Cumulative Advantage and the Incentive to Commit Fraud in Science

Remco Heesen

This article investigates how the credit incentive to engage in questionable research practices (up to and including fraud) interacts with cumulative advantage, the process whereby high-status academics more easily increase their status than low-status academics. I use a mathematical model to highlight two dynamics that have not yet received much attention. First, because of cumulative advantage, questionable research practices may pay off over the course of an academic career even if they are not attractive at the level of individual publications. Second, because of the role of bottleneck moments in academic careers, questionable research practices may be selected for even if they do not provide a benefit in expectation. I also observe that within the model, the most successful academics are the most likely to have benefited from fraud.

1. Introduction

Trust in academic science consists at least partially in trust in academics. It is a cause of concern, then, when that trust appears to have been misplaced, as happens when cases of fraud are revealed.

Cases of data fabrication, plagiarism, and other forms of outright fraud attract a lot of attention when they are uncovered, but these are perceived by many observers as being 'rare' (Merton [1957], p. 651). In contrast, so-called questionable research practices (including *p*-hacking and salami publishing, among others) are perceived as less bad but also 'more frequent' (Merton [1957], p. 651). For evidence that these perceptions are widely shared and accurate, see (John et al. [2012]).

Here I focus on what fraud and questionable research practices have in common. Practices that fall under either of these labels are ways of enhancing an individual academic's productivity and prestige (which mutually reinforce one another in a

Electronically published July 12, 2024.

The British Journal for the Philosophy of Science, volume 75, number 3, September 2024.

Published by The University of Chicago Press for The British Society for the Philosophy of Science.
https://doi.org/10.1086/716235

process known as cumulative advantage) at some epistemic cost. As a result, the same reasons may attract academics to either of them.

This article investigates these reasons. I consider the consequences for an academic's productivity and her career of engaging in these epistemically costly practices, focusing on the role of cumulative advantage. I highlight two dynamics in particular. First, even when the chance of being caught and associated penalties are sufficiently high that epistemically dubious research outputs do not individually confer a net benefit, academics may gain a career advantage from them. Second, even when these practices are not rewarded on average, they may spread in academic communities. I will provide some tentative reasons to think that the most successful members of an academic community are more likely to have benefited from fraud or questionable research practices.

The article proceeds as follows: Section 2 provides a more careful introduction to key concepts: cumulative advantage, the credit economy, fraud, and questionable research practices. In section 3, I develop a simplified model of cumulative advantage in which an academic's productivity is represented as a non-homogeneous Poisson process. Section 4 adds a key downside of questionable research practices to the model, namely, the possibility of being exposed. Section 5 shows how the two dynamics mentioned above may arise in the model, while section 6 considers more systematically when this happens by varying the parameters of the model. Section 7 concludes.

2. Cumulative Advantage and Fraud

The academic world is strongly hierarchical (Cole and Cole [1973]). There is a small group of professors who have it all: a chair at a prestigious university, time and money for research, graduate students, frequently cited publications in highly regarded journals, prizes, media appearances, and so on. In contrast, there is a much larger group with few or none of these status markers, including lesser-known tenured or tenurable professors and academics without secure employment (postdocs, adjunct professors, lab technicians, and graduate students). These differences in status are often keenly felt, for example, through prestige bias in hiring or publishing: status markers such as individual reputation, institutional affiliation, or publishing track record influence one's chances of being hired (Clauset et al. [2015]; de Cruz [2018]) or navigating a paper through peer review (Lee et al. [2013], p. 7; Tomkins et al. [2017]).

Since the institution where one is hired and the journals one publishes in are themselves status markers, prestige bias has a self-reinforcing effect. Those who manage to obtain some of these markers, especially early in their academic career, have an easier time being hired into a prestigious job, acquiring research grants, and more generally increasing their reputation. Meanwhile, those who struggle more at the start of their career and fall behind in the prestige hierarchy find it more difficult to catch up. This general pattern, where early success begets more success, is known as cumulative advantage (DiPrete and Eirich [2006]), or, in the academic context, the Matthew effect (Merton [1968]).

Given the central importance of prestige, academics have reason to pursue it. We study this as 'the incentive structure of academic science', or alternatively 'the credit economy'. Academics receive credit first and foremost for making (and publishing) academic contributions, with originality being particularly prized (Merton [1957]; Strevens [2003]). As indicated above, this form of credit (or prestige; I treat these as synonyms) interacts in a mutually reinforcing way with other forms, such as citations, prizes, and prestigious appointments and grants.

The only form of credit that academics have significant individual control over is the production of academic contributions, and the submission of these contributions to journals. Given the importance of credit, this leads to an intense pressure to produce and publish research output, the 'publish or perish' culture (Fanelli [2010]; Brischoux and Angelier [2015]).

Academics facing this pressure might look to take shortcuts to increase their productivity. I use the term 'questionable research practices' (QRPs) for such shortcuts. For my purposes, I regard all of the following behaviours as QRPs (some of which aim to improve productivity directly, while others aim to increase the impact of publications): fabricating data and other forms of outright fraud (Bruner [2013]; Bright [2017]); using multiple model specifications but only reporting those in which a result is statistically significant, known as *p*-hacking (Simmons et al. [2011]); hypothesizing after the results are known (Kerr [1998]); distributing findings from a single study over multiple papers, called salami publishing (Abraham [2000]); general sloppiness due to the desire to complete projects quickly and thus 'rushing into print' (Heesen [2018]); and being named as author on work where one has made no substantial intellectual contribution, that is, honorary co-authorship (Flanagin et al. [1998]).

