Communism and the Incentive to Share in Science*

Remco Heesen†

August 24, 2015

Abstract

The communist norm requires that scientists widely share the results of their work. Where did this norm come from, and how does it persist? Michael Strevens provides a partial answer to these questions by showing that scientists should be willing to sign a social contract that mandates sharing. However, he also argues that it is not in an individual credit-maximizing scientist's interest to follow this norm. I argue against Strevens that individual scientists can rationally conform to the communist norm, even in the absence of a social contract or other ways of socially enforcing the norm, by proving results to this effect in a game-theoretic model. This shows that the incentives provided to scientists through the priority rule are sufficient to explain both the origins and the persistence of the communist norm, adding to previous results emphasizing the benefits of the incentive structure created by the priority rule.

*Thanks to Kevin Zollman, Michael Strevens, Stephan Hartmann, Teddy Seidenfeld, Thomas Boyer, Lee Elkin, Liam Bright, and audiences at the Bristol-Groningen Conference in Formal Epistemology and the Logic Colloquium in Helsinki 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

The social value of scientific work is highest when it is widely shared. Work that is shared can be built upon by other scientists, and utilized in the wider society. Work that is not shared can only be built upon or utilized by the original discoverer, and would have to be duplicated by others before they can use it, leading to inefficient double work.[1]

To put the point more strongly, it can be argued that work that is not widely shared is not really scientific work. Insofar as science is essentially a social enterprise, representing the cumulative stock of human knowledge, work that other scientists do not know about and cannot build upon is not science (cf. the distinction between Science and Technology in Dasgupta and David 1994). The sharing of scientific work is thus a necessary condition not merely for the success of science, but in an important sense for its very existence.

The sociologist Robert Merton first noticed that there exists an institutional norm in science that mandates widely sharing results. He called this the *communist norm*, according to which "[t]he substantive findings of science. . . are assigned to the community. . . The scientist's claim to 'his' intellectual 'property' is limited to that of recognition and esteem" (Merton 1942, p. 121). Subsequent empirical work by Louis et al. (2002) and Macfarlane and Cheng (2008) confirms that over ninety percent of scientists recognize this norm of sharing. Moreover, most scientists (if not as many as ninety percent) consistently conform to the communist norm.

The existence of this norm raises two questions. Where did it come from? And how does it persist? In light of what I said above, these are important questions. A good understanding of what makes the communist norm persist tells us which aspects of the institutional (incentive) structure of science can

[1] Of course scientific work is often duplicated by others even when it is shared (so-called replications). But this is not inefficient in the same way, as after the replication is shared the work is known by all to be more certainly established than if only one or the other instance was shared.

be changed without affecting the communist norm. Understanding its origins might allow us to reinstate the communist norm if it disappeared for whatever reason. Insofar as we value the existence and success of science, these are things we should want to know.

There must be some sense in which it is in scientists' interests to uphold the communist norm and conform to it, or else it would disappear.[2] One such sense is given by Strevens (forthcoming). He gives what he calls a "Hobbesian vindication" of the communist norm by showing that scientists should be willing to sign a contract that enforced sharing. The claim is that, from a credit-maximizing perspective, it is not beneficial for an individual scientist to share her work (which would help other scientists more than her), but every scientist is better off if everyone shares than if no one shares.

As Strevens is well aware, this only partially answers the question of the persistence of the communist norm, and says little about its origins. In contrast, I argue that *sharing is rational from a credit-maximizing perspective for an individual scientist*. If my argument is successful, it provides a much more detailed account of both the origins and the persistence of the communist norm. It also adds to a tradition of work in philosophy and economics that has emphasized the power of the priority rule to incentivize scientists to organize themselves in ways that further the aims of science (e.g., Kitcher 1990, Dasgupta and David 1994, Strevens 2003).

Because the existence of a norm can itself change what is in scientists' interests to do, the sense in which sharing is or is not rational or beneficial to scientists needs to be clarified. For this purpose, I rely on the terminology for social norms developed by Bicchieri (2006). I explain this terminology in section 2 and use it to state Strevens' position more precisely than I did above.

Section 3 sets out my own position by explaining how the idea that sci-

[2]It is perhaps debatable whether there would be a norm worth speaking of if scientists recognized an obligation to share but never acted on it, but since that is not the case nothing in this paper turns on that definitional question.

entists can publish and claim credit for intermediate results can be used to establish the rationality of sharing. Section 4 makes this more precise by describing a game-theoretic model of scientists working on a research project needing to decide whether to share their intermediate results.[3]

I then show that rational credit-maximizing scientists should indeed be expected to share in two versions of the model (sections 5 and 6). In section 7 I use these results to give an explanation of the persistence of the communist norm, and I consider some objections. I extend my explanation to include the origins of the norm in section 8, which involves considering boundedly rational scientists and some historical evidence. A brief conclusion wraps up the paper.

2 Social Norms and Communism

The question that this paper focuses on is whether it is in a scientist's interest to behave in accordance with the communist norm. Here, the crucial turning point is what is meant by a scientist's "interest". The specific question I want to raise is whether it would be in scientists' interest to share their work even in the absence of a norm telling them to do so. To clarify the distinction I have in mind, I use some terminology defined by Bicchieri (2006). She defines a *social norm* as follows:

> Let R be a *behavioral rule* for situations of type S, where S can be represented as a mixed-motive game. We say that R is a social norm in a population P if there exists a sufficiently large subset $P_{\text{cf}} \subseteq P$ such that, for each individual $i \in P_{\text{cf}}$:
>
> *Contingency*: i knows that a rule R exists and applies to situations of type S;

[3]The idea of using game theory to get a better understanding of norms in science goes back at least as far as Bicchieri (1988).

> *Conditional preference*: i prefers to conform to R in situations of type S on the condition that:
>
> (a) *Empirical expectations*: i believes that a sufficiently large subset of P conforms to R in situations of type S;
>
> and either
>
> (b) *Normative expectations*: i believes that a sufficiently large subset of P expects i to conform to R in situations of type S;
>
> or
>
> (b′) *Normative expectations with sanctions*: i believes that a sufficiently large subset of P expects i to conform to R in situations of type S, prefers i to conform, and may sanction behavior. (Bicchieri 2006, p. 11)

The crucial feature of this definition is the requirement of normative expectations. This says that an individual's preference to conform to the norm is conditional on others' expectations (possibly enforced by sanctions). For example, norms surrounding the sharing of food are plausibly social norms: in the absence of others expecting them to share, many people might prefer not to share even if they knew that a norm existed and most people followed it. In contrast, if an individual knows that in a particular country there exists a norm to drive on the right side of the road which is followed by most people, she would probably prefer to conform to that norm even if others had no expectations about her behavior.

In other words, a social norm actively works to change people's preferences: the norm makes it such that people expect each other to conform to it, and this expectation from others is itself necessary to make individuals prefer to conform. Other kinds of norms, such as descriptive norms and conventions, do not have this feature. They merely work with already existing preferences to help coordinate behavior.

The language of game theory is useful to sharpen these ideas. Recall that a *(Nash) equilibrium* is a situation in which each individual involved in the situation behaves in such a way that no individual has an incentive to deviate unilaterally. That is, keeping everyone else's behavior fixed, it is not in an individual's interest to change her behavior.

Consider a situation of type S and a putative norm R. If knowledge of R and empirical expectations (that others will conform to R) are sufficient to make an individual prefer to conform to R, then by definition R is an equilibrium of the underlying game that is being played in situations of type S. But if normative expectations are required, that is, if individuals only prefer to conform to R if others expect them to conform (and, possibly, are willing to back this up with sanctions), then R is not an equilibrium of the "original" game: it is only made into an equilibrium by the existence of the norm itself. So the existence of a social norm—unlike other kinds of norms—transforms the underlying game by changing people's preferences, thus creating a new equilibrium (Bicchieri 2006, pp. 25–27).

Is the communist norm a social norm in this sense, i.e., are normative expectations a necessary ingredient to make it in scientists' interest to share their work? In order to answer this question, one needs to know what scientists' interests are. In particular, an account of their motivations is needed that is independent of the communist norm, so that the question can be asked whether a self-interested scientist would share her work in the absence of a normative expectation that she do so.

A scientist's achievements create for her a stock of *credit*. This credit is the means by which she advances her career, which determines both her income and her status in the profession. Insofar as a scientist is someone who is interested in building a career in science, it is then in her interest to maximize credit. This claim has been defended by philosophers and sociologists as diverse as Hull (1988, chapter 8), Kitcher (1990), Strevens (2003), Merton (1957, 1969), and Latour and Woolgar (1986, chapter 5).

This is not to deny that the scientist may have other interests, either as a scientist (e.g., to advance human knowledge) or apart from being a scientist (e.g., to have time for other pursuits). But these are idiosyncratic.

I aim to show that sharing is beneficial to scientists as a consequence of an interest that all scientists share. Credit maximization is, in my view, the only candidate here.

What kind of achievements does the scientist get credit for? The answer is simple: scientific discoveries. The institutions of science put a premium on originality. Credit is awarded to the first scientist to publish some particular result, and the amount of credit awarded is roughly proportional to the significance of the result. This feature of science is known as the *priority rule*, and the extent to which it shapes scientists' behavior is well-documented (Merton 1957, 1969, Kitcher 1990, Strevens 2003).

