Expediting the Flow of Knowledge Versus Rushing into Print*

Remco Heesen†‡

October 15, 2018

Abstract

Recent empirical work has shown that many scientific results may not be reproducible. By itself, this does not entail that there is a problem (or a "reproducibility crisis"). However, I argue that there is a problem: the reward structure of science incentivizes scientists to focus on speed and impact at the expense of the reproducibility of their work. I illustrate this using a well-known failure of reproducibility: Fleischmann and Pons' work on cold fusion. I then use a rational choice model to identify a set of sufficient conditions for this problem to arise, and I argue that these conditions plausibly apply to a wide range of research situations. In the conclusion I consider possible solutions and implications for how Fleischmann and Pons' work should be evaluated.

---

*This paper is a longer version of "Why the Reward Structure of Science Makes Reproducibility Problems Inevitable" (published in *the Journal of Philosophy*). Proofs of the theorems mentioned in that paper are found in appendix A. This longer version of the paper includes a case study on cold fusion and a section on a variation of the model with a three-way tradeoff between speed, reproducibility, and impact.

†Department of Philosophy, School of Humanities, University of Western Australia, Crawley, WA 6009, Australia. Email: remco.heesen@uwa.edu.au.

‡Faculty of Philosophy, University of Cambridge, Sidgwick Avenue, Cambridge CB3 9DA, UK.

1 Introduction

The reproducibility of scientific research is a cornerstone of the scientific method. If science is to discover general laws or principles, it should not matter who tests them, or when, or where. Thus it is a necessary condition for the acceptability of a particular scientific result that, if some (hypothetical or actual) scientist competently performs the same experiment, it produces the same result.[1]

Reproducibility has come under increased scrutiny, especially in the fields of medicine and psychology. There has long been "a general impression that many results that are published are hard to reproduce" (Prinz et al. 2011), which has recently begun to be empirically tested. Two studies by pharmaceutical companies reported that they could reproduce less than a quarter of results in cancer biology (Prinz et al. 2011, Begley and Ellis 2012). A more systematic study is currently underway (Nosek and Errington 2017). A large systematic study of results published in prominent psychology journals found that less than half could be reproduced (Open Science Collaboration 2015), while in a similar study of social science experiments slightly more than half could be reproduced (Camerer et al. 2018).

This has led to talk of *reproducibility problems* (or a "crisis"): the above results are taken to show that a smaller number of scientific results is reproducible than would be "expected" if science was being done "properly". However, empirically measured reproducibility rates cannot prove this by themselves as it is not obvious what reproducibility rates should be expected. The following toy model illustrates this (based on Ioannidis 2005).

Consider a scientific community engaging in various projects with the aim to make major scientific discoveries. Suppose that only one in every twenty-one projects leads to a genuine discovery, but there is a false discovery rate $\alpha = 0.05$. Then, on average, two out of twenty-one projects will claim a

[1]Determining whether a particular result should count as "the same" can be a complicated affair. The field of statistics is largely devoted to this task.

discovery, but only half of these will be genuine.

In such a world, finding that approximately half of all scientific discoveries can be reproduced would in fact be expected if science was being done properly. More generally, the unknown number of genuine scientific results waiting to be discovered (relative to the total number of scientific projects) influences the reproducibility rates that can be expected. So the empirical reproducibility rates cited above do not suffice to show that there is a crisis.

The aim of this paper is to show that there is in fact a problem, and to diagnose one source of the problem. The diagnosis is that due to the *reward structure of science* scientists have an incentive to produce research that is less likely to be reproducible than we should want. This diagnosis is more general than those that identify particular statistical practices or journal practices as the source of the problem, implying that the problem is harder to solve than some have thought.

The reward structure of science is centered around *credit*. Credit is acquired by receiving recognition for one's work, e.g., by having it published in a scientific journal, by having it cited, or by receiving a prestigious award for it. Because their careers may depend on it, scientists are very concerned about credit. This point has long been recognized by philosophers of science like Hull (1988, chapter 8), Kitcher (1993, chapter 8), Strevens (2003), and Zollman (2018) and sociologists such as Merton (1957, 1969) and Latour and Woolgar (1986, chapter 5).

This raises the question what behaviors scientists are likely to engage in to get credit. Philosophers and economists have used rational choice models to answer this question. Kitcher (1993, chapter 8) and Strevens (2003) argue that credit can incentivize scientists to distribute themselves over research programs in a way that is closer to optimal than if they were individually epistemically rational. Dasgupta and David (1994) and Zollman (2018) argue that credit incentives speed up the progress of science. Boyer-Kassem and Imbert (2015) argue that credit incentives encourage collaboration between scientists. And Boyer (2014), Strevens (2017), and Heesen (2017) argue that credit incentives can motivate scientists to share their work widely.

This literature largely focuses on the positive effects of credit incentives.[2] In contrast, it has been suggested that the incentive structure of science may be (at least partially) responsible for reproducibility problems—a negative effect. I distinguish four such claims.

Claim (Rushing into print)**.** *Scientists are incentivized to produce more results and/or more important results, at the expense of spending more time on the reproducibility of any given result.*

Claim (Publication bias)**.** *Scientists are incentivized to favor positive results at the expense of reproducibility.*

Claim (Novelty bias)**.** *Scientists are incentivized to favor novel results at the expense of reproducibility.*

Claim (Checking bias)**.** *Scientists are disincentivized to attempt to reproduce each other's work, reducing the incentive to make sure their own work is reproducible.*

Publication bias, novelty bias, and checking bias each pick out relatively specific features of the reward structure of science.[3] Publication bias originates from journals' preference for positive results (Easterbrook et al. 1991, Egger and Smith 1998).[4] A "positive" result usually means that a statistical

[2]Largely, but not exclusively. For example, it has been argued that credit incentives contribute to herding behavior (Strevens 2013) and a productivity gap between male and female scientists (Bright 2017b).

[3]I do not mean to take a strong stand here on whether these phenomena are appropriately called biases. The name "publication bias" is well established in the literature and I am simply naming novelty bias and checking bias by analogy to this existing terminology.

[4]I focus on journals here because journal publications are an important way to be rewarded for scientific work. This includes both a preference by editors and reviewers for positive results and a preference by scientists to only submit positive results in anticipation of such a preference (note that the latter can happen even if journals do not actually have such a preference).

hypothesis that a certain experimental condition has “no effect” (a null hypothesis) is rejected. As a result, evidence favoring a null hypothesis is not published, biasing the scientific literature (the so-called “file drawer problem” Rosenthal 1979).

Novelty bias and checking bias originate from journals’ preference for novel results. In particular, journals are generally not interested in publishing *direct replications*, i.e., studies that follow (as much as possible) exactly the same experimental design as a previously published study (Neuliep and Crandall 1990). This incentivizes scientists to skew (the presentation of) their results to focus on novel findings (novelty bias). It also means scientists do not expect that anyone will attempt to reproduce their work, weakening their incentive to make sure their work is reproducible (checking bias).

Publication bias, novelty bias, and checking bias each suggest their own solution. Publication bias would be prevented or seriously reduced if null results were regularly published and rewarded (Ioannidis 2006). Novelty bias and checking bias would be seriously reduced if replication studies were regularly published and rewarded (Koole and Lakens 2012).

The incentive to rush into print, on which I focus in this paper, is different in two ways. First, as I will argue, it does not depend on fairly specific features of the reward structure of science, but rather on the general facts that scientific work is rewarded and that these rewards are determined at least partially before it is known whether the work is reproducible. Second (and as a consequence of the first), it is much less obvious how the incentive to rush can be reduced or eliminated.

Rushing into print describes the incentive that scientists have to focus on the speed with which they can produce results and/or the impact of those results, and the corresponding lack of focus on the reproducibility of these results. This is usually described with the phrase “pressure to publish” (Fanelli 2010, Prinz et al. 2011), although I argue the phenomenon is not essentially tied to journal publications.

I illustrate rushing into print in section 2 with a case study in which a concern for credit led to the publication of research that other scientists were unable to reproduce: Fleischmann and Pons' work on cold nuclear fusion.

I then provide a rational choice model of scientists' decision how much time to spend on a particular study before trying to publish it[5] and I compare the credit-maximizing choice to the choice that is optimal from a social perspective. I show that three ingredients are sufficient to create a systematic incentive to rush into print (section 3). First, the fact that speed and reproducibility trade off against each other. Second, the fact that scientists get rewarded for publications. And third, the fact that the system of peer review fails to predict perfectly which work will be successfully reproduced.

The point of the rational choice model is to show that scientists have an incentive to produce work that is less likely to be reproducible than is socially desirable. As I argued above, this is not something that can be shown by empirical studies of reproducibility rates. I also show that the particular journal preferences responsible for publication bias, novelty bias, and checking bias are not necessary for reproducibility problems to arise. They may, however, exacerbate such problems.

In section 4 I expand the model by allowing the scientist to also choose a desired level of impact. High-impact work has greater scientific value and yields more credit, but this trades off against speed and/or reproducibility. I show the robustness of my earlier results in this expanded model, and I consider how it gives rise to different types of scientists: "impact-seekers" and "safety-seekers".

In the conclusion (section 5) I summarize my results. I also discuss possible ways to disincentivize rushing into print. And finally, I discuss the extent

[5]The model, as well as this paper's title, reflects a tension in scientists' motivations with regard to this decision. This tension was, to my knowledge, first noted by Merton (1969, p. 209) and expressed dramatically by Fuller (2002, p. 201): "[T]he scientist is supposed to *both* expedite the flow of knowledge *and* not rush into print. But how can he "expedite" without also "rushing"?"

to which my results support an interpretation of Fleischmann and Pons' cold fusion research as a case of rushing into print.

My focus in this paper is exclusively on the incentives that scientists have to do their work in a way that is less likely to be reproducible. This *incentive-based* approach to reproducibility problems may be contrasted with a *methodological* approach. A methodological approach focuses on identifying practices that scientists engage in that may lead to irreproducible research.

For example, various choices have to be made in conducting a research study: Should outliers be excluded? What statistical test should be used? And so on. These choices have been called "researcher degrees of freedom" (Simmons et al. 2011). If scientists run their analysis multiple times (varying how these choices are made) and report only cherry-picked[6] results, this creates a biased publication record (Ioannidis 2005, Simmons et al. 2011).

Focusing on methodology invites different suggestions for improving reproducibility than focusing on incentives, e.g., pre-registration of studies and more complete reporting of results. However, such suggestions will not eliminate researcher degrees of freedom entirely. It is both impossible and undesirable to reduce the scientific method to a fixed mechanical protocol, as Feyerabend (1975) famously argued.

This suggests that if a scientist has an incentive to produce biased results, it will always be possible for her to do so without straying from the methodological norms of her field. For this reason I think methodological suggestions for improving reproducibility are unlikely to be as effective as their proponents hope unless the incentives leading to irreproducible research are also addressed.

[6] The discussion above has suggested what kind of cherry-picking scientists may engage in. For example, results that reject a null hypothesis would be favored over those that fail to reject a null, results that have a "novelty" factor would be favored, and so on. This can be done consciously (for careerist reasons) or unconsciously ("this result fails to support our hypothesis, so something must have gone wrong").

2 Cold Fusion

In this section I use a case study to argue that the pressure to publish can lead to the publication of research that cannot be reproduced. The next section aims to show that this is a structural rather than an incidental problem.

On March 23, 1989, two established and respected chemists named Martin Fleischmann and Stanley Pons gave a remarkable press conference at the University of Utah. They claimed that by loading a palladium rod with deuterium through electrolysis, they had turned the rod into a source of energy, producing up to four times as much heat as they put in.

They hypothesized that the deuterium atoms might be packed together so closely within the palladium as to force pairs of them together in an energy-producing process known as *nuclear fusion*. Conventional wisdom held that a sustained, controlled fusion reaction—the kind needed for a viable source of energy—requires temperatures over a hundred million degrees (among other things). Now two chemists claimed to be able to achieve the same thing at room temperature. Hence the phenomenon came to be known as *cold fusion*.

A media hype ensued, as cold fusion held the promise of a clean and nearly boundless source of energy. Given these implications, and Fleischmann and Pons' impeccable credentials as experimentalists, scientists around the world dropped what they were doing to attempt to reproduce the experiment.

Within the first few weeks after the press conference, a number of announcements were made (usually also via press conference) by researchers seeing similar phenomena. But as time passed their claims came under heavy criticism. The excess heat measurements were attributed to mistakes in accounting for the potential recombination of gases released during the experiment. The neutron measurements (Fleischmann and Pons' other important piece of evidence) could not be replicated with more sophisticated equipment. After the meeting of the American Physical Society (APS) in May 1989, the tide shifted from a mixture of excitement and skepticism to a consensus that Fleischmann and Pons had been mistaken; the phenomenon was deemed irreproducible.

The current scientific consensus, then, is that it is not possible to achieve cold fusion at meaningful rates. Fleischmann and Pons' claim to the contrary on March 23, 1989, has been heavily criticized by scientists. In their books on the case, Close and Huizenga judge that they "went public too soon with immature results" (Close 1991, p. 328) and that their "gamble to go public... is the scientific fiasco of the century" (Huizenga 1993, p. 214). What led Fleischmann and Pons to make this fateful decision to "go public"?

At the nearby Brigham Young University, physics professor Steven Jones and his team had been working on a very similar project. The main differences were that Jones was primarily interested in explaining the heat at the center of the Earth (rather than creating a new source of energy) and that he focused on measuring neutron production rather than excess heat.

The two teams first became aware of each other in September of 1988 and interacted a number of times. In February of 1989, Jones announced that he was going to present his data at the APS meeting in May and was planning to submit an article to a journal soon. Fleischmann and Pons were not ready for this yet. They were confident in their evidence that some experiments produced excess heat, but much remained to be investigated. They indicated that they wanted to do another eighteen months of research before going public (Pool 1989, Huizenga 1993, p. 18).

The two groups agreed to a compromise: they would submit their results to *Nature* simultaneously on March 24. Jones has claimed that there was a further agreement not to publicize the work until that time, but Pons has denied this (Pool 1989, Close 1991, p. 94). Either way, Fleischmann and Pons did publicize their work: they sent a manuscript to the *Journal of Electroanalytical Chemistry* on March 11, and they held the above-mentioned press conference on March 23.

Their goal in going public was to establish priority—and thus claim credit—for the cold fusion research, especially relative to Jones (Huizenga

1993, p. 19). This may seem unnecessary in hindsight, as Jones' experimental results were quite different (measuring neutrons rather than heat) and of such a different order of magnitude as to hold no promise for a viable source of energy. Jones' publication would thus not appear to be a threat to the originality or importance of Fleischmann and Pons' work. But this was not so clear at the time, as Fleischmann explained later.

> We could not tell whether Jones had heat data or was planning to look for this. How could one tell? He was certainly thinking about fusion as a source of heat in the Earth. If he was going to say *that* in the paper, which was surely his intention to do, it would almost certainly destroy any possibility of patent protection (quoted in Close 1991, pp. 99–100).

Thus, both the decision to agree to publish simultaneously with Jones and the later decisions to submit to a different journal before Jones and hold a press conference were made out of a concern for credit. Fleischmann and Pons were aware that their results were still preliminary (they wanted to do another eighteen months of research before publishing anything) but went public anyway to establish priority.

So a concern for credit led to the publication of research that other scientists were unable to reproduce. The model presented next aims to establish that this is a structural problem: scientists have a credit incentive to rush into print. In doing so, the model also lends some support to the claim that Fleischmann and Pons did nothing irrational by going public when they did, despite Close and Huizenga's criticism of this decision (as I argue in more detail in section 5).

3 A Tradeoff Between Speed and Reproducibility

Here I develop a rational choice model to evaluate decisions to go public with results of scientific research. By giving a model, I aim to show the existence of a (structural) credit incentive to rush into print. This section considers only the tradeoff between speed and reproducibility, while section 4 also allows the potential importance or impact of the result to vary.

Consider a scientist—or a team of scientists, such as Fleischmann and Pons—working on a research study. When should she attempt to publicize her work, say in the form of a journal article? As the case of Fleischmann and Pons illustrates, getting credit for the work is an important consideration.

As I mentioned in the introduction, scientists' concern for credit is well-documented and understandable, given its importance to scientific careers (Merton 1957, 1969). Because I am interested in what the scientist has a credit incentive to do, I assume that credit is her only concern. This is a *methodological* assumption to isolate the credit incentive.

This is represented in the model by assuming that the scientist aims to maximize the amount of credit she accrues per unit time. As a consequence, the scientist prefers to get her work published faster rather than slower (all else being equal): if the amount of credit per publication is some fixed number c (this assumption is relaxed in section 4), publishing twice as fast will double credit per unit time. So the concern for credit entails a concern for *speed* (to be defined more formally below).

At the same time, publishing faster reduces *reproducibility*. By reproducibility I mean, loosely speaking, the likelihood that the result of the research study (e.g., "cold fusion is a viable source of energy") is reproduced if someone attempts to do so. But this loose definition has two problems. First, what if no one attempts to reproduce the result? And second, what if multiple attempts to reproduce it are made, with some succeeding and some failing?