A few quick clarifications. First, not all QRPs are equally bad. Most would agree that outright fraud is the worst one, and for some the jury is still out on whether they should be regarded as bad at all. But when it comes to the incentive to engage in them, these QRPs may be treated equally. That is, whenever this article identifies scenarios in which academics have an incentive to engage in QRPs, this applies to all of the foregoing behaviours. Emphasizing the worst-case outcomes that my argument supports, I sometimes summarize my findings in terms of an incentive to commit fraud.

Second, when I say that an academic has an incentive to commit fraud or would be rational to do so, I do not mean to imply that academics go through a conscious reasoning process anything like the analysis I provide here. This would be unrealistic for all or most academics. For my purposes, an incentive to commit fraud exists whenever an academic would in fact be better off (in expectation or probability) if they committed fraud as opposed to making some other salient choice, regardless of whether the academic is aware of this. As I discuss in more detail in section 5, selection effects entail that incentives have predictive power even if no one knows they exist, as long as we assume some diversity of heuristics among agents (this could include some random experimentation, some success-based imitation, and some diversity of risk tolerance).

3. Modelling Cumulative Advantage

I now construct a relatively simple model of the credit economy with cumulative advantage built in at its core. I take the simplified nature of the model to be a strength, as it focuses attention on a small number of features of the credit economy and their consequences. Combining this with empirical data regarding academics' incentives and behaviours yields a reasonable degree of confidence that the patterns of incentives identified here also operate in the real world.

My model differs in a number of ways from others in the literature. First, it is explicitly dynamic, in contrast to earlier static models of the credit economy (Kitcher [1990]; Dasgupta and David [1994]; Strevens [2003]; Zollman [2018]), although by now plenty of dynamic models exist as well (Smaldino and McElreath [2016]; O'Connor [2019]; O'Connor and Bruner [2019]; Zollman [unpublished]). Second, it uses continuous time rather than discrete time units. Where relevant, previous models have tended to use discrete time (Boyer [2014]; Zollman [unpublished]). Continuous time is more realistic and, perhaps surprisingly, also more mathematically tractable (compare Heesen [2017b] to Boyer [2014] and the present article to Zollman [unpublished]). Third, rather than assuming that academics maximize expected credit, I take academics' aim to be to satisfice relative to particular credit thresholds (partially inspired by tournament theory; see Lazear and Rosen [1981]; Hvide [2002]). I will motivate this modelling choice in section 5. This makes my model unique among those that look at academic incentives in a rational choice model (as opposed to an evolutionary model, where an analogous move has been made by Smaldino and McElreath [2016] and O'Connor [2019]).

As already noted, publications play a central role in the academic world and in the way credit is distributed. The basis of the present model is a stochastic counting process that keeps track of the publications of a given academic over a period of time. The idea is that an academic's productivity (both the number of publications and their distribution in time) has both a random component and a systematic component. The random component stands in for all factors affecting productivity that are not explicitly modelled, such as extraneous circumstances in the academic's life or the difficulty of the particular scientific problem she is working on. The systematic component consists of the academic's talent and skill, as well as the amount of time and resources she has available. This includes cumulative advantage: an academic who has already been productive is more likely to be given time and resources that boost productivity.

To model the random component I use a Poisson process (see any textbook on stochastic models, such as Norris [1998], sec. 2.4; Tijms [2003], chap. 1). In a Poisson process, the time between two publications is an exponentially distributed random variable. Moreover, the time between any two publications is probabilistically independent of all events before the first of these two publications. In other words, the 'interarrival' times of publications are independent and identically distributed.

Why model the random component in this way? The Poisson process has the following important feature: looking backwards the publications generated by it will appear to be randomly scattered in time. More precisely, given that a Poisson process produced a specific number of publications in a given time interval, the conditional probability distribution for when each of these publications arrived is the uniform distribution (Tijms [2003], theorem 1.1.5). So there is a precise sense in which this is a truly random model. As Norris ([1998], p. 73) puts it, 'a Poisson process is the natural probabilistic model for any uncoordinated stream of discrete events in continuous time'. That said, this feature will no longer hold once I add the systematic component to the model.

I have modelled academic productivity as a Poisson process in previous work. Viewing the present article in this light highlights another virtue of the Poisson model: it is very flexible. It can be used to analyse a number of trade-offs that academics face: sharing as against secrecy (Heesen [2017b]), speed as against accuracy (Heesen [2018]), high-risk high-reward as against incremental progress (Heesen [2019]), and the risk–reward trade-off of questionable or fraudulent methods. The shared modelling framework of credit maximization under Poisson productivity provides some confidence that the conclusions drawn in these articles cohere with each other (compared to the potentially contradictory assumptions made if a new model is designed from scratch for each trade-off).

A Poisson process has one parameter, usually denoted λ. It is interpreted as the rate of publication, that is, the expected number of publications per unit time. This can be used to add a systematic component to the model. For example, the publication output of two academics might be modelled using two Poisson processes with parameters λ_1 and λ_2, with $\lambda_1 > \lambda_2$ to indicate that the first academic has more time and resources and is therefore expected to be more productive.

Using the parameter in this way allows me to model persistent productivity differences between academics (see also Heesen [2017b]). However, I also want to capture cumulative advantage—the effect of earlier output on later productivity. This requires the systematic component to vary dynamically and endogenously. For this purpose I use a non-homogeneous Poisson process (also known as a non-stationary Poisson process; see Tijms [2003], sec. 1.3), which is like a regular Poisson process except that the rate of publication $\lambda(t)$ is a function of time.

I assume that publications generate credit over time (it takes time for a new publication to have its influence as word spreads, other academics start citing it, and so on). The credit accumulated by an academic may be turned into time and resources for research (Latour and Woolgar [1986], chap. 5). This is a multi-faceted process involving higher chances of winning research grants, a job at a more prestigious university, more graduate students, and more. Rather than attempt to model this explicitly, I assume that credit buys time and resources directly.

