Communism and the Incentive to Share in Science*

Remco Heesen†

October 18, 2016

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 in many cases sufficient to explain both the origins and the persistence of the communist

*Thanks to Kevin Zollman, Michael Strevens, Stephan Hartmann, Teddy Seidenfeld, Thomas Boyer-Kassem, Lee Elkin, Liam Bright, Jan Sprenger, Arif Ahmed, two anonymous referees, and audiences at the Bristol-Groningen Conference in Formal Epistemology, the Logic Colloquium in Helsinki, and EPSA15 in Düsseldorf for valuable comments and discussion. This work was partially supported by the National Science Foundation under grant SES 1254291 and by an Early Career Fellowship from the Leverhulme Trust.

†Faculty of Philosophy, University of Cambridge, Sidgwick Avenue, Cambridge CB3 9DA, UK. Email: rdh51@cam.ac.uk.

norm, adding to previous results emphasizing the benefits of the incentive structure created by the priority rule.

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, 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?

[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.

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 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.

Strevens (forthcoming) gives what he calls a "Hobbesian vindication" of the communist norm by showing that scientists should be willing to sign a contract that enforces 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 in many circumstances *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 how individual scientists' "selfish" desire to receive credit for their work furthers 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 of "rational" in the above 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.

Section 3 sets out my own position by explaining how the idea that scientists 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.[2]

I then show that rational credit-maximizing scientists should indeed be expected to share under a range of conditions (section 5). Section 6 discusses the assumptions and limitations of this formal result. 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. More specifically, would it be in scientists' interest to share their work even in the absence of a norm telling them to do so? To clarify the question, 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;
>
> *Conditional preference*: i prefers to conform to R in situations of type S on the condition that:

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

> (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 most other people shared. In contrast, if an individual knows that in a particular country most people drive on the right side of the road, she would probably prefer to do the same even if others had no expectations about her behavior.

The language of game theory is useful to sharpen these ideas. Recall that conforming to a behavioral rule R constitutes a *(Nash) equilibrium* if no individual has an incentive to deviate unilaterally, i.e., everyone prefers to conform given that everyone else does.

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 according to this 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 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, an account of scientists' interests 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.

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 a 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. Credit maximization is an interest that all scientists share. This makes it a particularly powerful tool to explain scientists' behavior.

The institutions of science put a premium on originality. Credit is awarded to the first scientist to publish some particular result or discovery. 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, Dasgupta and David 1994, 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)

Thus, at the crucial stage at which scientific progress 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. Putting this all together, Strevens has effectively claimed 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. Claims like this are also made by Dasgupta and David (1994, p. 500), 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 often in a scientist's own interest to share her work widely. In other words, in many realistic situations 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 least not primarily). It merely describes a rule of behavior that it is often 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. Boyer (2014, p. 18 and p. 21) gives some 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.

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 the scientists involved were acting in their own best interest. While credit can be claimed when a smaller discovery is published, the advantage that the smaller discovery gives on the way toward the major discovery is thereby lost. 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. In his model the benefits of sharing *intermediate results* outweigh the costs, with costs and benefits both measured in credit assigned via the priority rule. Although Boyer does not specifically discuss the communist norm, his result could be used to argue that normative expectations are not necessary to explain it: the priority rule encourages wide sharing of scientific work even before the potential of future discoveries based on this work has been exhausted.

The key claim is that, *contra* Strevens (forthcoming), a social contract may not be needed to enforce sharing. The reason for this is the possibility to claim credit for intermediate results.

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 many if not most research situations are such that there is a credit incentive to publish intermediate results, which requires a more general model. I relax Boyer's assumptions that there are only two scientists, that the scientists are equally productive, that different intermediate results are equally hard to achieve, and that scientists share either all or no intermediate reuslts.

The second worry questions the relevance of equilibria. This worry has two sides. One side claims that showing that sharing 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). The other side claims that showing that sharing is an equilibrium is not necessary; observed behavioral patterns need not be the equilibrium of some underlying game. I alleviate both of these worries by showing that sharing is an equilibrium that one should expect to be realized by both fully rational and boundedly rational scientists. Thus, 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.[3] 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.[4] The number of intermediate stages of the project is denoted k.

Competition plays a central role in the model. 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 "scientist" may refer to someone working in the natural sciences, the social sciences, the humanities, or any other 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.[5] 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. The amount of credit is given by the parameter $c_j > 0$ for each stage j, with $C = \sum_{j=1}^{k} c_j$ denoting the total credit available. Publishing also benefits the scientific community: other scientists

[3]Although it was developed independently, the model turns out to be essentially identical to the model studied by Banerjee et al. (2014). In section 6 I discuss their results, which are roughly speaking weaker results in a more general model. Banerjee et al. do not, however, give a detailed defense of the assumptions, or the application to explaining the communist norm.

[4]This ordering assumption may seem restrictive and unrealistic, but in section 6 I give reasons to think relaxing it will not affect my results.

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

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.