Since credit is conferred socially, what really matters is the standing of a result in the eyes of other scientists. So I call a scientific result *accurate* if it holds up in the relevant scientific community in the mid-term, i.e., either no one attempts to reproduce it, or any subsequent studies are taken on balance to have reproduced the result. Conversely, I call a result *erroneous* if it does not hold up in the community in the mid-term, i.e., the community deems the result irreproducible. The *reproducibility* of the result is then the scientist's subjective probability, given the evidence gathered at the time of publication, that the result is accurate.[7] This definition should be interpreted broadly, applying to both experimental and non-experimental contributions (e.g., a mathematical theorem is considered reproducible by my definition if no one discovers a mistake in it).

In the model, the scientist chooses the desired reproducibility $p \in [0, 1]$ *ex ante*. I assume this to be fixed for the duration of the study. That is, the scientist works on her research project until she obtains a result that she thinks has at least probability p of holding up in the community, at which time she publishes.

Reproducibility takes time (think of the eighteen more months of research Fleischmann and Pons wanted to do). This is reflected in the model by the *speed function* λ. The value $\lambda(p)$ reflects the scientist's expected speed if the desired reproducibility is p (see figure 3.1). More specifically, $\gamma(p) = 1/\lambda(p)$ is the expected time until completion of the research project (so $\lambda(p)$ is the number of projects "like this one" that the scientist would expect to complete per unit time). This reflects the scientist's *ex ante* expectation about the duration of the study.

[7] Note that it follows from these definitions that reproducibility goes up if fewer attempts to reproduce scientific results are made, because results that are never tested count as accurate. For this reason, the present model is not suitable to study strategic decisions scientists might make regarding whether to reproduce others' work. In particular, I do not aim to capture the phenomenon of novelty bias, discussed in the introduction. See Bruner (2013) for a model in which the incentives to reproduce others' work are considered.

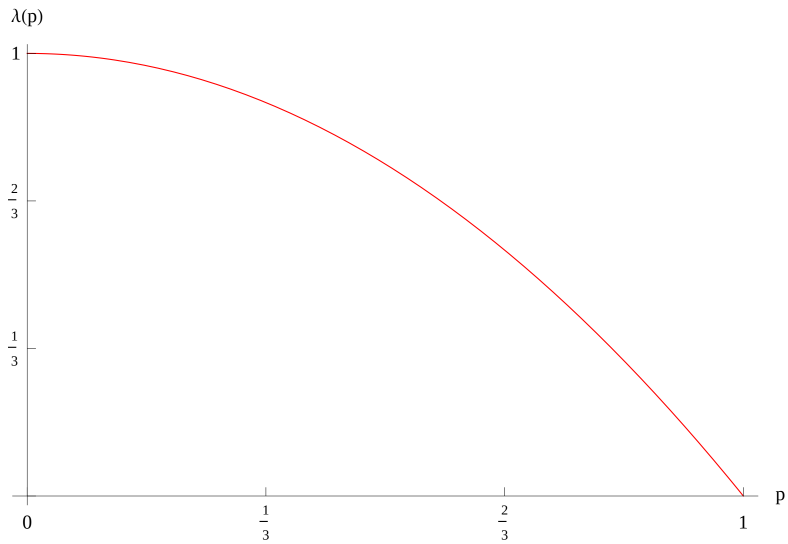

Figure 3.1: p and λ trade off against each other. In this example, $\lambda(p) = 1 - p^2$, satisfying assumptions 3.1–3.3.

Reducing reproducibility (lowering p) allows the scientist to publish faster. "Rushing" the work in this way could mean that the scientist ends the study sooner (gathering less evidence), or it could mean that the scientist tries to gather the same amount of evidence more quickly (potentially making mistakes). The present model is not intended to investigate incentives related to deliberate fraud, such as when data is misreported or fabricated, or when publication time is reduced through (self-)plagiarism. For formal work on credit-based incentives for fraud, see Bruner (2013) and Bright (2017a).

I make a number of assumptions on the way speed and reproducibility trade off against each other, as reflected in the speed function λ.

Assumption 3.1 (The speed function is decreasing). *For all* $p, q \in [0, 1]$, *if* $p < q$, *then* $\lambda(q) < \lambda(p)$.

Assuming that λ is decreasing means that the scientist expects to take more time to do research that is less likely to be erroneous. For example, this may involve collecting more data or being more thorough, either of which takes time.

One might object that in some situations (e.g., the scientist discovers a mistake in her previous work) the scientist's confidence in the reproducibility of her work might go down instead of up over time, seemingly in violation of this assumption. But this misinterprets the function λ. This function gives, for each desired reproducibility p, the scientist's *ex ante* expectation of how long it would take for her confidence in her result to reach at least level p. So if the scientist's confidence was at p or above before discovering the mistake, she would already have published, but if her confidence was below p she will have to work until it finally reaches p before publishing.

Thus the model does not capture the dynamics of a scientist's expectations about the duration of the project as they change over time. However, I suspect that similar conclusions might be reached in a suitable dynamic model by evaluating the scientist's expectations at any given time.

Assumption 3.2 (The speed function is concave)**.** *For every* $p, q, t \in [0, 1]$*,*

$$t\lambda(p) + (1 - t)\lambda(q) \leq \lambda(tp + (1 - t)q).$$

This assumption may be described as a kind of decreasing marginal returns. As the reproducibility p is lowered, the expected speed λ is increased by assumption 3.1, but it increases ever slower as p approaches zero: writing the paper itself takes time, which becomes relatively more significant if the scientist spends relatively little time on the research content. Conversely, if the scientist aims for higher reproducibility (increasing p), the speed λ drops off ever faster. More time is required, e.g., to increase p from 0.8 to 0.9 than from 0.7 to 0.8. Peirce (1967 [1879]) makes the same observation.

To put the point in more statistical terms: the size of the standard error of a parameter estimate gives an indication of the reproducibility of the result. As n (the amount of data collected) increases, this width decreases at a rate proportional to $\sqrt{n}$. So there are decreasing marginal returns (in terms of the certainty with which the result is established) from investing the time to collect more data, such that, e.g., doubling the dataset does not halve the margin of error.

Assumption 3.3 (No perfect work)**.** $\lim_{p\to 1} \lambda(p) = 0$.

This assumption asserts that the scientist cannot deliver perfect work (in the sense of zero probability of errors), no matter how slowly she works. This reflects the fact that there is no certainty in science: for any fact or discovery, it is always possible that it will later be overturned, as Lakatos (1978) and Quine (1951, section VI) have argued.

Note that these assumptions imply the following restrictions on the expected completion time $\gamma(p) = 1/\lambda(p)$: the expected completion time is increasing, convex, and diverges to infinity as p approaches one. These restrictions can be given analogous justifications to the above.

The above assumptions also imply that the expected speed is a continuous function of reproducibility, which may be unrealistic when (say) experimental results arrive in batches, leading to discontinuous jumps in reproducibility. This is only a problem for my model if such discontinuities are sufficiently common and predictable that the scientist can anticipate them (since the speed function reflects her *ex ante* expectations). This requires not only that the scientist knows in advance that experimental results arrive in batches, but also that she can predict fairly accurately what level of reproducibility she will reach with the first batch.

Such cases may arise; my claim here is not to capture all scientists everywhere, but a large range of cases. The types of cases excluded from the model are those in which evidence is gathered in discrete amounts, with relatively predictable effects on the scientist's confidence in her results, and where the scientist is in a position to decide whether or not to gather more evidence after seeing some initial results.

How does all this affect the scientist's credit? For reasons I outlined above, I assume the scientist gets credit only for published work. Whether or not the scientist's work is published is determined through *peer review*. The purpose of peer review is to determine the accuracy of the scientist's work, i.e., whether her results are likely to stand up to the scrutiny of the scientific community.

Suppose it does this "pre-screening" perfectly: all and only those papers that are in fact accurate are accepted (I will drop this assumption momentarily). The scientist does not know whether her paper is accurate; she only knows the reproducibility p, i.e., her own credence that it is accurate. So from the scientist's perspective, if she produces a paper with reproducibility p, there is a probability p that the journal publishes it.

Suppose that the average amount of credit for a published accurate result is c_a. Then the scientist's expected credit per unit time is a function C of the chosen reproducibility p and the speed λ (which is itself a function of p) given by $C(p) = c_a p \lambda(p)$.

In reality the peer review system cannot perfectly predict the success of future attempts to reproduce present results. Some accurate results get rejected, while some erroneous results get accepted. An example of the latter is Fleischmann and Pons' paper in the *Journal of Electroanalytical Chemistry*: it passed peer review but was thoroughly discredited within a year of publication (thus satisfying my definition of erroneous).

The acceptance of an erroneous result is called a *false positive* and the rejection of an accurate result a *false negative*. Following common usage in statistics I write α for the probability of a false positive and β for the "power" (the probability that a false negative is avoided, i.e., that an accurate result is accepted). The case of "perfect peer review" described above would be one where $\beta = 1$ and $\alpha = 0$.

Here I assume instead that peer review is imperfect in the sense of a positive probability of false positives ($\alpha > 0$). Note that I remain agnostic on the possibility of false negatives (β may or may not be equal to one) although it seems reasonable to assume that those occur as well. I do assume that accurate results, like erroneous results, have a non-negligible chance of acceptance ($\beta > 0$).

Assumption 3.4 (Imperfect peer review)**.** *The peer review acceptance probabilities are such that $\alpha > 0$ and $\beta > 0$.*

I write c_e for the average amount of credit for a published erroneous result. While such results are eventually "discredited" (by my definition of erroneous), this does not necessarily equate to zero credit. Research that could not be reproduced frequently still gets cited as if it was accurate (Tatsioni et al. 2007), even after a formal retraction (Budd et al. 1998). In other cases the fact that the proposed hypothesis has fallen out of favor does not prevent it from being a credit-worthy contribution to science, e.g., Priestley's work on phlogiston. This suggests that erroneous publications are worth some credit, i.e., $c_e > 0$.[8]

Putting all of this together yields the following. The scientist works on the research project at expected speed $\lambda(p)$. The result is accurate with probability p. In this case it gets published with probability β and this publication is worth c_a units of credit. With probability $1 - p$ the result is erroneous, which leads to a publication worth c_e units of credit with probability α. Thus the scientist's expected credit per unit time, as a function of p, is given by

$$C(p) = c_a \beta p \lambda(p) + c_e \alpha (1 - p) \lambda(p).$$

To compare the individually optimal (i.e., credit-maximizing) tradeoff between speed and reproducibility to the socially optimal tradeoff, it is important to be explicit about what is meant by the *social value* of a research study. Here I have in mind the contribution that it makes to science as a

[8] On the other hand, some discredited research can actively harm a scientist's career (more so than publishing nothing at all would have done), suggesting that $c_e < 0$. These are usually cases of fraud rather than honest mistakes and so they are not my primary concern here. However, the point here is not to argue that c_e is necessarily positive, but that erroneous publications can influence a scientist's credit stock.

social enterprise, which in turn benefits society. This is reflected in the first place by the extent to which the work is utilized by other scientists, and in the second place by the extent to which it or work based on it finds its way into society, e.g., in the form of a new medicine.

What is the expected social value V of the scientist's research? I assume that research can have social value only when it is published. The probabilities of publication α and β, the reproducibility p, and the expected speed $\lambda(p)$ are all as above.[9] Hence

$$V(p) = v_a \beta p \lambda(p) + v_e \alpha (1-p) \lambda(p),$$

where v_a is the average social value of an accurate result, and v_e the average social value of an erroneous result. The social value function looks very similar to the credit function, but I argue below that there is reason to expect v_e to differ systematically from c_e.

Before stating the first result I require the following relatively modest assumption on the parameters.

Assumption 3.5 (Positive value)**.** *Accurate results have positive credit value ($c_a > 0$) and social value ($v_a > 0$).*

The first result follows from the assumptions made so far. It states that the functions C and V have unique maxima, i.e., there is a particular reproducibility that a rational credit-maximizing scientist would choose, and there is a particular reproducibility that maximizes the social value of the scientist's contribution.

Theorem 3.1 (Unique maxima)**.** *If assumptions 3.1–3.5 are satisfied, then there exist unique values $p^*_C < 1$ and $p^*_V < 1$ that maximize the functions C*

[9] Hence the social value V of the scientist's research is more precisely the scientist's own subjective estimate of the expected social value of the research (because p is a subjective probability). This may seem problematic when I use the function V below to argue that credit incentivizes scientists to make choices that are not socially optimal. I address this point below.

and V respectively, that is,

$$C(p_C^*) = \max_{p\in[0,1]} C(p) \qquad \textit{and} \qquad V(p_V^*) = \max_{p\in[0,1]} V(p).$$

Proofs for the results in this section are given in appendix A.

Note that even with these fairly minimal assumptions, it follows that $p_V^* < 1$. This means that even from the social perspective perfect reproducibility is not a goal worth striving for. Or in other words, even if the scientist was "high-minded" in the sense that she only cared about maximizing the social value of her scientific work, she should not strive to avoid error at all cost.

That $p_V^* < 1$ is a more or less direct consequence of the "no perfect work" assumption and hence reflects the insight of Lakatos and Quine that there is no certainty in science. It means that even in a science functioning perfectly, a tradeoff between speed and reproducibility must be made, and hence errors should be expected. This reflects back on the discussion of peer review: it is designed on the basic premise that there will be errors, and science must attempt to catch them as early as possible.

Other than the fact that perfect accuracy is not to be expected, however, theorem 3.1 does not say very much about the credit-maximizing reproducibility p_C^* or the social value-maximizing reproducibility p_V^*. The main result requires further assumptions on the parameter values, which I discuss next.

Credit is awarded for (accurate) scientific work proportional to its social value ($v_a = c_a$). Since, for all I have said so far, credit and social value are measured on unspecified interval scales, this may be viewed merely as fixing these scales (without loss of generality). A more substantive argument may be obtained from the literature on rewards in science. Merton enumerates the various kinds of rewards that exist in science—from the Nobel Prize to a journal publication—and concludes that "rewards are to be meted out in accord with the measure of accomplishment" (Merton 1957, p. 659). Strevens compares rewards in science to those given out in other areas and concludes that in general “society accords prestige and other rewards... in proportion to the social good resulting from [the achievement]” (Strevens 2003, p. 78).

If a more exact measure of the amount of credit awarded to a specific publication (as opposed to a scientist) is wanted, a good candidate is the number of times it is cited. But at the same time the number of citations provides a measure of the extent to which the publication has been utilized by other scientists, which I argued reflects its social value. So all three of these lines of reasoning support my assumption that $v_a = c_a$.

How about the social value of an erroneous result v_e? While errors can sometimes be instructive, I take it that the case in which they are distracting or actively misleading is more common. Take for instance a study which erroneously (in hindsight) finds that a particular medicine helps cure some disease. Perhaps the error was in the design of the study, or perhaps it was simply bad luck, i.e., the data were acquired properly but they just happened to suggest a misleading conclusion. Either way, once the conclusion that the medicine is effective is published, it takes more time and effort to set the record straight than it would have to establish that the medicine is ineffective in the absence of the erroneous publication.[10] Moreover, before the error is corrected (and perhaps after as well, see Budd et al. 1998 and Tatsioni et al. 2007) the scientific community and society will proceed as if the medicine is effective, with potentially negative consequences for future research and public health.

So it seems to me that erroneous results are, on average, at best socially neutral, if not socially harmful: $v_e \leq 0$. And I suggested above that they may still yield positive credit: $c_e > 0$. However, I need not insist on these conclusions. The weaker assumption that the social value of erroneous results is less than the credit given for them ($v_e < c_e$) suffices for my argument.

Once again reasoning in terms of citations yields a similar conclusion.

[10]Recall that I defined an erroneous result as one that is later shown to be irreproducible. Thus it is impossible by definition for an error to go uncorrected.

As mentioned above an erroneous result may still receive plenty of citations (Budd et al. 1998, Tatsioni et al. 2007). But here it does not seem so plausible that this a direct measure of its social value. Some of these citations may be actively criticizing the result. Others may be utilizing it under the assumption that it is accurate, possibly causing them to make errors in turn.

At the same time, citations to erroneous results are still worth credit, regardless of whether they are supportive or critical: they recognize the publication and its author as worth engaging with. The enormous effort undertaken by physicists to attempt to replicate Fleischmann and Pons' results is testament to Fleischmann and Pons' authority as competent electrochemists (Kitcher 1993, section 8.2). In contrast, work by subsequent cold fusion researchers has largely been ignored (Huizenga 1993, p. 208). So a citation to an erroneous publication is generally a positive marker of credit, but need not be a positive marker of social value.

Assumption 3.6 summarizes what I have argued are reasonable constraints on the parameter values that reflect the credit and social value of scientific work in typical cases (including, in particular, the case that Fleischmann and Pons found themselves in).