Suppose that each publication generates c_g units of credit per unit time (the subscript stands for 'good'; a notion of 'bad' credit will be introduced below). For an academic with p_g publications this means she accumulates credit at a rate of $c_g \cdot p_g$ per unit time. I use $c(t)$ to denote the credit accumulation rate and $C(t)$ for the total amount of credit accumulated at time t, so

$$c(t) = c_g \cdot p_g \quad \text{and} \quad C(t) = \int_0^t c(u)du. \tag{1}$$

By turning this into time and resources, the academic increases her rate of publication. The following formula captures this effect:

$$\frac{d}{dt}\lambda(t) = \log(c(t) + 1). \tag{2}$$

That is, the rate of publication increases (over time) proportionally to the logarithm of the credit rate (the $+1$ makes sure that if $c(t) = 0$ then $(d/dt)\lambda(t) = 0$). The logarithm reflects a type of decreasing marginal returns: if the credit accumulation rate is already high then the effect (on the publication rate) of increasing it further is smaller. The underlying idea is that large amounts of credit are harder to effectively turn into resources.

The credit rate $c(t)$ is a step function: it is constant between publications and jumps instantaneously from $c_g \cdot p_g$ to $c_g(p_g + 1)$ when a new publication appears. As a result, the rate of publication changes linearly between publications: if T_0 and T_1 are consecutive publication arrival times then for $T_0 < t < T_1$ the rate of publication is

$$\lambda(t) = \lambda(T_0) + \log(c(T_0) + 1)(t - T_0). \tag{3}$$

It is useful to see an example of how the rate of publication develops. To this end I have implemented the model described above in R (R Core Team [2020]). A sample trajectory for λ is shown in figure 1.[1]

The rate of publication stays flat until the first publication comes in just before $t = 1$, then increases in steepness as more publications arrive. Publications arrive closer together as the rate of publication increases. Thus publications and the rate of publication mutually reinforce one another, producing the cumulative advantage effect.

In this example, the academic accumulates $C(5) = 22.74$ units of credit (one unit per publication from the moment of publication until the end of the simulation at $t = 5$). This may vary because of the stochastic nature of the process, but repeated

[1] All code used in this article is available at (Heesen [2021]).

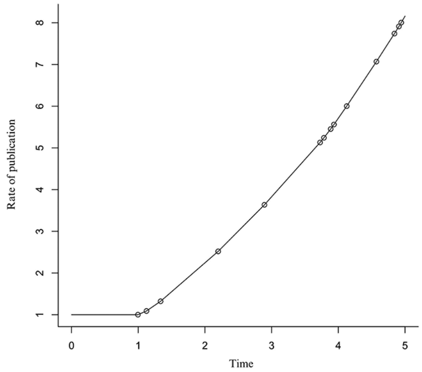

Figure 1. The rate of publication increases faster as more publications (indicated by dots) come in (parameters: $c_g = 1$ and $\lambda(0) = 1$).

simulation runs give a sense for the typical outcomes. In this case average credit at $t = 5$ is $\mathbb{E}[C(5)] \approx 27.93$. There is significant variation though: in 10,000 runs the minimum credit was zero, the maximum 112.77, and the standard deviation 15.74.

4. Replications, Exposures, and Negative Credit

The model described above assumes that each publication generates credit indefinitely. There are a number of respects in which this may be unrealistic. First, the impact of most publications fades over time. Since my aim is to model relatively short intervals of time (for example, from being hired to going up for tenure, or from starting graduate school to going on the job market), I will ignore this factor—though my model is an instance of a more general model, known as a Hawkes process or self-exciting process, which can incorporate this factor. Second, and more immediately relevant, fraudulent or shoddy work may be exposed, and even research of the highest standard may fail to replicate. As recent studies have shown, significant proportions of published results in various sciences fail to replicate (Open Science Collaboration [2015]; Nosek and Errington [2017]; Camerer et al. [2018]). When this happens, it changes how the original work and its author(s) are perceived and thus the credit associated with that publication.

To incorporate this in the model, assume that each publication has a chance of being 'exposed'. This may mean a failure to replicate, a discovery that the work was fraudulent, or any other event with significant negative impact on the perception of the work. In particular, I introduce a new parameter, the publication exposure rate μ, and assume that for each publication the time between it being published and it being exposed is exponentially distributed with rate μ. The probability of a publication never being exposed equals the probability that this exponential distribution fails to trigger in the time window under consideration (this assumption is unrealistic if the window is long, as it entails that every publication gets exposed eventually and

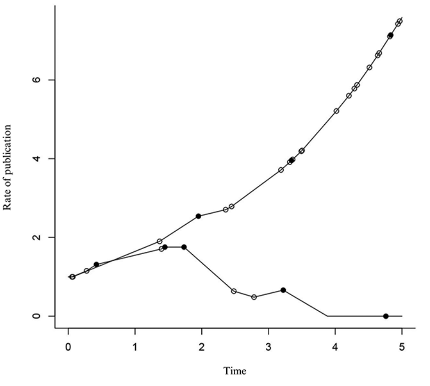

Figure 2. Two simulation runs (one in which publications outrun exposures and one in which they do not) with parameters $c_g = 1$, $c_b = 1/2$, $\lambda(0) = 1$, and $\mu = 1/4$. Publication events are marked as white dots and exposure events as black dots.

that the chance of exposure for an unexposed paper does not depend on its age). As a result, the stochastic process that counts exposure events for a given academic is a non-homogeneous Poisson process, with a total exposure rate at any given time of $p_g \cdot \mu$ (the number of publications available to be exposed times the exposure rate).