If the scientist keeps her result 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. 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 she always publishes, claiming credit for all unpublished stages and completing the research project.

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."[6]

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 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

[6]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. 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.

only candidate that has significant empirical support behind it (Huber 2001, more on this below).

To be precise, the time scientist i takes to complete stage j follows an exponential distribution with parameter $\lambda_{ij} > 0$. The parameter can be interpreted as the speed at which the scientist works. In particular, $1/\lambda_{ij}$ is the expected time scientist i needs to complete stage j. The speed parameter may vary by scientist and by stage, allowing for differences in difficulty between stages and differences in talent, skill, resources, or specialization between scientists.

The assumption that waiting times are exponential is equivalent to the assumption that scientists' productivity is a (nonstationary) Poisson process. Empirical work has shown that scientists' productivity fits a Poisson distribution quite well. 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 psychology.

Under this assumption the probability that it will take scientist i more than t time units to complete stage j is $\exp\{-t\lambda_{ij}\}$.[7] 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

[7] 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. To address this, suppose each time unit is divided into x equal parts, and in each part the scientist completes the stage with probability λ/x. The probability that the scientist has not completed the stage at time t (where t is measured in the original time units) is $(1-\lambda/x)^{tx}$. A continuous-time model is obtained by taking the limit as x goes to infinity. Then 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 with continuous time.

remaining waiting time is equal to the original distribution of the waiting time. Second, if scientist i is working on stage j_i then the waiting time until one of the scientists finishes the stage she is working on is exponentially distributed with parameter

$$\sigma_{j_1 j_2 \cdots j_n} = \sum_{i=1}^{n} \lambda_{i j_i}.$$

In the special case where all scientists are working on the same stage I will write $\sigma_j = \sum_{i=1}^{n} \lambda_{ij}$. Third, the probability that scientist i is the first one to finish the stage she is working on is $\lambda_{ij_i}/\sigma_{j_1 j_2 \cdots j_n}$.

In general, whether there is an incentive to share intermediate results in this model depends on the amount of credit given for each stage and the speed with which the different scientists can solve them. The results presented in the next section require the following assumption.

Assumption 4.1 (Proportional Credit)**.** *The speed parameters and the credit rewards stand in the following relation: for every scientist i and for each pair of stages $j < j'$,*

$$c_j \lambda_{ij} \geq c_{j'} \lambda_{ij'}.$$

This assumption states that the credit given for each stage is either proportional to its difficulty, or earlier stages are awarded more credit than later ones (relative to their difficulty). I discuss the import of this assumption, as well as other limitations of the model, in section 6.

5 The Incentive to Share in the Model

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 5.1.

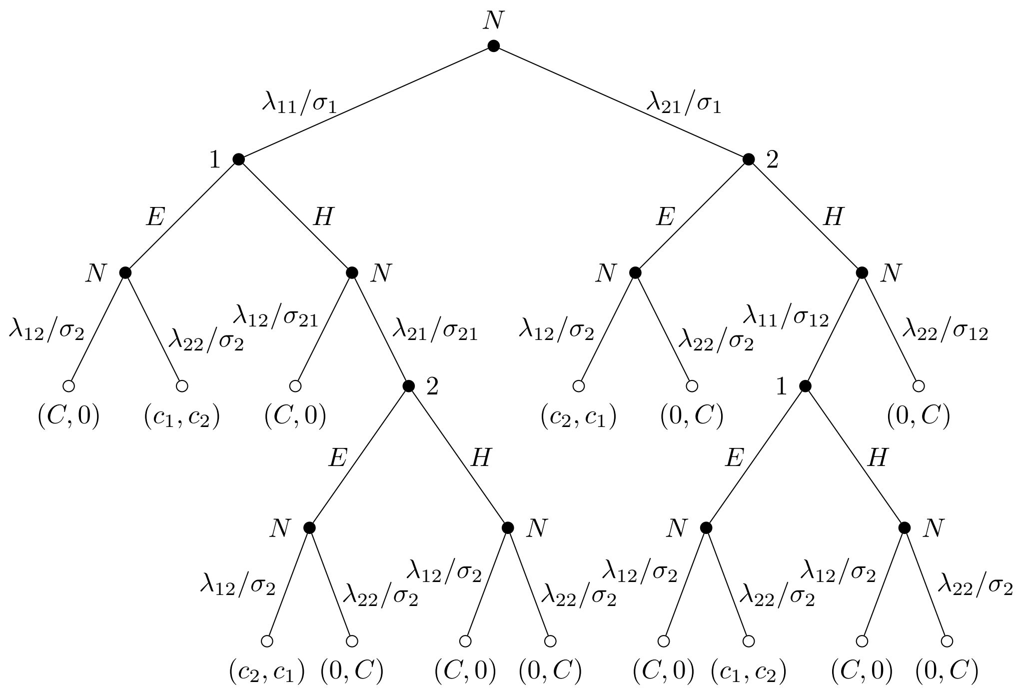

Figure 5.1: Extensive form of the game of perfect information with $n = 2$ and $k = 2$.

