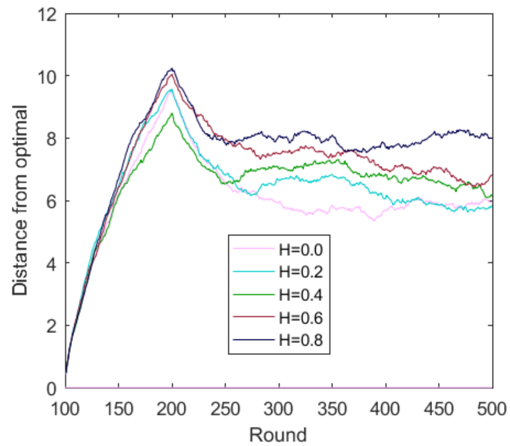


Figure 3: Distance from the optimal distribution over time, for different levels of homophily, where the HUG is 10% of the population.

made within that research program. In fact, we find that differences in w_1 and w_2 for each type predict whether that type will make up a greater proportion of research program 1 or 2. Let p_i refer to the proportion of program i's members which are HEGs. For the HEGs, the covariance between what we might call their 'credit differential' $(w_1 - w_2)$ and 'research program differential' $(p_1 - p_2)$ 25 rounds later (after new scientists have joined programs using the estimates of w_1 and w_2) is .79. The covariance between similarly defined 'differentials' for the HUGs is the same. Scientists tend to go where they think its more likely they will get credit for their discoveries, in which case the optimal division of cognitive labor is not incentivized.

4 Concluding remarks

Strevens [2003] interprets the priority rule in science as a behavior regulator for the scientific community, part of a grand reward scheme which benefits society by adequately structuring the distribution of intellectual labor across pre-existing research programs. However, prestige does not necessarily go to who deserves it. People are not always rewarded based on the benefit they confer, and so it is difficult to regard the priority rule as part of a grand reward scheme that resonates with our notions of fairness.

Further, if scientists instead believe the correct picture of awarding prestige is closer to our discussion of the priority rule (and consequently that news of a scientist's discovery must spread through the scientific network), Strevens's optimal division of labor is likely not what is incentivized. Section 3.3 gives one example of how social identity and network structure may affect scientists' estimations of how likely they are to receive credit for a discovery, leading to deviations from optimal division of labor, but there are many other factors that scientists may consider. For instance, in the context of our basic model (section 3.1), scientists might be incentivized to pick a research program with fewer well-connected people, so as to decrease the likelihood that the prestige associated with their potential discovery will be "scooped" by another who is in a better position to capture it.

If less-established scientists' contributions are more likely to be recognized in a less promising research program, the scientific community may end up with an inefficient division of cognitive labor, with more scientists working in a less promising program than what would be optimal. This is because not only do less-established scientists consider how likely it is that a research program will lead to a discovery, they also consider how likely it is that their achievements will be recognized and rewarded with the commensurate prestige. In other words, the disincentive of being scooped may often be in tension with a scientist's considerations of the "intrinsic potentials" of the competing programs, as considered by Strevens.[21]

Absent further study, it is difficult to discern which of these various factors (or others) clearly dominates in the decision-making of scientists,[22] and even more difficult to suss out the epistemic consequences of such decision-making; we leave this investigation for future work. Our point is that these considerations are absent in current 'credit economy' models of science, and that the extent to which any such strategies are motivated at all can only be evaluated in a model of the priority rule which takes into account individual scientists' credit attributions. Altogether, the epistemic benefits of the priority rule that Strevens sees are far from guaranteed when we take seriously the implications of a distributive reading of the statement: The scientific community disburses prestige.

References

A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

A.-L. Barabási, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, and T. Vicsek. Evolution of the social network of scientific collaborations. *Physica A: Statistical mechanics and its applications*, 311(3-4):590–614, 2002.

A. Boschini and A. Sjögren. Is team formation gender neutral? evidence from coauthorship patterns. *Journal of Labor Economics*, 25(2):325–365, 2007.

[21] Worse still, it is worth flagging that this state of affairs may also eventually lead to clustering into sub-disciplines according to social identity. Following the arguments given by Schneider et al. [2019], there is some historical precedent to suggest that such clustering is ultimately detrimental to the general state of our scientific knowledge across disciplines.

[22] One might also want to consider subtleties arising from different reference classes valuing contributions differently [Lee, 2020].