By rewarding only the first scientist, the priority rule encourages scientists to work and publish quickly (Dasgupta and David 1994). In this way, it seems that the priority rule creates an incentive for scientists to share their work. However, "the same considerations give you a powerful incentive not to share your results before you have extracted every last publication from them" (Strevens forthcoming, p. 2). If results were shared before publication, this would improve other scientists' chances of scooping important discoveries for which those results are relevant. So, Strevens argues, there is a split in the motivations provided by the priority rule:

> The priority rule motivates a scientist to keep all data, all technology of experimentation, all incipient hypothesizing secret before discovery, and then to publish, that is to share widely, anything and everything of social value as soon as possible after discovery (should a discovery actually be made). The interests of society and the scientist are therefore in complete alignment after discovery, but before discovery, they appear to be diametrically opposed. (Strevens forthcoming, pp. 2–3)

Of course, any sharing that happens after a discovery has been made does not help science in coming to that discovery faster. Thus, at the crucial stage at which science can be sped up by sharing, the priority rule provides no incentive to do so, according to Strevens.

Strevens then goes on to show that a social contract, in which all scientists agree to widely share their work (even before discovery), would be beneficial to all scientists. In doing so, he shows that the problem of sharing has the structure of a Prisoner's Dilemma: every scientist would be better off if every scientist shared, but each individual scientist has an incentive not to share. The communist norm is thus a social norm on Strevens' view: without normative expectations to transform the game (into something that looks more like a Stag Hunt), widely sharing scientific work is not an equilibrium.

Strevens is not the only one to make this claim. For example, Resnik (2006, p. 135) observes that "the desire to protect priority, credit, and intellectual property" can motivate scientists to keep scientific results secret. Similarly, "[the priority rule] sets up an immediate tension between cooperative compliance with the norm of full disclosure (to assist oneself and colleagues in the communal search for knowledge), and the individualistic competitive urge to win priority races" (Dasgupta and David 1994, p. 500).[4] Claims like these are also made by Arzberger et al. (2004, p. 146), Borgman (2012, p. 1072), and Soranno et al. (2015, p. 70), among others.

3 Communism and Intermediate Results

In this paper I argue that, given the priority rule, it is in a scientist's own interest to share her work widely, at least whenever she expects other scientists to do the same. In other words, sharing widely is an equilibrium of the relevant game even in the absence of normative expectations. The problem of sharing is thus not like a Prisoner's Dilemma: the role of the communist norm is not to change scientists' preferences to make sharing attractive (at

[4]Dasgupta and David (1994, p. 502) go on to semi-formally characterize a situation very similar to the model of Boyer (2014) and this paper, but they draw the opposite conclusion: they agree with Strevens that conforming to the communist norm is structurally similar to cooperating in a Prisoner's Dilemma.

least not primarily).[5] It merely describes a rule of behavior that it is in scientists' own best interests to follow.

An important part of my argument is the insight that major discoveries can often be split into multiple smaller discoveries that were made along the way. Newton's famous comment "If I have seen further it is by standing on the shoulders of giants" illustrates this accumulative nature of science. Boyer (2014, p. 18 and p. 21) gives some more detailed examples: the construction of the first laser can be split into a theoretical development and the actual construction based on that theory, and the experimental test of the EPR thought experiment by Aspect et al. (1982) was preceded by a number of papers defining and refining the experiment.

It is noteworthy that in these cases each of the smaller discoveries was published as soon as it was done, rather than after the major discovery was completed. It is not obvious that it is always in an individual scientist's best interest to behave as these scientists did. On the one hand, credit can be claimed for the smaller discovery. On the other hand, the advantage that the smaller discovery gives on the way toward the major discovery is lost by publishing (and hence widely sharing) it. In fact, Schawlow and Townes seem to have lost the race to build the first working laser at least partially because their publication of the theoretical idea spurred on other teams.

Boyer (2014) provides a model to analyze this tradeoff. He shows that in some idealized circumstances the benefits of sharing these *intermediate results* outweigh the costs, with costs and benefits both measured in credit assigned via the priority rule. Although Boyer does not use these terms, his result could be used to argue that normative expectations are not necessary for the communist norm to arise: the priority rule encourages wide sharing of scientific work even before the potential of future discoveries based on this work has been exhausted, i.e., "before you have extracted every last

[5] I do not deny that normative expectations calling on scientists to share their work exist, as they in fact appear to do (Louis et al. 2002, Macfarlane and Cheng 2008). The point is rather that these are not required to explain the origins or persistence of the norm. I return to this point in section 7.

publication” (Strevens forthcoming, p. 2).

Unfortunately, things are not that simple. A number of objections can be made. The remainder of this section describes two such objections, which motivate the construction and analysis of a formal model in sections 4–6. In section 7 I flesh out the explanation of the communist norm suggested by this model, and I consider some further objections.

One may worry that Boyer’s result is not general enough to support claims about the origins or persistence of the communist norm. By his own admission, he only shows that “there exist simple and plausible research situations for which the [credit] incentive to publish intermediate steps is sufficient” (Boyer 2014, p. 29). I aim to show that in fact all or most research situations are such that there is a credit incentive to publish intermediate results, which requires a more general model. The results I obtain may be viewed as generalizations of Boyer’s—relaxing the assumptions that there are only two scientists, that the scientists are equally productive, and that scientists share either all or no intermediate results—although speaking strictly mathematically they are not (because Boyer uses discrete time steps and I use continuous time).

The second worry questions the relevance of equilibria. The worry may be either that showing that the communist norm is an equilibrium is not sufficient to show that one should expect real scientists to share, especially when there are also other equilibria (this is known as the equilibrium selection problem). Or alternatively one may disagree with Bicchieri that any observed behavioral rule has to be the equilibrium of some underlying game. I alleviate both of these worries by showing that the communist norm is not merely an equilibrium, but an equilibrium that one should expect to be realized by both fully rational and boundedly rational scientists. Thus, regardless of what one thinks of the general relevance of equilibria, the particular equilibrium considered here has behavioral implications.

4 A General Game-Theoretic Model of Intermediate Results

The game-theoretic model I develop in this section is intended to investigate scientists' incentives when they are working on a project that can be divided into a number of intermediate stages. An *intermediate stage* is a part of the project which, when completed successfully, yields a publishable intermediate result in the sense of Boyer (2014, section 2). I assume that stages can only be completed in one order.[6] The number of intermediate stages of the project is denoted k.

Competition plays a central role in the model. I assume that scientists are aware that other scientists are working on the same project (or at least believe this to be the case). Merton (1961) argued for the ubiquity of multiple discoveries in science, which suggests that scientists should almost always expect other scientists to be working on the same project. I thus assume that $n \geq 2$, where n is the number of scientists (or research groups) working on the project. Note that by "scientist" I mean not just one working in the natural sciences, but also the social sciences, the humanities, or any other creative field where the priority rule applies.

Whenever a scientist completes an intermediate stage, she has to make a choice: she can either publish the result, or keep it to herself.[7] Publishing benefits the scientist, because she thereby claims credit for completing that intermediate stage as well as any preceding stages that remain unpublished, in accordance with the priority rule. I assume that all stages are equally valuable, so the amount of credit obtained is equal to the number of stages

[6] This linearity assumption may seem restrictive and unrealistic. However, any alternative assumption would only make sharing more attractive by improving the chance that a scientist can claim credit for an intermediate result without hurting her chances of future credit (because, e.g., other scientists are following a different path within the research project and thus are not helped by the publication of the intermediate result).

[7] By assumption, the result is publishable, i.e., if she decides to publish it, it will be accepted by a journal.

published. Publishing also benefits the scientific community: other scientists no longer need to work independently on the stages that have been published. Publishing thus “expedites the flow of knowledge”. I use E to denote this strategy.

The way that the scientific community benefits from publications is a potential downside to the individual scientist: if she keeps her results secret instead, she can start working on the next stage before anyone else can. This improves her chance of being the first to successfully complete the next stage, thus allowing her to claim credit for more stages later (at the risk that someone else claims credit for the one she did not publish). Holding onto a discovery until a more expedient time might thus be beneficial to the scientist. Call this strategy H.

When a scientist completes the last stage there is no incentive (within the model) to keep her from publishing. So when a scientist completes stage k she always publishes, claiming credit for all unpublished stages and concluding this instance of the model.

Note that I assume that scientists care only about credit[8], and that the only way to get credit is by publishing. Scientists are thus not assumed to have any inherent preference for or against sharing their work. In particular, expectations (normative or otherwise) from other scientists are not built into the individual scientist’s preferences.

An interesting feature of the priority rule is its uncompromising nature. According to the priority rule, there are no second prizes, even if the time interval between the two discoveries is very small. This feature was noted by Merton (1957, p. 658), who quotes the French scientist François Arago as saying: “‘about the same time’ proves nothing; questions as to priority may depend on weeks, on days, on hours, on minutes.”[9]

[8]More precisely: I investigate the incentives provided to scientists through credit, independent of any other interests or incentives they might have.

[9]Merton (1957, pp. 658–659) goes on to argue that this is a pathological extreme: when the interval between two discoveries is so small, “priority has lost all functional significance.” I agree with Strevens (2003, section IV.1) that this is not obviously correct.

To incorporate this feature in the model, it needs to be able to distinguish arbitrarily small time intervals. This suggests a continuous-time model: a model using discrete time units might place two discoveries in the same time unit even though in reality one of them happened (slightly) earlier than the other. This would fail to adequately model the uncompromising nature of the priority rule.

This means that a continuous-time probability distribution is needed to model the *waiting time*: the time it takes a given scientist to complete an intermediate stage. For this purpose I use the exponential distribution, the only candidate that has significant empirical support behind it (Huber 2001, more on this below). In particular, I assume that the time scientist i takes to complete any intermediate stage follows an exponential distribution with parameter λ_i. The parameter can be interpreted as the average number of stages completed by the scientist per unit time. The parameter may be different for different scientists, indicating the possibility that some scientists work faster than others, or are part of a larger or more efficient research group.

