Expediting the Flow of Knowledge Versus Rushing into Print*

Remco Heesen†

July 23, 2015

Abstract

I show that the conflicting pressures to publish quickly and to publish accurately (which requires careful checking) allow for a rational tradeoff from the perspective of credit maximization. Under some plausible assumptions, the balance that is optimal for the individual scientist will be suboptimal from a social perspective. In particular, I demonstrate in two models (one in which accuracy trades off against speed and one in which accuracy trades off against speed and impact) that scientists will tend to publish work of a lower accuracy level than is socially desirable. I illustrate the relevant phenomena using a case study of Fleischmann and Pons’ research on cold fusion. I consider philosophical and practical implications of my results as well as consequences for how Fleischmann and Pons’ work should be evaluated.

*Thanks to Kevin Zollman, Michael Strevens, Stephan Hartmann, Teddy Seidenfeld, Jan Sprenger, Liam Bright and an audience at the University of Tilburg for valuable comments and discussion. This work was partially supported by the National Science Foundation under grant SES 1254291.

†Department of Philosophy, Baker Hall 161, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA. Email: rheesen@cmu.edu
1 Introduction

The desire to be first and receive credit for their work are important motivations for scientists. This is not to deny that scientists may have other motivations, such as advancing the state of human knowledge. But such motivations are idiosyncratic, while credit is necessary to have a career in science, and so professional scientists have to care about it to some extent. This point has long been recognized by sociologists such as Merton (1957, 1969) and Latour and Woolgar (1986, chapter 5) and philosophers of science like Hull (1988, chapter 8), Kitcher (1993, chapter 8), and Strevens (2003).

Recognizing scientists’ desire for credit may seem problematic on a naive picture of science in which scientists selflessly strive for truth (see Kitcher 1993, chapter 1, for a caricature with references). A trend in the social epistemology of science has been to argue that things are not so bad. Kitcher (1993, chapter 8) and Strevens (2003) have argued 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) have argued that credit incentives speed up the progress of science. And Boyer (2014), Strevens (forthcoming) and chapter ?? of this dissertation argue that credit incentives can motivate scientists to share their work widely.

Given this trend, it would be tempting to conclude that credit incentives act like an infallible “invisible hand” that can solve all problems in the social organization of science. Here, I go against this trend in identifying one way in which credit incentives may lead to bad outcomes at the social level. The central claim of this paper is that credit incentives create a pressure to publish that, when combined with a system of peer review that is necessarily imperfect, leads to reproducibility problems.

The issue of reproducibility has come under increased scrutiny (Ioannidis 2005, Pashler and Wagenmakers 2012). Recent studies have shown that

---

1 I am not the first to do so. For example, Strevens (2013) discusses ways in which credit incentives may lead to herding behavior.
published work in fields such as medicine and psychology frequently fails to be reproducible (Prinz et al. 2011, Begley and Ellis 2012). Some have argued that the pressure to publish is a cause of this phenomenon (Fanelli 2010, Prinz et al. 2011). I illustrate this claim in section 2 with a case study in which the pressure to publish led to the publication of research that failed to reproduce: Fleischmann and Pons’ work on cold nuclear fusion.

I then go on to argue that reproducibility problems arise as a structural problem of credit incentives rather than as (a series of) incidents. In section 3 I construct a model to show that two crucial ingredients are sufficient for reproducibility problems to arise. First, the tradeoff between speed and accuracy: the scientist must choose how quickly to work, knowing that the faster she works the greater the chance of errors. Second, the system of peer review which aims to publish accurate results while rejecting erroneous results.

I show that given these two ingredients, the scientist has a unique credit-maximizing way to trade off speed against accuracy. But, in a way to be made precise, the tradeoff favors speed over accuracy. Scientists are given too much of an incentive to “rush into print”, compared to what would be optimal from a social perspective. This shows that credit incentives are a possible cause of reproducibility 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 accuracy. I show the robustness of my earlier results in this expanded model, and I consider how it gives rise to different types of scientists: “mavericks” and “followers”.

In the conclusion (section 5) I summarize my results. I also discuss possible ways to diminish or remove the incentive to produce work that is structurally less reproducible than is socially desirable. And finally, I discuss whether my results support an interpretation of Fleischmann and Pons’ cold fusion research as a case of “rushing into print”.