At the root node Nature decides which of the two scientists is the first one to complete the first stage of the project with the indicated probabilities. This leads to one of two decision nodes marked with a number indicating which scientist makes a decision at this node.

The scientist can choose one of two strategies (E or H), then Nature decides who is the next scientist to complete the stage she is working on, and so on until one of the scientists completes the second stage. At this point the game ends, with payoff pairs indicating credit awarded to each scientist.

It is implicitly assumed in figure 5.1 that each scientist knows when another scientist completes a stage, even when she keeps the result secret. Is it realistic to assume that scientists have this kind of information? It depends. 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 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.

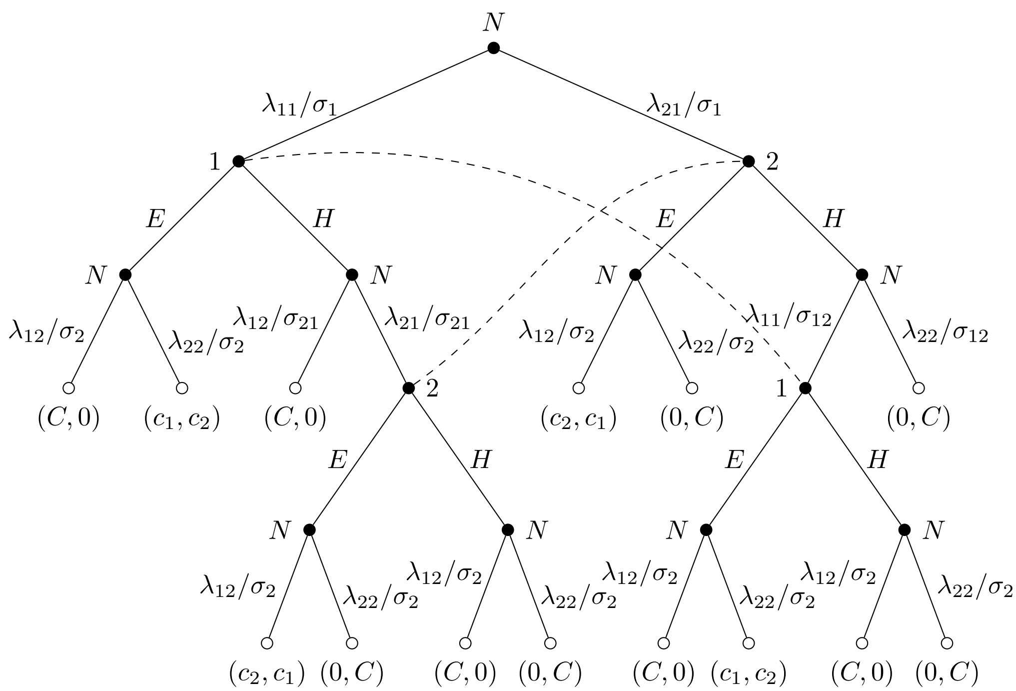

Figure 5.2: Extensive form of the game of imperfect information with $n = 2$ and $k = 2$.

But in other fields this kind of information might not be available. If this assumption is dropped scientists are unable to distinguish between certain decision nodes, indicated by so-called *information sets* (see figure 5.2). This yields a *game of imperfect information*. In contrast, the version of the game in which scientists can make these distinctions (as in figure 5.1) is a *game of perfect information*. I analyze both versions of the game.

Recall that I am interested in finding equilibria of these games. One way to find an equilibrium in a game of perfect information is by *backwards induction*. This involves identifying what a rational scientist will do at a terminal decision node, and then going backwards through the tree, identifying rational actions for the scientists by assuming other scientists will play rationally downstream.

In figure 5.1 it is rational for the scientists at the two lower decision nodes to play strategy E: this yields either the same payoff or a higher payoff than playing strategy H. Assuming that the scientists play E at the lower nodes, and assuming Proportional Credit, it is also rational for the scientists at the two higher nodes to play strategy E. Thus, under Proportional Credit the backwards induction solution of this game is for both scientists to play E at both of their decision nodes.

The following theorem shows that this backwards induction analysis also goes through when there are more than two scientists and/or more than two stages (for a proof, see appendix A). Moreover, any other equilibrium of the game is behaviorially indistinguishable from the backwards induction solution. 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.

Theorem 5.1. *Consider the game with perfect information with $n \geq 2$ scientists and $k \geq 1$ stages, and assume Proportional Credit.*

(a) This game has a (unique) backwards induction solution in which all scientists play strategy E at every decision node.

(b) *There are no equilibria (in pure or mixed strategies) that are behaviorally distinct from the backwards induction solution.*

An equilibrium analysis thus yields a unique prediction for the game of perfect information. How about the game of imperfect information? Backwards induction does not apply to this type of game. But equilibria can still be identified by analyzing the normal form of the game. Table 5.1 gives the expected credit for each scientist in two examples, one in which the first scientist is thrice as fast as the second, and one in which the second stage can be completed thrice as quickly as the first stage. Note that because the scientists cannot distinguish between their two decision nodes, only two (pure) strategies are available to them.