Once a publication is exposed, it is removed from the set of publications available to be exposed (p_g is reduced by one) and stops generating c_g units of credit per unit time. Instead it is added to the set of exposed ('bad') publications (p_b is increased by one) and starts losing c_b units of credit per unit time. The idea is that as news of the exposure spreads through the academic community, credit is taken away from the academic whose paper has been exposed, adjusting the total amount of credit generated by this publication downwards. The credit accumulation rate function c is adjusted to reflect this:

$$c(t) = c_g \cdot p_g - c_b \cdot p_b. \tag{4}$$

The credit accumulation rate may now decrease or even become negative. This requires adjusting the formula for changes in the rate of publication as well, since the logarithm is not defined for negative numbers:

$$\frac{d}{dt}\lambda(t) = \begin{cases} \log(c(t) + 1) & \text{if } c(t) \geq 0, \\ c(t) & \text{otherwise.} \end{cases} \tag{5}$$

Figure 2 illustrates this increased range of possibilities (see Heesen [2021]). It shows the development of the rate of publication for two simulation runs. In the former, the occasional exposure slows down the rate of publication, but on the whole publications come in fast enough so the rate of publication continues to increase. In the latter, there are fewer publications and more exposures, with the publication rate eventually dropping down to zero. In the former, the academic accumulates

$C(5) = 19.21$ units of credit, whereas in the latter, the academic ends up with negative credit ($C(5) = -2.11$). With these parameters, the former outcome is more typical: after 10,000 runs average credit is $\mathbb{E}[C(5)] \approx 14.65$ (standard deviation 10.50), with only 6.16% of runs accumulating zero credit or less.

5. What Are Academics' Incentives?

So far, questionable research practices have not taken centre stage. Such practices allow the academic to work more quickly and will lead to higher impact publications as she is able to achieve more surprising or more newsworthy results. Under what circumstances might academics have an incentive to engage in fraud or QRPs?

Suppose an early career academic faces the choice whether to engage in QRPs. Regular (non-fraudulent or 'honest') publications yield credit at a rate of $c_g^* = 1$ whereas publications obtained using fraud or QRPs yield more: $c_g^\dagger = 1.25$. This captures the fact that the latter tend to have higher impact. Indirectly, it also captures the fact that QRPs allow the academic to work more quickly, as the higher credit obtained is converted into an increased rate of publication.

The use of QRPs comes at a cost. Publications acquired in this way are more likely to be exposed: μ is higher. Once exposed, they are also punished more harshly as people recognize not only that the result is wrong, but that bad methods were used to obtain it: c_b is higher. There are important nuances here: sometimes academics are falsely accused of fraud (as was probably the case with Fisher's accusations against Mendel, or Newton's against Leibniz) and sometimes fraudulent work is recognized as irreproducible but not fraud. Moreover, large collaborations make it harder to detect fraud and to adjust individual credit in response (Huebner and Bright [2020], p. 364; see also Wray [2017], p. 129). Still, a publication is more likely to be exposed as fraudulent if it is in fact fraudulent, so the assumption that c_b is higher if QRPs are used seems apt. To capture this with (fairly arbitrary) numbers, suppose $\mu^* = 1/6$ whereas $\mu^\dagger = 1/4$, and $c_b^* = 0$ whereas $c_b^\dagger = 1/2$.

These choices of parameter values apparently favour honest academic work over QRPs: a relatively modest (25%) increase in credit from publications seems to be more than offset by the increased exposure rate and the associated negative credit. To substantiate this, I have worked out (see the appendix) the credit expected to accrue to a single publication up to a given time t_1. For an academic with only a single publication at time t_0 (the first publication arrives at $T_0 = t_0$ and the second publication arrival time T_1 is after t_1), expected credit is

$$\mathbb{E}[C(t_1)|T_0 = t_0, T_1 > t_1] = \frac{c_g + c_b}{\mu}\left(1 - e^{-\mu(t_1 - t_0)}\right) - c_b(t_1 - t_0). \tag{6}$$

With the parameters above and publication at time $t_0 = 0$ (I will continue to use $t_1 = 5$ in all examples), the honest academic expects to get more credit than the fraudulent one:

$$\mathbb{E}[C^*(5)\big|T_0 = 0, T_1 > 5] \approx 3.39 > 2.49 \approx \mathbb{E}[C^\dagger(5)\big|T_0 = 0, T_1 > 5].$$

Similarly, if the publication arrives at $t_0 = 1$ the honest academic expects 2.92 units of credit from it and the fraudulent one merely 2.42. So from the perspective of expected credit per publication, it is better to be honest (although this result eventually flips if the publication arrives close to t_1, as such publications are less likely to be exposed in the remaining time; but such publications contribute less overall as they are around less long).

Considering only credit for a single publication is misleading, however, as this credit is not evenly distributed over time. Whereas the honest academic expects a steady stream of credit that lasts for a while, the fraudulent academic expects more credit initially, but also for a shorter period, followed by a period of negative credit. While the fraudulent academic ends up with less credit at time t_1, she can use the higher early credit to increase her rate of publication. She might be able to offset the later negative effects by producing (many) more publications overall. The pattern of credit accumulation by the fraudulent academic then resembles a Ponzi scheme (Zollman [unpublished]).

So because of cumulative advantage, QRPs might pay off when the expected credit of all publications is considered, even though the honest strategy is better from the perspective of credit per publication. Once again I estimate expected credit by simulating the process 10,000 times. In the running example this yields

$$\mathbb{E}[C^*(5)] \approx 21.52 \quad \text{and} \quad \mathbb{E}[C^\dagger(5)] \approx 21.28.$$

Despite cumulative advantage, with these parameters it is still slightly better to be an honest academic from the perspective of maximizing expected credit.