The assumption that waiting times are exponential is equivalent to the assumption that scientists' productivity is a Poisson process with a parameter that is constant over time. Empirical work has shown that scientists' productivity fits a Poisson distribution quite well, and the percentage of authors who experience significant trends or surges over time is small. Huber (1998a,b) has established this for the rate at which patents are produced by inventors, Huber and Wagner-Döbler (2001a) for publications in mathematical logic, Huber and Wagner-Döbler (2001b) for publications in 19th century physics, and Huber (2001) for publications in modern physics, biology, and

---

A version of the priority rule which gives shared credit when the time interval between discoveries is below a certain threshold would create a different incentive structure for scientists, and it is an open question whether that incentive structure would be better or worse. In any case, here I simply take the uncompromising version of the priority rule as given.

psychology.[10]

The assumption that waiting times are exponential means that the probability that it will take scientist i more than t time units to complete a given stage is $\exp\{-t\lambda_i\}$.[11] This distribution has some formal features that I will make use of (Norris 1998, section 2.3). First, it is "memoryless". This means that after a certain amount of time has passed and the waiting time has not ended yet, the distribution of the remaining waiting time is equal to the original distribution of the waiting time. Second, the minimum of n independent exponential random variables with parameters λ_i $(i = 1, \ldots, n)$ is itself exponentially distributed with parameter $\lambda = \lambda_1 + \ldots + \lambda_n$. Thus, the waiting time until at least one of the scientists completes a stage of the project is exponentially distributed with parameter λ. Third, the probability that scientist i is the first one to finish the stage she is working on is λ_i/λ.

The memorylessness property may seem odd, as it suggests that the scien-

[10]The fact that scientists' total career productivity follows a Poisson distribution (if accepted) does not imply exponential waiting times. One could generate Poisson distributions in other ways. But the evidence regarding trends and surges, as well as the fact that the evidence includes scientific careers cut short, suggests the stronger claim that at any given time in a scientist's career the Poisson distribution is a good model for her productivity up to that point. On this interpretation it is a simple mathematical consequence that the waiting times are exponential.

[11]Compare this with Boyer's assumption that there is a fixed probability λ that a given scientist will solve a given stage in a given time unit. As noted above, by using discrete time units this model provides no way of applying the priority rule when two scientists finish the same stage in the same time unit. This problem can be addressed by using smaller time units. Suppose that what was previously one time unit is now x time units, and in each unit the scientist completes the stage with probability λ/x. Unfortunately, the same problem may still arise, and this will be true for any (finite) magnification factor x. The problem is solved by taking the limit as x goes to infinity. For any finite x, the probability that the scientist has not completed the stage at time t (measured in the original time units before magnification) is $(1 - \lambda/x)^{tx}$. In the limit the probability that the scientist has not completed the stage at time t is $\lim_{x\to\infty}(1 - \lambda/x)^{tx} = \exp\{-t\lambda\}$. So, in addition to being independently empirically justified, exponential waiting times naturally arise as the limiting case of Boyer's model where the priority rule can be applied unambiguously.

tist herself never knows whether she is making any progress on the problem. Moreover, if she starts working on a given stage much later than another scientist she has the same chance of completing it first as she would have had if both had started at the same time (conditional on the fact that the other scientist does not complete the stage before she starts).

While these features of the exponential distribution do not seem to mesh well with the subjective experience of working on a research project, I want to insist that Huber's empirical evidence should be given more weight than subjective experience. The following consideration may help to smooth the apparent conflict.

In the model, scientists only make decisions after they have just completed a stage. So for the model it only matters that when a scientist completes a stage, she views the time she needs to complete future stages and the time other scientists need to complete stages as exponentially distributed. I do not need to insist that the scientist views the time needed to complete stages as exponentially distributed while she is in the middle of one.

How does my model compare to the one given by Strevens (forthcoming)? Perhaps the key difference is that contrary to Strevens I have described a zero-sum game. In my model it is implicitly assumed that the scientists will always eventually complete the entire research project.[12] Since each stage is worth one unit of credit, and the first scientist to complete stage k always claims credit for it and any unclaimed stages, this means that at the end of the game the scientists have always divided k units of credit between them. So any change in strategy that leads to one scientist improving her (expected) credit must always lead to a decrease for at least one other scientist.

In contrast, a key component of Strevens' model is the chance each scientist has of successfully completing the research project "in isolation", which leaves room for the scenario in which the research project is never completed by anyone. By sharing their progress, Strevens assumes, the scientists improve each other's chances of completing the research project. In fact this

[12]More precisely: the scientists complete all k stages in finite time with probability one.

appears to be the main driving force behind his result that scientists should be willing to sign a social contract that enforces sharing: in his model sharing "creates" expected credit (by improving the overall chance that any credit is awarded at all), and as long as this "extra" credit is divided in such a way that everyone benefits at least a little (in expectation), it is clear that everyone will be better off if everyone shares.

By allowing for a chance that no scientist completes the research project, Strevens' model is arguably more realistic than mine. But I claim that this is a strength rather than a weakness of my model. Working with a zero-sum game reflects a strictly more pessimistic assumption about the benefits of sharing than working with a model like Strevens'. The result that sharing is incentive-compatible which I state and prove below is thus a somewhat surprising result: it is stronger than the result Strevens proved, while his model makes a more optimistic assumption about the benefits of sharing. Insofar as I show that the priority rule is sufficient for a communist norm to arise (without a need for normative expectations) in my model, this result should hold *a fortiori* in a more realistic (not zero-sum) model.

There are other ways to change the model that would make it no longer zero-sum. For example, Boyer and Imbert (forthcoming, section 4) argue that the relevant notion to consider is credit per unit time (rather than "total credit" which I use). This incorporates the idea that if the research project finishes faster the competing scientists will be free to work on other (potentially credit-worthy) projects sooner. Then sharing benefits all scientists to some extent by decreasing the expected completion time of the research project; Boyer and Imbert call this a "speedup effect". So considering credit per unit time instead of total credit also invalidates the zero-sum property. Since, as above, it does so in a way that makes sharing more attractive, the result I get in my model holds *a fortiori* when credit per unit time is used.

5 A Backwards Induction Analysis

The previous section described a game-theoretic model of scientists working on a project that requires some number of intermediate stages to be completed. The game consists of a sequence of (probabilistic) events, in which the scientists can intervene at specific points through their choice of strategy by publishing their work (E) or keeping it secret (H). Each scientist attempts to maximize her credit.

In the simplest version of the game there are two scientists ($n = 2$) and the research project has two stages ($k = 2$). The extensive form of the game is given in figure 1.

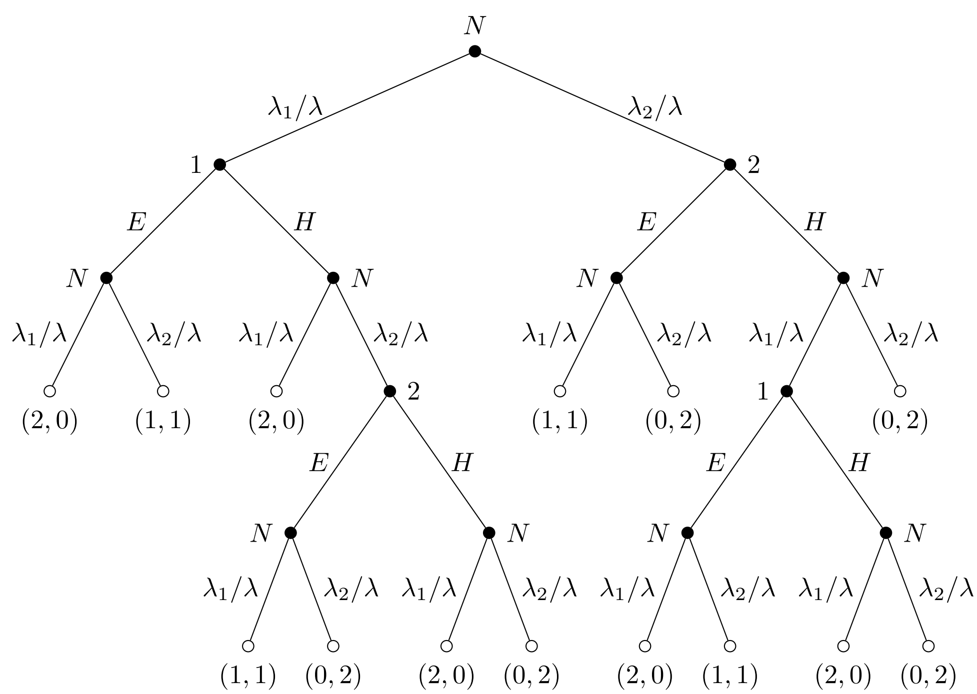

Figure 1: Extensive form of the game with $n = 2$ and $k = 2$

At the root node (marked "N") Nature decides which of the two scientists is the first one to complete the first stage of the project. As indicated, Nature picks scientist 1 with probability λ_1/λ and scientist 2 with probability λ_2/λ.

Suppose Nature picks scientist 1. This leads to a decision node marked "1", indicating that scientist 1 is the one to make a decision at this node. If scientist 1 publishes (strategy E), she collects one unit of credit. Both scientists now know the solution to stage 1 of the project, so they start working on stage 2.

Nature again decides with the indicated probabilities which of the two scientists completes the second stage first. In either case the game ends. If Nature picks scientist 1, she gets credit for completing both stages of the project and scientist 2 gets nothing (as indicated by the payoff pair $(2, 0)$ in the figure). If Nature picks scientist 2, she gets credit for completing the second stage, and since scientist 1 had already claimed credit for the first stage, both scientists end up with one unit of credit.

What if scientist 1 chooses not to publish her solution to the first stage of the project (strategy H at the node marked "1")? Then scientist 1 does not collect a unit of credit, and scientist 2 does not learn the solution to stage 1. So now scientist 1 starts working on stage 2, while scientist 2 continues to work on stage 1.