3
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 fails to reproduce. 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 (UU). 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.

Although it seemed impossible given existing theories, physicists and chemists alike were initially willing to give Fleischmann and Pons the benefit of the doubt. Experimental results take precedence over theory, and the two’s credentials as experimentalists were impeccable. Given the potential implications, and the media hype, scientists around the world dropped what they were doing to attempt to replicate 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 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”?

While in the press conference they claimed to have been working on this project for five years, in reality most of the work had been done in the last six months (Close 1991, p. 82). Before then, they had done some exploratory work that seemed promising. In August 1988, they requested funding for their cold fusion research from the Department of Energy (Huizenga 1993, p. 16). This brought them into contact with Steven Jones, a professor of physics at Brigham Young University (BYU).

Unbeknownst to Fleischmann and Pons, Jones and his team had been working on a very similar project. The main differences were that Jones was primarily interested in explaining some of 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.

Jones noted that their work seemed complementary. The two teams exchanged information and Fleischmann and Pons visited Jones’ lab on February 23, 1989. By now Jones had obtained his best data, which gave some evidence of neutrons above background levels at the right energy level to be potentially due to fusion. 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, in contrast, were not ready to go public. Pons’ graduate student Marvin Hawkins had been running experiments since October 1988. They were confident in their evidence that some experiments produced excess heat, but they had only just started measuring neutrons, and
their apparatus was much less sophisticated than Jones’. Much remained to be investigated: why did some experiments produce excess heat while others did not, and how could they explain the discrepancy between the heat measurements and the neutron counts (which were many orders of magnitude too low compared to the heat if fusion was occurring)? Fleischmann and Pons indicated that they wanted to do another eighteen months of research before going public [Pool 1989, Huizenga 1993, p. 18].

With Jones about to go public, Fleischmann and Pons felt that they had to inform the university. This led to another meeting at BYU on March 6, this time with the presidents of both universities present. Especially at UU there was a clear sense of the potential impact of cold fusion at this point, with the president of UU suggesting that billions of dollars and Nobel prizes were at stake [Close 1991, p. 93]. Moreover, those associated with the UU group accused Jones of stealing their ideas on at least two occasions in mid-February and mid-March [Close 1991, chapter 6]. After Jones insisted that he was ready to publish his results, the two groups agreed to 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 hastily written manuscript to the *Journal of Electroanalytical Chemistry* on March 11, and they held the above-mentioned press conference, a day before the scheduled simultaneous submission.

Why was this press conference held? Ostensibly it was to correct “rumors, leaks, questions, and false information” that were already doing the rounds [Pool 1989, Huizenga 1993, p. 19]. But it seems clear that the real reason was to establish priority for the cold fusion research, especially relative to Jones [Huizenga 1993, p. 19]. This may seem unnecessary, as Jones’ experimental

---

2The claim by UU officials that the reason for the press conference was “leaks” does not hold up. The evidence for the leak was an article on cold fusion in the *Financial Times* on the morning of the press conference. This was neither a real leak (Fleischmann and Pons had given permission for it), nor could it have caused the press conference to be held, as
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.

In the situation we were then in, we were obliged to tell the university of the work we had done and they perceived that they were obliged to go for patent protection at that time. 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 priority. 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, under pressure from university officials.

So a concern for priority led to the publication of research that turned out to be erroneous. While this is a particularly high-profile case, it has recently been suggested that erroneous research is quite prevalent in the scientific literature (Prinz et al. 2011, Begley and Ellis 2012, ?). The model of the next two sections aims to establish that this is a structural problem stemming from scientists’ (credit) incentives. 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).

this had been scheduled two days earlier (Close 1991 pp. 101–102).
3 A Tradeoff Between Speed and Accuracy

Here I develop a decision-theoretic model to evaluate decisions to go public with results of scientific research. By giving a model, I aim to show that the problem of erroneous results arises structurally rather than incidentally. This section considers only the tradeoff between speed and accuracy, while section 4 also allows the potential importance or impact of the result to influence the decision.

Consider a scientist—or a team of scientists, such as Fleischmann and Pons—working on a research project. Why would she decide to publicize her work, say in the form of a journal article? As the case of Fleischmann and Pons illustrates, an important reason is to establish priority for the work.