In models like this one, it is typically assumed that academics' goal is to maximize expected credit. This makes a certain amount of sense, given the close analogy between credit (in motivating academics) and utility (in motivating arbitrary rational agents), and the role of expected utility theory as the standard model of rational choice. For most academics, however, it is arguably more important to meet specific credit thresholds. The competitive aspects of academic life are felt most keenly at a few pivotal career moments, such as when the academic is on the job market or going up for tenure.

On the job market, the credit the academic has accumulated is likely to play an important role in her prospects. What matters in that moment is whether the academic has accumulated enough credit: enough to be competitive, enough to land that dream job, enough to achieve whatever goal she has set for herself. Simplifying significantly (and ignoring many other factors such as teaching competence, personality, or how fashionable her research area is), the academic's goal may be formulated as a target amount of credit. Meeting or exceeding this target constitutes success, and falling short failure.

Going up for tenure is similar, though typically not explicitly competitive. The academic is given a set of (possibly vague) criteria she needs to meet. As far as the research component of these criteria goes, the reputation the academic has built (her accumulated credit) is the most important factor, if not the only one. In order to get tenure, she needs to meet or exceed a credit threshold (here and in the next section, I consider a single threshold; expanding the time horizon slightly, the academic might need to consider two thresholds: first to get a job and then tenure; I consider this variation in the appendix).

If getting a job or tenure is far more important to the academic than any other credit-related goals, then it would be rational to aim to maximize the probability of meeting the relevant credit threshold. We can capture this in a utility function in a number of ways depending on what 'far more important' means. If the academic only cares about having a job, her utility function would be a step function: one if the credit threshold is met, zero otherwise. If the academic only cares about maximizing credit if she has a job, her utility function would be lexicographic, with the precise amount of credit only being relevant if the threshold is met. If the academic cares a little bit about credit even if she does not have a job, her utility function might be a weighted average of the step function and accumulated credit (weighted heavily towards the step function). In the first two cases, maximizing expected utility equates to maximizing the probability of meeting the threshold, and in the third case, this holds approximately.

Setting aside considerations of rationality momentarily, the job market and the tenure process are important points at which it is decided who stays in academia and who leaves (other points may be just as important in determining one's place in the academic hierarchy, but having any such place at all is only possible if one stays in academia). If those who meet credit thresholds stay and those who fail to meet them leave, then academia effectively selects those academics who maximize their chances of meeting thresholds. This has been noted by Smaldino and McElreath ([2016]) and O'Connor ([2019], p. 27). Analogous phenomena have shown up elsewhere in evolutionary models (for example, Robson [2002], especially note 8).

The strategy that maximizes the probability of meeting a threshold need not be the same as the strategy that maximizes expected credit. In the model, an academic who chooses QRPs introduces more random variation in the amount of credit she accumulates than if she did honest work. Even when QRPs decrease expected credit, they increase the variance. It is possible that QRPs increase the probability of meeting a threshold despite lowering the mean, provided the threshold is relatively high.

This is exactly what happens in the running example. If the threshold the academic aims to meet is, say, twenty-five, then it is better to use QRPs:

$$\Pr(C^*(5) \geq 25) \approx 0.3655 \quad \text{and} \quad \Pr(C^\dagger(5) \geq 25) \approx 0.3708.$$

Here, doing honest academic work gives her a chance of meeting the threshold of about 36.5%, but with QRPs her chances are just over 37%. The difference increases if the threshold is raised: with a threshold of thirty, the honest academic's chances are 24.2%, but QRPs raise this to 26.6%. With a threshold of fifty, the respective chances are 2.1% and 3.5% (see Heesen [2021]).

A high threshold corresponds to a low probability of success. This may be a feature of the competitive process (a job market where few academics get a job, or a tenure process in which only a fraction of candidates are given tenure), or a feature of the specific academic (given her own dispositions, background, and training, she is from the outset relatively unlikely to get a job or tenure), or both.

In the model, other academics are not explicitly represented: this is a decision-theoretic rather than a game-theoretic model. In a game, one might represent competition by assuming that academics have to beat each other's credit totals instead of a fixed threshold. This yields a 'tournament model' (Lazear and Rosen [1981]). The phenomenon that agents might rationally choose to reduce their expected output in favour of increasing its variance has been observed in tournament models as well (Hvide [2002]).

6. How Common Is the Incentive to Commit Fraud?

Above I considered a single academic facing the choice to commit fraud (or more generally use QRPs) or do honest research. I highlighted two phenomena. First, the possibility that fraud may be rational for the academic even when the expected credit of individual honest publications is higher, due to cumulative advantage. Second, the possibility that fraud may be rational when the expected credit of honest research is higher (even after taking cumulative advantage into account), if the academic's goals require beating a relatively high credit threshold. The previous section may be interpreted as a proof of possibility: it suggests these phenomena may occur, but says nothing about how often they do.

To say something a bit stronger, I now investigate these phenomena a little more systematically as they arise (or fail to) under different parameter settings. If they turn out to arise robustly in the model, this is not sufficient evidence to conclude they commonly arise outside of the model as well, but it is suggestive, especially if one has been persuaded by the preceding sections that the model captures important qualitative features of the incentive structure of academic science.