Once again Nature decides which of the two scientists finishes the stage she is working on first (due to the memorylessness of the exponential distribution, scientist 2 is not more likely to finish fast despite having already spent some time working on stage 1). If Nature picks scientist 1, she completes the project. The game ends and scientist 1 gets both units of credit.

If Nature picks scientist 2, she now has a decision to make (at the node marked "2"). She can claim one unit of credit by playing strategy E, or defer by playing H. In either case, both scientists can now work on stage 2.

Nature makes its final decision by picking a scientist who completes the second stage first. That scientist gets both units of credit (and the other gets nothing) if scientist 2 chose strategy H, whereas if scientist 2 chose E she gets one unit of credit for sure and the scientist picked by Nature gets the other unit.

The right-hand side of the figure (associated with Nature picking scientist 2 at the root node) works similarly.

If the first scientist to complete stage 1 plays H, and the other scientist completes stage 1 before the first scientist finishes the project, it is rational for the other scientist to play E: this makes it certain that she will get one unit of credit, without reducing either her probability of completing the second stage or her payoff if she does so, and without giving the first scientist any information she does not already have.

This is a so-called "backwards induction" argument: if a certain node is reached, then it is rational for the scientist who has to make a decision at that node to choose x; therefore, other scientists may assume that if that node is reached, x will be played. Applying this argument to the terminal decision nodes in figure 1 leads to a truncated game tree, as shown in figure 2.

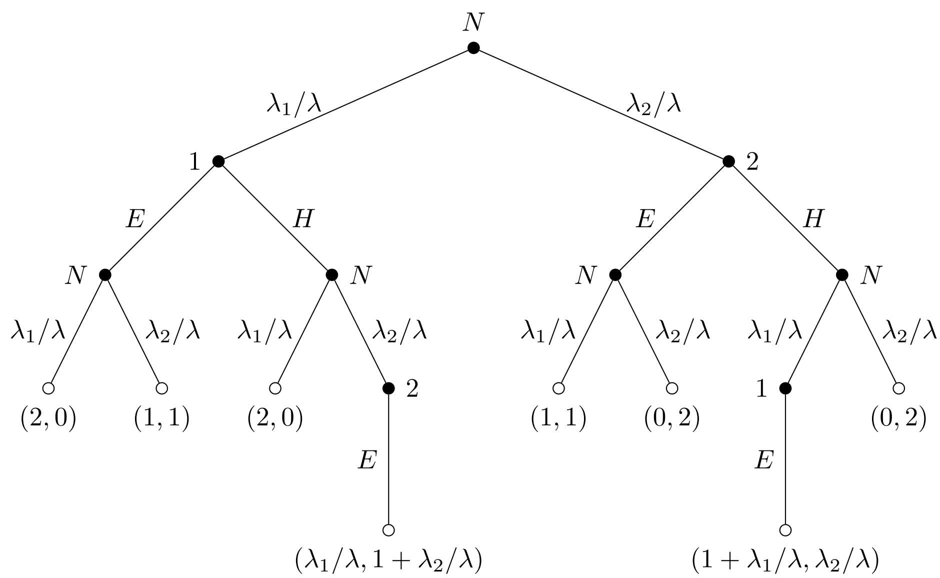

Figure 2: Truncated game tree for the game with $n = 2$ and $k = 2$

Here it is assumed that the second scientist to complete stage 1 always plays strategy E. The (expected) payoff of that strategy is one plus the probability of being the first to complete stage 2 for the scientist who just completed stage 1, and just the probability of being the first to complete stage 2 for the other scientist.

Now consider the decision scientist 1 has to make if she completes stage 1 first. If she plays strategy E, her payoff is 2 with probability λ_1/λ and 1 with probability λ_2/λ, so her expected payoff is $2 \cdot \lambda_1/\lambda + 1 \cdot \lambda_2/\lambda$.

If she plays strategy H instead, her payoff is 2 with probability λ_1/λ and λ_1/λ with probability λ_2/λ. So in this case her expected payoff is $2 \cdot \lambda_1/\lambda + \lambda_2/\lambda \cdot \lambda_1/\lambda$.

Since $\lambda_1/\lambda < 1$ and $\lambda_2/\lambda > 0$, the expected payoff of E is strictly greater than the expected payoff of H. So scientist 1 should play E if she is the first to complete stage 1 (and a similar argument applies to scientist 2).

Thus, the backwards induction solution of this game is for both scientists to play E at both of their decision nodes. Like any backwards induction solution, this is an equilibrium. The expected payoff if this equilibrium is played is $2\lambda_1/\lambda$ to scientist 1 and $2\lambda_2/\lambda$ to scientist 2.

Nothing in the above analysis depended on the values of λ_1 and λ_2. Moreover, it can be shown that a similar analysis goes through when the number of scientists or the number of stages is changed, as stated in the following theorem (see appendix A for a proof).

Theorem 1. *In the (unique) backwards induction solution to this game with $n \geq 2$ scientists and $k \geq 1$ stages, every scientist plays E at every decision node. Moreover, there are no equilibria that are behaviorally distinct*[13] *from the backwards induction solution.*

Backwards induction thus yields a unique prediction for this game. But under what circumstances should scientists be expected to behave according to the backwards induction solution? A sufficient condition is that all scientists are rational (maximizing expected credit) and that this fact is common knowledge among the scientists (Aumann 1995).

[13]That is, while there may be other equilibria, these differ only in that some scientists make different decisions at decision nodes that will not actually be reached in the game (given the strategies of the other scientists).

Common knowledge of rationality is a very strong assumption. It requires that scientists expect each other to behave rationally, even when they have already been seen to behave irrationally (making it a common belief rather than common knowledge). In the next section, I relax this assumption as well as an unrealistic assumption about the information that is available to the scientists in this game.

6 A Game of Imperfect Information

The analysis in section 5 uses backwards induction, in which one works through the game tree from the end of the game back to the beginning. This type of analysis relies on the scientists having very specific knowledge about the state of the game.

For example, in the case with two scientists and two stages I argued that it is rational for a scientist to play strategy E if she completes the first stage after the other scientist has already done so. This argument relies on the assumption that she can distinguish between the situation in which the other scientist has already completed the first stage but decided not to publish this information and the situation in which the other scientist does not have a solution to the first stage yet. Without this assumption the backwards induction analysis never gets off the ground.

Is it realistic to assume that scientists know the results their peers have obtained even when they have not published them? I think this differs from field to field. In small fields where everyone knows what everyone else is working on word gets around when one of the labs has solved a particular problem, even when they manage to keep the details to themselves. Or, with pre-registration of clinical trials becoming more and more common, scientists might know that some other scientist knows, say, whether a particular drug is effective, without knowing whether the answer is yes or no.

But in other fields this kind of information might not be available. In this section I analyze a version of the model in which scientists do not know if other scientists have any unpublished results. I retain the assumption that once a result is published all scientists know about it. This yields a *game of imperfect information*.[14]

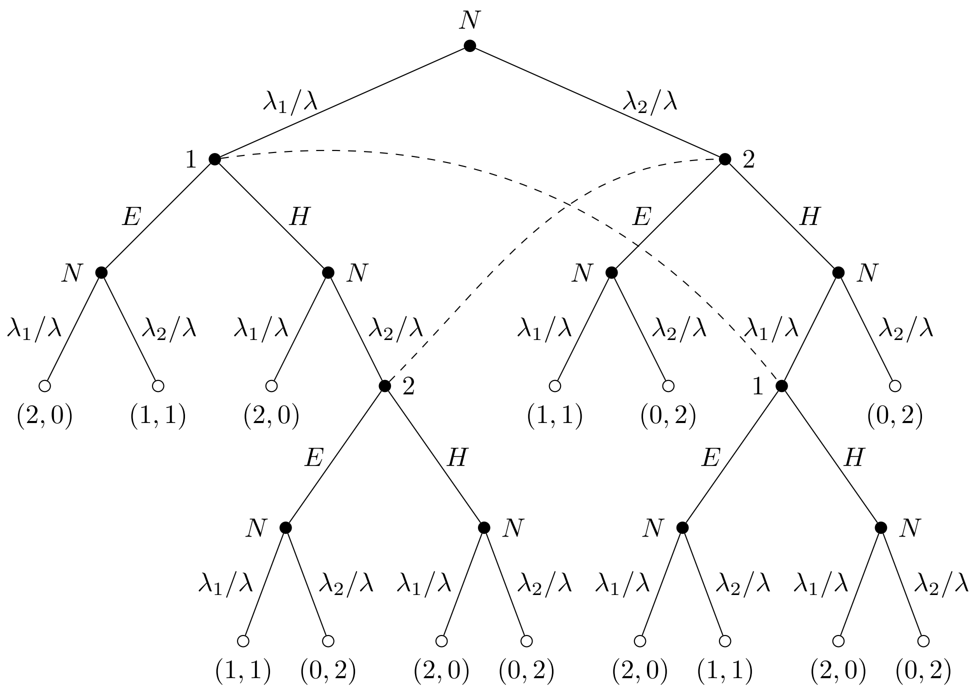

Figure 3: Extensive form of the game of imperfect information with $n = 2$ and $k = 2$

Figure 3 shows the extensive form of the game of imperfect information in its simplest form ($n = 2$ and $k = 2$). The only difference compared to figure 1 is the appearance of the dashed lines between decision nodes. These indicate so-called *information sets*: sets of decision nodes that the scientist who has to make a decision cannot distinguish between (i.e., she must play the same strategy at each node in the set).

[14]This is a technical term for a game in which players cannot distinguish certain decision nodes. Not to be confused with a game of incomplete information, where the players may not know each other's preferences or possible strategies.