	E	H
E	$(24, 8)$	$(26\frac{1}{4}, 5\frac{3}{4})$
H	$(23\frac{1}{4}, 8\frac{3}{4})$	$(27, 5)$

	E	H
E	$(16, 16)$	$(15, 17)$
H	$(17, 15)$	$(16, 16)$

Table 5.1: Normal form of the game of imperfect information with scientist 1's strategy as the rows and scientist 2's strategy as the columns. On the left, $\lambda_{11} = \lambda_{12} = 3$, $\lambda_{21} = \lambda_{22} = 1$, and $c_1 = c_2 = 16$. On the right, $\lambda_{11} = \lambda_{21} = 1$, $\lambda_{12} = \lambda_{22} = 3$, and $c_1 = c_2 = 16$.

Since the credit given for each stage is equal in both cases, the example on the left satisfies Proportional Credit while the example on the right does not. On the left, the only equilibrium is the one in which both scientists play strategy E, and this is a *strict equilibrium* (a scientist who deviates is strictly worse off). On the right, both scientists play strategy H in the unique and strict equilibrium.

The features of the example on the left generalize for different numbers of scientists and stages (see appendix A for a proof).

Theorem 5.2. *Consider the game with imperfect information with $n \geq 2$ scientists and $k \geq 1$ stages and assume Proportional Credit.*

(a) This game has an equilibrium in which all scientists play strategy E at every information set.[8]

(b) There are no other equilibria (in pure or mixed strategies).

(c) The equilibrium is strict.

What do theorems 5.1 and 5.2 say about what it is rational for a scientist to do when working on a research project where Proportional Credit is satisfied? They say 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.

6 Limitations of the Formal Results

I have shown using a formal model that scientists have a credit incentive to share their intermediate results in a range of circumstances. In order to show this I had to make certain assumptions. This section discusses how these assumptions limit the applicability of my theorems. Sections 7 and 8 discuss how the theorems may explain the communist norm.

A key assumption is Proportional Credit. It states that, relative to their difficulty, earlier stages must be rewarded with at least as much credit as later stages. As one of the examples in section 5 illustrated, if this assumption is violated there may not be an equilibrium in which scientists share.

Is Proportional Credit likely to be violated in practice? It may seem reasonable to reward scientists proportional to the difficulty of their contributions. On the other hand, Strevens (2003) and Heesen (2016) have argued

[8] This result is a corollary of Banerjee et al. (2014, theorem 2.1). See the discussion in section 6.

that scientific contributions should be rewarded based on their social value, which may not always correlate with difficulty. Additionally, in practice it may happen that the scientist who finishes the last stage ("puts it all together") gets a relatively large share of the credit.

From a descriptive perspective, these might be the kinds of cases where scientists do not share their intermediate results, and my model suggests why. From a normative perspective, perhaps the appropriate conclusion is that Proportional Credit should be enforced. If scientists are rewarded proportionally to difficulty, without extra credit for completing the last stage of a research project, then sharing is incentivized.

Enforcing Proportional Credit may be hard if the difficulty of scientific contributions is unclear. Banerjee et al. (2014) show that there may still be an equilibrium in which scientists share if Proportional Credit is enforced with some limited degree of error, depending on the scientists' speed parameters.[9] But Banerjee et al. show neither uniqueness[10] nor strictness of the equilibrium. Since my explanation of the communist norm relies on these features, it is not clear how helpful this limited degree of robustness of the equilibrium is.

A different limitation is the assumption that the stages can only be completed in one order. In reality, there might be different ways to complete a research project. This could be incorporated by modeling the stages as a directed graph with different "paths" to complete the project. Banerjee et al. (2014, theorem 3.3) show that under Proportional Credit sharing intermediate results is an equilibrium of this generalized version of the game. The same caveat applies—Banerjee et al. do not show uniqueness or strictness—but here I think there is some reason to believe that the equilibrium they

[9] More precisely, Banerjee et al. (2014, theorem 2.1) show that in the game of imperfect information there is an equilibrium in which every scientist plays strategy E at every information set if for every scientist i and for each pair of stages $j < j'$, $\frac{c_j(\sigma_j - \lambda_{ij})}{c_{j'}(\sigma_{j'} - \lambda_{ij'})} \geq \frac{\lambda_{ij'}}{\sigma_{j'}}$.

[10] Banerjee et al. only prove uniqueness for a case in which some of the scientists commit to sharing before the game starts (so-called Stackelberg agents). Theorem 5.2 above does not require this.

find is indeed unique and strict. This is because this adaptation of the model seems to only make sharing more attractive, because sharing is less likely to help a scientist's competitors (who may be on a different path).

