1 Introduction

On a naive view, science involves a set of practices that unerringly march toward the truth. Scientists use the best methods available to gather evidence. They reason dispassionately about this evidence and change their beliefs and theories accordingly. They share their data freely and widely and listen carefully and fairly to the findings of other scientists. When they discover that their methods or theories are flawed, they abandon them for better ones.

Of course, the reality is messy and imperfect. Scientists are humans, and, like all human enterprises, science has successes and failures, good practices and poor ones. Foibles of human psychology impact science at every stage of the process, from grant seeking, to hypothesis choice, to evidence gathering, to theory generation, to argumentation and publication.

This messy reality means two things. First, to understand the workings of science, researchers must study it as a human enterprise. Second, this study has the potential to improve scientific practice. While science is imperfect, it is also often self-reflective and self-correcting. By studying science, it is possible to make discoveries about which features of the scientific process are the most successful (or the most problematic) and make changes accordingly.

Starting in the mid-twentieth century, theorists have engaged in just this sort of study under the headings of "philosophy of science," "sociology of science," "the science of science," and, more recently, "metascience." Researchers across the social sciences, philosophy, and even STEM disciplines like engineering, biology, and computer science have investigated their own practices, and the practices of their colleagues, to see how science works and how it might work better.

The goal of this Element will not be to overview this broad ranging literature but to focus on one part of it – research using models to understand scientific communities. In recent decades, the use of models to study human behavior has become increasingly popular. Especially as researchers gain access to more and more computational power, it has become clear that mathematical and computational representations of human groups have the capacity to elucidate a wide range of phenomena. In particular, models are useful for reasoning about groups and processes that are complicated and distributed across time and space, that is, those that are difficult to study using empirical methods alone. Science fits this picture. Scientific theory change, for example, often happens over significant time spans and involves thousands of interactions between hundreds of researchers performing hundreds of experiments. For this reason, it is no surprise that researchers have turned to models to study various features of science. As we will see, these models can play many roles in the study of scientific communities.

This Element is a short overview. For most of the models described, I will not go into much mathematical detail, instead focusing on general descriptions and take-aways. Notably, this Element will not give an in-depth survey of other theoretical and empirical work on science, except inasmuch as this work is relevant to the models of science discussed. It also will not survey the large literature using models to study social epistemology – the spread and development of ideas, beliefs, opinions, and knowledge in human groups – more generally. This literature, which ranges across the social sciences and philosophy, has yielded many important insights about human knowledge production. Only those insights especially germane to thinking about scientific communities will be covered here.[1]

The different sections of this Element will mostly be organized around different modeling approaches. Section 2, The Credit Economy, looks at models where scientists seek academic credit. These models are derived from game- and decision-theoretic approaches that treat humans as utility maximizers in order to explain and predict human behavior. As we will see, this work on the material incentives that scientists face yields insights on topics ranging from the division of scientific labor to sharing of academic research, to fraud. Section 3, The Natural Selection of Science, looks at models with a slightly different assumption – various people and practices in scientific communities undergo variations of selective processes similar to those seen in biological populations. By focusing on selective processes, these models elucidate emergent phenomena that go beyond the credit-seeking choices of individual scientists. These include the persistence of poor research methods, the effects of interdisciplinarity on progress, and industry influence on science via strategic funding. In Section 4, Social Networks and Scientific Knowledge, we see models that focus on the social connections between scientists and consider how these social connections impact things like theory change and belief spread. As will become clear, the way information flows in scientific communities deeply impacts the progress of science. Section 5, Epistemic Landscapes, focuses on problem choice in science and how strategies for problem choice can benefit or slow discovery. In particular, this section considers "landscape models" that represent a space of problems scientists can move through and explore. These models shed light, in particular, on the role of cognitive diversity in scientific communities.

[1] Readers may also notice that this Element puts extra focus on the literature coming out of philosophy of science. Throughout, I try to incorporate a broad range of work modeling scientific communities. But I am a philosopher after all.

Section 6, The Replication Crisis and Methodological Reform, considers a set of models with formal similarities that are also topically unified. This is because the models in question were developed alongside the metascience movement in response to the replication crisis. For the most part, these models consider various statistical practices and how they impact data gathering and inference in scientific communities. Central questions include: Why have so many findings failed to replicate? What are the main incentives and practices leading to this failure? What responses and interventions can improve scientific practice in the future? The conclusion of the Element summarizes and synthesizes policy recommendations from models throughout the Element.

Before beginning, I want to take a little more space for a (brief) discussion of model epistemology. What can models like those presented here tell us about science? And how should we take them to inform our understandings of it? Science is complicated and multifaceted. Science is diverse and varied. Any attempt to yield general theories of the workings of "science" will necessarily fail. The models presented in this Element are in keeping with an approach that works piecewise to improve the understanding of certain features, processes, and parts of the scientific enterprise. As such, none of the modeling results here should be taken as the be-all-end-all on some topic. Rather they are just one set of investigations contributing to our understanding of this complex and long-standing human enterprise.