What is rational for the scientists to do in this version of the game? On the one hand the information sets have made the problem harder, because backwards induction can no longer be used. But on the other hand they have also made the problem easier by reducing the number of possible strategies. Previously, each scientist had four possible strategies: they could play either E or H independently at either of their decision nodes. As they can no longer distinguish between their decision nodes, conditional strategies are no longer allowed, so each scientist only has two possible strategies: E and H.

	E	H
E	$\left(2\frac{\lambda_1}{\lambda}, 2\frac{\lambda_2}{\lambda}\right)$	$\left(2\frac{\lambda_1}{\lambda} + \frac{\lambda_1^2}{\lambda^2}\frac{\lambda_2}{\lambda}, 2\frac{\lambda_2}{\lambda} - \frac{\lambda_1^2}{\lambda^2}\frac{\lambda_2}{\lambda}\right)$
H	$\left(2\frac{\lambda_1}{\lambda} - \frac{\lambda_1}{\lambda}\frac{\lambda_2^2}{\lambda^2}, 2\frac{\lambda_2}{\lambda} + \frac{\lambda_1}{\lambda}\frac{\lambda_2^2}{\lambda^2}\right)$	$\left(2\frac{\lambda_1^2}{\lambda^2} + 4\frac{\lambda_1^2}{\lambda^2}\frac{\lambda_2}{\lambda}, 2\frac{\lambda_2^2}{\lambda^2} + 4\frac{\lambda_1}{\lambda}\frac{\lambda_2^2}{\lambda^2}\right)$

Table 1: Expected credit for each scientist as a function of scientist 1's strategy (row) and scientist 2's strategy (column)

Table 1 gives the expected credit for each scientist as a function of the scientists' choice of strategy and the values of λ_1 and λ_2. This game only has one equilibrium (including mixed-strategy equilibria) regardless of the values of λ_1 and λ_2: both scientists play strategy E. This can be seen by noting that strategy E is the unique best response if the other scientist plays E as well, and for at least one of the two scientists (possibly both, depending on the values of λ_1 and λ_2)[15] strategy E is also the unique best response if the other scientist plays H.

Moreover, this is a *strict equilibrium*. An equilibrium is strict if, keeping the other scientists' strategies fixed, deviating from the equilibrium strictly

[15]Strategy E is the unique best response to strategy H for scientist 1 if $2\frac{\lambda_1}{\lambda} + \frac{\lambda_1^2}{\lambda^2}\frac{\lambda_2}{\lambda} > 2\frac{\lambda_1^2}{\lambda^2} + 4\frac{\lambda_1^2}{\lambda^2}\frac{\lambda_2}{\lambda}$, which happens if and only if $2\lambda_2 > \lambda_1$. Similarly, strategy E is the unique best response to strategy H for scientist 2 if $2\lambda_1 > \lambda_2$. At least one of these conditions always holds, and both of them hold whenever $\frac{1}{2}\lambda_2 < \lambda_1 < 2\lambda_2$, i.e., when the values of λ_1 and λ_2 are "close".

decreases a scientist's expected credit. This is a stronger requirement than that for an equilibrium because that definition allows for cases in which a deviating scientist is equally well off.

In the general version of the game (with n and k possibly greater than 2) each scientist has to formulate a strategy (E or H) for each information set. At an information set, the scientist knows which stage was the last one to be completed and shared by some scientist, and how many stages she has since completed herself. However, she does not know how many stages have been completed but not shared by other scientists. As a result, the number of possible strategies is smaller than in the game of perfect information of section 5 (but greater than two if $k > 2$). It turns out that the general version of the game also has only one equilibrium.

Theorem 2. *The game with imperfect information with* $n \geq 2$ *scientists and* $k \geq 1$ *stages has a unique equilibrium in which all scientists play strategy* E *at every information set. Moreover, this is a strict equilibrium.*

A proof of this fairly strong theorem in favor of the sharing of intermediate results is given in appendix A.

What does this result say about what it is rational for a scientist to do? It says that if not every scientist immediately shares any stage that she completes, there is at least one scientist who is irrational in the sense that she would have had a higher expected credit if she had played a different strategy. So the only way these scientists can all be rational is if they all share every stage. In other words, if all scientists are rational expected credit maximizers they will all share.

7 Explaining the Persistence of the Communist Norm

Above I have shown in a game-theoretic model that it is rational for credit-maximizing scientists subject to the priority rule to share their intermediate results. I take this result to give an explanation for the *persistence* of the communist norm.

The explanation runs as follows. Suppose the communist norm is in place, i.e., scientists are sharing their intermediate results. If a given scientist deviates by not sharing an intermediate result, she thereby lowers her expected credit (this is just what it means for sharing to be a strict equilibrium). Hence the scientist has a credit incentive to return to conforming to the norm. So credit incentives can correct small deviations from the norm (and, since there are no other equilibria, arguably also larger deviations).

Note that I do not claim that real scientists are rational credit-maximizers. This is not necessary for my explanation. I have shown that rational credit-maximizing scientists would conform to the norm. All that follows for real scientists is that they have a credit incentive to conform to the norm. This fact, combined with the fact that real scientists are at least somewhat sensitive to credit incentives (more on this in section 8), constitutes my explanation of the persistence of the norm.

Here I want to point out a number of peculiar features of my explanation and consider some objections based on those features.

My explanation relies on only three basic principles: scientists' sensitivity to credit incentives, the credit-worthiness of intermediate results, and the priority rule as the mechanism for assigning credit. These ingredients are sufficient to explain the persistence of the norm. In particular, there is no need for a social contract, normative expectations, or altruism.

This leads to a potential objection. On my construal, the communist norm is not strictly a social norm in Bicchieri's sense, as normative expectations have no role in the explanation. But the sociological evidence cited above seems to refute this: scientists do view the communist norm as a social norm, they (normatively) expect other scientists to conform to it, and they feel the weight of this expectation when making their own decisions (Louis et al. 2002, Macfarlane and Cheng 2008). This appears to be at odds with my model: since the game is zero-sum, other scientists benefit when a given scientist deviates from the norm, so from a credit-maximizing perspective they should actually be encouraging each other to deviate.[16]

To answer this objection I note that the model considers only those scientists who are directly competing on a given research project. While those scientists may stand to gain if their competitors fail to share their intermediate results, the wider scientific community stands to lose, as it will take longer to complete the research project. It is this wider community, I argue, that is the source of any normative expectations regarding sharing behavior. The normative expectations can then also be explained from self-interest, as the completion of the research project may benefit other scientist' research.[17]

This yields an empirical prediction that might be used to help decide between Strevens' explanation and mine. On Strevens' explanation a deviation from the communist norm is a breach of a social contract which most directly impacts the immediate competitors of the scientist within the research project, who may legitimately regard it as unfair. On my explanation a deviation actually benefits the immediate competitors; the most direct negative impact is on those scientists who work on nearby projects. An examination of which scientists (direct competitors or those working on nearby projects) tend to most vocally object to deviations from the communist norm may thus shed light on the question which of these explanations is closer to the truth.

Because my explanation depends on the claim that it is rational for credit-maximizing scientists to share their intermediate results, which is supported by a game-theoretic model, the explanation depends on the generality of that model. While my model is more general than Boyer's in allowing an arbitrary number of competing scientists, arbitrary differences in productivity among the scientists, and in considering a large strategy space, still some assump-

[16] I thank Michael Strevens and Liam Bright for pressing me on this point.

[17] Alternatively or additionally, normative expectations may arise simply because everyone in the community is in fact behaving in a certain way. Bicchieri points out that "[s]ome conventions may not involve externalities, at least initially, but they may become so well entrenched that people start attaching value to them" (Bicchieri 2006, p. 40).

tions had to be made. Which of these assumptions are truly restrictive?

A number of assumptions are, perhaps surprisingly, not restrictive. The reason is that realistic ways of relaxing these assumptions would actually make sharing more attractive rather than less, and thus would not affect the results obtained in theorems 1 and 2. These are: (1) the assumption that the scientists always finish the project, (2) the assumption that scientists maximize (total) credit rather than credit per unit time, (3) the assumption that the project can only be completed by finishing these particular intermediate stages, and (4) the assumption that the scientists know in advance which intermediate stages need to completed.

This leaves two crucial assumptions: exponential waiting times and equal credit for different stages. It is possible that using a different distribution for the waiting times would lead to a model in which the equivalent of theorems 1 and 2 does not hold. But I claim that any such deviation would actually make the model less realistic, citing once again the empirical evidence obtained by Huber (1998a,b, 2001) and Huber and Wagner-Döbler (2001a,b).

I have given no real defense of the assumption that each intermediate stage has the same value (in terms of credit). In fact it seems quite realistic that the scientist to finish the last stage ("puts it all together") might get more credit. And this is exactly the circumstance in which my results may fail: if later stages are worth more credit than earlier ones, there may be an incentive not to share.[18]

Here I have little to add to Boyer (2014, section 4.3.1). From a descriptive perspective, this might be the kind of cases where scientists do not share their intermediate results, and with good reason. From a normative perspective, this could be viewed as an argument against giving extra credit to the

[18]Boyer (2014, theorem 3) suggests (for the case where $n = 2$ and $k = 2$) that if the second stage is worth up to twice as much credit as the first, there may still be an incentive to share. This would indicate some fairly significant robustness of the result. However, my own investigations suggest that this is an artifact of Boyer's assumption that the scientists have equal productivity: the larger the differences in productivity among the scientists the less robust the incentive to share.

scientist who finishes the last stage (because the more equal the division of credit, the more incentive scientists have to share).