Assumption 3.6 (Credit and social value)**.** *Accurate results are awarded credit proportional to their social value ($c_a = v_a$), while the social value of erroneous results is less than the credit given for them ($v_e < c_e$).*

The main result of this paper can now be stated. It says that the imperfections in the peer review system and the way credit is awarded systematically favor lower levels of reproducibility. That is, a scientist who maximizes expected credit will set her reproducibility target no higher than the optimal level from the perspective of maximizing social value.

Theorem 3.2 (Rushing into print)**.** *Let assumptions 3.1–3.6 be satisfied, and define p^*_C and p^*_V as in theorem 3.1. Then $p^*_C \leq p^*_V$.*

I interpret this result as showing that, given imperfect peer review, there is a credit incentive to produce research at a systematically lower reproducibility than is socially optimal. This result depends crucially on the imperfections in the peer review system, and in particular the possibility of false positives: if $\alpha = 0$ and $\beta > 0$ then assumptions 3.1–3.3 and 3.5 are sufficient to show that the functions C and V have unique maxima, and that these maxima are equal. Intuitively, given imperfect peer review it makes sense for scientists to quickly produce lots of papers and "see what sticks" rather than spending too much time perfecting any one paper, and any resulting errors hurt society more than the scientist.

Hence I interpret theorem 3.2 as showing that imperfections in the peer review system create a systematic bias that leads credit-maximizing scientists to favor speed over reproducibility relative to the social optimum: the claim I called rushing into print in the introduction. What does this result mean for real scientists, who may care about other things than maximizing credit, and who may be less than fully rational? It means that whenever they face a research situation that satisfies the assumptions of my model (which I have argued to apply to typical cases of scientific research, including the one Fleischmann and Pons found themselves in) they either rush into print or they could have improved their expected credit if they had rushed into print. Insofar as credit acts as a selection mechanism in science this means scientists who rush into print are more likely to succeed than scientists who do not, and one should expect rushing into print to increase over time (cf. Smaldino and McElreath 2016). Thus there is a *structural* misalignment of incentives, the effect of which is to push scientists in the direction of rushing into print.

I think this misalignment is worth addressing, but one might object that there might be countervailing motivations (goals of scientists other than credit) or systematic irrationalities that make scientists choose socially optimal reproducibility levels despite my argument. It would be a surprising coincidence if other motivations or irrationalities balanced out the incentive to rush into print exactly, but I do not have an argument to rule this out. The objection does illustrate the more general point that in evaluating potential policy responses we should consider not just their effect on the issue at hand (here, the credit incentive to rush into print) but also what the potential side effects might be (here, effects on other motivations or irrationalities) and how they can be managed. This is one reason why I stop short of recommending any particular action in section 5.

I now briefly consider two objections to my interpretation of theorem 3.2. The first one points out that the theorem establishes an inequality between the credit-maximizing reproducibility and the social value-maximizing reproducibility, but not a strict inequality. So the theorem leaves open the possibility that $p_C^* = p_V^*$, the happy case in which individual and social incentives align exactly.

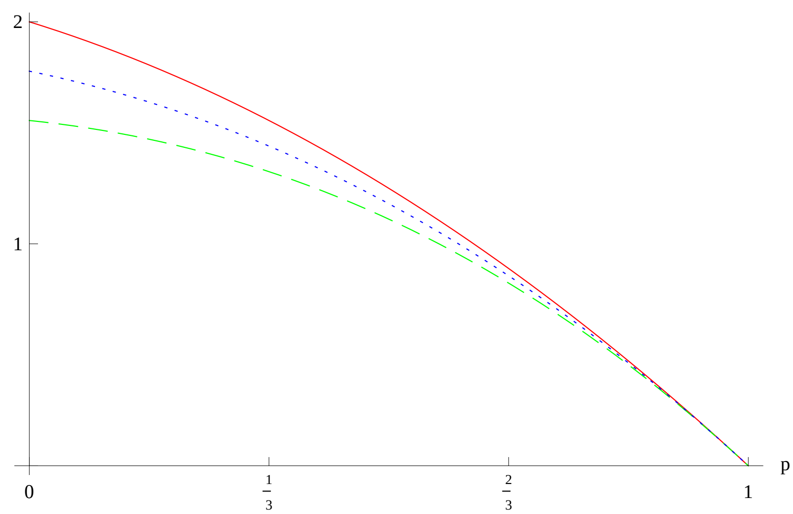

Figure 3.2: If $\lambda(p) = 2 - p - p^2$ (the solid red line) and v_e is relatively high, it may be that $p_C^* = p_V^* = 0$. In this example, the function C is shown as a dotted blue line (with $c_a\beta = 1$ and $c_e\alpha = 8/9$) and the function V is shown as a dashed green line (with $v_a\beta = 1$ and $v_e\alpha = 7/9$).

However, the happy case can only arise in one of two situations. First, if the value of erroneous results is so high that it is both individually and socially optimal to have no concern whatsoever for reproducibility ($p^*_C = p^*_V = 0$—not actually a very happy case, see figure 3.2). Second, if the speed function is not differentiable at the point of optimality (see figure 3.3).

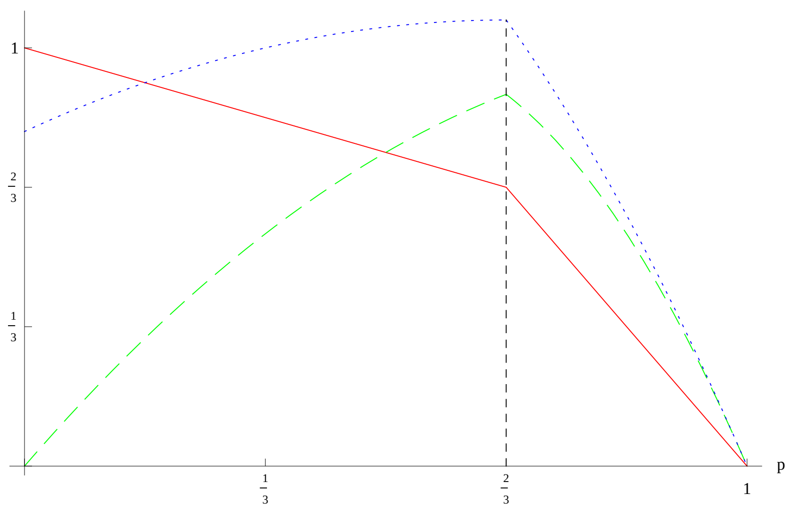

Figure 3.3: If $\lambda(p) = 1 - p/2$ for $p \leq 2/3$ and $\lambda(p) = 2(1-p)$ for $p > 2/3$ then λ (solid red) is not differentiable at $p = 2/3$. Then the functions C (dotted blue, with $c_a\beta = 2$ and $c_e\alpha = 4/5$) and V (dashed green, with $v_a\beta = 2$ and $v_e\alpha = 0$) may both be maximized there: $p^*_C = p^*_V = 2/3$.

I take these two cases to be highly exceptional. If they are ruled out, a strict inequality can be shown to hold.

Assumption 3.7 (Limited social value of errors)**.** *The social value of erroneous results (weighted by the chance of acceptance) is less than half that of accurate results:* $\alpha v_e < \beta v_a/2$.

Assumption 3.8 (The speed function is differentiable)**.** *The function* λ *is differentiable on the interior of its domain, i.e., for all* $p \in (0, 1)$.

Theorem 3.3 (Strict inequality)**.** *Let assumptions 3.1–3.8 be satisfied, and define* p_C^* *and* p_V^* *as in theorem 3.1. Then* $p_C^* < p_V^*$.

The second objection is based on the fact that in the model the reproducibility p is a subjective probability. While this is reasonable from the perspective of the scientist's choice of when to go public (this will be based on her own subjective estimate of the reproducibility of the result), it does not seem so reasonable from the perspective of assessing the social value of the scientist's contribution. When it comes to social value, what matters is the actual reproducibility of the result, not the scientist's estimate (cf. footnote 9).

I have two responses to this objection. First, I think it is reasonable to expect the scientist's estimate of reproducibility to be quite good, so that the subjective and the objective probability should be roughly equal. An important part of scientists' training, after all, involves learning how to assess evidence as objectively as possible. Some empirical support for this assertion is provided by a recent systematic study of the reproducibility of social science experiments: scientists (as a group) were remarkably accurate in predicting which experiments would be reproduced successfully and which would not (Camerer et al. 2018).

Second, if there is a discrepancy between the scientist's estimate of the reproducibility of the result and its objective reproducibility, the scientist is going to be overconfident more often than underconfident. This is also a result of scientists' training: scientists learn to view the methods they use in their research as the best ones to address the problems they work on (and/or they self-select into working with the methods they think are best).

To capture this formally, note that the scientist's choice of (subjective) reproducibility determines the objective reproducibility. So I introduce an *objective reproducibility function* o, where $o(p)$ is interpreted as the objective reproducibility that results if the scientist's choice of (subjective) reproducibility is p. Then the foregoing suggests that either $o(p) = p$ or $o(p) < p$.

I also assume that the objective reproducibility function is surjective. This means that any objective reproducibility level is achievable in the sense that there exists a subjective reproducibility level corresponding to it.

Assumption 3.9 (Confident scientist)**.** *The objective reproducibility function $o : [0, 1] \to [0, 1]$ is surjective, i.e., for all $p \in [0, 1]$ there is a $q \in [0, 1]$ such that $o(q) = p$. Moreover, $o(p) \leq p$ for all $p \in [0, 1]$.*

A credit-maximizing scientist chooses reproducibility p_C^*, the (subjective) probability that maximizes the credit function C. Social value is maximized if the scientist chooses her reproducibility p such that the objective probability $o(p)$ maximizes the social value function V. It follows from theorem 3.1 that social value is maximized if p is chosen such that $o(p) = p_V^*$. But then it follows that $p \geq p_V^*$. So by theorem 3.2 $p \geq p_C^*$, i.e., the rushing into print result extends to the case with objective reproducibility.

Corollary 3.1. *Let assumptions 3.1–3.6 and 3.9 be satisfied, and define p_C^* as in theorem 3.1. Let q_V^* be any value such that*

$$V(o(q_V^*)) = \max_{p \in [0,1]} V(o(p)).$$

Then $p_C^* \leq q_V^*$.

Corollary 3.2. *Let assumptions 3.1–3.9 be satisfied, define p_C^* as in theorem 3.1, and q_V^* as in corollary 3.1. Then $p_C^* < q_V^*$.*

4 A Tradeoff Between Speed, Reproducibility, and Impact

One feature of Fleischmann and Pons' work that presumably played a role in their decision to go public but did not appear in the model so far is the potential *impact* of their work. As the media emphasized in the days after the press conference, if cold fusion worked it held the promise of an energy revolution.

Fleischmann and Pons could perhaps be described as "impact-seekers", scientists who go in for risky research in relatively unexplored areas that promises to yield great rewards if successful (Close 1991, p. 71, describes Fleischmann in this way). In contrast, many scientists are "safety-seekers", content to make small contributions that are likely to be correct and accepted and/or can be made relatively quickly. The distinction is analogous to that between mavericks and followers (Weisberg and Muldoon 2009) or explorers and extractors (Thoma 2015) and has a long history in philosophy of science (e.g., Hull 1988, p. 474). I use different terminology to avoid the implication that this is a binary distinction rather than a graded one, or that there is necessarily a psychological explanation for it.

In this section I expand the model to include research studies with differential potential impact. The scientist now has to make a three-way tradeoff. She chooses both the reproducibility and the impact, but choosing either or both of these too highly comes at the expense of speed (compare the old business saying "You can have it good, fast, or cheap; pick two").

The first question I aim to investigate here is whether the rushing into print phenomenon also shows up in this more general model. The second question is to what extent the different "types" of scientists—impact-seekers and safety-seekers—show up in the model. More specifically, can credit incentives explain the existence of both types?[11]

In the model of this section the scientist chooses both a desired reproducibility p and a desired impact level c. Since p is interpreted as a probability, its domain is constrained to the interval $[0, 1]$. The impact c is not

[11]This question is raised by Thoma (2015, section 4.4). She points out that, from the purely epistemic perspective taken by Weisberg and Muldoon (2009), this cannot be explained: "In their model, it was unclear why anybody would choose to be a [safety-seeker], given their lack of productivity. In [Thoma's model], the question is why anybody would choose to be an [impact-seeker]" (Thoma 2015, p. 470).

similarly constrained. However, I assume that, at least for a given value of p, there is a maximum impact that can be achieved $\mu(p)$. For any admissible choice of p and c, $\lambda(p, c)$ gives the scientist's speed. The following definitions formalize this setup.

Definition 4.1. The *maximum impact function* is a function $\mu : [0, 1] \to [0, \infty)$. The *set of admissible choices* is the set $D = \{(p, c) \mid p \in [0, 1], c \in [0, \mu(p)]\}$. The *speed function* has D as its domain: it is a function $\lambda : D \to \mathbb{R}$.

I make a number of assumptions on the shape of λ. These assumptions are very similar to the ones I made before. Although they have to be adapted to the new context, their justification is as before.

First I assume that the speed function is decreasing in each of its arguments. That is, at a fixed level of reproducibility, increasing the impact decreases speed, and at a fixed level of impact, increasing reproducibility decreases speed.

Assumption 4.1 (The speed function is decreasing)**.**

4.1.a. *For all $p, p' \in [0, 1]$, if $p < p'$ and $c \le \min\{\mu(p), \mu(p')\}$, then $\lambda(p', c) < \lambda(p, c)$.*

4.1.b. *For all $p \in [0, 1)$, if $c < c' \le \mu(p)$, then $\lambda(p, c') < \lambda(p, c)$.*[12]

Assumption 4.2. *The function λ vanishes as p or c approaches the edge of its domain D.*

4.2.a. $\lim_{p \to 1} \lambda(p, 0) = 0$.

4.2.b. *For all $p \in [0, 1]$, $\lim_{c \to \mu(p)} \lambda(p, c) = 0$.*

This assumption has a role similar to assumption 3.3 (the "no perfect work" assumption). Assumption 4.2.a is in fact identical to that assumption

[12]This assumption excludes the case where $p = 1$. This is because subsequent assumptions entail that $\lambda(1, c) = 0$ for all c, which would contradict this assumption if $\mu(1) > 0$.

(although for technical reasons I only need to make the assumption for the case $c = 0$) and has the same justification. Assumption 4.2.b formalizes the idea that $\mu(p)$ represents the maximum impact that can be achieved at a given reproducibility p, by requiring that the scientist's speed becomes negligible as this value is approached.

Assumption 4.3 (The speed function is concave)**.** *For any* $(p, c), (p', c') \in D$ *and* $t \in [0, 1]$*,*

4.3.a. $(tp + (1-t)p', tc + (1-t)c') \in D$;[13]

4.3.b. $t\lambda(p, c) + (1-t)\lambda(p', c') \leq \lambda(tp + (1-t)p', tc + (1-t)c')$.

As before, this assumption says that there are decreasing marginal returns from decreasing reproducibility to gain speed. This more general version says that there are also decreasing marginal returns from decreasing the impact level, which is justified for the same reason.

What does the credit function look like in this more general setting? The main difference is that the credit for an accurate result is no longer an exogenously fixed parameter c_a, but a variable c whose value is chosen by the scientist. As for the credit for an erroneous result, there is a modeling choice to be made. Either it is a fixed absolute value, independent of the impact the result would have had if it was accurate, or it is proportional to the impact. Here I choose the latter option (although I suspect that similar results could be proven if the former option was used).

So credit for erroneous results (c_e in the previous section) is now given by $r_c c$, where r_c is a proportionality constant. The effect of making credit for erroneous results proportional is as follows. If $r_c > 0$, erroneous high-impact results get more credit than erroneous low-impact results ("at least you tried

[13] It does not follow from the definition of the domain D of λ or the assumptions made so far that $(tp + (1-t)p', tc + (1-t)c') \in D$, but this is required for the idea of a concave function to make sense, hence this assumption. It is equivalent to the assumption that μ is a concave function.

something ambitious"). If $r_c < 0$, erroneous high-impact results are penalized more harshly than erroneous low-impact results ("the bigger they are, the harder they fall"). This seems right at least for the case of Fleischmann and Pons: the amount of attention given to proving them wrong, and the effect on their personal reputations, seems to have been bigger exactly because of the potential impact their work could have had.

So the scientist's expected credit, as a function of p and c (and defined whenever $(p,c) \in D$), is

$$C(p,c) = \beta pc\lambda(p,c) + \alpha(1-p)r_c c\lambda(p,c).$$

Now consider the social value of the scientist's work. I assume that the impact level c chosen by the scientist reflects not only the potential reward but also the potential social value of the work. So the variable c replaces not only the parameter c_a but also the parameter v_a. This is equivalent to one of the assumptions made above ($c_a = v_a$) but for notational convenience I build this assumption into the definition of the function V rather than stating it separately.