The parameter settings considered in this section are as follows. First, I fix the time scale by ending all simulations at $t = 5$. For the intended interpretation of a graduate school education or a tenure clock, this means one unit is roughly equal to a year. This is a harmless assumption, as I could set the simulations to end, say, at $t = 60$, interpret this in months, rescale the other parameters appropriately, and get exactly the same results. For the other variables I pick a range. For the initial publication rate $\lambda(0)$, this runs from 0.5 to 2 in increments of 0.5 (interpretation: academics vary in their initial average productivity between half a paper and two papers per year). For honest academics, the credit accumulation rate for non-exposed papers c_g^* is either one or two, the exposure rate μ^* ranges from 0 to 0.25, and the negative credit for exposed papers c_b^* from 0 to 0.2.

The fraudulent academic expects to get more credit from her papers in the short run, but they are more likely to be exposed and accrue more negative credit when this happens. The former is implemented by increasing the credit accumulation rate $c_g^\dagger$ by a percentage (ranging from 10% to 60%) relative to the honest academic's rate c_g^*. The exposure rate $\mu^\dagger$ is set to be between 0.05 and 0.25 higher than that of the honest academic. And the negative credit $c_b^\dagger$ varies from 0.3 to 0.5 (so all possible values for the fraudulent strategy are higher than all possible values for the honest one). This is summarized in table 1. Results are estimated based on 10,000 runs for each setting.

I focus on the effect of the (extra) credit for non-exposed fraudulent papers. The first result is that even when this is at its lowest ($c_g^\dagger = 1.1 \cdot c_g^*$, a 10% credit premium for fraud), there is a non-negligible range of values of the other parameters for which fraud is a better strategy than honesty in expectation. To state this more precisely, note that if we fix $c_g^\dagger = 1.1 \cdot c_g^*$ there are 864 possible combinations of values of the other parameters listed in table 1 (though this does involve some double counting because when $\mu^* = 0$ the value of c_b^* has no effect). Of these combinations, there are 147 (about 17%) for which fraud is the best strategy in expectation ($\mathbb{E}[C^\dagger(5)] > \mathbb{E}[C^*(5)]$). In the remaining 717 cases, the honest strategy is better in expectation when $c_g^\dagger = 1.1 \cdot c_g^*$.

As the value of $c_g^\dagger$ increases, fraud becomes more attractive: the number of parameter settings for which the fraudulent academic expects higher credit than the honest academic gradually increases from 17% when $c_g^\dagger = 1.1 \cdot c_g^*$ to over 99% when $c_g^\dagger = 1.6 \cdot c_g^*$. This is shown by the solid lines and black dots in figure 3. Similarly, the number of parameter settings where the fraudulent strategy has a greater probability of exceeding a credit threshold of thirty increases as $c_g^\dagger$ increases (dashed lines and grey dots in fig. 3), as it does when the credit threshold is fifty (dot-dashed lines and white dots in fig. 3). When the credit threshold is fifty and $c_g^\dagger = 1.6 \cdot c_g^*$, the fraudulent strategy has a greater probability of exceeding the threshold than the honest strategy in all 864 cases (Heesen [2021]).

Table 1. Parameter values used in this section

Parameter	Values
$\lambda(0)$	$\{0.5, 1, 1.5, 2\}$
c_g^*	$\{1, 2\}$
c_b^*	$\{0, 0.1, 0.2\}$
μ^*	$\{0, 0.05, 0.15, 0.25\}$
$c_g^\dagger$	$c_g^* \cdot \{1.1, 1.2, 1.3, 1.4, 1.5, 1.6\}$
$c_b^\dagger$	$\{0.3, 0.4, 0.5\}$
$\mu^\dagger$	$\mu^* + \{0.05, 0.15, 0.25\}$

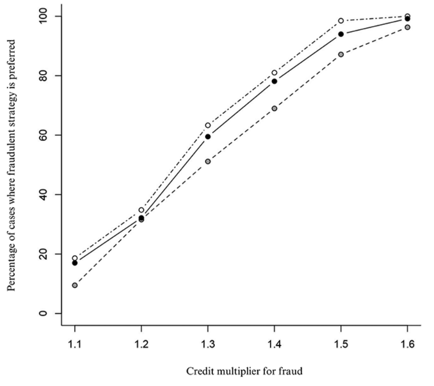

Figure 3. Percentage of parameter settings for which the fraudulent strategy is preferred over the honest strategy as a function of the ratio $c_g^\dagger / c_g^*$. Solid lines and black dots indicate the percentage of cases where $\mathbb{E}[C^\dagger(5)] > \mathbb{E}[C^*(5)]$; dashed lines and grey dots indicate the percentage where $\Pr(C^\dagger(5) \geq 30) > \Pr(C^*(5) \geq 30)$; dot-dashed lines and white dots indicate the percentage where $\Pr(C^\dagger(5) \geq 50) > \Pr(C^*(5) \geq 50)$. Table 2 in the appendix shows the same data numerically.

This idea generalizes. Regardless of the value of the other parameters, if the credit premium for non-exposed fraudulent papers is large enough, the fraudulent academic is better off than the honest academic. This holds regardless of whether 'better off' is cashed out in terms of expected credit or in terms of the probability of meeting a credit threshold (a proof is given in the appendix).

Theorem 1: Let $\lambda(0) > 0$ and $t_1 > 0$. For all values of c_g^*, c_b^*, μ^*, $c_b^\dagger$, and $\mu^\dagger$, there exist values of $c_g^\dagger$ large enough such that $\mathbb{E}[C^\dagger(t_1)] > \mathbb{E}[C^*(t_1)]$. Moreover, for any credit threshold $\theta > 0$ there exist values of $c_g^\dagger$ large enough such that $\Pr(C^\dagger(t_1) \geq \theta) > \Pr(C^*(t_1) \geq \theta)$.

The simulation results and the theorem are suggestive. They show the credit incentive for fraud is not an isolated phenomenon, at least within this model. Rather, such an incentive arises systematically whenever the credit premium for non-exposed fraudulent papers is large enough. The theorem shows this in principle, and the simulation results show that the values of $c_g^\dagger$ required are not always unrealistically high.