Another feature of my explanation is that it explains sharing behavior only for "intermediate results", i.e., results that are significant enough to be publishable in their own right. Strevens points out that on this view, "nothing will be shared until something relevant is ready for publication, and worse, it is only what characteristically goes into the journals that gets broadcast, so details of experimental or computational methods and raw data will remain hidden" (Strevens forthcoming, p. 5). This constitutes an objection to my explanation, as according to Strevens the communist norm requires that any and all results should be shared, regardless of their credit-worthiness.

To this worry I reply that it is not clear that the communist norm makes such strong requirements. When the material under consideration is too little or too detailed to be considered publishable, scientists' actual compliance with a putative norm of sharing drops off steeply (Louis et al. 2002, Tenopir et al. 2011).[19] If Strevens' aim is to explain a norm of sharing for these cases, he may be trying to explain something that does not exist.

Strevens may reply to this that regardless of the content of the norm currently in place, it would be good to have a maximally inclusive communist norm. After all, scientists would benefit most from each other's work (thus speeding up the overall progress of science) if they shared results even before they had achieved publishable size and without hiding crucial details. By using the framework of a social contract to point out the benefits of more widespread sharing, Strevens could argue, it might be possible to help the scientific community get to such an improved norm.

That would be both a fair point and a laudable goal. However, the re-

[19]If it is assumed that material that cannot be published in a journal is worth zero credit when shared, then my model would of course predict that nothing would be shared. This prediction is not borne out empirically: while there is much less sharing of this kind of material, there is still some sharing. Perhaps this behavior is simply unexplainable from a pure credit-maximizing perspective. However, the assumption that this material is worth zero credit may not need to be granted. See Piwowar (2013) and the discussion below.

sults from my model can do the same. They suggest a clear way to make it incentive-compatible for scientists to share work below publishable size: allow smaller publications. And sharing crucial details can similarly be made incentive-compatible just by giving credit for it (Tenopir et al. 2011, Goring et al. 2014). If getting scientists to share these minor results or crucial details is a goal that scientists and policy makers consider important, the model gives clear directions on how to get there (but it may not be possible or desirable to do this, cf. Boyer 2014, section 4.4). Modern information technology readily suggests ways in which this can be done without overburdening existing scientific journals. Developments in this area are already underway (Piwowar 2013). In this sense, the results from this paper are more actionable than Strevens'.

8 Explaining the Origins of the Communist Norm

Above I argued that the results from the game-theoretic model explain the persistence of the communist norm. It could be argued that they also explain the *origins* of the norm: the uniqueness clauses in theorems 1 and 2 guarantee that behavior in accordance with the communist norm is the only pattern that rational credit-maximizing scientists could settle on.

But such an argument would make stringent demands on the scientists' rationality which real scientists are unlikely to satisfy. This section investigates the question whether less than perfectly rational scientists would also learn to share their intermediate results, thus giving a more robust account of the origins of the communist norm.

To answer this question I consider a boundedly rational learning rule that makes only minimal assumptions on the cognitive abilities of the scientists. In particular, it requires only that the scientists know which strategies are available to them and that they can compare the credit earned on the previous round to that earned on the current round (where a "round" is one instance of the game of imperfect information; to evaluate this bounded rationality rule one needs to assume the game is played repeatedly).

The rule I consider is *probe and adjust*. A scientist using probe and adjust follows the following simple procedure: on each round, play the same strategy as the round before with probability $1 - \varepsilon$, or "probe" a new strategy with probability ε (with $0 < \varepsilon < 1$; ε is usually "small"). In case of a probe, she picks a new strategy uniformly at random from all possible strategies. After playing this strategy for one round, the probe is evaluated: if the payoff for the round in which she probed is higher than the payoff in the previous round, keep the probed strategy (at least until the next probe); if the payoff is lower, return to the old strategy; if payoffs are equal, return to the old strategy with probability q and retain the probe with probability $1 - q$ ($0 < q < 1$).

Note that this is not quite the same as asking whether the probed strategy is a better reply to the other scientists' strategy than the old strategy, as other scientists may have changed their strategy as well. In particular, if all scientists are using probe and adjust, simultaneous probes and probes on subsequent rounds prevent this rule from necessarily always picking the better reply.

Consider a population of $n \geq 2$ scientists using probe and adjust to determine their strategy in repeated plays of the game of imperfect information with the number of stages $k \geq 1$ fixed. Assume all scientists use the same values of ε and q (this assumption can be relaxed, see Huttegger et al. 2014, pp. 837–838). Then the following result can be proven (see appendix A).

Theorem 3. *For any probability $p < 1$, if the probe probability $\varepsilon > 0$ is small enough there exists a T such that on an arbitrary round t with $t > T$, all scientists play strategy E at every information set with probability at least p.*

If, on a given round, all scientists play strategy E at every information set, they may be said to have learned to share their intermediate results. The theorem says that the probability of this happening can be made arbitrarily high by choosing a small enough probe probability and a long enough waiting time. Moreover, the theorem says that once the scientists learn to share their intermediate results they continue to do so on most subsequent rounds. So even on this very cognitively simple learning rule both the origins and the persistence of the communist norm can be explained on the basis of credit incentives.

Having already shown the same to be the case for highly rational scientists in sections 5 and 6, I suggest that similar results should be expected for intermediate levels of rationality.[20] Conforming to the communist norm is then shown to be incentive-compatible for credit-maximizing scientists regardless of their level of rationality.

I have suggested that credit incentives are responsible for the origins of the communist norm. How historically plausible is this? It is not entirely clear how one should evaluate this question. But a necessary condition for my explanation to be the correct one is that credit for scientific work, and in particular credit awarded in accordance with the priority rule, predates the communist norm. The remainder of this section argues that this condition is satisfied.

As Merton (1957) points out, scientists' concern for priority goes back at least as far as Galileo. In 1610, he used an anagram to report seeing Saturn as a "triple star" (the first sighting of the rings of Saturn). The device of the anagram served "the double purpose of establishing priority of conception and of yet not putting rivals on to one's original ideas, until they had been further worked out" (Merton 1957, p. 654). If Galileo was concerned about establishing priority for his ideas, it seems that the priority rule must already have been in effect in 1610. Priority disputes also go back at least as far, as Galileo wrote multiple polemics to defend his priority on various discoveries (Galilei 1607, 1623).

The communist norm, on the other hand, was not established as a norm of science until the 1660s, in the controversy between Boyle and Hobbes over

[20]Because the equilibrium in the game of imperfect information is both strict and unique, various other learning rules and evolutionary dynamics can easily be shown to converge to it. Examples include fictitious play, the best-response dynamics, and the replicator dynamics.

the air-pump (Shapin and Schaffer 1985). Part of what was at stake in this controversy were the norms for establishing a "matter of fact", i.e., a scientific fact. Boyle (who ended up "winning" the controversy) argued that

> An experience, even of a rigidly controlled experimental performance, that one man alone witnessed was not adequate to make a matter of fact. If that experience could be extended to many, and in principle to all men, then the result could be constituted as a matter of fact. In this way, the matter of fact is to be seen as both an epistemological and a social category. The foundational item of experimental knowledge, and of what counted as properly grounded knowledge generally, was an artifact of communication and whatever social forms were deemed necessary to sustain and enhance communication. (Shapin and Schaffer 1985, p. 25)

Scientific facts are attributed to the community rather than the individual, echoing Merton's definition of the communist norm, and this leads directly to a call for enhanced communication, i.e., sharing. If I am right that here the communist norm is being first established, the necessary condition that the priority rule predates the communist norm is satisfied.

9 Conclusion

In the introduction I argued that the sharing of scientific results (mandated by the communist norm) is important to the success of science and indeed to the existence of science as we know it. Theorems 1, 2, and 3 show that the priority rule gives scientists an incentive to share any and all intermediate results. These results can be used to explain both the origins and the persistence of the communist norm, answering the questions I raised in the introduction.

If my explanation is accepted, the crucial features of the social structure of science that maintain the communist norm are seen to be the fact that scientists respond to credit incentives, the priority rule, and the credit-worthiness of intermediate results. Tinkering with these features thus risks undercutting one of the most central aspects of science as a social enterprise.

By emphasizing credit incentives moderated by the priority rule, this paper falls in the tradition of Kitcher (1990), Dasgupta and David (1994), and Strevens (2003). Like those papers, I have picked one aspect of the social structure of science, and shown how the priority rule has the power to shape that aspect to science's benefit.

I take my results to show that no special explanation (using, e.g., normative expectations and/or a social contract) is required for the communist norm, *contra* Strevens (forthcoming). However, this only applies to whatever is publishable (or otherwise credit-worthy) in a given scientific community. Sharing scientific work that is too insignificant to be published is not incentivized in the same way. But insofar as this is a problem it suggests its own solution: give credit in accordance with the priority rule for whatever one would like to see shared, and scientists will indeed start sharing it.

A A Unique Nash Equilibrium

Let $n \geq 2$ be the number of scientists and $k \geq 1$ the number of stages. Let $G^p_{n,k}$ denote the game with perfect information described in section 5 and let $G^m_{n,k}$ denote the game with imperfect information described in section 6.

As is commonly done in game theory, I use $u_i(s_i, s_{-i})$ to denote the payoff (expected units of credit at the end of the game) to scientist i if she plays strategy s_i and s_{-i} gives the strategies of all scientists other than scientist i (call this an "incomplete strategy profile").

One strategy is of particular interest. Let s_i^E denote the strategy for scientist i in which she plays E (that is, shares and claims credit for her most recently completed stage) at every decision node in $G^p_{n,k}$ or at every information set in $G^m_{n,k}$ (so technically, s_i^E denotes two strategies, one for each game, but they share a lot of features which I use below). Let s_{-i}^E denote the incomplete strategy profile (in either game) where every scientist i' other than scientist i plays strategy $s_{i'}^E$. Let S^E denote the strategy profile (in either game) in which every scientist i plays strategy s_i^E.