As I mentioned in the introduction, scientists’ concern for priority is well-documented and understandable, given its importance to scientific careers (Merton 1957, 1969). But even a scientist who only cares about the progress of science altruistically would be concerned about priority. After all, the primary way to contribute to the progress of science is to do original work that causes other scientists to learn something earlier than they otherwise would have (cf. Strevens 2003, p. 55). Thus, while I use the notion of credit to represent the concern for speed in the model, I claim that this model can still capture the motivations of scientists not primarily concerned about credit.

For these reasons, the scientist prefers to work and publish faster rather than slower (all else being equal). This is represented in the model by assuming that the scientist aims to maximize the amount of credit she accrues per unit time. E.g., if the amount of credit per publication is some fixed number $c$ (this assumption is relaxed in section 4), working and publishing twice as fast will double credit per unit time, all else being equal.

But of course all else is not equal. Working faster reduces accuracy. By accuracy I mean the amount of certainty with which the result of the research project (e.g., “cold fusion is a viable source of energy”) is established. Loosely speaking I have in mind the probability that the result is true, given the
amount of evidence the scientist has gathered at the time of publication.\footnote{What kind of probability is this? For the scientist’s decision whether or not to go public, what matters is her own assessment of this probability, i.e., the accuracy level is a subjective probability. When I am concerned with the social value of scientific research below, an objective interpretation of the probability seems more relevant. As long as there are no structural biases (i.e., on average scientists’ subjective assessment equals the objective probability) this presents no problems for the interpretation of my results, which are only concerned with averages anyway. Since in a competitive environment like science such biases would carry serious penalties, it seems reasonable to assume them away.}

But this will not do in a credit-maximization model, since credit is conferred socially by other scientists, and thus cannot be determined directly by (objective) truth.\footnote{I could get around this problem by simply defining to be true that which a given scientific community accepts at a given time, as some sociologists of science have done (e.g., \cite{collins1981, latourandwoolgar1986, bloor1991}. I do not use this way of speaking to avoid the impression that I am committed to this as a substantive theory of truth.} So to be more precise: I call a scientific result \textit{accurate} if it holds up as “true” (i.e., is not discredited) in the relevant scientific community in the mid-term, say for ten years after publication. Conversely, I call a result inaccurate or \textit{erroneous} if it does not hold up in the community in the mid-term. The \textit{accuracy level} of a scientific result is the probability, given the evidence gathered at the time of publication, that the result is accurate.

In the model, the scientist chooses the desired accuracy level \( p \in [0, 1] \). That is, the scientist works on her research project until she obtains a result that has at least probability \( p \) of holding up as “true” in the community, at which time she publishes.

Accuracy takes time. This is reflected in the model by the \textit{speed function} \( \lambda \). The value \( \lambda(p) \) reflects the speed at which the scientist works if the desired level of accuracy is \( p \). More specifically, \( 1/\lambda(p) \) is the expected time until completion of the research project (so \( \lambda(p) \) is the number of research projects “like this one” that the scientist would expect to complete per unit time). If \( \lambda \) is a decreasing function of \( p \), as I assume, the expected time until completion indeed goes up if \( p \) is increased (see figure 1).
Figure 1: $p$ and $\lambda$ trade off against each other. In this example, $\lambda(p) = 1 - p^2$, satisfying assumptions 1, 2, 3, and 4.

On the other hand, reducing the accuracy level (lowering $p$) allows the scientist to work more quickly. This represents the idea of “rushing into print”, leading to a higher probability of errors slipping in. I assume that these errors represent honest mistakes by the scientist. 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 (2014).

I make a number of assumptions on the way speed and accuracy trade off against each other, as reflected in the speed function $\lambda$. The first assumption simply reflects the idea that they in fact trade off, as discussed above.