As I did for the case of credit, I assume that the social value of an erroneous result is determined in proportion to the value of an accurate result, i.e., v_e is replaced by $r_v c$, where r_v is the proportionality constant for the social value of erroneous results.[14] So the social value of the scientist's research, as a function of p and c (for $(p,c) \in D$), is

$$V(p,c) = \beta pc\lambda(p,c) + \alpha(1-p)r_v c\lambda(p,c).$$

The assumption on the peer review parameters α and β is as before. I restate it here merely as a reminder.

[14]If the social value of erroneous results is usually negative, as I suggested in section 3, this means that the social cost of erroneous high-impact results is higher than the social cost of erroneous low-impact results. Again the case of Fleischmann and Pons seems to illustrate this phenomenon, with lots of resources being devoted to proving them wrong.

Assumption 4.4 (Imperfect peer review)**.** *The peer review acceptance probabilities are such that* $\alpha > 0$ *and* $\beta > 0$*.*

An assumption similar to assumption 3.5 is not needed in this version of the model, as the parameters c_a and v_a have been replaced by the variable c, which is restricted to be nonnegative by definition.

The first result says that there are unique choices of reproducibility and impact level that maximize expected credit and that maximize social value.

Theorem 4.1 (Unique maxima (redux))**.** *If assumptions 4.1–4.4 are satisfied, then there exist unique points* (p_C^*, c_C^*) *and* (p_V^*, c_V^*) *that maximize the functions* C *and* V *respectively, that is,*

$$C(p_C^*, c_C^*) = \max_{(p,c)\in D} C(p,c) \qquad \textit{and} \qquad V(p_V^*, c_V^*) = \max_{(p,c)\in D} V(p,c).$$

Moreover, $p_C^* < 1$ *and* $0 < c_C^* < \mu(p_C^*)$*; and* $p_V^* < 1$ *and* $0 < c_V^* < \mu(p_V^*)$*.*

Proofs for the results in this section are given in appendix B.

As in section 3, an additional assumption is needed to get the rushing into print result. Once again I assume that erroneous results yield less social value on average than accurate results ($r_v < r_c$). The additional assumption that the credit and social value of accurate results are equal is no longer needed because it is built into the definition of the functions C and V.

Assumption 4.5 (Credit and social value)**.** *The social value of erroneous results is less than the credit given for them:* $r_v < r_c$*.*

This yields the result that I referred to as rushing into print in section 3, namely that the credit-maximizing reproducibility p_C^* is no higher than the social value maximizing reproducibility p_V^*.

Theorem 4.2 (Rushing into print (redux))**.** *Let assumptions 4.1–4.5 be satisfied, and define* (p_C^*, c_C^*) *and* (p_V^*, c_V^*) *as in theorem 4.1. Then* $p_C^* \leq p_V^*$*.*

As before, one may ask under what circumstances the inequality $p_C^* \leq p_V^*$ can be guaranteed to be strict. This involves ruling out cases in which $p_C^* = p_V^* = 0$ and cases in which the speed function is not differentiable. For technical reasons, the former requires a slightly stronger condition than before.

Assumption 4.6 (Limited social value of errors)**.** *The social value of erroneous results (weighted by the chance of acceptance) is less than a third that of accurate results:* $\alpha r_v < \beta/3$*.*

Assumption 4.7 (The speed function is differentiable (in p))**.** *The partial derivative of the function* λ *with respect to its first argument exists on the interior of its domain, i.e.,* $\frac{\partial}{\partial p}\lambda(p, c)$ *exists whenever* $0 < p < 1$ *and* $0 < c < \mu(p)$*.*

Theorem 4.3 (Strict inequality (redux))**.** *Let assumptions 4.1–4.7 be satisfied, and define* (p_C^*, c_C^*) *and* (p_V^*, c_V^*) *as in theorem 4.1. Then* $p_C^* < p_V^*$*.*

How do these results shed light on the two questions I raised above?

First, imperfections in the peer review system give the scientist an incentive to favor speed and/or impact over reproducibility, relative to what she would do if she were trying to maximize the social value of her work. In other words there is a credit incentive to rush into print.

The incentive to rush exists in the model under essentially the same conditions as above. So the results expressed in theorems 3.2 and 3.3 are seen to be robust against the introduction of the dimension of impact.

Second, theorem 4.1 rules out the possibility that a scientist could switch from being a safety-seeker to an impact-seeker (increasing impact at the expense of reproducibility) or vice versa, while remaining at a global maximum of either C or V. For a credit-maximizing scientist, there is just one rational choice, not a range of admissible values between which an independent preference for being an impact-seeker or a safety-seeker might act as a tie-breaker. This consequence of the model may be seen as surprising in light of the way philosophers like Hull (1988), Weisberg and Muldoon (2009), and Thoma (2015) discuss these types.

This does not rule out the existence of different "types" of scientists. But it suggests that these types are the result of differences in the shape of the speed function of different scientists. If the speed function describes the tradeoff between reproducibility, impact, and speed for a given scientist, the location of the optimum given that particular speed function determines the type of scientist she will be (or at least has a credit-incentive to be). If the speed function is more or less fixed over the course of a career[15] and outside the scientist's control, theorem 4.1 can be interpreted as showing that different types of scientists are the result of differences in aptitude rather than choice. The following example illustrates this.

Example 4.1. Consider two scientists. For scientist 1, the tradeoff between reproducibility, impact, and speed is given by the speed function λ_1, where

$$\lambda_1(p, c) = -\frac{3}{4}p^4 - \frac{1}{4}p^2 - \frac{1}{2}pc - \frac{1}{4}c^2 - \frac{3}{4}c + 1,$$

for all $0 \leq p \leq 1$ and $0 \leq c \leq \frac{1}{2}(\sqrt{25 + 12p - 12p^4} - 3 - 2p)$ (see figure 4.1). Note that this function satisfies assumptions 4.1–4.7. Suppose further that $r_c = 0$.[16] Then the credit-maximizing choice for scientist 1 is $p \approx 0.52$ and $c \approx 0.38$.

In contrast, scientist 2's speed function is given by

$$\lambda_2(p, c) = -\frac{1}{4}p^2 - \frac{1}{2}pc - \frac{1}{4}c^2 - \frac{3}{4}c^4 - \frac{3}{4}p + 1,$$

for all $0 \leq c \leq 1$ and $0 \leq p \leq \frac{1}{2}(\sqrt{25 + 12c - 12c^4} - 3 - 2c)$ (see figure 4.1). This function also satisfies assumptions 4.1–4.7. But the credit-maximizing

[15]See Huber (2001, and citations therein) for evidence that the productivity of scientists is, on average, constant over the course of a career.

[16]This is a convenient value of r_c for illustrative purposes since it entails that I need to make no specific assumption on the values of α and β: the maximum of C will not depend on this as long as $\beta > 0$.

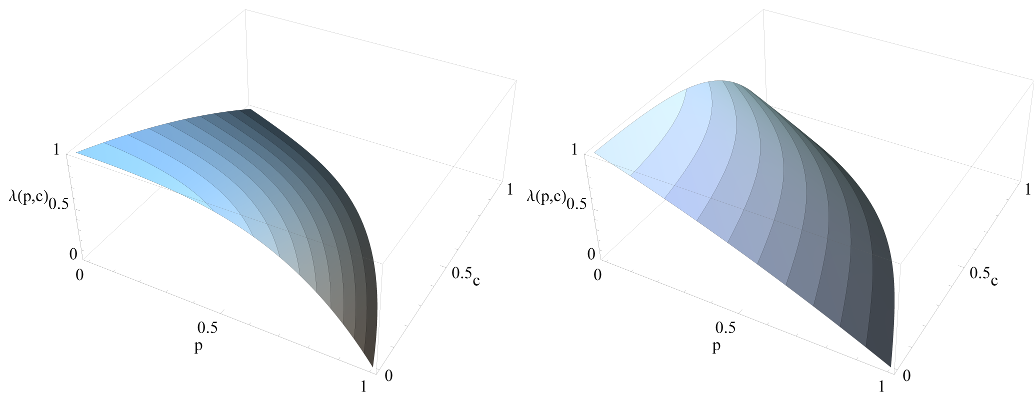

Figure 4.1: Graphs of λ_1 (on the left) and λ_2 (on the right).

choice for scientist 2 (assuming $r_c = 0$) is $p \approx 0.38$ and $c \approx 0.52$.

Scientist 1's speed function shows an aptitude for reproducibility compared to scientist 2's, which shows an aptitude for impact. This is because λ_1 is closer to linear in c—having only a small quadratic component—while it is a fourth-degree polynomial in p (λ_2 is simply its mirror image). So if the scientists are responsive to credit incentives, scientist 1 will behave more like a safety-seeker, doing relatively safe, low-impact research. Scientist 2 on the other hand will behave more like an impact-seeker, doing more risky, high-impact research.

In general, given that the speed function is concave and decreasing, the less linear it is in one of its variables the higher the optimal value for that variable will be. So a function that is more linear in c and less linear in p produces safety-seekers, and the reverse impact-seekers.

Moreover, if scientists are credit-maximizers, and assumptions 4.1, 4.2, 4.3, and 4.4 are justified, then theorem 4.1 guarantees that differences in the shape of the speed function are the *only* way different types of scientists can arise.

This is potentially a problem. Weisberg and Muldoon (2009) and Thoma (2015) have investigated whether there is an epistemically optimal distribution of safety-seekers and impact-seekers in a scientific community. Thoma

(2015, section 4.4) points out that credit may play a crucial role in motivating scientists to distribute themselves over the types. But if differences in aptitude are required for credit to play this role, there is no reason to expect the resulting distribution of safety-seekers and impact-seekers to be anywhere close to optimal.

5 Conclusion

The following five conclusions can be drawn from the work presented in this paper. First, I have argued that under a wide range of plausible conditions scientists have a credit incentive to publish work that is unlikely to be successfully reproduced (relative to the socially optimal reproducibility level). Three key ingredients are responsible for this misalignment of incentives: the tradeoff between speed and reproducibility (and impact), the fact that scientists are rewarded for publications, and imperfections in the peer review system.

This misalignment hurts science and society: by definition, any deviation from the social optimum hurts the progress of science and the social benefits of that progress. More specifically, I have shown how credit incentives may contribute to the reproducibility problems that have recently attracted significant attention. In particular, I have argued that credit incentives may do so even in the absence of some of the particular phenomena that have previously been identified as culprits (e.g., publication bias, novelty bias, and checking bias).

What can be done about this? One solution is to eliminate imperfections in the peer review system. Without those imperfections credit incentives are perfectly aligned with the social optimum in my model. But this is a lot to ask: it requires reviewers at scientific journals to be perfect predictors of whether future work will successfully replicate a result.

However, I noted that the misalignment of incentives in the model is exclusively caused by false positives (accepting erroneous results for publication). So reducing those can bring the credit-maximizing optimum closer to the social optimum.[17] This seems to recommend conservative editorial practices: rejecting papers even based on fairly minimal doubts about their reproducibility. But if reducing false positives leads to more false negatives (rejecting accurate results) the effect will be that the maximum social value is itself lowered, even if the credit-maximizing optimum is brought closer to it. Investigating this further tradeoff is beyond the scope of this paper.

A different way to eliminate imperfections in the peer review system would be to get rid of peer review (and perhaps even scientific journals) altogether, possibly replacing it with post-publication peer review. But even such a drastic rethinking of the way scientific research is disseminated would not avoid this problem. The problem arises because scientific work needs to be evaluated in some way or other in the short run, before it is known whether it will in fact be reproduced (e.g., scientists need to decide what to read and what to cite). Hence, while I have focused discussion on imperfections in the peer review system, the existence of peer review in its current form is not essential to the incentive to rush into print.

Another solution focuses on the amount of credit given for irreproducible results. I referred repeatedly to Budd et al. (1998) and Tatsioni et al. (2007), who showed that scientists continue to give credit (in the form of citations) to research that has been refuted. If the credit given to irreproducible results matched the social value of those results more closely, the gap between the credit-maximizing optimum and the social optimum would be reduced. It would help if there was a broader general awareness of which research has been refuted, but this may be hard to achieve in practice. More specifically, one might aim to make hiring and promotion committees more aware of candidates' refuted results.

[17]More specifically, it can be shown (under the assumptions of theorems 3.3 and 4.3) that reducing the value of α reduces the difference between p_C^* and p_V^*. In the limiting case where $\alpha = 0$ they are equal.

A third solution aims to somehow compensate for the misalignment. For example, Nelson et al. (2012) have suggested limiting the number of papers scientists may publish per unit time. This would create an incentive to favor reproducibility over speed that could in principle balance out the incentive to rush. But this suggestion comes with its own problems. The limit on the number of papers would have to be just right to balance out the incentive to favor speed over reproducibility without overshooting the optimum in the other direction, needlessly harming the timely publication of accurate results. This problem is exacerbated by the fact that different scientists may have different speed functions, which may require different publication limits to create the best incentive structure.

In this paper I have focused on rushing into print, without denying that publication bias, novelty bias, and checking bias may also capture important ways in which scientists are incentivized to do research that is less than optimally reproducible. But whereas these biases wear their corresponding solutions on their sleeve (scientists should be rewarded for negative results and replications), the above discussion suggests that the solution to rushing into print is much less clear, if one exists at all.

The second conclusion is that certainty of reproducibility is neither to be expected nor to be desired. The reason for this is that if scientists were too demanding in perfecting their research before publishing it, nothing would ever get published. The point is hardly new (it goes back at least to Lakatos and Quine), but since philosophers of science and epistemologists have said a lot about error avoidance but relatively little about how to achieve this in a reasonable time frame (cf. Friedman 1979, Heesen 2015), it is worth emphasizing.

The third conclusion concerns next steps in applying this work. One may ask whether there is a way to validate the model. In particular one may want to calibrate the parameter values, for example to see whether there is reason to be worried about rushing into print in a particular case. A good starting point for this kind of work may be in medicine. Here we find relatively well-defined problems (particular diseases) with well-defined solutions (particular treatments). Moreover, measures of the social value of some treatment (e.g., survival rates or recovery rates) can be separated relatively cleanly from measures of credit (e.g., citations, prestigious publications, or prizes).

Fourth, I considered the difference between scientists who pursue high-impact research that is risky with regard to reproducibility and/or speed ("impact-seekers") and scientists who pursue more mundane research relatively likely to be reproducible and/or fast ("safety-seekers"). My model suggests that the existence of these types of scientists reflects a difference in aptitude rather than a preference for certain kinds of research.

Considering the tradeoff between speed, reproducibility, and impact explicitly shows that high-impact research (or "transformative" research in modern terms) is likely to be less reproducible. Example 4.1 illustrates this. Thus it is perhaps unreasonable to hold impact-seekers to the same standards of evidence as safety-seekers. In this way my model justifies to some extent the practice at institutions like the NSF and the NIH to consider a grant proposal's "potential to be transformative" separately from its likelihood to succeed. By considering the criteria separately, these institutions aim to prevent biasing their evaluation process for or against impact-seekers or safety-seekers.

Finally, the work in this paper suggests a reevaluation of Fleischmann and Pons' decision to go public with their work on cold fusion. That decision has been much maligned for being premature. The rejection of cold fusion by the scientific establishment and the subsequent decline of cold fusion research would seem to vindicate the judgment of prematurity. But Fleischmann and Pons could not know this at the time. The question is whether their decision was irrational, given the information available to them.

Two of the above conclusions suggest that it may not have been. First, imperfections in the peer review system may make it rational for a credit-maximizing scientist to submit work the reproducibility of which is not yet firmly established. Second, scientists who are pursuing high-impact research should be given more leeway to produce work that is relatively less likely to be reproduced.

Fleischmann and Pons were well aware of the uncertainties surrounding cold fusion at the time they went public. They also knew that if they did not go public, the risk of being scooped was high. The above considerations suggest (without proving of course) that under these circumstances it may well have been rational to go public despite the uncertainties.

Fleischmann and Pons went out on a limb, as every scientist does when she publishes her work. On this occasion, they got burned. But I submit that this was not primarily the result of poor judgment, although it may be easy to come to the opposite conclusion with the benefit of hindsight. Rather, they did exactly what other scientists have done on countless occasions: they weighed the risk of going public against the potential reward. That they are now maligned rather than celebrated is largely the result of bad luck.[18]

A The Tradeoff Between Speed and Reproducibility

Assumptions and theorems whose labels start with a 3 are restated from section 3. Assumptions, lemmas, and theorems labeled with an "A" are original to this appendix.

I begin by briefly restating the key features of the model. Define the

[18]Thanks to Kevin Zollman, Michael Strevens, Stephan Hartmann, Teddy Seidenfeld, Jan Sprenger, Liam Bright, Cailin O'Connor, Seamus Bradley, Conor Mayo-Wilson, Adrian Currie, Shahar Avin, Rory Švarc, and audiences at Tilburg University, the National University of Singapore, the Congress of Logic, Methodology and Philosophy of Science in Helsinki, the Formal Epistemology Workshop in Groningen, and Risk and the Culture of Science in Cambridge 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 and the Isaac Newton Trust.

following functions for all $p \in [0, 1]$:

$$\begin{aligned} C(p) &= c_a \beta p \lambda(p) + c_e \alpha (1-p) \lambda(p), \\ V(p) &= v_a \beta p \lambda(p) + v_e \alpha (1-p) \lambda(p), \\ c_{a,e}(p) &= a p \lambda(p) + e(1-p)\lambda(p). \end{aligned}$$

The function C reflects the scientist's expected credit as a function of p: she aims to choose a value of p that maximizes C. The function V reflects the expected social value of the scientist's work. The family of functions $c_{a,e}$ is defined in such a way that the functions C and V are members of the family. This family captures the common structure of C and V, which will be useful in the analysis below.