Lemma 4. *In both $G_{n,k}^p$ and $G_{n,k}^m$, for any scientist i, the payoff when every scientist always shares any stages she completes immediately is*

$$u_i\left(s_i^E, s_{-i}^E\right) = k\frac{\lambda_i}{\lambda}.$$

Proof. Scientist i is the first to complete stage 1 with probability λ_i/λ. If she does she immediately claims one unit of credit. If any other scientist completes stage 1 before scientist i, that scientist immediately claims one unit of credit. Thus scientist i's expected credit from the first stage is λ_i/λ. Then all scientists simultaneously start working on the next stage. So by the same reasoning, scientist i's expected credit from any given stage is λ_i/λ. The result follows. □

Lemma 5. *In both $G_{n,k}^p$ and $G_{n,k}^m$, if scientist i plays strategy s_i^E but not every other scientist always shares any stages she completes, the payoff to scientist i is strictly higher than the payoff given in lemma 4. More precisely: let s_{-i} denote any incomplete strategy profile such that at least one scientist i' plays some strategy other than $s_{i'}^E$ (this can be either a different pure strategy, or any mixed strategy which plays strategy $s_{i'}^E$ with probability less than one). In the case of $G_{n,k}^p$, add the further assumption that this involves a deviation on the equilibrium path, i.e., there is at least one scientist i' who plays a strategy $s_{i'}$ (or a mixed strategy in which $s_{i'}$ is played with positive probability) such that if every other scientist i'' plays strategy $s_{i''}^E$ then there is a positive probability of reaching a decision node at which strategy $s_{i'}$ plays strategy H. Then*

$$u_i\left(s_i^E, s_{-i}\right) > k\frac{\lambda_i}{\lambda}.$$

Proof. Note that in the case described by lemma 4, i.e., when the strategy profile S^E is being played, the outcome of a single instance of the game can be described by a sequence $(i_1, i_2, \ldots, i_k)$, where the first member denotes the first scientist who completes a stage, the second member the second scientist to complete a stage (not necessarily a different scientist than the first), and so on. Because every scientist i plays strategy s_i^E, each member of the sequence also denotes the claiming of one unit of credit by that scientist. The probability of such a sequence describing the outcome of the game is

$$\frac{\lambda_{i_1}}{\lambda} \cdot \frac{\lambda_{i_2}}{\lambda} \cdots \frac{\lambda_{i_k}}{\lambda}.$$

Now suppose that there is at least one scientist i' playing a strategy different from $s_{i'}^E$. Let $s_{i'} \neq s_{i'}^E$ be some strategy that scientist i' plays with some positive probability p (where $p = 1$ if scientist i' plays a pure strategy), and assume that $s_{i'}$ involves a deviation on the equilibrium path in the case of $G_{n,k}^p$.

A sequence like $(i_1, i_2, \ldots, i_k)$ can still be used to describe the first k scientists to complete a stage, but because not everyone always claims credit, this may not completely describe the outcome of the game: if a scientist completed a stage but did not claim credit for it either immediately or later, it is possible that not all k units of credit have been claimed after k scientists have completed a stage.

However, regardless of whether credit is being claimed, the probability of the sequence remains unchanged due to the memorylessness property of the exponential distribution. Moreover, because scientist i plays strategy s_i^E, she is still claiming a unit of credit whenever she occurs in the sequence. Thus, all possible sequences $(i_1, i_2, \ldots, i_k)$ still occur with the same probability, and scientist i claims the same amount of credit in them. So scientist i now expects to accrue $k\lambda_i/\lambda$ units of credit during the time it takes for k scientists to complete a stage.

But, by assumption, there is at least one sequence $(i_1, i_2, \ldots, i_k)$ in which i' occurs and (with probability p) plays strategy H at the corresponding decision node or information set, and i' does not occur in the remainder of that sequence. As a result, at the end of that sequence at most $k - 1$ units of credit have been claimed. In the remainder of that game, there is a positive probability (at least λ_i/λ, the probability that she is the very next one to complete a stage) that scientist i gains more credit, credit that she would not have obtained if scientist i' had played strategy $s_{i'}^E$. Since $p > 0$ and $\lambda_{i''} > 0$ for all i'', it follows that

$$u_i\left(s_i^E, s_{-i}\right) \geq k\frac{\lambda_i}{\lambda} + \frac{\lambda_{i_1}}{\lambda} \cdot \frac{\lambda_{i_2}}{\lambda} \cdots \frac{\lambda_{i_k}}{\lambda} \cdot p \cdot \frac{\lambda_i}{\lambda} > k\frac{\lambda_i}{\lambda}. \qquad \square$$

Theorem 6. *Let S be any strategy profile for $G_{n,k}^m$ other than S^E, or let S be any strategy profile for $G_{n,k}^p$ that involves deviations on the equilibrium path relative to S^E. Then there exists at least one scientist i playing strategy $s_i \neq s_i^E$ such that she would be strictly better off playing strategy s_i^E:*

$$u_i\left(s_i^E, s_{-i}\right) > u_i\left(s_i, s_{-i}\right).$$

Proof. Note that the game is zero-sum: regardless of strategies, there are k units of credit to be divided, and so if one scientist increases her payoff, that of another must decrease. This fact, combined with lemmas 4 and 5, yields the theorem. Distinguish three cases:

1. There is only one scientist i playing a (pure or mixed) strategy $s_i \neq s_i^E$. Then every scientist i' other than scientist i is playing strategy $s_{i'}^E$ and so by by lemma 5 is getting a payoff greater than $k\lambda_{i'}/\lambda$. Because the game is zero-sum, it follows that $u_i(s_i, s_{-i}) < k\lambda_i/\lambda$. By lemma 4, $u_i(s_i^E, s_{-i}) = k\lambda_i/\lambda$, and the result follows.

2. There is at least one scientist i' playing strategy $s_{i'}^E$ and at least two scientists playing some other strategy. Then any scientist i' who is playing strategy $s_{i'}^E$ is getting a payoff greater than $k\lambda_{i'}/\lambda$ by lemma 5. Because the game is zero-sum, at least one of the remaining scientists, say scientist i, must be getting a payoff less than $k\lambda_i/\lambda$. But if scientist i changed her strategy to s_i^E, by lemma 5 she would get a payoff greater than $k\lambda_i/\lambda$. So $u_i(s_i^E, s_{-i}) > k\lambda_i/\lambda > u_i(s_i, s_{-i})$.

3. Every scientist i' is playing some strategy $s_{i'} \neq s_{i'}^E$. Because the game is zero-sum, it is impossible for every scientist i' to be getting a greater payoff than $k\lambda_{i'}/\lambda$. So there is at least one scientist, say scientist i, such that $u_i(s_i, s_{-i}) \leq k\lambda_i/\lambda$. By lemma 5, $u_i(s_i^E, s_{-i}) > k\lambda_i/\lambda$, and the result follows. □

Theorem 6 plays an important role in the proofs of the results in the main text.

Proof of theorem 1. Consider the game $G_{n,k}^p$. In any profile (of pure or mixed strategies) at least one scientist has an incentive to change her strategy, unless every scientist i plays strategy s_i^E or a strategy that deviates from s_i^E only off the equilibrium path. Thus no profile is a Nash equilibrium unless every scientist i plays strategy s_i^E or a strategy that deviates from s_i^E only off the equilibrium path. But since the backwards induction solution is a Nash equilibrium, it follows that in the backwards induction solution (which is guaranteed to exist for any finite game of perfect information) every scientist i must play strategy s_i^E or a strategy that deviates from s_i^E only off the equilibrium path. So in the backwards induction solution every scientist immediately shares and claims credit for any stage she completes.

A direct proof using backwards induction is also possible. This proof yields the slightly stronger result that in the backwards induction solution every scientist plays strategy E at every decision node (including those off the equilibrium path) and is available from the author upon request. □

Proof of theorem 2. Let S be any profile (of pure or mixed strategies) for the game $G_{n,k}^m$. If $S \neq S^E$, then at least one scientist has an incentive to change her strategy, and so S is not a Nash equilibrium. Thus there is at most one Nash equilibrium: S^E.

That S^E is indeed a Nash equilibrium, and in fact a strict Nash equilibrium, also follows from theorem 6 by considering the special case where $s_{-i} = s_{-i}^E$. This shows that a scientist i who deviates unilaterally makes herself strictly worse off. □

To prove theorem 3, some terminology and a result from Huttegger et al. (2014) are needed. Define a *weakly better reply path* to be a sequence of profiles $(S^1, \ldots, S^\ell)$ such that for any $j < \ell$, profile S^j differs from profile S^{j+1} only in one scientist's strategy, say scientist i (so $s^j_{-i} = s^{j+1}_{-i}$), and $u_i(S^{j+1}) \geq u_i(S^j)$, i.e., scientist i changes to a strategy that is a (weakly) better reply to the other scientists' strategies. Define a *weakly better reply game* to be a game in which for every profile S there exists a weakly better reply path from S to a strict Nash equilibrium.

Let G be a weakly better reply game with n scientists. Assume the scientists play G repeatedly, adjusting their strategy using probe and adjust and using the same values of ε and q. Let S^t be the profile of strategies played on round t.

Theorem 7 (Huttegger et al. (2014))**.** *For any probability $p < 1$, if the probe rate $\varepsilon > 0$ is sufficiently small, then the profile S^t is a strict Nash equilibrium of G for all sufficiently large t with probability at least p.*

Theorem 3 is a corollary of theorems 2, 6 and 7.