Even so, we need to be careful about what we take away. Whenever simplified models are used to study complex social realities, there is room for error. Sometimes models fail to account for important real-world factors and, thus, support mistaken conclusions. Sometimes models abstract away from their target systems so severely that it is hard to assess their value.[2] That said, as we will see in this Element, models can nonetheless play a variety of important roles in the study of science. They can suggest new hypotheses for future study, challenge impossibility claims, suggest interventions that might not have been obvious, identify ways that proposed interventions might go wrong, and so on. In addition, they can act as an aid to ordinary reasoning or theorizing. Whether any particular model will be appropriate for some epistemic role will depend on the details of that model and the ways it is used. Many of the models presented here can play successful roles in some sorts of argumentation and inference,

[2] I do not really think that models are importantly different from many empirical investigations in this regard. Most studies involve abstracted representations of a full reality (say, twenty subjects answering questions in a lab). Like models, empirical data can only support inferences about the world that are appropriately tuned to the data gathered and the system it targets. That said, with highly simplified social models, there are often many ways that such inferences can go wrong.

even if they are not appropriate for others. While I will not be able to assess the quality and applicability of every model discussed, I will often describe the ways I think they are successfully used in argumentation about science.

As an example, many of the models overviewed focus on assessing various policy proposals. They are useful tools for doing so because it can be costly and/or difficult to implement new social policies. Models are a relatively cheap and easy way to start exploring how some new policy might impact practice. But it is risky to go directly from modeling outcomes to policy proposals for the reasons just mentioned. Instead, some of the models presented generate new (sometimes unexpected) hypotheses about what sorts of outcomes can follow policy interventions. When they do, it is often worthwhile empirically testing these hypotheses. In this way, a simple model does not tell us directly about what will happen in a complex reality, but opens up possibilities for study. While the model in question would not be an appropriate tool to directly shape policy, it is an appropriate tool to spur further exploration.

This example includes a generalizable lesson. The models described in this Element sometimes yield take-aways that are straightforward. More often, the best way to employ them to understand and improve science is in combination with empirical methods and other sorts of theorizing. Empirical studies of science help us build good models. Models help shape theory. Theory directs empirical research, which sometimes prompts further modeling. Via this sort of back and forth, models and empirical tools can work hand-in-glove to improve our understanding of the complex processes underway in scientific communities and help us shape the future of science.

2 The Credit Economy

Zihan works in astrophysics and had been planning to investigate a certain, exciting pattern in nebula formation. When she hears that another very prominent team is working on the same problem, though, she worries that they will publish first and get credit for the discovery. She decides to switch her group to another more modest project.

Jerome studies emotions in infants. After preparing his latest manuscript, he spends a long time deciding which journal to send it to. The most prominent ones would really boost his reputation but take a long time to review. The chance of his work getting accepted is low, and he might waste his time submitting. In the end, he tries for a lower level journal since his tenure review is coming up the following year.

In graduate school, Firuzeh's mentors were highly critical of her work unless it was absolutely stellar. What she did not know was that these critical reactions were shaped, at least in part, by the fact that she was a Muslim woman. Over time, she developed an expectation that finished academic work should be of extremely high quality if she wanted to get it accepted for publication, and, as a result, she started taking a very long time to perfect her work before submission.

Alice and Andy are two co-PIs working on human gene sequencing in competition with a number of other labs. They develop a new technique that will allow them to yield gene sequences much more quickly. If they share this technique, they will be credited for having discovered it, but other labs will be able to use it. If they wait, they risk another team developing the same technique and getting credit for it. But in the meantime, their research will go more quickly than their competitors. In the end, they decide to wait to share their new technique until they are further along in the project.

• • •

All the aforementioned examples involve scientists who are making decisions for strategic reasons. The branches of mathematics typically used to model this sort of decision-making are game and decision theory. In such models, agents are usually treated as utility maximizers. These models assume agents' actions yield different payoffs depending on either the behaviors of interactive partners or the structure of the world. By supposing that agents will prefer whatever actions maximize their expected payoffs, the models help explain and predict strategic behavior in humans.