A difference between my model and the one given by Strevens (forthcoming) is that my model is zero-sum. This is because I have implicitly assumed that the scientists eventually complete the research project.[11] Hence when the game ends a total of C units of credit have been divided among the scientists, and so any change in strategy that leads to some scientist improving her (expected) credit must lower another's.

In contrast, Strevens' model explicitly 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 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 improves 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.

On this point, Strevens's model is arguably more realistic, as research projects sometimes fail to reach their goal. It would be interesting to study a model which incorporates both a positive probability of failure and credit for intermediate results. Whether there would be an incentive to share intermediate results under conditions similar to those I have found here is a question I leave for future research.

There are other ways to change the model that would make it no longer zero-sum. For example, Boyer-Kassem and Imbert (2015, section 4) argue that one should consider credit per unit time (rather than "total credit" which I use). Then sharing benefits all scientists to some extent by decreasing the

[11]More precisely: the scientists complete all k stages in finite time with probability one. The models of Banerjee et al. (2014), Boyer (2014), and Boyer-Kassem and Imbert (2015) have the same feature.

expected completion time of the research project; Boyer-Kassem and Imbert call this a "speedup effect". For present purposes it makes no difference: my theorems still hold if credit is measured per unit time (see appendix A).

7 Explaining the Persistence of the Communist Norm

I take the results from section 5 to give an explanation for the *persistence* of the communist norm, i.e., the fact that scientists do in fact publish their intermediate results in a large range of cases. The explanation runs as follows.

Suppose scientists are generally sharing their intermediate results. If a given scientist withholds 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.

Note that I do not claim that real scientists are rational credit-maximizers. All that follows for real scientists is that they have a credit *incentive* to conform to the norm (even when they fail to act on it). This fact, combined with the fact that real scientists are at least somewhat sensitive to credit incentives, constitutes my explanation of the persistence of the norm.

In the remainder of this section I point out a number of peculiar features of my explanation and consider some objections based on those features.

My explanation relies on three basic principles: scientists' sensitivity to credit incentives, intermediate results being given sufficient credit as specified in Proportional Credit, 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 a social norm in Bicchieri's sense, as normative expectations have no role in the explanation. But the available evidence seems to refute this: scientists (normatively) expect other scientists to conform to the communist norm (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 actually benefit when a given scientist deviates from the norm, so from a credit-maximizing perspective they should be encouraging each other to deviate.

But 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. I claim that this wider community 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.[12]

This yields an empirical prediction that might be used to help decide between Strevens' explanation and mine. On Strevens' explanation withholding an intermediate result 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 withholding 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 when other scientists fail to share may thus help decide, in a given case, whether sharing is happening out of self-interested credit-maximization or as the result of a social contract.

The scope of my explanation is restricted to the sharing of "intermediate results", i.e., results that are significant enough to be publishable in their

[12] 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).

own right. Strevens points out a limitation of 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.

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). If Strevens’ aim is to explain a norm of sharing for these cases, he may be trying to explain something that does not exist.

This leaves the question of what to do if one wants to encourage sharing work below publishable size, especially in the kinds of cases where sharing is currently not standard practice. Strevens’ contribution is to show that scientists’ have a common interest in establishing a norm of sharing for such cases.

My contribution, in contrast, is in providing a suggestion for *how* such a norm could be established. 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: give credit for smaller publications and for sharing crucial details (Tenopir et al. 2011, Goring et al. 2014). Modern information technology readily suggests ways in which this can be done without overburdening existing scientific journals (Piwowar 2013).

8 Explaining the Origins of the Communist Norm

Above I argued that the results from section 5 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 5.1 and 5.2 guarantee that behavior in accordance with the communist norm is the only pattern that rational credit-maximizing scientists could settle on (in cases where their assumptions are satisfied).

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 under different strategies.

The rule I consider is *probe and adjust*. Suppose the game of imperfect information is played repeatedly. A scientist using probe and adjust follows a simple procedure: on each round (one instance of the game), 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, pick a new strategy uniformly at random from all possible strategies. After playing this strategy for one round, the probe is evaluated: if the payoff in the probing round is higher than the payoff in the previous round, keep the probed strategy; if the payoff is lower, return to the old strategy; if payoffs are equal, return to the old strategy with probability $q \in (0, 1)$ and retain the probe with probability $1 - q$.

Consider a population of $n \geq 2$ scientists using probe and adjust to determine their strategy in 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 8.1. *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. 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 cognitively simple learning rule both the origins and the persistence of the communist norm can be explained on the basis of credit incentives in a large range of cases.

Having already shown the same to be the case for highly rational scientists in section 5, I suggest that similar results should be expected for intermediate levels of rationality.[13] Conforming to the communist norm is then shown to be incentive-compatible for credit-maximizing scientists regardless of their level of rationality.

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

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

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).

The communist norm, on the other hand, was not established as a norm of science until around 1665. At the time, "many men of science still set a premium upon secrecy" (Zuckerman and Merton 1971, p. 69). The first scientific journals—the *Journal des Sçavans* and the *Philosophical Transactions*, both founded in 1665—were instrumental "for the emergence of that component of the ethos of science which has been described as 'communism': the norm which prescribes the open communication of findings to other scientists" (Zuckerman and Merton 1971, p. 69).