**Assumption 1** (The speed function is decreasing). *The speed function $\lambda : [0, 1] \rightarrow \mathbb{R}$ is such that for all $p, p' \in [0, 1]$, if $p < p'$, then $\lambda(p') < \lambda(p)$.*

As I indicated above, assuming that $\lambda$ is decreasing simply reflects the fact that it takes more time to do research that is less likely to be erroneous.
**Assumption 2** (The speed function is concave). For every $p, p', t \in [0, 1]$,

$$t\lambda(p) + (1 - t)\lambda(p') \leq \lambda(tp + (1 - t)p').$$

This assumption may be described as a kind of decreasing marginal returns: as the accuracy level $p$ is lowered, the speed $\lambda$ is increased by assumption 1 but it increases ever slower as $p$ approaches zero: writing the papers itself takes time, which becomes relatively more significant if the scientist spends relatively little time on the research content. Conversely, if the scientist aims to be more accurate (increasing $p$), the speed $\lambda$ drops off ever faster. In other words, more and more extra research time is required to increase $p$ as $p$ approaches one.

**Assumption 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 and Quine have argued.

**Assumption 4** (The speed function is differentiable). The function $\lambda$ is differentiable on the interior of its domain, i.e., for all $p \in (0, 1)$.

This assumption says something about the “smoothness” of the tradeoff between speed and accuracy. I have no opinion on whether this assumption is justified. In fact, it is hard to imagine what evidence for or against this assumption would even look like. For this reason, when I state the results that can be proven based on these assumptions below, I consider both the case with and without assumption 4.

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 whether the results of the scientist’s work are likely to stand up to the scrutiny of the scientific community.

In the simplest possible case it does this “pre-screening” perfectly: all and only those papers that are in fact accurate are accepted. The scientist does not know whether her paper is accurate; she only knows the accuracy level \( p \), i.e., the probability that it is accurate. So from the scientist’s perspective, if she produces a paper with accuracy level \( p \), there is a probability \( p \) that the journal publishes it.

Suppose that the 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 level of accuracy \( p \) and the speed \( \lambda \) (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 future. 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).

Adapting the terminology from statistics (following Laband and Piette 1994) I call the acceptance of an erroneous result a type I error, and the rejection of an accurate result a type II error. Again following statistics I write \( \alpha \) for the probability of a type I error and \( \beta \) for the “power” (the probability that a type II error 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 type I errors (\( \alpha > 0 \)). Note that I remain agnostic on the possibility of type II errors (\( \beta \) may or may not be equal to one) although it seems reasonable to assume that those exist as well. I do assume that there is some discernment in the peer review system, i.e., accurate results are more likely to be accepted than erroneous results (\( \beta > \alpha \)).
Assumption 5 (Imperfect peer review). The peer review acceptance probabilities \( \alpha \in [0, 1] \) and \( \beta \in [0, 1] \) are such that \( \beta > \alpha > 0 \).

For the amount of credit for a published erroneous result I write \( c_e \). While this reflects results that are “discredited”, that does not necessarily equate to zero credit. Discredited research 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, although no more than accurate publications, i.e., \( 0 < c_e \leq c_a \).

Putting all of this together yields the following. The scientist works on the research project at speed \( \lambda(p) \). The result is accurate with probability \( p \). In this case it gets published with probability \( \beta \) 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 \( \alpha \). 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 accuracy to the socially optimal tradeoff, it is important to be explicit about what is meant by the social value of scientific research. Here I have in mind the contribution that it makes to science as a 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

\(^5\)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 in the main text is not to argue that \( c_e \) is necessarily positive, but that erroneous publications can influence a scientist’s credit stock.
What is the social value $V$ of the scientist’s research, as a function of her choice of accuracy level $p$? I assume that research can have social value only when it is published. The probabilities of publication $\alpha$ and $\beta$, the accuracy level $p$, and the speed $\lambda(p)$ are all as above. The only difference is in the value of the research:

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

where $v_a$ is the social value of an accurate result, and $v_e$ the social value of an erroneous result.

Credit is awarded for (accurate) scientific work proportional to its social value, as has been argued in the literature. 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. Based on these two lines of reasoning, I assume 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

\[\text{Work in the natural sciences is perhaps more likely to produce new technology than work in the social sciences or the humanities. But all kinds of research can find its way into society. For example, research in economics may affect policy decisions, and philosophical research may affect the public debate.}\]
erroneously (in hindsight) finds 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. 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: $v_e \leq 0$. However, I need not insist on this conclusion. For my purposes here it suffices that (a) the social value of erroneous results is less than the credit given for them ($v_e < c_e$), and (b) the social value of erroneous results is substantially less than that of accurate results ($v_e \leq v_a/2$). Both (a) and (b) follow immediately if my argument that $v_e \leq 0$ is accepted, but if one thinks that erroneous results have some positive social value, my argument below still goes through as long as that value is relatively small.