Credit-economy models of science apply game- and decision-theoretic models to science but with a twist. Instead of maximizing payoff generally, these models assume that scientists attempt to maximize "credit." While credit in this sense is not a perfectly defined concept, it approximately tracks reputation and status in science, and attending benefits: fancy jobs, good pay, prestigious talk invitations, and so on.[3] In each of the aforementioned examples, the scientists in question made decisions not because they wanted to increase their production of useful knowledge but because they wanted successful careers. The sociologist of science Robert Merton was one of the first to clearly describe the credit motives of scientists (Merton, 1973). In his, and subsequent, work, it has been well established that many of the decisions scientists make day to day are indeed driven by credit motives. These motives, in turn, are shaped by credit structures of science – norms like the "priority rule," which gives credit

[3] Dasgupta and David (1994) describe the credit system as follows: "the greater the [scientific] achievement, the larger the rewards – which may come eventually, if not immediately, in the form of salary increases, subsequent research grants, scientific prizes, eponymy, and, most generally, peer-group esteem" (499).

only to the first scientist to make a discovery, journal practices, like publication bias (only publishing positive results), grant-giving rules, and so on.

In this section of the Element, we consider models that start with the assumption that scientists are motivated by credit and see how these motives might shape outcomes in science. There are a significant number of discussions that describe and defend this general approach. We will not address these arguments in detail, but interested readers should see Dasgupta and Maskin (1987), Goldman and Shaked (1991), Dasgupta and David (1994), Stephan (1996), Polanyi et al. (2000), Leonard (2002), Hull (1988), Strevens (2011), and Zollman (2018).

2.1 The Division of Scientific Labor

One might instinctively think of credit motives as a bad thing in science. Should we not expect that greedy or impure motives will drive scientists toward bad practices? And won't scientists with "purer" motivations, related to finding the truth, do better work? These questions actually go as far back as Du Bois (1898) and drive much of the literature described in this section of the Element.[4]

In an early credit-economy model, Kitcher (1990) argues that credit incentives can actually help scientific progress by improving the division of labor. It is typically desirable for members of a scientific community to work on an array of different topics or approaches. By doing so, they ensure that important discoveries are not missed. A community that is too uniform with respect to problem choice/theory adoption risks settling on theories that are suboptimal or failing to make potential breakthroughs. This is sometimes referred to as the "division of scientific labor."[5] But suppose that all scientists are purely motivated by a desire to discover true things. And suppose further that they have access to the same sorts of information and evidence. If so, they may agree on what topics for exploration are most promising and fail to divide labor effectively.

In Kitcher's model, scientists choose between projects, each with some intrinsic quality or tendency to succeed. He assumes scientists share an objective assessment of which projects are most promising. Thus, if they choose a project based on epistemic merit alone, they fail to divide labor. When scientists are motivated by credit, however, they are attracted to projects that fewer

[4] Du Bois (1898) argues for epistemic motives. In an even earlier work, Adam Smith argues that mathematicians and natural philosophers, unlike poets and fine writers, are not subject to credit-type motives and takes this to be a good thing (1759, part III, chapter 2). Many others have weighed in on this general debate, but we do not overview this literature for space reasons.

[5] Later, in Section 4, we will see a similar topic glossed as "transient diversity of practice" in science.

peers are currently working on. This is because they are more likely to be the one to make important discoveries on such projects and, thus, to receive credit (like Zihan who decided not to work on nebula formation). As Kitcher argues, "The very factors that are frequently thought of as interfering with the rational pursuit of science – the thirst for fame and fortune, for example – might actually play a constructive role in our community epistemic project" (16).[6]

Strevens (2003) uses a similar model to argue for the benefits of a specific credit-allocation rule in science – the priority rule. As noted, this rule specifies that only the first scientist to make a discovery receives credit, even if another scientist is unaware of the previous finding and even if the discoveries are nearly concurrent (Merton, 1957; Strevens, 2003). In Strevens's model, researchers again choose between projects and receive credit incentives either in line with (1) the priority rule, or some alternative, including rules that (2) give credit based on marginal contributions to research and (3) give credit to all scientists who make a discovery. He shows that all these incentive schemes can drive the division of labor but that, in doing so, the priority rule puts extra incentive on the most promising projects.[7] Strevens takes this result to help explain why science has adopted the priority rule. In science, a discovery need only be made once for its benefits to be conferred on society. In such a scheme, the division of labor yielded by the priority rule is particularly efficient on his model.[8]

Some have shed doubt on the usefulness of these models. Zollman (2018) points out that if scientists are motivated by a pure desire that the truth be discovered, they are already incentivized to divide labor in the ideal way to facilitate this discovery. Division of labor is the best way to ensure this discovery after all. Credit will only help if they are only motivated by a desire to discover the truth *themselves* and do not want another researcher to make the discovery. (But, one might ask, why would a truly truth-motivated scientist care who makes a discovery?) Bedessem (2019) argues that these and other models representing division of labor in science fail to track the complexity and variability of scientific problems/theories. Reijula and Kuorikoski (2019) criticize Strevens's model by pointing out that he fails to provide a mechanism for how credit incentives might emerge to effectively divide scientific labor. Goldman (1999) and Viola (2015) point out that there may be better ways to