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. My results show that the priority rule gives scientists an incentive to share intermediate results whenever these are awarded credit proportional to their difficulty. 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 intermediate results being awarded sufficient credit. 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 sufficiently rewarded with credit. 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 sufficient credit for whatever one would like to see shared, and scientists will indeed start sharing it.

A A Unique Nash Equilibrium

The remainder of this section is used to prove theorem A.4, which is then used to prove the results in the main text.

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 of perfect information and let $G^m_{n,k}$ denote the game of imperfect information, as described in sections 4 and 5.

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 s_i gives her strategy and s_{-i} gives the strategies of all scientists other than 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 $G^m_{n,k}$.[14] Let s_{-i}^E denote the incomplete strategy profile (in

[14]This is an abuse of notation because I use s_i^E to denote two distinct strategies: one for each game. But the proofs below rely almost exclusively on features they share. I will

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.

Let W_j denote the waiting time until some scientist shares the solution to stage j. In particular, W_k is the waiting time until the research project is completed. These waiting times may depend on the scientists' strategies and so I write $W_j(S)$ for the waiting time until stage j is shared under strategy profile S.

Lemma A.1. *Let $S \neq S^E$ be some arbitrary strategy profile. In the case of $G_{n,k}^p$, add the further assumption that this involves a deviation on the equilibrium path relative to S^E, i.e., there is a positive probability of reaching a decision node where some scientist plays strategy H. Then in both $G_{n,k}^p$ and $G_{n,k}^m$*

$$\mathbb{E}\left(W_{j'}(S)\right) \geq \mathbb{E}\left(W_{j'}(S^E)\right) = \sum_{j=1}^{j'} \frac{1}{\sigma_j},$$

with strict inequality if $j' = k$.

Proof. Under S^E, the time it takes the scientists to solve and share stage j is exponentially distributed with parameter σ_j (cf. section 4). So the expected time to solve and share stage j is $1/\sigma_j$. This establishes the equality.

Under S, scientists may make discoveries that are not shared. This may reduce the waiting time for particular stages (for example, if all scientists play strategy H at stage $j-1$ then the waiting time between stage $j-1$ and stage j is zero with probability one) but this is always compensated by an increase in the waiting time for one or more preceding stages. On average, $W_{j'}$ can only increase if some scientists are keeping results secret (cf. Banerjee et al. 2014, section 4).

By the assumptions of the lemma, there is a positive probability of reaching a decision node where some scientist, say scientist i, plays strategy H.

explicitly note when this is not the case.

Then there is a positive probability that scientist $i' \neq i$ is the next scientist to share the solution to some stage. It follows that scientist i' has shared at least one stage that scientist i had already solved but not shared. Since events like this occur with positive probability,

$$\mathbb{E}\left(W_k(S)\right) > \mathbb{E}\left(W_k(S^E)\right). \qquad \square$$

Lemma A.2. *In both $G^p_{n,k}$ and $G^m_{n,k}$, under the strategy profile S^E (i.e., every scientist always shares any stages she completes immediately) the payoff to scientist i is*

$$u_i\left(s_i^E, s_{-i}^E\right) = \sum_{j=1}^{k} \frac{c_j \lambda_{ij}}{\sigma_j}.$$

Proof. If scientist i plays strategy s_i^E she always immediately shares her intermediate results. Hence, she is always working on the stage immediately after the most recently shared stage and so always only takes credit for one stage at a time. Under these circumstances, scientist i can be viewed as a nonstationary reward process producing payoff at a rate of $c_j\lambda_{ij}$ units of payoff per unit time if j is the current stage.[15] Since the expected amount of time spent in each stage is $1/\sigma_j$ by lemma A.1, this gives the desired result. $\square$

Note that the preceding lemmas do not require Proportional Credit; they are true for all (positive) productivity rates and for all (positive) credit rewards. The next lemma shows that under Proportional Credit, if not every

[15]Note that this is true despite the fact that scientist i may not be the first to solve stage j and hence may not end up getting those particular c_j units of credit. This is due to the separability of independent Poisson processes. This way of looking at the game is crucial to my proof of lemma A.3 below, but the present lemma can also be proven without relying on this separability:

Scientist i is the first to complete stage 1 with probability λ_{i1}/σ_1. If she does she immediately claims c_1 units of credit. If any other scientist completes stage 1 before scientist i, that scientist immediately claims c_1 units of credit. Thus scientist i's expected credit from the first stage is $c_1\lambda_{i1}/\sigma_1$. Then all scientists simultaneously start working on the next stage. By the same reasoning, scientist i's expected credit from stage j is $c_j\lambda_{ij}/\sigma_j$.

scientist always shares, scientists who always share get a higher payoff than they do in lemma A.2.