In section 5, I highlighted the possibility that fraud can be incentivized even when the expected credit of honesty is higher, as career success may require beating a threshold. While the theorem does not speak to this directly, the simulation results provide some support for this. Across 5,184 combinations of parameter settings, there were 233 instances (about 4.5%) where $\Pr(C^\dagger(5) \geq 50) > \Pr(C^*(5) \geq 50)$ even though $\mathbb{E}[C^*(5)] > \mathbb{E}[C^\dagger(5)]$ (see Heesen [2021]).

What does this tell us outside the model? There is first the general question whether the model captures the right dynamics to have any relevance to real academics. I have argued for this throughout the construction of the model. But there is a second, more specific question: what should we make of the condition that the credit premium for non-exposed fraudulent papers is 'large enough'?

My claim is that in a given community, there are at least some academics for whom there is a credit incentive to use QRPs or fraud, because the parameters of the model are different for each academic. While some relevant factors are largely fixed within an academic community (for example, the chance that a fraudulent paper is exposed or the amount of credit one needs to have a chance at a job or tenure), others depend on the skills, dispositions, and luck of specific academics (such as an academic's productivity with a fixed level of resources). Importantly, the credit premium for non-exposed fraudulent papers has aspects of both. It partially depends on the community (through the value this community assigns to 'surprisingness' or 'flashiness'), but also on the academic's ability to dress up shoddy work, advertise its virtues, hide its weaknesses, and quickly convert credit into productive resources.

If there is variety in the parameter values experienced by different members of an academic community, the credit premium for non-exposed fraudulent papers will be relatively high for some. The simulation results and the theorem suggest that these academics may have an incentive to use QRPs or fraud.

Moving beyond a proof of possibility, this section provides tentative evidence for a stronger claim. The claim is that in most (if not all) academic communities, some academics have an incentive to use QRPs or fraud as a direct result of the need to accumulate credit to get a job or tenure.

As noted, QRPs tend to increase the variance in how much credit an academic accumulates. Roughly, this means that academics using QRPs are more likely to do either very well or very poorly. These academics are then likely to be over-represented at the bottom and the top of the credit distribution. So in those communities where some academics have an incentive to use QRPs, the most famous academics are likely to be the ones using them.

The simulation results illustrate this phenomenon. Suppose that the parameter ranges in table 1 describe the variety among academics of a given community. The most famous academics in that community are those who accumulate the largest amount of credit. We can get a sense for who this might be from the highest credit totals realized across all simulation runs (since there are 5,184 parameter settings with 10,000 simulation runs each for the fraudulent and the honest strategies, this involves more than a hundred million data points).

The maximum of 458.0 units of credit is realized under the fraudulent strategy, with $\lambda(0) = 2$, $c_g^\dagger = 3.2$, $c_b^\dagger = 0.5$, and $\mu^\dagger = 0.05$ (see Heesen [2021]). In contrast, the highest credit achieved across all simulation runs using the honest strategy is 302.1 units. This is arguably not surprising given that the fraudulent strategy does better than the honest strategy by most measures when $c_g^\dagger = 1.6 \cdot c_g^*$. The following result is more surprising: even among settings with $c_g^\dagger = 1.1 \cdot c_g^*$, there is a simulation run that achieves 327.9 units of credit (incidentally, this suggests that the fraudulent strategy would also be attractive in an explicitly game-theoretic tournament model, with all or most of the payoff going to the academic achieving the highest credit total, as mentioned in section 5).

So, with these parameter ranges, even if the credit premium for non-exposed fraudulent papers is restricted to 10%, the most famous academics are likely fraudulent. These extremely successful academics are those who have gotten lucky, in that few or none of their papers have been exposed. (Elsewhere, I reach a similar conclusion in a different model with a different notion of luck; see Heesen [2017a].)

While I have argued that the model gives a reasonable approximation of the pressures faced by real academics, caution remains advisable in applying these findings. One should arguably always check the robustness of the results (Frey and Šešelja [2018]; Heesen et al. [2019]). This can be done by changing the parameter values, changing specific modelling assumptions, or with an entirely different modelling framework. I suspect the first two approaches will yield few surprises, as the key results hinge only on the fact that the fraudulent strategy has higher variance than the honest strategy. Nevertheless, one should be particularly cautious with accusations or suggestions of misconduct in the absence of specific empirical evidence (Smith [2018], sec. 1).

It is difficult to test the hypothesis that more famous academics are more likely to be fraudulent empirically. This is because we have no independent means to detect undiscovered fraud, and samples based on discovered fraud are likely to exhibit selection bias (one reason why modelling work in this area is particularly valuable). Still, limited support for the hypothesis is provided by Stroebe et al. ([2012]), who look at a 'convenience sample' of cases of exposed fraud, mostly in psychology and biomedicine. They write: '[The following] pattern is quite typical for all of these cases. Either the researchers committing the fraud were highly respected, or in the case of young researchers, they published their work with highly respected senior colleagues' (Stroebe et al. [2012], p. 672).

7. Conclusion

I have highlighted two phenomena that favour the use of QRPs that become apparent in a dynamic model of the credit economy. First, cumulative advantage may allow a fraudulent academic to be successful even if the fraud does not pay off at the level of individual publications (Zollman [unpublished]). Second, selection events in an academic's career may lead fraudulent academics to be successful even if fraud does not pay off in expectation.