Once again reasoning in terms of citations yields a similar conclusion. 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. This suggests that the social value of erroneous results is substantially less than that of accurate results, in support of (b).  

7. Recall that I defined an erroneous result as one that does not hold up as true in the scientific community in the mid-term. Thus it is impossible by definition for an error to go uncorrected.

8. This argument assumes that erroneous results do not get cited more than accurate results. It seems unlikely that an erroneous result would, on average, get cited more than
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 (and when that failed to show that the mistake was with them, see [Close 1991, chapter 12] is testament to Fleischmann and Pons’ authority as competent electrochemists ([Kitcher 1993, section 8.2). In contrast, work by subsequent cold fusion researchers has been largely ignored ([Huizenga 1993, p. 208), as has work by creation scientists that challenges mainstream paleontology ([Kitcher 1993 section 8.2). So the “credit value” of a citation to an erroneous publication is higher than its social value, in support of (a).

Assumption 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 6** (Credit and social value).

6.a. Accurate results have positive value: \( c_a > 0 \) and \( v_a > 0 \).

6.b. Accurate results are awarded credit proportional to their social value: \( c_a = v_a \).

6.c. Erroneous results are awarded no more credit than accurate results: \( c_e \leq c_a \).

6.d. The social value of erroneous results is less than the credit given for them: \( v_e < c_e \).

6.e. The social value of erroneous results is at most half that of accurate results: \( v_e \leq v_a/2 \).

---

an accurate result. If it gets cited less, this provides further support for my conclusion. [Tatsioni et al. 2007] provide some empirical support for this assumption.
I now state three results that can be proven based on the assumptions I have made. The first result states that the functions $C$ and $V$ have unique maxima, i.e., there is a particular accuracy level that a rational credit-maximizing scientist would choose, and there is a particular accuracy level that maximizes the social value of the scientist’s contribution (which may be different or the same as the value that maximizes credit). This result does not use the controversial assumption 4 (which concerns the smoothness of the speed function) or the potentially controversial parts of assumption 6.

**Theorem 7.** If assumptions 1, 2, 3, 5, and 6.a are satisfied, then there exist unique values $p_C^* < 1$ and $p_V^* < 1$ that maximize the functions $C$ and $V$ respectively, i.e.,

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

Proofs of this and subsequent results are provided in a separate document, entitled “Analysis and Results For ‘Expediting the Flow of Knowledge Versus Rushing into Print’”.

Note that even with these fairly minimal assumptions, it follows that $p_V^* < 1$. This means that even from the social perspective perfect accuracy 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 assumption 3 (“no perfect work”) 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 accuracy 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. If errors are expected, it also seems wrong to hold it against individual scientists too much when they make mistakes, which gives some additional philosophical support for assumption 6.d ($v_e < c_e$).
The second result says that the imperfections in the peer review system and the way credit is awarded systematically favor lower levels of accuracy. That is, a scientist who maximizes expected credit will choose an accuracy level no higher than the optimal level from the perspective of maximizing social value. This result does not use the controversial assumption 4 or the assumption that the social value of erroneous results is at most half that of accurate results (assumption 6.e).

**Theorem 8.** Let assumptions 1, 2, 3, 5, and 6.a–6.d be satisfied, and define \( p^*_{C} \) and \( p^*_{V} \) as in theorem 7. 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 accuracy level than is socially optimal. But it could easily be objected that I have not shown this: theorem 8 leaves open the possibility that \( p^*_{C} = p^*_{V} \), the happy case in which individual and social incentives align exactly.

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 accuracy (\( p^*_{C} = p^*_{V} = 0 \), see figure 2). Second, if the speed function is not differentiable at the point of optimality (see figure 3). The third result shows that if these two situations are ruled out there is a definite misalignment between individual and social incentives.

**Theorem 9.** Let assumptions 1, 2, 3, 4, 5, and 6 be satisfied, and define \( p^*_{C} \) and \( p^*_{V} \) as in theorem 7. Then \( p^*_{C} < p^*_{V} \).

So when all the assumptions are brought into play, credit-maximization gives the scientist an incentive to work faster, but less accurately, than is optimal from the perspective of maximizing social value.