S. Currarini, M. O. Jackson, and P. Pin. An economic model of friendship: Homophily, minorities, and segregation. *Econometrica*, 77(4):1003–1045, 2009.

A. del Carmen and R. L. Bing. Academic productivity of African Americans in criminology and criminal justice. *Journal of Criminal Justice Education*, 11 (2):237–249, 2000.

M. A. Ferber and M. Teiman. Are women economists at a disadvantage in publishing journal articles? *Eastern Economic Journal*, 6(3/4):189–193, 1980.

G. Ghiasi, P. Mongeon, C. Sugimoto, and V. Larivière. Gender homophily in citations. In *23rd International Conference on Science and Technology Indicators (STI 2018)(September 2018)*, pages 1519–1525, 2018.

R. Heesen. Communism and the incentive to share in science. *Philosophy of Science*, 84(4):698–716, 2017.

R. Heesen. Why the reward structure of science makes reproducibility problems inevitable. *The Journal of Philosophy*, 115(12):661–674, 2018.

R. Hegselmann. Thomas C. Schelling and James M. Sakoda: The intellectual, technical, and social history of a model. *Journal of Artificial Societies and Social Simulation*, 20(3), 2017.

F. Karimi, M. Génois, C. Wagner, P. Singer, and M. Strohmaier. Homophily influences ranking of minorities in social networks. *Scientific reports*, 8, 2018.

K. Kim and J. Altmann. Effect of homophily on network formation. *Communications in Nonlinear Science and Numerical Simulation*, 44:482–494, 2017.

P. Kitcher. The division of cognitive labor. *The journal of philosophy*, 87(1): 5–22, 1990.

R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal. Stochastic models for the web graph. In *Foundations of Computer Science, 2000. Proceedings. 41st Annual Symposium on*, pages 57–65. IEEE, 2000.

K. N. Laland, K. Sterelny, J. Odling-Smee, W. Hoppitt, and T. Uller. Cause and effect in biology revisited: is Mayr's proximate-ultimate dichotomy still useful? *Science*, 334(6062):1512–1516, 2011.

C. J. Lee. The reference class problem for credit valuation in science. *Philosophy of Science*, 87(5):1026–1036, 2020. doi: 10.1086/710615.

K. Ludwig. *From individual to plural agency: collective action I*. Oxford University Press, 2016.

J. M. McDowell and J. K. Smith. The effect of gender-sorting on propensity to coauthor: Implications for academic promotion. *Economic Inquiry*, 30(1): 68–82, 1992.

R. K. Merton. Priorities in scientific discovery: a chapter in the sociology of science. *American sociological review*, 22(6):635–659, 1957.

M. E. Newman. Clustering and preferential attachment in growing networks. *Physical review E*, 64(2):025102, 2001.

W. F. Ogburn and D. Thomas. Are inventions inevitable? a note on social evolution. *Political Science Quarterly*, 37(1):83–98, 1922.

G. Paris, G. De Leo, P. Menozzi, and M. Gatto. Region-based citation bias in science. *Nature*, 396(6708):210–210, 1998.

F. Romero. Novelty versus replicability: Virtues and vices in the reward system of science. *Philosophy of Science*, 84(5):1031–1043, 2017.

M. D. Schneider, H. Rubin, and C. O'Connor. Promoting diverse collaborations. In G. Ramsey and A. De Block, editors, *The Dynamics of Science: Computational Frontiers in History and Philosophy of Science*. 2019.

M. Seevinck. Challenging the gospel: Grete Hermann on von Neumann's no-hidden-variables proof. In *Grete Hermann-Between Physics and Philosophy*, pages 107–117. Springer, 2016.

M. Strevens. The role of the priority rule in science. *The Journal of philosophy*, 100(2):55–79, 2003.

Y. S. Wang, C. J. Lee, J. D. West, C. T. Bergstrom, and E. A. Erosheva. Gender-based homophily in collaborations across a heterogeneous scholarly landscape. *arXiv preprint arXiv:1909.01284*, 2019.

D. A. Wardle. Journal citation impact factors and parochial citation practices. *Bulletin of the Ecological Society of America*, 76(2):102–104, 1995.

J. D. West, J. Jacquet, M. M. King, S. J. Correll, and C. T. Bergstrom. The role of gender in scholarly authorship. *PloS one*, 8(7):e66212, 2013.

K. J. Zollman. The credit economy and the economic rationality of science. *The Journal of Philosophy*, 115(1):5–33, 2018.