Lemma A.3. *Let i be a scientist and assume Proportional Credit. Let s_{-i} denote any incomplete strategy profile such that at least one scientist i' plays some strategy other than $s^E_{i'}$ (this can be either a different pure strategy, or any mixed strategy which plays strategy $s^E_{i'}$ with probability less than one). In the case of $G^p_{n,k}$, add the further assumption that this involves a deviation on the equilibrium path, i.e., on the (complete) strategy profile (s^E_i, s_{-i}) there is a positive probability of reaching a decision node where some scientist plays strategy H. Then in both $G^p_{n,k}$ and $G^m_{n,k}$*

$$u_i\left(s^E_i, s_{-i}\right) > u_i\left(s^E_i, s^E_{-i}\right).$$

Proof. Let S denote the strategy profile (s^E_i, s_{-i}). Just like in the proof of lemma A.2, it is useful to view scientist i as a nonstationary reward process. When she is working on stage j, she produces payoff at a rate of $c_j \lambda_{ij}$ units of payoff per unit time. Call this the reward rate, denoted $r_{i,j} = c_j \lambda_{ij}$. The overall expected payoff to scientist i is found by multiplying these rates by the expected time spent on each stage (setting $W_0(S) = 0$ for notational convenience):

$$u_i(S) = \sum_{j=1}^{k} r_{i,j} \mathbb{E}\left(W_j(S) - W_{j-1}(S)\right).$$

By Proportional Credit, $r_{i,j'} \geq r_{i,j'+1} \geq \ldots \geq r_{ik} > 0$, so the differences between them are nonnegative: $r_{i,j'} - r_{i,j'+1} \geq 0$. For notational convenience, write $r_{i,k+1} = 0$. Then

$$u_i(S) = \sum_{j=1}^{k} \sum_{j'=j}^{k} \left(r_{i,j'} - r_{i,j'+1}\right) \mathbb{E}\left(W_j(S) - W_{j-1}(S)\right).$$

Interchanging the sums yields

$$\begin{aligned}
u_i(S) &= \sum_{j'=1}^{k} \sum_{j=1}^{j'} (r_{i,j'} - r_{i,j'+1}) \, \mathbb{E}\left(W_j(S) - W_{j-1}(S)\right) \\
&= \sum_{j'=1}^{k} (r_{i,j'} - r_{i,j'+1}) \, \mathbb{E}\left(W_{j'}(S)\right) \\
&> \sum_{j'=1}^{k} (r_{i,j'} - r_{i,j'+1}) \sum_{j=1}^{j'} \frac{1}{\sigma_j},
\end{aligned}$$

where the inequality uses the following facts: $r_{i,j'} - r_{i,j'+1} \geq 0$ for all j' (by Proportional Credit), $r_{i,k} - r_{i,k+1} = r_{i,k} > 0$, $\mathbb{E}(W_{j'}(S)) \geq \sum_{j=1}^{j'} 1/\sigma_j$ for all j' by lemma A.1, and $\mathbb{E}(W_k(S)) > \sum_{j=1}^{k} 1/\sigma_j$ by lemma A.1.

Interchanging the sums again yields the desired result:

$$u_i(S) > \sum_{j=1}^{k} \sum_{j'=j}^{k} (r_{i,j'} - r_{i,j'+1}) \frac{1}{\sigma_j} = \sum_{j=1}^{k} \frac{r_{i,j}}{\sigma_j} = u_i(S^E). \qquad \square$$

Theorem A.4. *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. Under Proportional Credit 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. Recall that the game is zero-sum: regardless of strategies, there are C units of credit to be divided, and so if one scientist's payoff increases, another's decreases. Combined with lemmas A.2 and A.3 this 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 A.3 is getting a payoff greater than $u_{i'}(s_{i'}^E, s_{-i'}^E)$.

Because the game is zero-sum, it follows that $u_i(s_i, s_{-i}) < u_i(s_i^E, s_{-i}^E)$. By lemma A.2, $u_i(s_i^E, s_{-i}) = u_i(s_i^E, s_{-i}^E)$, 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 $u_{i'}(s_{i'}^E, s_{-i'}^E)$ by lemma A.3. Because the game is zero-sum, at least one of the remaining scientists, say scientist i, must be getting a payoff less than $u_i(s_i^E, s_{-i}^E)$. But if scientist i changed her strategy to s_i^E, by lemma A.3 she would get a payoff $u_i(s_i^E, s_{-i}) > u_i(s_i^E, s_{-i}^E)$, establishing the result.

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 $u_{i'}(s_{i'}^E, s_{-i'}^E)$. So there is at least one scientist, say scientist i, such that $u_i(s_i, s_{-i}) \leq u_i(s_i^E, s_{-i}^E)$. By lemma A.3, $u_i(s_i^E, s_{-i}) > u_i(s_i^E, s_{-i}^E)$, and the result follows. □

Theorem A.4 may be used to prove theorems 5.1 and 5.2.