---

9Some of the assumptions can be weakened: I have chosen to present them in a way that allows me to focus on their plausibility, rather than stating them as generally as possible. For example, in theorems 7 and 8 assumption 5 (imperfect peer review) can be
Figure 2: If $\lambda(p) = 2 - p - p^2$ (the solid red line) and $v_c$ 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 $v_e \alpha = 7/9$).

This result depends crucially on the imperfections in the peer review system, and in particular the type I error. If $\alpha = 0$ and $\beta > 0$ then assumptions 1, 2, 3, and 6.a are sufficient to show that the functions $C$ and $V$ have unique maxima, and that these maxima are equal. Hence I interpret theorems 8 and 9 as showing that imperfections in the peer review system create a systematic bias that leads credit-maximizing scientists to favor speed over accuracy relative to the social optimum.

weakened: $\beta > 0$ and $\beta \geq \alpha$ suffices for theorem 8 and just $\beta > 0$ suffices for theorem 7.

In theorems 8 and 9 assumption 6.b ($v_a = c_a$) can be weakened: if $c_e \geq 0$ then $v_a \geq c_a$ suffices and if $c_e \leq 0$ then $v_a \leq c_a$ suffices.
Figure 3: If $\lambda(p) = 1 - p/2$ for $p \leq 2/3$ and $\lambda(p) = 2(1 - p)$ for $p > 2/3$ then $\lambda$ (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: $p^*_C = p^*_V = 2/3$

4 A Tradeoff Between Speed, Accuracy, 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 “mavericks”, 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 “followers”, content to make small contributions that are likely to be correct and accepted. I adopt this terminology from Weisberg and Muldoon (2009), although the distinction
goes back much further (e.g., [Hull 1988, p. 474]).

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

Above I showed that if there are imperfections in the peer review system, scientists tend to favor speed over accuracy (relative to the social optimum). The first question I aim to investigate here is what the consequences of imperfections in the peer review system are in this more complicated model. The second question is to what extent the different “types” of scientists—mavericks and followers—show up in the model.

In the model of this section the scientist chooses both a desired accuracy level $p$ and a desired impact level $c$. Since $p$ is interpreted as a probability, its domain is naturally constrained to the interval $[0, 1]$. The impact $c$ is not 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 10.** The maximum impact function is a function $\mu : [0, 1] \rightarrow [0, \infty)$. 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 $\lambda : D \rightarrow \mathbb{R}$.

I make a number of assumptions on the shape of $\lambda$. 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 accuracy level, increasing the impact decreases speed, and at a fixed level of impact, increasing accuracy decreases speed.

**Assumption 11** (The speed function is decreasing).
11.a. For all \( p, p' \in [0, 1] \), if \( p < p' \) and \( c \leq \min\{\mu(p), \mu(p')\} \), then \( \lambda(p', c) < \lambda(p, c) \).

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

**Assumption 12** (The speed function is concave). For any \((p, c), (p', c') \in D\) and \( t \in [0, 1] \),

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

12.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 the accuracy level 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.

**Assumption 13.** The function \( \lambda \) vanishes as \( p \) or \( c \) approaches the edge of its domain \( D \).

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

13.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 (the “no perfect work” assumption). Assumption 13.a is in fact identical to that assumption (although for technical reasons I only need to make the assumption for the case \( c = 0 \)) and has the same justification. Assumption 13.b formalizes the idea that \( \mu(p) \) represents the maximum impact that can be achieved at a given accuracy level \( p \), by requiring that the scientist’s speed becomes negligible as this value is approached.

---

10 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 \( \mu(1) > 0 \).

11 It does not follow from the definition of the domain \( D \) of \( \lambda \) 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 \( \mu \) is a concave function.
Assumption 14 (The speed function is differentiable (in $p$)). The partial derivative of the function $\lambda$ 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 $0 < c < \mu(p)$.

This assumption requires that the speed function is “smooth” (at least in one direction). As before, I consider results both with and without this assumption.