The variable p is the desired level of reproducibility, which can range from zero to one, and is chosen by the scientist. The parameters have the following interpretations: c_a is the expected credit to the scientist for an accurate paper, c_e is the expected credit for an erroneous paper, v_a is the expected social value of an accurate paper, v_e is the expected social value of an erroneous paper, β is the probability that an accurate paper passes peer review, and α is the probability that an erroneous paper passes peer review.

The function $\lambda : [0, 1] \to [0, \infty)$ reflects the tradeoff between speed and reproducibility: $\lambda(p)$ is the speed at which the scientist works given that the desired reproducibility is p, and $\gamma(p) = 1/\lambda(p)$ is the expected completion time of the project. In the paper I made the following assumptions about the shape of this function.

Assumption 3.1 (The speed function is decreasing)**.** *For all $p, q \in [0, 1]$, if $p < q$, then $\lambda(q) < \lambda(p)$.*

Assumption 3.2 (The speed function is concave)**.** *For every $p, q, t \in [0, 1]$,*

$$t\lambda(p) + (1-t)\lambda(q) \leq \lambda(tp + (1-t)q).$$

Assumption 3.3 (No perfect work)**.** $\lim_{p\to 1} \lambda(p) = 0$.

This assumption asserts that the scientist cannot deliver perfect work (in the sense of zero probability of errors), no matter how slowly she works. She can, however, get arbitrarily close: due to assumption 3.1, $\lambda(p) > 0$ for all $p < 1$.

It is worth noting that assumptions 3.1–3.3 entail the following related characteristics of the expected completion time function γ.

Lemma A.1. *If assumptions 3.1–3.3 are satisfied, the function γ is increasing and convex. Moreover, $\lim_{p\to 1} \gamma(p) = \infty$.*

Proof. Since $\lambda(p) > 0$ for all $p < 1$, $\lambda(q) < \lambda(p)$ if and only if $\gamma(q) > \gamma(p)$. Since λ gets arbitrarily small (while remaining positive) as $p \to 1$, γ gets arbitrarily large as $p \to 1$. Since λ is strictly positive and concave on $[0, 1)$, γ is convex on $[0, 1)$ (see Kantrowitz and Neumann 2005, proposition 2 for a proof). □

Assumption 3.1 entails that λ is bounded, as $\lambda(p) \leq \lambda(0) < \infty$ for all $p \in [0, 1]$. Assumption 3.2 entails that λ is continuous on $(0, 1)$ (but not necessarily at the endpoints). Combined with the other two assumptions, however, it follows that λ is continuous on its entire domain, as the following lemma shows.

Lemma A.2. *If assumptions 3.1–3.3 are satisfied, λ is continuous on $[0, 1]$.*

Proof. Due to assumption 3.2, λ is continuous on $(0, 1)$. By assumption 3.3, $\lim_{p\to 1} \lambda(p) = 0$. By definition this means that for every $\varepsilon > 0$ there exists a choice of $p < 1$ such that $\lambda(p) < \varepsilon$. It follows that $\lambda(1) = 0$ (if $\lambda(1) > 0$ then one could choose $\varepsilon = \lambda(1)$ to find a p such that $\lambda(p) < \varepsilon = \lambda(1)$, contradicting assumption 3.1). So $\lim_{p\to 1} \lambda(p) = 0 = \lambda(1)$, i.e., λ is continuous at $p = 1$.

It remains to show that λ is continuous at $p = 0$. Because λ is monotone and bounded, $\ell = \lim_{p\to 0} \lambda(p)$ exists. Because λ is decreasing, $\ell \leq \lambda(0)$. Suppose for reductio that $\ell < \lambda(0)$. By definition of the limit there exists a choice of $p > 0$ small enough such that $\ell - \lambda(p) < (\lambda(0) - \ell)/2$. Then $\lambda(0)/2+\lambda(p)/2 > \ell > \lambda(p/2)$, contradicting assumption 3.2. Hence $\ell = \lambda(0)$, i.e., λ is continuous at $p = 0$. $\square$

The next lemma depends on a result by Kantrowitz and Neumann (2005), which I state first.

Theorem A.1 (Kantrowitz and Neumann 2005)**.** *Let the functions $f_1, \ldots, f_n$ be concave, non-negative, and not identically equal to zero on the closed bounded interval $[a, b]$. Then the product $h := f_1 \cdots f_n$ has the following properties:*

(i) *$h(x) > 0$ for all $x \in (a, b)$;*

(ii) *there exist numbers α and β with $a \leq \alpha \leq \beta \leq b$ for which h is strictly increasing on $[a, \alpha)$, constant on (α, β), and strictly decreasing on $(\beta, b]$;*

(iii) *if one of the functions $f_1, \ldots, f_n$ has at most one global maximum point in $[a, b]$, then so does h;*

(iv) *the product h is constant on $[a, b]$ if and only if each of the functions $f_1, \ldots, f_n$ is constant on $[a, b]$;*

(v) *if $f_1, \ldots, f_n$ are differentiable at a point $x \in (a, b)$, then $h'(x) > 0$ if $x \in (a, \alpha)$, while $h'(x) < 0$ if $x \in (\beta, b)$.*

Lemma A.3. *If assumptions 3.1–3.3 are satisfied, there exists $p_{a,e}$ such that $c_{a,e}(p_{a,e}) = \max_{p\in[0,1]} c_{a,e}(p)$. Moreover, if $a > 0$, then $p_{a,e} < 1$ uniquely maximizes $c_{a,e}$.*

Proof. By lemma A.2, λ is continuous on $[0, 1]$. It follows that $c_{a,e}$ is continuous. By the extreme value theorem, $c_{a,e}$ attains its maximum, i.e., there exists $p_{a,e} \in [0, 1]$ such that $c_{a,e}(p_{a,e}) = \max_{p\in[0,1]} c_{a,e}(p)$.

Note that $a > 0$ implies that there is at least some value of p for which $c_{a,e}(p) > 0$: if $e \geq 0$ this is true for all $p \in (0,1)$ and if $e < 0$ this is true because

$$\frac{a-2e}{2(a-e)} \in (0,1), \qquad \text{and}$$
$$c_{a,e}\left(\frac{a-2e}{2(a-e)}\right) = \frac{1}{2}a\lambda\left(\frac{a-2e}{2(a-e)}\right) > 0.$$

It follows that $c_{a,e}(p_{a,e}) > 0$. Since $\lambda(1) = 0$, $c_{a,e}(1) = 0 < c_{a,e}(p_{a,e})$. So $p_{a,e} \neq 1$.

To see that $a > 0$ implies uniqueness of the maximum, write $c_{a,e}$ as the product of two concave functions:

$$c_{a,e}(p) = (e + (a-e)p)\lambda(p),$$

where λ is concave by assumption 3.2 and $e + (a-e)p$ is concave because it is linear. Uniqueness of the maximum is established by checking that the conditions of theorem A.1.iii are satisfied.

As a result of assumptions 3.1 and 3.3 the function λ is nonnegative and not identically zero on $[0,1]$. If $a > 0$ and $e \geq 0$ then the function $e+(a-e)p$ is also nonnegative and not identically zero on $[0,1]$. If on the other hand $a > 0$ and $e < 0$ then $e+(a-e)p$ is only nonnegative whenever $p \geq \frac{-e}{a-e}$, where $\frac{-e}{a-e} < 1$. So I restrict attention to the nonempty interval $[\max\{0, \frac{-e}{a-e}\}, 1]$. In this interval both λ and $e + (a-e)p$ are nonnegative and not identically zero.

Moreover, on the interval $[\max\{0, \frac{-e}{a-e}\}, 1]$ the function λ has a unique maximum at $\max\{0, \frac{-e}{a-e}\}$. So by theorem A.1.iii, the function $c_{a,e}$ has a unique maximum on the interval $[\max\{0, \frac{-e}{a-e}\}, 1]$. Since $c_{a,e} < 0$ whenever $0 \leq p < \frac{-e}{a-e}$, it follows that $c_{a,e}$ has a unique maximum on $[0,1]$. □

Theorem A.2. *If assumptions 3.1–3.3 are satisfied, and $c_a\beta > 0$ and $v_a\beta > 0$, then there exist unique values $p_C < 1$ and $p_V < 1$ that maximize the* *functions C and V respectively, that is,*

$$C(p_C) = \max_{p\in[0,1]} C(p) \qquad \textit{and} \qquad V(p_V) = \max_{p\in[0,1]} V(p).$$

Proof. Note that C and V are special cases of $c_{a,e}$, with $C = c_{c_a\beta, c_e\alpha}$ and $V = c_{v_a\beta, v_e\alpha}$. Because $c_a\beta > 0$ and $v_a\beta > 0$ the conditions of lemma A.3 apply to C and V. The result follows immediately. □

The above result is a somewhat more general version of theorem 3.1 in the paper. To see this, first recall the following assumptions.

Assumption 3.4 (Imperfect peer review)**.** *The peer review acceptance probabilities are such that $\alpha > 0$ and $\beta > 0$.*

Assumption 3.5 (Positive value)**.** *Accurate results have positive credit value ($c_a > 0$) and social value ($v_a > 0$).*

Theorem 3.1 now follows as a corollary of theorem A.2.

Theorem 3.1 (Unique maxima)**.** *If assumptions 3.1–3.5 are satisfied, then there exist unique values $p_C < 1$ and $p_V < 1$ that maximize the functions C and V respectively, that is,*

$$C(p_C) = \max_{p\in[0,1]} C(p) \qquad \textit{and} \qquad V(p_V) = \max_{p\in[0,1]} V(p).$$

Proof. From $\beta > 0$, $c_a > 0$, and $v_a > 0$, it follows that $c_a\beta > 0$ and $v_a\beta > 0$. Hence theorem A.2 applies. □

Having established the existence of a unique maximum for each of the functions C, V, and $c_{a,e}$, I now prove a number of lemmas that are instrumental in establishing the main rushing into print result.

Lemma A.4. *If assumptions 3.1–3.3 are satisfied, $a > 0$ and $a > 2e$, then $p_{a,e} > 0$ (where $p_{a,e}$ uniquely maximizes $c_{a,e}$).*

Proof. Since $a > 0$, the function $c_{a,e}$ has a unique maximum at $p_{a,e}$ by lemma A.3. It follows from $a > 2e$ that $\frac{a-2e}{2(a-e)} \in (0,1)$. Using the definition of concavity (see assumption 3.2) with $t = \frac{a-2e}{2(a-e)}$, $p = 1$, and $q = 0$ yields:

$$\lambda\left(\frac{a-2e}{2(a-e)}\right) \geq \frac{a}{2(a-e)}\lambda(0) + \frac{a-2e}{2(a-e)}\lambda(1) = \frac{a}{2(a-e)}\lambda(0),$$

where the equality uses the fact that $\lambda(1) = 0$. Hence,

$$c_{a,e}\left(\frac{a-2e}{2(a-e)}\right) = \frac{1}{2}a\lambda\left(\frac{a-2e}{2(a-e)}\right) \geq \frac{a^2}{4(a-e)}\lambda(0).$$

Now note that $(a-2e)^2 > 0$ and therefore $a^2 > 4e(a-e)$. Since $a - e > 0$,

$$c_{a,e}\left(\frac{a-2e}{2(a-e)}\right) \geq \frac{a^2}{4(a-e)}\lambda(0) > e\lambda(0) = c_{a,e}(0).$$

This shows that $c_{a,e}$ is not maximized at $p = 0$. □

Lemma A.5. *If assumptions 3.1–3.3 are satisfied, and $a > 0$, the function $c_{a,0}$ (that is, $c_{a,e}$ with $e = 0$) is uniquely maximized at $p^* \in (0,1)$, where the value of p^* does not depend on the value of a, and $c_{a,0}$ is increasing on $[0, p^*]$ and decreasing on $[p^*, 1]$.*

Proof. The conditions of this lemma entail that the conditions of lemmas A.3, and A.4 are satisfied. Hence there exists $p^* \in (0,1)$ that uniquely maximizes $c_{a,0}$. Because $c_{a,0}(p) = ap\lambda(p)$, a is merely a scaling constant, so the maximum is unchanged when a changes.

It remains to show that $c_{a,0}$ is monotonically increasing on $[0, p^*]$ and decreasing on $[p^*, 1]$. Note (as in the proof of lemma A.3) that $c_{a,0}$ can be written as the product of two concave functions (ap and λ) that are nonnegative and not identically zero on $[0,1]$. So by theorem A.1.ii, there exist x_1 and x_2 such that $c_{a,e}$ is increasing on $[0, x_1)$, constant on (x_1, x_2), and decreasing on $(x_2, 1]$. Since the maximum is unique, it follows that $x_1 = x_2 = p^*$. $\square$

Lemma A.6. *If assumptions 3.1–3.3 are satisfied and moreover $a > 0$ and $e \geq 0$, then $p_{a,e} \leq p^*$ (where $p_{a,e}$ uniquely maximizes $c_{a,e}$ and p^* is as defined in lemma A.5).*

Proof. The existence and uniqueness of $p_{a,e}$ and p^* follow from lemmas A.3 and A.5 respectively. Suppose for reductio that $p_{a,e} > p^*$. By definition p^* maximizes the function $c_{a,0}$ (see lemma A.5), so $ap^*\lambda(p^*) > ap_{a,e}\lambda(p_{a,e})$. Since λ is decreasing (hence $\lambda(p^*) \geq \lambda(p_{a,e})$) and $e \geq 0$ it also follows that $e(1-p^*)\lambda(p^*) \geq e(1-p_{a,e})\lambda(p_{a,e})$. So

$$\begin{aligned} c_{a,e}(p^*) &= ap^*\lambda(p^*) + e(1-p^*)\lambda(p^*) \\ &> ap_{a,e}\lambda(p_{a,e}) + e(1-p_{a,e})\lambda(p_{a,e}) = c_{a,e}(p_{a,e}). \end{aligned}$$

But $p_{a,e}$ maximizes $c_{a,e}$ by definition, which entails $c_{a,e}(p_{a,e}) \geq c_{a,e}(p^*)$. Contradiction. So $p_{a,e} \leq p^*$. $\square$

Lemma A.7. *If assumptions 3.1–3.3 are satisfied and moreover $a > 0$ and $e \leq 0$, then $p_{a,e} \geq p^*$ (where $p_{a,e}$ uniquely maximizes $c_{a,e}$ and p^* is as defined in lemma A.5).*

Proof. The existence and uniqueness of $p_{a,e}$ and p^* follow from lemmas A.3 and A.5 respectively. Suppose for reductio that $p_{a,e} < p^*$. By definition p^* maximizes the function $c_{a,0}$ (see lemma A.5), so $ap^*\lambda(p^*) > ap_{a,e}\lambda(p_{a,e})$. Since λ is decreasing (hence $\lambda(p^*) \leq \lambda(p_{a,e})$) and $e \leq 0$ it also follows that $e(1-p^*)\lambda(p^*) \geq e(1-p_{a,e})\lambda(p_{a,e})$. So

$$\begin{aligned} c_{a,e}(p^*) &= ap^*\lambda(p^*) + e(1-p^*)\lambda(p^*) \\ &> ap_{a,e}\lambda(p_{a,e}) + e(1-p_{a,e})\lambda(p_{a,e}) = c_{a,e}(p_{a,e}), \end{aligned}$$

But $p_{a,e}$ maximizes $c_{a,e}$ by definition, which entails $c_{a,e}(p_{a,e}) \geq c_{a,e}(p^*)$. Contradiction. So $p_{a,e} \geq p^*$. $\square$

Assumption A.1. *One of the following conditions holds: either*

$$v_a\beta \geq c_a\beta > 0 \qquad \textit{and} \qquad c_e\alpha \geq 0 \qquad \textit{and} \qquad c_e\alpha \geq v_e\alpha, \tag{A.1}$$

or

$$c_a\beta \geq v_a\beta > 0 \qquad \textit{and} \qquad c_e\alpha \leq 0 \qquad \textit{and} \qquad c_e\alpha \geq v_e\alpha. \tag{A.2}$$

Theorem A.3. *Let assumptions 3.1–3.3 and A.1 be satisfied. Define p_C and p_V as in theorem A.2. Then $p_C \leq p_V$.*

Proof. Both sets of conditions imply that $c_a\beta > 0$ and $v_a\beta > 0$, so theorem A.2 applies. Define p_C and p_V to be those choices of p that uniquely maximize the functions C and V respectively.