[6] Remarkably, in 1879, C. S. Peirce developed a model of division of labor in science with many similarities to Kitcher's. He was not interested in credit motives, though (Peirce, 1967).
[7] See also Kleinberg and Oren (2011).
[8] Strevens (2013), though, points out that if scientists tend to overestimate their likelihood to contribute to a research program, the priority rule in his model will drive too many of them to work on the more promising project.

coordinate division of labor that take into account centralized funding bodies (a topic we will return to in Sections 4 and 5). Muldoon and Weisberg (2011) criticize the assumptions that (1) scientists know how other scientists are distributing their labor and (2) can calculate the likelihood of success for different projects (thus calculating how they should best distribute their own labor). They develop an agent-based version of the model and find that when agents know only the research choices of a few community members, credit incentives do not work to divide labor. This is because agents do not have the proper information to incentivize them to choose the less promising alternative. And De Langhe (2014) points out that these models focus on dividing labor between existing options, rather than the exploration of new possibilities in science. He develops a credit model where agents can either explore new theories or test existing ones. He argues that the priority rule incentivizes exploration, while the fact that scientists tend to credit those working on similar topics to themselves incentivizes the study of existing theories. This addresses a different sort of division of labor in science – between exploiting the known and exploring the unknown. We will return to this issue at greater length when looking at epistemic landscape models in Section 5.[9]

In the end, do the models support the claim that credit incentives improve scientific division of labor? The evidence is mixed. A further observation is that scientists generally are complex and different. They do not typically assess the potential quality of theories or research topics in similar ways. They have different training and different interests that shape their research choices. Division of labor in science is often driven by these sorts of factors as much as credit motives. In assessing whether credit incentives in science are beneficial, and how we should shape them, there are other, arguably more pressing, issues than division of labor, which we turn to now.

2.2 Replication

Romero (2017) points out that Strevens (2003), and others advocating the benefits of credit motivations for division of labor, fail to consider the importance of replication. A hallmark of scientific knowledge is that it is replicable – re-running a scientific test should generate the same outcome. But the "replication crisis" has created massive upheaval as researchers in a number of disciplines have discovered that many core findings fail to replicate (Begley and Ellis, 2012; Open Science et al., 2015; Baker, 2016). (We return to this

[9] In contention with the models from Kitcher and Strevens, see also Dasgupta and Maskin (1987) who argue that credit incentives will instead push scientists to herd onto the same problems and approaches.

issue at length in Section 6.) As a result, many have advocated for researchers to spend more time replicating extant results. The priority rule, though, strongly disincentivizes replications, by assigning credit only to new discoveries. In other words, when we look at yet another sort of division of scientific labor – between seeking out new findings and verifying old ones – the priority rule causes problems. In support of this claim, Higginson and Munafò (2016) develop a model showing how the priority rule will tend to disincentivize the replication of existing results in favor of novelty, for just the reasons described.

In response to these sorts of issues, several authors argue that scientific communities should shape credit incentives to directly promote replication. Begley and Ellis (2012) argue that replications should always be required alongside new findings in order to publish. Romero (2018, 2020) advocates creating groups of scientists whose careers are entirely devoted to reproducing extant work. For these scientists, all credit then derives from attempting to replicate other experiments. On these proposals, credit incentives are re-engineered to avoid issues with priority.[10]

2.3 Fraud and Corner Cutting

Another worry about credit motivations is that they drive scientists to commit fraud or else to cut corners and engage in sloppy or imprecise research practices (Merton, 1973; Casadevall and Fang, 2012). Scientists who seek "fame and fortune" might be more likely to fabricate data supporting an impactful result. Likewise, scientists who aim to publish a lot of research quickly to gain credit, or win priority races, may be more likely to do sloppy work. Studies suggest that serious types of fraud are relatively rare but not insignificant in science. Most estimates put the percentage of researchers who have committed fraud at 1–3 percent (Fanelli, 2009; Bauchner et al., 2018; Xie et al., 2021) though some estimates are significantly higher (Gopalakrishna et al., 2022), especially those derived from reports estimating fraud among colleagues rather than oneself (Fanelli, 2009). In addition, estimates of the prevalence of less serious questionable research practices (QRPs) are much higher (Fanelli, 2009; Xie et al., 2021; Gopalakrishna et al., 2022).

[10] Lewandowsky and Oberauer (2020) use a simulation to explore another topic related to replication – the costs and benefits of either requiring replication of studies before publication or replicating only "high interest" (i.e., highly cited) papers after publication. The former has been advocated as a way to prevent the spread of false claims. They argue that both approaches are equally successful and advocate for the latter as more efficient in that fewer replications are required. However, their model does not consider practical costs to holding temporary false beliefs or the difficulties of retracting false research. Thus, the efficiency benefits they outline may be outweighed by other concerns.