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 $r_c c$: a proportionality constant $r_c$ times the scientist’s chosen impact level $c$. If $r_c > 0$, this means that erroneous results that would have had a high impact get more credit (“at least you tried something ambitious”). If $r_c < 0$, this means that erroneous results of potentially high impact are penalized more harshly (“the bigger they are, the harder they fall”). This seems accurate 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 $c$, 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 (credit) 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 assumption 6.b ($c_a = v_a$) but for notational convenience here 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. So the social value of the scientist’s research, as a function of $p$ and $c$, 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 $\alpha$ and $\beta$ is exactly like before. I restate it here as a reminder.

**Assumption 15** (Imperfect peer review). The peer review acceptance probabilities $\alpha \in [0, 1]$ and $\beta \in [0, 1]$ are such that $\beta > \alpha > 0$.

The assumptions on the parameters $r_c$ and $r_v$ are similar to the assumptions on $c_e$ and $v_e$ I made in assumption 6, with the following changes. First, assumptions 6.a and 6.b are no longer needed because they are built into the definition of the functions $C$ and $V$. Second, assuming that $r_c \leq 1$ is interpretationally equivalent to the assumption that $r_c \leq c_a$ which I made above (because $r_c \leq 1$ if and only if $r_c c \leq c$). Third, for technical reasons, a slightly stronger assumption on the value of erroneous results is needed: $r_v \leq 1/3$ instead of $v_e \leq v_a/2$.

**Assumption 16** (Credit and social value).

16.a. Erroneous results are awarded no more credit than accurate results: $r_c \leq 1$.

16.b. The social value of erroneous results is less than the credit given for them: $r_v < r_c$. 

24
16.c. The social value of erroneous results is at most a third that of accurate results: \( r_v \leq 1/3 \).

I present three results that use some or all of the above assumptions. The first result says that there are unique choices of accuracy level and impact level that maximize expected credit and that maximize social value. Assumptions 14 and 16 are not needed.

**Theorem 17.** If assumptions 11, 12, 13, and 15 are satisfied, then there exist unique points \((p^*_C, c^*_C)\) and \((p^*_V, c^*_V)\) that maximize the functions \(C\) and \(V\) respectively, i.e.,

\[
C(p^*_C, c^*_C) = \max_{(p, c) \in D} C(p, c) \quad \text{and} \quad V(p^*_V, c^*_V) = \max_{(p, c) \in D} V(p, c).
\]

Moreover, \(p^*_C < 1\) and \(0 < c^*_C < \mu(p^*_C)\); and \(p^*_V < 1\) and \(0 < c^*_V < \mu(p^*_V)\).

If assumptions 16.a and 16.b are added the credit-maximizing accuracy level \(p^*_C\) is at most the social value maximizing accuracy level \(p^*_V\).

**Theorem 18.** Let assumptions 11, 12, 13, 15, 16.a, and 16.b be satisfied, and define \((p^*_C, c^*_C)\) and \((p^*_V, c^*_V)\) as in theorem 17. Then \(p^*_C \leq p^*_V\).

And, finally, if assumptions 14 and 16.c are added the inequality is strict.

**Theorem 19.** Let assumptions 11, 12, 13, 14, 15, and 16 be satisfied, and define \((p^*_C, c^*_C)\) and \((p^*_V, c^*_V)\) as in theorem 17. 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 accuracy, relative to what she would do if she were trying to maximize the social value of her work. This is true under essentially the same conditions as above. So the results expressed in theorems 8 and 9 are seen to be robust against the introduction of the dimension of impact.
Second, theorem 17 rules out the possibility that a scientist could switch from being a follower to a maverick (increasing impact at the expense of accuracy) 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 a maverick or a follower might act as a tie-breaker. This consequence of the model is somewhat surprising.

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 accuracy, 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 and outside the scientist’s control, theorem 17 can be interpreted as showing that different types of scientists are the result of differences in aptitude rather than choice.

Whether a scientist is likely to be a maverick or a follower is thus determined by the shape of her speed function. The following example illustrates this.

**Example 20.** Consider two scientists. For scientist 1, the tradeoff between accuracy, impact, and speed is given by the speed function $\lambda_1$, 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 $0 \leq c \leq \frac{1}{2}(\sqrt{25 + 12p - 12p^2} - 3 - 2p)$ (see figure 4). Note that this function satisfies assumptions 11, 12, 13, and 14. Suppose further that $r_c = 0$. Then the credit-maximizing choice for scientist 1 is

---

12See Huber (2001) and citations therein) for evidence that the productivity of scientists is, on average, constant over the course of a career.