Assumption A.1 specifies two sets of conditions. I prove the result for the two sets separately. Consider the first set of conditions, as specified in equation (A.1). Because $c_a\beta > 0$ and $c_e\alpha \geq 0$, lemma A.6 applies, so $p_C \leq p^*$, where p^* is as defined in lemma A.5. Suppose for reductio that $p_V < p_C$. Since p_C uniquely maximizes the function C, $C(p_C) > C(p_V)$. Since $0 \leq p_V < p_C \leq p^*$ and since the function $p\lambda(p)$ is increasing on the interval $[0, p^*]$ by lemma A.5, $p_C\lambda(p_C) > p_V\lambda(p_V)$. Finally, since λ is decreasing by assumption 3.1, $(1 - p_C)\lambda(p_C) < (1 - p_V)\lambda(p_V)$. Putting this all together yields

$$\begin{aligned}
V(p_C) &= C(p_C) + (v_a\beta - c_a\beta)p_C\lambda(p_C) + (v_e\alpha - c_e\alpha)(1 - p_C)\lambda(p_C) \\
&> C(p_V) + (v_a\beta - c_a\beta)p_C\lambda(p_C) + (v_e\alpha - c_e\alpha)(1 - p_C)\lambda(p_C) \\
&\geq C(p_V) + (v_a\beta - c_a\beta)p_V\lambda(p_V) + (v_e\alpha - c_e\alpha)(1 - p_V)\lambda(p_V) \\
&= V(p_V).
\end{aligned}$$

But this contradicts the supposition that p_V maximizes V. So $p_C \leq p_V$.

Now consider the second set of conditions, as specified in equation (A.2). Because $v_a\beta > 0$ and $v_e\alpha \leq 0$, lemma A.7 applies, so $p_V \geq p^*$. Suppose for reductio that $p_V < p_C$. Then $C(p_C) > C(p_V)$ because p_C uniquely maximizes C. Since $p^* \leq p_V < p_C \leq 1$ and $p\lambda(p)$ is decreasing on the interval $[p^*, 1]$ by lemma A.5, $p_C\lambda(p_C) < p_V\lambda(p_V)$. And, since λ is decreasing by assumption 3.1, $(1 - p_C)\lambda(p_C) < (1 - p_V)\lambda(p_V)$. So

$$\begin{aligned} V(p_C) &= C(p_C) + (v_a\beta - c_a\beta)p_C\lambda(p_C) + (v_e\alpha - c_e\alpha)(1 - p_C)\lambda(p_C) \\ &> C(p_V) + (v_a\beta - c_a\beta)p_V\lambda(p_V) + (v_e\alpha - c_e\alpha)(1 - p_V)\lambda(p_V) \\ &= V(p_V). \end{aligned}$$

But this contradicts the supposition that p_V maximizes V. So $p_C \leq p_V$. □

The above theorem implies the main result from the paper (in particular, assumption A.1 is strictly weaker than assumptions 3.4–3.6). Recall assumption 3.6.

Assumption 3.6 (Credit and social value)**.** *Accurate results are awarded credit proportional to their social value ($c_a = v_a$), while the social value of erroneous results is less than the credit given for them ($v_e < c_e$).*

Theorem 3.2 (Rushing into print)**.** *Let assumptions 3.1–3.6 be satisfied, and define p_C and p_V as in theorem 3.1. Then $p_C \leq p_V$.*

Proof. Assumptions 3.4–3.6 imply that $c_a\beta = v_a\beta > 0$ and that $c_e\alpha \geq v_e\alpha$. Thus, if $c_e\alpha \geq 0$ the set of conditions (A.1) is satisfied, and if $c_e\alpha \leq 0$ the set of conditions (A.2) is satisfied. So assumptions 3.4–3.6 imply that assumption A.1 holds, which means that theorem A.3 applies. □

In order to get strict inequality some additional assumptions are needed.

Assumption 3.7 (Limited social value of errors)**.** *The social value of erroneous results (weighted by the chance of acceptance) is less than half that of accurate results: $\alpha v_e < \beta v_a/2$.*

This assumption guarantees that the function V is not maximized at zero (using lemma A.4).

Assumption 3.8 (The speed function is differentiable)**.** *The function λ is differentiable on the interior of its domain, i.e., for all $p \in (0,1)$.*

With these additional assumption the previous result can be strengthened to a strict inequality.

Theorem A.4. *Let assumptions 3.1–3.3, 3.7–3.8, and A.1 be satisfied, and assume moreover that $v_e < c_e$. Define p_C and p_V as in theorem A.2. Then $p_C < p_V$.*

Proof. Since the assumptions of theorem A.3 are satisfied, $p_C \le p_V$. Because $v_a\beta > 2v_e\alpha$ (by assumption 3.7) and $v_a\beta > 0$ (by assumption A.1), the conditions of lemmas A.3 and A.4 are satisfied, so $0 < p_V < 1$. If $p_C = 0$ this completes the proof, so assume that $p_C > 0$. Then $0 < p_C \le p_V < 1$ so the maximum of C is achieved in the interior of its domain. Because λ is differentiable on $(0,1)$, C and V are differentiable on $(0,1)$. In particular, since p_C maximizes C, $C'(p_C) = 0$. To show that $p_C \neq p_V$ it suffices to show that $V'(p_C) \neq 0$.

Consider first the case where $v_a\beta \ge c_a\beta$ and $c_e\alpha \ge 0$ (condition (A.1) of assumption A.1). Because $c_e\alpha \ge 0$, lemma A.6 applies, so $p_C \le p^*$, where p^* is as defined in lemma A.5. Because λ is differentiable, the derivative of $p\lambda(p)$ exists and is given by $p\lambda'(p) + \lambda(p)$. By lemma A.5, $p\lambda(p)$ is increasing on $[0, p^*]$, which means its derivative is nonnegative, so in particular $p_C\lambda'(p_C) + \lambda(p_C) \ge 0$. By assumption 3.1, $\lambda'(p_C) < 0$. Putting all of this together yields

$$\begin{aligned} V'(p_C) = C'(p_C) &+ (v_a\beta - c_a\beta)(p_C\lambda'(p_C) + \lambda(p_C)) \\ &+ (v_e\alpha - c_e\alpha)(1 - p_C)\lambda'(p_C) - (v_e\alpha - c_e\alpha)\lambda(p_C) \\ &\ge (v_e\alpha - c_e\alpha)(1 - p_C)\lambda'(p_C) - (v_e\alpha - c_e\alpha)\lambda(p_C) > 0. \end{aligned}$$

Now consider the case where $c_a\beta \ge v_a\beta$ and $c_e\alpha \le 0$. Because $c_e\alpha \le 0$, lemma A.7 applies, so $p_C \ge p^*$. By lemma A.5, $p\lambda(p)$ is decreasing on $[p^*, 1]$, so $p_C\lambda'(p_C) + \lambda(p_C) \leq 0$. Hence

$$\begin{aligned} V'(p_C) &= C'(p_C) + (v_a\beta - c_a\beta)(p_C\lambda'(p_C) + \lambda(p_C)) \\ &\qquad + (v_e\alpha - c_e\alpha)(1-p_C)\lambda'(p_C) - (v_e\alpha - c_e\alpha)\lambda(p_C) \\ &\geq (v_e\alpha - c_e\alpha)(1-p_C)\lambda'(p_C) - (v_e\alpha - c_e\alpha)\lambda(p_C) > 0. \end{aligned} \qquad \square$$

This yields the result from the main text.

Theorem 3.3 (Strict inequality)**.** *Let assumptions 3.1–3.8 be satisfied, and define p_C and p_V as in theorem 3.1. Then $p_C < p_V$.*

Proof. As in the proof of theorem 3.2, note that assumptions 3.4–3.6 imply that $c_a\beta = v_a\beta > 0$ and that $c_e\alpha \geq v_e\alpha$. Thus, if $c_e\alpha \geq 0$ the set of conditions (A.1) is satisfied, and if $c_e\alpha \leq 0$ the set of conditions (A.2) is satisfied. So assumptions 3.4–3.6 imply that assumption A.1 holds.

Moreover, $v_e < c_e$ holds by assumption 3.6. Hence all assumptions of theorem A.4 hold, so $p_C < p_V$. $\square$

Next I consider a version of the model in which the scientist's estimated reproducibility level p (a subjective probability) is replaced by the actual (objective) reproducibility level. To capture this formally, note that the scientist's choice of (subjective) reproducibility determines the objective reproducibility. So I introduce an objective reproducibility function o, where $o(p)$ is interpreted as the objective reproducibility that results if the scientist's choice of (subjective) reproducibility is p.

As I argued in the main text, it seems reasonable to assume that the scientist is usually pretty accurate in her estimations of reproducibility ($o(p) \approx p$) and that when she errs she is more likely to be overconfident than underconfident ($o(p) < p$). So I assume that $o(p) \leq p$, which should capture a large range of cases.

I also assume that the objective reproducibility function is surjective. This means that any objective reproducibility level is achievable in the sense that there exists a subjective reproducibility level corresponding to it.

Assumption 3.9 (Confident scientist)**.** *The objective reproducibility function $o : [0,1] \to [0,1]$ is surjective, i.e., for all $p \in [0,1]$ there is a $q \in [0,1]$ such that $o(q) = p$. Moreover, $o(p) \leq p$ for all $p \in [0,1]$.*

A credit-maximizing scientist chooses reproducibility p_C, the (subjective) probability that maximizes the credit function C. Social value is maximized if the scientist chooses her reproducibility such that the objective probability maximizes the social value function V (i.e., $V(o(\cdot))$ is maximized). The main results can now be restated for this expanded version of the model.

Theorem A.5. *Let assumptions 3.1–3.3, A.1, and 3.9 be satisfied. Define p_C as in theorem A.2. Let $q_V \in [0,1]$ be any value such that*

$$V(o(q_V)) = \max_{q \in [0,1]} V(o(q)).$$

Then $p_C \leq q_V$.

Proof. Because the assumptions of theorem A.3 are satisfied, there exist unique values p_C and p_V that maximize the functions C and V respectively, with $p_C \leq p_V$. Let $q_V \in [0,1]$ be any value that maximizes $V(o(\cdot))$. Given that $V(p_V) = \max_{p \in [0,1]} V(p)$ and given that the objective reproducibility function o is surjective, it follows that $V(o(q_V)) = V(p_V)$. Since p_V uniquely maximizes V it follows further that $o(q_V) = p_V$. By assumption 3.9 $o(q_V) \leq q_V$. So $p_C \leq p_V = o(q_V) \leq q_V$. □

Corollary 3.1. *Let assumptions 3.1–3.6 and 3.9 be satisfied, and define p_C as in theorem 3.1. Let q_V be any value such that*

$$V(o(q_V)) = \max_{p \in [0,1]} V(o(p)).$$

Then $p_C \leq q_V$.

Proof. As noted in multiple previous proofs assumptions 3.4–3.6 imply assumption A.1. The result follows immediately. □

Theorem A.6. *Let assumptions 3.1–3.3, 3.7–3.9, and A.1 be satisfied, and assume moreover that $v_e < c_e$. Define p_C as in theorem A.2 and q_V as in theorem A.5. Then $p_C < q_V$.*

Proof. Because the assumptions of theorem A.4 are satisfied, there exist unique values p_C and p_V that maximize the functions C and V respectively, with $p_C < p_V$. Let $q_V \in [0, 1]$ be any value that maximizes $V(o(\cdot))$. By the same reasoning as above, $p_C < p_V = o(q_V) \leq q_V$. □

Corollary 3.2. *Let assumptions 3.1–3.9 be satisfied, define p_C as in theorem 3.1, and q_V as in corollary 3.1. Then $p_C < q_V$.*

Proof. As above. □

B The Tradeoff Between Speed, Reproducibility, and Impact

In this appendix I investigate a model in which there is a three-way tradeoff between speed, reproducibility, and impact. The scientist chooses the minimal acceptable reproducibility p, and the level of impact she wishes to achieve c (equated with the amount of credit she will be given if she is successful), and her speed λ is determined as a function of p and c. Assumptions and theorems whose labels start with a 4 are restated from section 4.

As before, p is interpreted as a probability, so its domain is naturally constrained to the interval $[0, 1]$. The impact or credit c is not similarly constrained. However, I assume that, at least for a given reproducibility p, there is a maximum impact that can be achieved. The following definitions formalize this setup.

Definition B.1. Let $\alpha, \beta \in [0, 1]$ and $r_c, r_v \in \mathbb{R}$ be fixed parameters.

B.1.a. The *maximum impact function* is a function $\mu : [0,1] \to [0,\infty)$.

B.1.b. The *domain (of admissible choices)* is the set $D = \{(p,c) \mid p \in [0,1], c \in [0,\mu(p)]\}$.

B.1.c. The *speed function* is a function $\lambda : D \to [0,\infty)$.

B.1.d. The *credit function* is the function $C : D \to \mathbb{R}$ given by

$$C(p,c) = \beta pc\lambda(p,c) + \alpha(1-p)r_c c\lambda(p,c)$$

for all $(p,c) \in D$.

B.1.e. The *(social) value function* is the function $V : D \to \mathbb{R}$ given by

$$V(p,c) = \beta pc\lambda(p,c) + \alpha(1-p)r_v c\lambda(p,c)$$

for all $(p,c) \in D$.

I make a number of assumptions on the shape of λ. These assumptions are very similar to the ones I made before, although they have to be adapted to the new three-dimensional context.

Assumption 4.1 (The speed function is decreasing)**.**

4.1.a. For all $p, p' \in [0,1]$, if $p < p'$ and $c \le \min\{\mu(p), \mu(p')\}$, then $\lambda(p',c) < \lambda(p,c)$.

4.1.b. For all $p \in [0,1)$, if $c < c' \le \mu(p)$, then $\lambda(p,c') < \lambda(p,c)$.

Note that assumption 4.1.b excludes the case where $p = 1$. This is because assumption 4.2 below entails that $\lambda(1,c) = 0$ for all c, which is not decreasing if $\mu(1) > 0$.

Lemma B.1. *If assumption 4.1 is satisfied,*

$$\lambda(p,c) \le \lambda(p,0) \le \lambda(0,0) < \infty,$$

for any $(p, c) \in D$.

Proof. The first inequality follows from assumption 4.1.b and the second from assumption 4.1.a. $\square$

The next assumption has a role similar to assumption 3.3. It requires that as the scientist gets close to perfect reproducibility ($p \to 1$), her speed vanishes ($\lambda \to 0$). This is required only when $c = 0$ (but see lemma B.6).

Additionally, the assumption requires (for fixed reproducibility) that as the scientist gets close to maximum impact ($c \to \mu(p)$), her speed vanishes ($\lambda \to 0$). This formalizes the intended interpretation of μ as the maximum impact that can be achieved at a given level of reproducibility.

Assumption 4.2. *The function λ vanishes as p or c approaches the edge of its domain D.*

4.2.a. $\lim_{p \to 1} \lambda(p, 0) = 0$.

4.2.b. For all $p \in [0, 1]$, $\lim_{c \to \mu(p)} \lambda(p, c) = 0$.

Lemma B.2. *If assumptions 4.1.b and 4.2.b are satisfied, $\lambda(p, \mu(p)) = 0$ for all $p \in [0, 1)$.*

Proof. Let $p \in [0, 1)$ and $\varepsilon > 0$. By assumption 4.2.b, there exists a $c < \mu(p)$ such that $\lambda(p, c) < \varepsilon$. By assumption 4.1.b and nonnegativity of λ, $0 \leq \lambda(p, \mu(p)) < \lambda(p, c) < \varepsilon$. So $\lambda(p, \mu(p)) = 0$. $\square$

Lemma B.3. *If assumptions 4.1 and 4.2.b are satisfied, μ is decreasing on $[0, 1)$.*

Proof. Let $p < p' < 1$ and suppose for reductio that $\mu(p) \leq \mu(p')$. Note that it follows that $\mu(p) \leq \min\{\mu(p), \mu(p')\}$. By lemma B.2, $\lambda(p, \mu(p)) = \lambda(p', \mu(p')) = 0$. But by assumption 4.1,

$$\lambda(p, \mu(p)) > \lambda(p', \mu(p)) \geq \lambda(p', \mu(p')).$$

Contradiction. So $\mu(p') < \mu(p)$. $\square$

Lemma B.4. *If assumptions 4.1 and 4.2.b are satisfied, μ is bounded from above on $[0, 1)$ by $\mu(0) < \infty$.*

Proof. Immediate from lemma B.3. □

Lemma B.5. *If assumptions 4.1 and 4.2.b are satisfied, $\lambda(p, c) \leq \lambda(0, c) < \infty$ for any $(p, c) \in D$.*

Proof. Let $(p, c) \in D$. By definition $c \leq \mu(p)$. By lemma B.4 $\mu(p) \leq \mu(0)$. So $c \leq \min\{\mu(p), \mu(0)\}$. So by assumption 4.1.a $\lambda(p, c) \leq \lambda(0, c)$. □

Lemma B.6. *If assumptions 4.1 and 4.2 are satisfied, $\lim_{p\to 1} \mu(p)$ exists and $\lim_{p\to 1} \lambda(p, c) = 0$ for all $c \in [0, \lim_{p\to 1} \mu(p)]$.*

Proof. By lemma B.3, μ is decreasing, and by definition, μ is bounded from below (by 0). Hence $\lim_{p\to 1} \mu(p)$ exists.