Proof of theorem 5.1. 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. (A direct proof that in the backwards induction solution every scientist plays strategy E at every decision node—including those off the equilibrium path—is available from the author upon request.) □

Proof of theorem 5.2. Let S be any profile (of pure or mixed strategies) for the game $G^m_{n,k}$. 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.

That S^E is a Nash equilibrium, and in fact a strict Nash equilibrium, also follows from theorem A.4 by considering the special case where $s_{-i} = s^E_{-i}$. This shows that a scientist i who deviates unilaterally makes herself strictly worse off. □

To prove theorem 8.1, 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 A.5 (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 8.1 is a corollary of theorems 5.2, A.4 and A.5.

Proof of theorem 8.1. By theorem 5.2, 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 A.5.

That the game is a weakly better reply game follows from theorem A.4. At any strategy profile, for at least one scientist i whose strategy differs from

s_i^E switching to strategy s_i^E 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_i^E, and improves her payoff by doing so. $\square$

Finally, I show that my results do not depend on the fact that payoff is measured in total credit rather than credit per unit time.

Let $G_{n,k}^{pt}$ denote the adapted version of the game with perfect information. That is, in $G_{n,k}^{pt}$ scientists know when other scientists solve stages even if the results are not shared and the (utility) payoff is measured in units of credit per unit time. Similarly, $G_{n,k}^{mt}$ is the adapted version of the game with imperfect information.

Theorem A.6. *Let S be any strategy profile for $G_{n,k}^{mt}$ other than S^E, or let S be any strategy profile for $G_{n,k}^{pt}$ that involves deviations on the equilibrium path relative to S^E. Under Proportional Credit 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. By theorem A.4 there exists a scientist i whose total expected credit from the research project is higher under strategy profile (s_i^E, s_{-i}) than under (s_i, s_{-i}). But (by reasoning similar to that given in the proof of lemma A.1) it is also clear that the expected time to complete the research project can only decrease if scientist i switches to strategy s_i^E, i.e.,

$$\mathbb{E}\left(W_k(s_i^E, s_{-i})\right) \leq \mathbb{E}\left(W_k(s_i, s_{-i})\right).$$

But then it follows immediately that scientist i's expected credit per unit time must also be higher under (s_i^E, s_{-i}):

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

Theorem A.6 may be used to prove analogous versions of the main results found above for the adapted version of the model. The proofs are exactly as given above.

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.

Alain Aspect, Jean Dalibard, and Gérard Roger. Experimental test of Bell's inequalities using time-varying analyzers. *Physical Review Letters*, 49: 1804–1807, 1982.

Siddhartha Banerjee, Ashish Goel, and Anilesh Kollagunta Krishnaswamy. Re-incentivizing discovery: Mechanisms for partial-progress sharing in research. In *Proceedings of the Fifteenth ACM Conference on Economics and Computation*, pages 149–166, New York, 2014.

Cristina Bicchieri. Methodological rules as conventions. *Philosophy of the Social Sciences*, 18(4):477–495, 1988.

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

Christine L. Borgman. The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 63(6): 1059–1078, 2012.

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.

Thomas Boyer-Kassem and Cyrille Imbert. Scientific collaboration: Do two heads need to be more than twice better than one? *Philosophy of Science*, 82(4):667–688, 2015.

Partha Dasgupta and Paul A. David. Toward a new economics of science. *Research Policy*, 23(5):487–521, 1994.

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, 2014.

Remco Heesen. Why the reward structure of science makes reproducibility problems inevitable. Manuscript, Jul 2016.

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.

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.

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.

John C. Huber and Roland Wagner-Döbler. Scientific production: A statistical analysis of authors in mathematical logic. *Scientometrics*, 50(2): 323–337, 2001a.

John C. Huber and Roland Wagner-Döbler. Scientific production: A statistical analysis of authors in physics, 1800-1900. *Scientometrics*, 50(3): 437–453, 2001b.

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

Simon M. Huttegger, Brian Skyrms, and Kevin J. S. Zollman. Probe and adjust in information transfer games. *Erkenntnis*, 79(4):835–853, 2014.

Philip Kitcher. The division of cognitive labor. *The Journal of Philosophy*, 87(1):5–22, 1990.

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.

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.

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. 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. Reprinted in Merton (1973, chapter 16).

Robert K. Merton. Behavior patterns of scientists. *The American Scholar*, 38(2):197–225, 1969. Reprinted in Merton (1973, chapter 15).

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

James R. Norris. *Markov Chains*. Cambridge University Press, Cambridge, 1998.

Heather Piwowar. Altmetrics: Value all research products. *Nature*, 493 (7431):159, 2013.

David B. Resnik. Openness versus secrecy in scientific research. *Episteme*, 2:135–147, 2006.

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.

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

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, 2011.

Harriet Zuckerman and Robert K. Merton. Patterns of evaluation in science: Institutionalisation, structure and functions of the referee system. *Minerva*, 9(1):66–100, 1971. Reprinted in Merton (1973, chapter 21).