13As a result I need to make no specific assumption on the values of $\alpha$ and $\beta$: the maximum of $C$ will not depend on this as long as $\beta > 0$. 

---

26
\[ p \approx 0.52 \text{ and } c \approx 0.38. \]

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 \( 0 \leq p \leq \frac{1}{2}(\sqrt{25 + 12c} - 12c^2 - 3 - 2c) \) (see figure 4). This function also satisfies assumptions 11, 12, 13, and 14. But the credit-maximizing choice for scientist 2 is \( p \approx 0.38 \) and \( c \approx 0.52 \).

Scientist 1’s speed function “favors” accuracy compared to scientist 2’s, which “favors” impact. This is because \( \lambda_1 \) is closer to linear in \( c \)—having only a small quadratic component—while it is a fourth-degree polynomial in \( p \) (\( \lambda_2 \) is simply its mirror image). So if the scientists are responsive to credit incentives, scientist 1 will behave more like a follower, doing relatively safe, low-impact research. Scientist 2 on the other hand will behave more like a maverick, 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 more linear behavior in \( c \) and less linear behavior in \( p \) produces followers, and the reverse mavericks.

Moreover, if scientists are credit-maximizers, and assumptions 11, 12, 13.
and that are justified, then theorem guarantees that differences in the shape of the speed function are the way different types of scientists can arise.

5 Conclusion

The following four conclusions can be drawn from the work presented in this paper.

First, imperfections in the peer review system create a misalignment between when it is optimal to “go public” from a credit-maximizing perspective and the socially optimal time to do so. This misalignment systematically sacrifices accuracy. So scientists have a credit-incentive to produce work that is less accurate than is socially optimal. This is true when only the tradeoff between speed and accuracy is considered as well as when a three-way tradeoff between speed, accuracy, and impact is considered (in each case, under some plausible assumptions on the way they trade off).

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, lower accuracy levels imply that more errors enter the scientific literature. The combination of imperfections in the peer review system and credit-maximization is thus a source of reproducibility problems.

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 never make mistakes in predicting the reception of the paper by other scientists in the short- to mid-term.

However, I noted that the misalignment of incentives in the model is exclusively caused by type I errors (accepting erroneous results for publication). So reducing those can bring the credit-maximizing optimum closer to the social optimum. This would seem to recommend conservative edito-
rial practices: rejecting papers even based on fairly minimal doubts about their accuracy. But if reducing type I errors leads to more type II errors (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 that particular 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. 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, whereas its accuracy is not known with certainty until at least the mid-term. A system that could predict accuracy perfectly seems in principle impossible, as it would need to predict the outcomes of future research that establishes the accurate or erroneous nature of the present work. 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 problem.

Another solution would focus on the amount of credit given for erroneous 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 shown to be erroneous. If the credit given to erroneous 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 be helpful, for example, if there was a broader awareness of which research has been shown to be erroneous. But again this may be hard to achieve in practice.

A third solution would be to try 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 accuracy over speed that could in principle balance out the misalignment I have shown. 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 accuracy 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.

As all of these suggestions have some problems associated with them, it is not clear which one(s) should be recommended. I leave a more detailed comparison of these and other possible solutions to future work.

The second conclusion is that perfect accuracy is neither to be expected nor to be desired. The reason for this is of course 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 forthcoming), it is worth emphasizing.

Third, I considered the difference between scientists who pursue high-impact research relatively likely to be erroneous (“mavericks”) and scientists who pursue more mundane research relatively likely to be accurate (“followers”). 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: mavericks are scientists with an aptitude for high-impact research at a relatively small cost in speed, while followers can pursue highly accurate research at a relatively small cost in speed.

Considering the tradeoff between speed, accuracy, and impact explicitly shows that high-impact research (or “transformative” research in modern terms) is likely to be less accurate. Example 20 illustrates this. Thus it seems unreasonable to hold mavericks to the same standards of evidence as followers. 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 consid-
ering the criteria separately, these institutions aim to prevent biasing their evaluation process for or against mavericks or followers.

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 of a relatively low accuracy level, i.e., with a relatively high chance of later being proven wrong. Second, scientists who are pursuing high-impact research should be given more leeway to produce results of a relatively low accuracy level.

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 extremely 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.
References


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