Proof of theorem 3. By theorem 2, the strategy profile in which every scientist plays strategy E at every information set is the only strict Nash equilibrium of the game. If $G^m_{n,k}$ is a weakly better reply game, the desired result follows from theorem 7.

That the game is a weakly better reply game follows straightforwardly from theorem 6. At any strategy profile, for at least one scientist i whose strategy differs from s^E_i switching to strategy s^E_i is a better reply for her. This switch leads to a profile which is either the strict Nash equilibrium or in which the same is true for some other scientist. The result is a path of length at most n from any profile to the strict Nash equilibrium, in which at each step along the path one scientist i switches her strategy to s^E_i, and improves her payoff by doing so. □

References

Peter Arzberger, Peter Schroeder, Anne Beaulieu, Geof Bowker, Kathleen Casey, Leif Laaksonen, David Moorman, Paul Uhlir, and Paul Wouters. Promoting access to public research data for scientific, economic, and social development. *Data Science Journal*, 3:135–152, 2004. doi: 10.2481/dsj.3.135. URL `http://dx.doi.org/10.2481/dsj.3.135`.

Alain Aspect, Jean Dalibard, and Gérard Roger. Experimental test of Bell's inequalities using time-varying analyzers. *Physical Review Letters*, 49:1804–1807, Dec 1982. doi: 10.1103/PhysRevLett.49.1804. URL `http://link.aps.org/doi/10.1103/PhysRevLett.49.1804`.

Robert J. Aumann. Backward induction and common knowledge of rationality. *Games and Economic Behavior*, 8(1):6–19, 1995. ISSN 0899-8256. doi: 10.1016/S0899-8256(05)80015-6. URL `http://www.sciencedirect.com/science/article/pii/S0899825605800156`.

Cristina Bicchieri. Methodological rules as conventions. *Philosophy of the Social Sciences*, 18(4):477–495, 1988. ISSN 0048-3931. doi: 10.1177/004839318801800403. URL `http://pos.sagepub.com/content/18/4/477.short`.

Cristina Bicchieri. *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press, Cambridge, 2006. ISBN 9780521574907.

Christine L. Borgman. The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 63(6):1059–1078, 2012. ISSN 1532-2890. doi: 10.1002/asi.22634. URL `http://dx.doi.org/10.1002/asi.22634`.

Thomas Boyer. Is a bird in the hand worth two in the bush? Or, whether scientists should publish intermediate results. *Synthese*, 191(1):

17–35, 2014. ISSN 0039-7857. doi: 10.1007/s11229-012-0242-4. URL http://dx.doi.org/10.1007/s11229-012-0242-4.

Thomas Boyer and Cyrille Thomas Imbert. Scientific collaboration: Do two heads need to be more than twice better than one? *Philosophy of Science*, forthcoming. ISSN 00318248.

Partha Dasgupta and Paul A. David. Toward a new economics of science. *Research Policy*, 23(5):487–521, 1994. ISSN 0048-7333. doi: 10.1016/0048-7333(94)01002-1. URL http://www.sciencedirect.com/science/article/pii/0048733394010021.

Galileo Galilei. *Difesa Contro alle Calunnie et Imposture di Baldessar Capra*. Baglioni, Venice, 1607.

Galileo Galilei. *Il Saggiatore*. Mascardi, Rome, 1623.

Simon J. Goring, Kathleen C. Weathers, Walter K. Dodds, Patricia A. Soranno, Lynn C. Sweet, Kendra S. Cheruvelil, John S. Kominoski, Janine Rüegg, Alexandra M. Thorn, and Ryan M. Utz. Improving the culture of interdisciplinary collaboration in ecology by expanding measures of success. *Frontiers in Ecology and the Environment*, 12(1):39–47, Feb 2014. ISSN 1540-9295. doi: 10.1890/120370. URL http://dx.doi.org/10.1890/120370.

John C. Huber. Invention and inventivity as a special kind of creativity, with implications for general creativity. *The Journal of Creative Behavior*, 32 (1):58–72, 1998a. ISSN 2162-6057. doi: 10.1002/j.2162-6057.1998.tb00806.x. URL http://dx.doi.org/10.1002/j.2162-6057.1998.tb00806.x.

John C. Huber. Invention and inventivity is a random, Poisson process: A potential guide to analysis of general creativity. *Creativity Research Journal*, 11(3):231–241, 1998b. doi: 10.1207/s15326934crj1103_3. URL http://dx.doi.org/10.1207/s15326934crj1103_3.

John C. Huber. A new method for analyzing scientific productivity. *Journal of the American Society for Information Science and Technology*, 52(13): 1089–1099, 2001. ISSN 1532-2890. doi: 10.1002/asi.1173. URL http://dx.doi.org/10.1002/asi.1173.

John C. Huber and Roland Wagner-Döbler. Scientific production: A statistical analysis of authors in mathematical logic. *Scientometrics*, 50(2): 323–337, 2001a. ISSN 0138-9130. doi: 10.1023/A:1010581925357. URL http://dx.doi.org/10.1023/A%3A1010581925357.

John C. Huber and Roland Wagner-Döbler. Scientific production: A statistical analysis of authors in physics, 1800-1900. *Scientometrics*, 50(3): 437–453, 2001b. ISSN 0138-9130. doi: 10.1023/A:1010558714879. URL http://dx.doi.org/10.1023/A%3A1010558714879.

David L. Hull. *Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science*. University of Chicago Press, Chicago, 1988. ISBN 0226360504.

Simon M. Huttegger, Brian Skyrms, and Kevin J. S. Zollman. Probe and adjust in information transfer games. *Erkenntnis*, 79(4):835–853, 2014. ISSN 0165-0106. doi: 10.1007/s10670-013-9467-y. URL http://dx.doi.org/10.1007/s10670-013-9467-y.

Philip Kitcher. The division of cognitive labor. *The Journal of Philosophy*, 87(1):5–22, 1990. ISSN 0022362X. URL http://www.jstor.org/stable/2026796.

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

Karen Seashore Louis, Lisa M. Jones, and Eric G. Campbell. Macroscope: Sharing in science. *American Scientist*, 90(4):304–307, 2002. ISSN 00030996. URL http://www.jstor.org/stable/27857685.

Bruce Macfarlane and Ming Cheng. Communism, universalism and disinterestedness: Re-examining contemporary support among academics for Merton's scientific norms. *Journal of Academic Ethics*, 6(1):67–78, 2008. ISSN 1570-1727. doi: 10.1007/s10805-008-9055-y. URL `http://dx.doi.org/10.1007/s10805-008-9055-y`.

Robert K. Merton. A note on science and democracy. *Journal of Legal and Political Sociology*, 1(1–2):115–126, 1942. Reprinted in Merton (1973, chapter 13).

Robert K. Merton. Priorities in scientific discovery: A chapter in the sociology of science. *American Sociological Review*, 22(6):635–659, 1957. ISSN 00031224. URL `http://www.jstor.org/stable/2089193`. Reprinted in Merton (1973, chapter 14).

Robert K. Merton. Singletons and multiples in scientific discovery: A chapter in the sociology of science. *Proceedings of the American Philosophical Society*, 105(5):470–486, 1961. ISSN 0003049X. URL `http://www.jstor.org/stable/985546`. Reprinted in Merton (1973, chapter 16).

Robert K. Merton. Behavior patterns of scientists. *The American Scholar*, 38(2):197–225, 1969. ISSN 00030937. URL `http://www.jstor.org/stable/41209646`. Reprinted in Merton (1973, chapter 15).

Robert K. Merton. *The Sociology of Science: Theoretical and Empirical Investigations*. The University of Chicago Press, Chicago, 1973. ISBN 0226520919.

James R. Norris. *Markov Chains*. Cambridge University Press, Cambridge, 1998. URL `http://dx.doi.org/10.1017/CBO9780511810633`.

Heather Piwowar. Altmetrics: Value all research products. *Nature*, 493(7431):159, Jan 2013. ISSN 1476-4687. doi: 10.1038/493159a. URL `http://dx.doi.org/10.1038/493159a`.

David B. Resnik. Openness versus secrecy in scientific research. *Episteme*, 2: 135–147, Oct 2006. ISSN 1750-0117. doi: 10.3366/epi.2005.2.3.135. URL http://journals.cambridge.org/article_S174236000000037X.

Steven Shapin and Simon Schaffer. *Leviathan and the Air-Pump: Hobbes, Boyle and the Experimental Life*. Princeton University Press, Princeton, 1985. ISBN 9780691024325.

Patricia A. Soranno, Kendra S. Cheruvelil, Kevin C. Elliott, and Georgina M. Montgomery. It's good to share: Why environmental scientists' ethics are out of date. *BioScience*, 65(1):69–73, 2015. doi: 10.1093/biosci/biu169. URL http://bioscience.oxfordjournals.org/content/65/1/69.abstract.

Michael Strevens. The role of the priority rule in science. *The Journal of Philosophy*, 100(2):55–79, 2003. ISSN 0022362X. URL http://www.jstor.org/stable/3655792.

Michael Strevens. Scientific sharing: Communism and the social contract. In Thomas Boyer-Kassem, Conor Mayo-Wilson, and Michael Weisberg, editors, *Scientific Collaboration and Collective Knowledge*. Oxford University Press, Oxford, forthcoming. URL http://www.strevens.org/research/scistruc/communicans.shtml.

Carol Tenopir, Suzie Allard, Kimberly Douglass, Arsev Umur Aydinoglu, Lei Wu, Eleanor Read, Maribeth Manoff, and Mike Frame. Data sharing by scientists: Practices and perceptions. *PLoS ONE*, 6(6):e21101, Jun 2011. doi: 10.1371/journal.pone.0021101. URL http://dx.doi.org/10.1371/journal.pone.0021101.