Let $0 \leq c \leq \lim_{p\to 1} \mu(p)$ and $\varepsilon > 0$. By assumption 4.2.a, there exists $p' < 1$ such that $\lambda(p, 0) < \varepsilon$ for all $p \in (p', 1)$. By assumption 4.1.b and nonnegativity of λ, $0 \leq \lambda(p, c) \leq \lambda(p, 0) < \varepsilon$. So $\lim_{p\to 1} \lambda(p, c) = 0$. □

Lemma B.7. *If assumptions 4.1 and 4.2 are satisfied, $\lambda(1, c) = 0$ for all $c \leq \min\{\mu(1), \lim_{p\to 1} \mu(p)\}$.*

Proof. Let $c \leq \min\{\mu(1), \lim_{p\to 1} \mu(p)\}$ and $\varepsilon > 0$. By lemma B.6, there exists $p < 1$ such that $\lambda(p, c) < \varepsilon$. By assumption 4.1.a and nonnegativity of λ, $0 \leq \lambda(1, c) < \lambda(p, c) < \varepsilon$. So $\lambda(1, c) = 0$. □

Assumption 4.3 (The speed function is concave)**.** *For any $(p, c), (p', c') \in D$ and $t \in [0, 1]$,*

4.3.a. $(tp + (1 - t)p', tc + (1 - t)c') \in D$;

4.3.b. $t\lambda(p, c) + (1 - t)\lambda(p', c') \leq \lambda(tp + (1 - t)p', tc + (1 - t)c')$.

It does not follow from the definition of the domain D of λ or the assumptions made so far that $(tp + (1 - t)p', tc + (1 - t)c') \in D$, but this is required for the idea of a concave function to make sense, hence assumption 4.3.a. The next lemma characterizes the meaning of assumption 4.3.a.

Lemma B.8. *The following are equivalent:*

1. *Assumption 4.3.a;*

2. *The set D is convex;*

3. *The function μ is concave.*

Proof. Note first that $tp + (1-t)p' \in [0,1]$ for all $p, p', t \in [0,1]$.

$1 \Rightarrow 2$: A set is convex if every convex combination of two points in the set is itself in the set. Assumption 4.3.a requires exactly that.

$2 \Rightarrow 3$: Let $p, p', t \in [0,1]$. By definition, $(p, \mu(p)), (p', \mu(p')) \in D$. By 2, it follows that $(tp + (1-t)p', t\mu(p) + (1-t)\mu(p')) \in D$. So by definition of D, $t\mu(p) + (1-t)\mu(p') \leq \mu(tp + (1-t)p')$. So μ is concave.

$3 \Rightarrow 1$: Let $(p, c), (p', c') \in D$ and $t \in [0,1]$. Then

$$tc + (1-t)c' \leq t\mu(p) + (1-t)\mu(p') \leq \mu(tp + (1-t)p'),$$

where the latter inequality follows from the concavity of μ. So $(tp + (1-t)p', tc + (1-t)c') \in D$. $\square$

Corollary B.1. *If assumptions 4.1, 4.2 and 4.3.a are satisfied, $\mu(1) \leq \lim_{p\to 1} \mu(p)$, and hence lemmas B.3 and B.4 apply also when $p = 1$.*

Proof. By lemma B.6, $\lim_{p\to 1} \mu(p)$ exists. By lemma B.8, μ is concave. $\mu(1) \leq \lim_{p\to 1} \mu(p)$ follows from this. $\square$

Lemma B.9. *If assumptions 4.1–4.3 are satisfied, then for all $p \in [0,1]$ $\lim_{c\to 0} \lambda(p, c) = \lambda(p, 0)$.*

Proof. Let $p \in [0,1]$. First consider the case $p = 1$. $\lambda(1, c) = 0$ for all $c \leq \mu(1)$ by lemma B.7, so $\lim_{c\to 0} \lambda(1, c) = 0 = \lambda(1, 0)$.

If $p < 1$, $\lim_{c\to 0} \lambda(p, c)$ exists because λ is decreasing in c (assumption 4.1.b) and bounded from above (lemma B.1). Lemma B.1 implies that $\lim_{c\to 0} \lambda(p, c) \leq \lambda(p, 0)$. This inequality cannot be strict because λ is concave (cf. the proof of lemma A.2). $\square$

Lemma B.10. *If assumptions 4.1, 4.2.b and 4.3 are satisfied, then for all* $c \in [0, \mu(0))$ $\lim_{p \to 0} \lambda(p, c) = \lambda(0, c)$.

Proof. Let $c \in [0, \mu(0))$. $\lim_{p \to 0} \lambda(p, c)$ exists because λ is decreasing in p (assumption 4.1.a) and bounded from above (lemma B.5). Lemma B.5 implies that $\lim_{p \to 0} \lambda(p, c) \leq \lambda(0, c)$. This inequality cannot be strict because λ is concave (cf. the proof of lemma A.2). □

Lemma B.11. *If assumptions 4.1–4.3 are satisfied,* λ *is continuous.*

Proof. Because λ is concave, it is continuous at any interior point of its domain. It remains to show that λ is continuous on the borders, that is at those points (p, c) with $p = 0$, $p = 1$, $c = 0$, or $c = \mu(p)$. I give a proof for the case $c = 0$ (the other cases are similar).

Note first that λ is continuous when one of its arguments is held fixed: the function $\lambda(\cdot, 0)$ (the restriction of λ along the p-axis) is continuous due to concavity (at least for non-extreme values of p), and the function $\lambda(p, \cdot)$ is continuous for any fixed value of p: for $0 < c < \mu(p)$ this follows from concavity, for $c = 0$ this follows from lemma B.9, and for $c = \mu(p)$ this follows from assumption 4.2.b and lemma B.2.

Let $p \in (0, 1)$ and $\varepsilon > 0$ (I set aside the cases where $p = 0$ or $p = 1$ to avoid having to worry about certain technicalities, but essentially the same proof works for those cases too.). By the foregoing there exists $\delta_1 > 0$ such that $\lambda(p', 0) < \lambda(p, 0) + \varepsilon$ for every $p' \in (p - \delta_1, p + \delta_1)$. By assumption 4.1.a, $\lambda(p', 0) < \lambda(p, 0) + \varepsilon$ for every $p' > p - \delta_1$. Similarly, there exists $\delta_2 > 0$ such that $\lambda(p, c') > \lambda(p, 0) - \varepsilon/2$ for every $c' < \delta_2$. And, given the particular value of δ_2 just chosen, there exists $\delta_3 > 0$ such that $\lambda(p', \delta_2) > \lambda(p, \delta_2) - \varepsilon/2 > \lambda(p, 0) - \varepsilon$ for every $p' < p + \delta_3$.

Choose $\delta = \min\{\delta_1, \delta_2, \delta_3\}$. Let $(p', c') \in D$ be such that $0 < \|(p', c') - (p, 0)\| < \delta$. It follows that $p - \delta_1 < p' < p + \delta_3$ and $0 \leq c' < \delta_2$. Hence, using assumption 4.1.b and the facts established in the previous paragraph,

$$\lambda(p, 0) - \varepsilon < \lambda(p', \delta_2) < \lambda(p', c') \leq \lambda(p', 0) < \lambda(p, 0) + \varepsilon.$$

So $|\lambda(p', c') - \lambda(p, 0)| < \varepsilon$. So λ is continuous at $(p, 0)$. □

Recall from topology the notion of the *interior* of a set. The interior of a set A, written $\operatorname{int} A$ is the set of all points $x \in A$ such that x is contained in an open subset of A. Any point $x \in A$ that is not in the interior of A is called a *boundary point* of A. The interior of the domain D is the set $\operatorname{int} D = \{(p, c) \mid p \in (0, 1), c \in (0, \mu(p))\}$.

Lemma B.12. *If assumptions 4.1–4.3 are satisfied, $\beta > 0$ and $\alpha r_c > 0$, there is a unique point $(p_C, c_C) \in D$ such that*

$$C(p_C, c_C) = \max_{(p,c) \in D} C(p, c).$$

Moreover, either $(p_C, c_C) \in \operatorname{int} D$ or $p_C = 0$.

Proof. Because λ is continuous (by lemma B.11), C is continuous as well. By the extreme value theorem, there exists a point $(p_C, c_C) \in D$ such that $C(p_C, c_C) = \max_{(p,c) \in D} C(p, c)$ (uniqueness will be shown below).

Note that $C(p, c) > 0$ for all $(p, c) \in \operatorname{int} D$. Conversely, $C(p, c) = 0$ if either $c = 0$, $c = \mu(p)$ (by lemma B.2), or $p = 1$ (by lemma B.7). Hence, either $(p_C, c_C) \in \operatorname{int} D$ or $p_C = 0$.

Let $(p', c') \neq (p_C, c_C)$ be any point in D. To show uniqueness of the maximum, it suffices to show that (p', c') does not maximize C.

Let $f : [0, 1] \to [0, \infty)$ be the function defined by

$$\begin{aligned}
f(t) &= C\left(tp_C + (1 - t)p', tc_C + (1 - t)c'\right) \\
&= \left(\alpha r_c + (\beta - \alpha r_c)\left(tp_C + (1 - t)p'\right)\right) \\
&\qquad \cdot \left(tc_C + (1 - t)c'\right) \lambda\left(tp_C + (1 - t)p', tc_C + (1 - t)c'\right)
\end{aligned}$$

for all $t \in [0, 1]$. Because C is maximized at (p_C, c_C), f is maximized at $t = 1$.

Note that f can be written as the product of three concave and nonnegative functions: λ is a concave function of t as a consequence of assumption 4.3, and $\alpha r_c+(\beta-\alpha r_c)(tp_C+(1-t)p')$ and $tc_C+(1-t)c'$ are linear functions of t and hence also concave. Moreover, since either $p_C \neq p'$ or $c_C \neq c'$, at least one of the functions $\alpha r_c + (\beta - \alpha r_c)(tp_C + (1-t)p')$ and $tc_C + (1-t)c'$ has a unique maximum (e.g., if $c_C > c'$, $tc_C + (1-t)c'$ is maximized at $t = 1$). Finally, none of the three functions are identically zero on $[0,1]$. So it follows from theorem A.1.iii that f has a unique maximum at $t = 1$. Hence $C(p', c') = f(0) < f(1) = C(p_C, c_C)$. □

Lemma B.13. *If assumptions 4.1–4.3 are satisfied and $\alpha r_c > \beta > 0$, then the unique point $(p_C, c_C) \in D$ that maximizes the function C satisfies $p_C = 0$.*

Proof. The assumptions of this lemma are a special case of the assumptions of lemma B.12, so there exists a unique point $(p_C, c_C) \in D$ that maximizes the function C, and either $(p_C, c_C) \in \operatorname{int} D$ or $p_C = 0$. Suppose for reductio that $(p_C, c_C) \in \operatorname{int} D$, i.e., $0 < p_C < 1$ and $0 < c_C < \mu(p_C) < \mu(0)$. Then

$$C(p_C, c_C) = (\alpha r_c + (\beta - \alpha r_c)p_C)c_C\lambda(p_C, c_C).$$

By assumption 4.1.a, $\lambda(p_C, c_C) < \lambda(0, c_C)$ and by the assumptions of this lemma, $\alpha r_c + (\beta - \alpha r_c)p_C < \alpha r_c$. Hence

$$C(p_C, c_C) < \alpha r_c c_C \lambda(0, c_C) = C(0, c_C),$$

which contradicts the supposition that $(p_C, c_C) \in \operatorname{int} D$ maximizes the function C. So $p_C = 0$. □

Lemma B.14. *If assumptions 4.1–4.3 are satisfied, $\beta > 0$ and $\alpha r_c \leq 0$, there is a unique point $(p_C, c_C) \in \operatorname{int} D$ such that*

$$C(p_C, c_C) = \max_{(p,c)\in D} C(p, c).$$

Proof. Because λ is continuous (by lemma B.11), C is continuous as well. By the extreme value theorem, there exists a point $(p_C, c_C) \in D$ such that $C(p_C, c_C) = \max_{(p,c)\in D} C(p, c)$ (uniqueness will be shown below).

Let $D_C^+ = \left\{(p, c) \in D \mid p \geq \frac{-\alpha r_c}{\beta - \alpha r_c}\right\}$. Because $\beta > 0$ and $\alpha r_c \leq 0$, $0 \leq \frac{-\alpha r_c}{\beta - \alpha r_c} < 1$. Hence $\operatorname{int} D_C^+ = \left\{(p, c) \mid \frac{-\alpha r_c}{\beta - \alpha r_c} < p < 1, 0 < c < \mu(p)\right\}$ is nonempty.

As the name suggests, the significance of D_C^+ is that it denotes the part of the domain where C is nonnegative. More precisely, $C(p, c) > 0$ if $(p, c) \in \operatorname{int} D_C^+$, $C(p, c) = 0$ if (p, c) is a boundary point of D_C^+, and $C(p, c) < 0$ if $(p, c) \in D \setminus D_C^+$. It follows that $(p_C, c_C) \in \operatorname{int} D_C^+$ and (since $\operatorname{int} D_C^+ \subseteq \operatorname{int} D$) that $(p_C, c_C) \in \operatorname{int} D$.

Let $(p', c') \neq (p_C, c_C)$ be any point in D. To show uniqueness of the maximum, it suffices to show that (p', c') does not maximize C. If $(p', c') \notin \operatorname{int} D_C^+$ then the proof is done because $C(p', c') \leq 0 < C(p_C, c_C)$. So suppose $(p', c') \in \operatorname{int} D_C^+$.

Let $f : [0, 1] \to [0, \infty)$ be the function defined by

$$\begin{aligned} f(t) &= C\left(tp_C + (1-t)p', tc_C + (1-t)c'\right) \\ &= \left(\alpha r_c + (\beta - \alpha r_c)\left(tp_C + (1-t)p'\right)\right) \\ &\qquad \cdot \left(tc_C + (1-t)c'\right) \lambda\left(tp_C + (1-t)p', tc_C + (1-t)c'\right) \end{aligned}$$

for all $t \in [0, 1]$. Because C is maximized at (p_C, c_C), f is maximized at $t = 1$.

Note that f can be written as the product of three concave and nonnegative functions: λ is a concave function of t as a consequence of assumption 4.3, and $\alpha r_c + (\beta - \alpha r_c)(tp_C + (1-t)p')$ and $tc_C + (1-t)c'$ are linear functions of t and hence also concave. Moreover, since either $p_C \neq p'$ or $c_C \neq c'$, at least one of the functions $\alpha r_c + (\beta - \alpha r_c)(tp_C + (1-t)p')$ and $tc_C + (1-t)c'$ has a unique maximum. Finally, none of the three functions are identically zero on $[0, 1]$. So it follows from theorem A.1.iii that f has a unique maximum at $t = 1$. Hence $C(p', c') = f(0) < f(1) = C(p_C, c_C)$. $\square$

Combining these lemmas yields the first theorem for this model.

Assumption 4.4 (Imperfect peer review)**.** *The peer review acceptance prob-*

abilities are such that $\alpha > 0$ and $\beta > 0$.

Theorem 4.1 (Unique maxima (redux))**.** *If assumptions 4.1–4.4 are satisfied, then there exist unique points (p_C, c_C) and (p_V, c_V) that maximize the functions C and V respectively, that is,*

$$C(p_C, c_C) = \max_{(p,c)\in D} C(p, c) \qquad \textit{and} \qquad V(p_V, c_V) = \max_{(p,c)\in D} V(p, c).$$

Moreover, $p_C < 1$ and $0 < c_C < \mu(p_C)$; and $p_V < 1$ and $0 < c_V < \mu(p_V)$.

Proof. For the function C, the result follows immediately from lemma B.12 if $r_c > 0$ (and hence $\alpha r_c > 0$) and from lemma B.14 if $r_c \leq 0$ (and hence $\alpha r_c \leq 0$).

Since the function V is identical to the function C except that r_c is replaced with r_v, the result for V similarly follows from lemma B.12 if $r_v > 0$ and from lemma B.14 if $r_v \leq 0$. □

Theorem B.1. *Let assumptions 4.1–4.4 be satisfied. Assume also that $r_v \leq r_c$. Define (p_C, c_C) and (p_V, c_V) as in theorem 4.1. Then either $p_C < p_V$ or $(p_C, c_C) = (p_V, c_V)$.*

Proof. Suppose for reductio that $p_C \geq p_V$ and that $(p_C, c_C) \neq (p_V, c_V)$ (that is, either $p_C \neq p_V$ or $c_C \neq c_V$). Because (p_C, c_C) and (p_V, c_V) are distinct, and (p_V, c_V) uniquely maximizes V,

$$\begin{aligned} V(p_V, c_V) &= (\alpha r_v + (\beta - \alpha r_v)p_V)c_V\lambda(p_V, c_V) \\ &> (\alpha r_v + (\beta - \alpha r_v)p_C)c_C\lambda(p_C, c_C) = V(p_C, c_C). \end{aligned}$$

I claim that it follows that $c_V\lambda(p_V, c_V) > c_C\lambda(p_C, c_C)$. To see this, it is useful to distinguish two cases.

If $\beta \geq \alpha r_v$ then the supposition that $p_C \geq p_V$ yields $\alpha r_v + (\beta - \alpha r_v)p_C \geq \alpha r_v + (\beta - \alpha r_v)p_V$ (moreover, the latter is positive since $V(p_V, c_V)$ is positive). But then the only way $V(p_V, c_V) > V(p_C, c_C)$ can be true is if $c_V\lambda(p_V, c_V) > c_C\lambda(p_C, c_C)$.

If, on the other hand, $\beta < \alpha r_v$ it follows that $\beta < \alpha r_c$. So $p_C = 0$ by lemma B.13 and hence also $p_V = 0$ (because $p_C \geq p_V$). But then the inequality established above reduces to $\alpha r_v c_V \lambda(p_V, c_V) > \alpha r_v c_C \lambda(p_C, c_C)$ which (given that $\alpha r_v > \beta > 0$) is equivalent to $c_V \lambda(p_V, c_V) > c_C \lambda(p_C, c_C)$.

Combining the fact $V(p_V, c_V) > V(p_C, c_C)$, $c_V \lambda(p_V, c_V) > c_C \lambda(p_C, c_C)$, $p_C \geq p_V$ and $r_c \geq r_v$ yields

$$\begin{aligned} C(p_V, c_V) &= V(p_V, c_V) + \alpha(r_c - r_v)(1 - p_V) c_V \lambda(p_V, c_V) \\ &> V(p_C, c_C) + \alpha(r_c - r_v)(1 - p_C) c_C \lambda(p_C, c_C) = C(p_C, c_C). \end{aligned}$$

But this contradicts the fact that (p_C, c_C) maximizes C. So the supposition is false, which means that either $(p_C, c_C) = (p_V, c_V)$ or $p_C < p_V$. □

Assumption 4.5 (Credit and social value)**.** *The social value of erroneous results is less than the credit given for them:* $r_v < r_c$.

Theorem 4.2 (Rushing into print (redux))**.** *Let assumptions 4.1–4.5 be satisfied, and define* (p_C, c_C) *and* (p_V, c_V) *as in theorem 4.1. Then* $p_C \leq p_V$.

Proof. Since assumption 4.5 implies that $r_v \leq r_c$, the assumptions of theorem B.1 are satisfied. Hence either $p_C < p_V$ or $(p_C, c_C) = (p_V, c_V)$. But this clearly implies $p_C \leq p_V$. □

In order to rule out the case that $(p_C, c_C) = (p_V, c_V)$ (and thus conclude that $p_C < p_V$) two additional assumptions are needed. First, in order to rule out that $p_C = p_V = 0$ I assume that the social value of erroneous results is significantly lower than the social value of accurate results. Whereas before a factor two sufficed ($\alpha v_e < \beta v_a / 2$), here a factor three is required.

Assumption 4.6 (Limited social value of errors)**.** *The social value of erroneous results (weighted by the chance of acceptance) is less than a third that of accurate results:* $\alpha r_v < \beta / 3$.

Lemma B.15. *Let assumptions 4.1–4.4, and 4.6 be satisfied. Define* (p_V, c_V) *as in theorem 4.1. Then* $(p_V, c_V) \in \operatorname{int} D$.

Proof. By theorem 4.1 either $(p_V, c_V) \in \operatorname{int} D$ or $p_V = 0$. So it suffices to show that $p_V \neq 0$.

Suppose for reductio that $(0, c)$ maximizes V for some $0 < c < \mu(0)$ and that $\alpha r_v > 0$ (the case $\alpha r_v \leq 0$ is handled by lemma B.14). By assumption 4.6 $\beta > 3\alpha r_v$ and so $t \in (0, \frac{1}{4})$, where t is defined by

$$t = \frac{\beta - 3\alpha r_v}{4(\beta - \alpha r_v)} \qquad \text{and hence} \qquad 1 - t = \frac{3\beta - \alpha r_v}{4(\beta - \alpha r_v)}.$$

Using assumption 4.3 and $\lambda(1, 0) = 0$ (by lemma B.7),

$$(1 - t)\lambda(0, c) = (1 - t)\lambda(0, c) + t\lambda(1, 0) \leq \lambda(t, (1 - t)c).$$

But then

$$\begin{aligned} V(t, (1-t)c) &= \left(\beta t(1-t)c + \alpha r_v (1-t)^2\right) \lambda(t, (1-t)c) \\ &\geq (1-t)^2 \left(\beta t c + \alpha r_v (1-t) c\right) \lambda(0, c) \\ &> \alpha r_v c \lambda(0, c) = V(0, c), \end{aligned}$$

where the second inequality follows because t was chosen such that

$$(1-t)^2 \left(\beta t + \alpha r_v (1-t)\right) - \alpha r_v = \frac{(\beta - 3\alpha r_v)^2 (9\beta - 7\alpha r_v)}{64(\beta - \alpha r_v)^2} > 0.$$

Since $V(t, (1-t)c) > V(0, c)$, V does not have a maximum at $(0, c)$. Contradiction. $\square$

Second, I assume that the function λ is differentiable in its first argument on the interior of its domain.

Assumption 4.7 (The speed function is differentiable (in p)). *The partial derivative of the function λ with respect to its first argument exists on the interior of its domain, i.e., $\frac{\partial}{\partial p}\lambda(p, c)$ exists whenever $0 < p < 1$ and $0 < c < \mu(p)$.*

Theorem 4.3 (Strict inequality (redux))**.** *Let assumptions 4.1–4.7 be satisfied, and define* (p_C, c_C) *and* (p_V, c_V) *as in theorem 4.1. Then* $p_C < p_V$*.*

Proof. Since all the conditions of theorem B.1 are satisfied, either $p_C < p_V$ or $(p_C, c_C) = (p_V, c_V)$. So it suffices to show that $(p_C, c_C) \neq (p_V, c_V)$.

Since the conditions of lemma B.15 are also satisfied, $(p_V, c_V) \in \operatorname{int} D$. If $(p_C, c_C) \notin \operatorname{int} D$ the proof is finished, so suppose $(p_C, c_C) \in \operatorname{int} D$.

Assumption 4.7 then entails that the partial derivatives of C and V with respect to their first argument exist at (p_C, c_C). Since (p_C, c_C) is an extremum of C achieved in the interior of its domain, $\frac{\partial}{\partial p} C(p_C, c_C) = 0$. By assumption 4.1.a, $\frac{\partial}{\partial p} \lambda(p_C, c_C) < 0$. Hence

$$\begin{aligned}\frac{\partial}{\partial p} V(p_C, c_C) = \frac{\partial}{\partial p} C(p_C, c_C) &+ (\alpha r_c - \alpha r_v) c_C \lambda(p_C, c_C) \\ &+ (\alpha r_v - \alpha r_c)(1 - p_C) c_C \frac{\partial}{\partial p} \lambda(p_C, c_C) > 0.\end{aligned}$$

So (p_C, c_C) does not maximize V. So $(p_C, c_C) \neq (p_V, c_V)$. □

References

C. Glenn Begley and Lee M. Ellis. Raise standards for preclinical cancer research. *Nature*, 483(7391):531–533, Mar 2012. doi: 10.1038/483531a. URL `http://dx.doi.org/10.1038/483531a`.

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-Kassem and Cyrille Imbert. Scientific collaboration: Do two heads need to be more than twice better than one? *Philosophy of Sci-*

ence, 82(4):667–688, 2015. ISSN 00318248. URL `http://www.jstor.org/stable/10.1086/682940`.

Liam Kofi Bright. On fraud. *Philosophical Studies*, 174(2):291–310, 2017a. ISSN 1573-0883. doi: 10.1007/s11098-016-0682-7. URL `http://dx.doi.org/10.1007/s11098-016-0682-7`.

Liam Kofi Bright. Decision theoretic model of the productivity gap. *Erkenntnis*, 82(2):421–442, 2017b. ISSN 1572-8420. doi: 10.1007/s10670-016-9826-6. URL `http://dx.doi.org/10.1007/s10670-016-9826-6`.

Justin P. Bruner. Policing epistemic communities. *Episteme*, 10(4):403–416, Dec 2013. ISSN 1750-0117. doi: 10.1017/epi.2013.34. URL `http://dx.doi.org/10.1017/epi.2013.34`.

John M. Budd, MaryEllen Sievert, and Tom R. Schultz. Phenomena of retraction: Reasons for retraction and citations to the publications. *Journal of the American Medical Association*, 280(3):296–297, 1998. doi: 10.1001/jama.280.3.296. URL `http://dx.doi.org/10.1001/jama.280.3.296`.

Colin F. Camerer, Anna Dreber, Felix Holzmeister, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Gideon Nave, Brian A. Nosek, Thomas Pfeiffer, Adam Altmejd, Nick Buttrick, Taizan Chan, Yiling Chen, Eskil Forsell, Anup Gampa, Emma Heikenstein, Lily Hummer, Taisuke Imai, Siri Isaksson, Dylan Manfredi, Julia Rose, Eric-Jan Wagenmakers, and Hang Wu. Evaluating the replicability of social science experiments in *Nature* and *Science* between 2010 and 2015. *Nature Human Behaviour*, 2(9):637–644, 2018. doi: 10.1038/s41562-018-0399-z. URL `http://dx.doi.org/10.1038/s41562-018-0399-z`.

Frank Close. *Too Hot to Handle: The Race for Cold Fusion*. Princeton University Press, Princeton, 1991.

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`.

P. J. Easterbrook, R. Gopalan, J. A. Berlin, and D. R. Matthews. Publication bias in clinical research. *The Lancet*, 337(8746):867–872, 1991. ISSN 0140-6736. doi: 10.1016/0140-6736(91)90201-Y. URL `http://dx.doi.org/10.1016/0140-6736(91)90201-Y`.

Matthias Egger and George Davey Smith. Bias in location and selection of studies. *BMJ*, 316(7124):61–66, Jan 1998. URL `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2665334/`.

Daniele Fanelli. Do pressures to publish increase scientists' bias? An empirical support from US states data. *PLoS ONE*, 5(4):e10271, Apr 2010. doi: 10.1371/journal.pone.0010271. URL `http://dx.doi.org/10.1371/journal.pone.0010271`.

Paul Feyerabend. *Against Method*. New Left Books, London, 1975.

Michael Friedman. Truth and confirmation. *The Journal of Philosophy*, 76(7):361–382, 1979. ISSN 0022362X. URL `http://www.jstor.org/stable/2025452`.

Steve Fuller. *Social Epistemology*. Indiana University Press, Bloomington, second edition, 2002.

Remco Heesen. How much evidence should one collect? *Philosophical Studies*, 172(9):2299–2313, 2015. ISSN 0031-8116. doi: 10.1007/s11098-014-0411-z. URL `http://dx.doi.org/10.1007/s11098-014-0411-z`.

Remco Heesen. Communism and the incentive to share in science. *Philosophy of Science*, 84(4):698–716, 2017. ISSN 0031-8248. doi: 10.1086/693875. URL `http://dx.doi.org/10.1086/693875`.

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 R. Huizenga. *Cold Fusion: The Scientific Fiasco of the Century*. Oxford University Press, Oxford, second edition, 1993.

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

John P. A. Ioannidis. Why most published research findings are false. *PLoS Medicine*, 2(8):e124, Aug 2005. doi: 10.1371/journal.pmed.0020124. URL `http://dx.doi.org/10.1371/journal.pmed.0020124`.

John P. A. Ioannidis. Journals should publish all "null" results and should sparingly publish "positive" results. *Cancer Epidemiology Biomarkers & Prevention*, 15(1):185, 2006. doi: 10.1158/1055-9965.EPI-05-0921. URL `http://cebp.aacrjournals.org/content/15/1/186.short`.

Robert Kantrowitz and Michael M. Neumann. Optimization for products of concave functions. *Rendiconti del Circolo Matematico di Palermo*, 54 (2):291–302, 2005. ISSN 0009-725X. doi: 10.1007/BF02874642. URL `http://dx.doi.org/10.1007/BF02874642`.

Philip Kitcher. *The Advancement of Science: Science without Legend, Objectivity without Illusions*. Oxford University Press, Oxford, 1993. ISBN 0195046285.

Sander L. Koole and Daniël Lakens. Rewarding replications: A sure and simple way to improve psychological science. *Perspectives on Psychological Science*, 7(6):608–614, 2012. doi: 10.1177/1745691612462586. URL http://pps.sagepub.com/cgi/content/abstract/7/6/608.

Imre Lakatos. *The Methodology of Scientific Research Programmes*. Cambridge University Press, Cambridge, 1978.

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

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. 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.

Leif D. Nelson, Joseph P. Simmons, and Uri Simonsohn. Let's publish fewer papers. *Psychological Inquiry*, 23(3):291–293, 2012. doi: 10.1080/1047840X.2012.705245. URL http://dx.doi.org/10.1080/1047840X.2012.705245.

James W. Neuliep and Rick Crandall. Editorial bias against replication research. *Journal of Social Behavior and Personality*, 5(4):85–90, 1990.

Brian A. Nosek and Timothy M. Errington. Making sense of replications. *eLife*, 6:e23383, Jan 2017. ISSN 2050-084X. doi: 10.7554/eLife.23383. URL `https://dx.doi.org/10.7554/eLife.23383`.

Open Science Collaboration. Estimating the reproducibility of psychological science. *Science*, 349(6251):aac4716, Aug 2015. doi: 10.1126/science.aac4716. URL `http://www.sciencemag.org/content/349/6251/aac4716.abstract`.

Charles Sanders Peirce. Note on the theory of the economy of research. *Operations Research*, 15(4):643–648, 1967 [1879]. URL `http://www.jstor.org/stable/168276`.

Robert Pool. Fusion followup: Confusion abounds. *Science*, 244(4900):27–29, 1989. doi: 10.1126/science.244.4900.27. URL `http://www.sciencemag.org/content/244/4900/27.short`.

Florian Prinz, Thomas Schlange, and Khusru Asadullah. Believe it or not: how much can we rely on published data on potential drug targets? *Nature Reviews Drug Discovery*, 10(9):712, Sep 2011. doi: 10.1038/nrd3439-c1. URL `http://dx.doi.org/10.1038/nrd3439-c1`.

W. V. Quine. Two dogmas of empiricism. *The Philosophical Review*, 60(1): 20–43, 1951. ISSN 00318108. doi: 10.2307/2181906. URL `http://dx.doi.org/10.2307/2181906`.

Robert Rosenthal. The “file drawer problem” and tolerance for null results. *Psychological Bulletin*, 86(3):638–641, 1979. doi: 10.1037/0033-2909.86.3.638. URL `http://psycnet.apa.org/doi/10.1037/0033-2909.86.3.638`.

Joseph P. Simmons, Leif D. Nelson, and Uri Simonsohn. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11):1359–1366,

2011. doi: 10.1177/0956797611417632. URL http://pss.sagepub.com/content/22/11/1359.abstract.

Paul E. Smaldino and Richard McElreath. The natural selection of bad science. *Royal Society Open Science*, 3(9):160384, 2016. doi: 10.1098/rsos.160384. URL http://rsos.royalsocietypublishing.org/content/3/9/160384.

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. Herding and the quest for credit. *Journal of Economic Methodology*, 20(1):19–34, 2013. doi: 10.1080/1350178X.2013.774849. URL http://dx.doi.org/10.1080/1350178X.2013.774849.

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*, chapter 1. Oxford University Press, Oxford, 2017. URL https://philpapers.org/rec/STRSSC-2.

Athina Tatsioni, Nikolaos G. Bonitsis, and John P. A. Ioannidis. Persistence of contradicted claims in the literature. *Journal of the American Medical Association*, 298(21):2517–2526, 2007. doi: 10.1001/jama.298.21.2517. URL http://dx.doi.org/10.1001/jama.298.21.2517.

Johanna Thoma. The epistemic division of labor revisited. *Philosophy of Science*, 82(3):454–472, 2015. ISSN 00318248. URL http://www.jstor.org/stable/10.1086/681768.

Michael Weisberg and Ryan Muldoon. Epistemic landscapes and the division of cognitive labor. *Philosophy of Science*, 76(2):225–252, 2009. ISSN 00318248. URL http://www.jstor.org/stable/10.1086/644786.

Kevin J. S. Zollman. The credit economy and the economic rationality of science. *The Journal of Philosophy*, 115(1):5–33, 2018. doi: 10.5840/jphil201811511. URL `http://dx.doi.org/10.5840/jphil201811511`.