I emphasize two takeaways in this conclusion. First, the two highlighted phenomena suggest that the incentive to engage in QRPs or fraud may be stronger than it appeared based on previous models of the credit economy (Bruner [2013]; Bright [2017]; Heesen [2018]). Second, within my model the most successful academics tend to be the most unscrupulous ones: those who are willing to gamble on fraud and manage to get away with it. If the model accurately captures the dynamics and incentives related to cumulative advantage and fraud, it raises the worry that those academics who are praised for being the best actually work according to worse than average epistemic standards.

Appendix

A.1. Multiple credit thresholds

Here I consider a variation of the model in which the academic must meet two credit thresholds at different times, first to get a job and then to get tenure, say. To maximize the ability to compare results, I consider the same ranges of parameter values used previously (see table 1). The simulation still ends at $t = 5$, but in this variation each simulation run is stopped early if the academic fails to beat a credit threshold of five halfway through, so if $C(2.5) < 5$. The threshold value of five is chosen such that approximately half of all simulation runs beat it.

Once again I consider whether the honest or the fraudulent strategy has a better chance of beating a credit threshold of thirty and a threshold of fifty. Table 2 shows the results.

The introduction of an additional threshold halfway through appears to favour the fraudulent strategy. The percentage of cases where fraud is more likely to meet a threshold of thirty than honesty is higher with the intermediate threshold for each of the six values of the credit multiplier for fraud (rows 2 and 3 of table 2). With a threshold of fifty the percentages are similar in the two variations of the model (rows 4 and 5). I hypothesize that there are almost no cases in the original model where the threshold of fifty is met at $t = 5$ without also meeting the intermediate threshold, which would explain the highly similar results. Probably for the same reason, looking at the highest credit totals realized across all simulation runs yields similar results with the intermediate threshold as without it: the overall maximum in this variation is 515.4, the maximum using the honest strategy is 297.1, and

Table 2. Percentage of parameter settings for which the fraudulent strategy is preferred over the honest strategy as a function of the ratio $c_g^{\dagger}/c_g^{*}$

$c_g^{\dagger}/c_g^{*}$	1.1	1.2	1.3	1.4	1.5	1.6
$\mathbb{E}[C^{\dagger}(5)] > \mathbb{E}[C^{*}(5)]$	17.01	32.18	59.49	78.13	93.98	99.19
$\Pr(C^{\dagger}(5) \geq 30) > \Pr(C^{*}(5) \geq 30)$	9.49	31.60	51.16	68.98	87.15	96.30
$\Pr(C^{\dagger}(5) \geq 30, C^{\dagger}(2.5) \geq 5) >$ $\Pr(C^{*}(5) \geq 30, C^{*}(2.5) \geq 5)$	19.21	38.89	63.66	85.07	97.22	99.77
$\Pr(C^{\dagger}(5) \geq 50) > \Pr(C^{*}(5) \geq 50)$	18.63	34.84	63.31	81.02	98.50	100
$\Pr(C^{\dagger}(5) \geq 50, C^{\dagger}(2.5) \geq 5) >$ $\Pr(C^{*}(5) \geq 50, C^{*}(2.5) \geq 5)$	18.98	37.04	63.77	82.75	97.80	100

Note.—Rows 1, 2, and 4 reproduce the data from figure 3; rows 3 and 5 show the data for the variation of the model with an intermediate threshold (Heesen [2021]).

the maximum with the fraudulent strategy under the restriction that $c_g^\dagger = 1.1 \cdot c_g^*$ is 310.3 (Heesen [2021]).

Note that I have omitted consideration of expected total credit in the variation of the model. This is because this concept has no unambiguous meaning here. One could calculate the expected total credit given that the intermediate threshold is met, or an overall expectation where runs that fail to meet the intermediate threshold are included, either with the total credit achieved at $t = 2.5$ or with some deemed value. None of these are straightforwardly comparable to the expectation calculated for the model without an intermediate threshold.

A.2. Proof of theorem 1

Let $\underline{C}(t)$ denote the credit directly associated with a single publication, or equivalently, the credit that would be accrued up to time t if we assumed that the publication rate λ drops to zero and stays there immediately after the first publication arrives. Let $T \sim \text{Exp}(\lambda(0))$ be the arrival time for that publication. Let $X \sim \text{Exp}(\mu)$ be the waiting time until the publication is exposed (so that exposure occurs at time $T + X$).

We first consider the expected credit conditional on $T = t < t_1$. The density function of X is given by $f_X(x) = \mu e^{-\mu x}$. If $X \geq t_1 - t$ the credit accrued is $c_g(t_1 - t)$, otherwise it is $c_g X - c_b(t_1 - t - X)$. So

$$\begin{aligned}\mathbb{E}[\underline{C}(t_1)|T = t] &= \int_0^\infty (c_g \min\{t_1 - t, x\} - c_b \max\{t_1 - t - x, 0\}) f_X(x) dx \\ &= \frac{c_g + c_b}{\mu}\left(1 - e^{-\mu(t_1 - t)}\right) - c_b(t_1 - t).\end{aligned}$$

A few observations. First, this justifies equation 6. Second, $\mathbb{E}[\underline{C}(t_1)|T = t]$ is a linearly increasing function of c_g. Consequently, we can guarantee that the expectation is positive by choosing c_g large enough:

$$\mathbb{E}[\underline{C}(t_1)|T = t] > 0 \quad \text{if and only if} \quad c_g > c_b \frac{\mu(t_1 - t)}{1 - e^{-\mu(t_1 - t)}} - c_b. \tag{7}$$

Third, if $\mathbb{E}[\underline{C}(t_1)|T = t_0] > 0$ then for all $t \in (t_0, t_1)$ also $\mathbb{E}[\underline{C}(t_1)|T = t] > 0$. This can be seen from equation 7 by noting that $x/(1 - e^{-x})$ is a strictly increasing function for all x.

Now we consider $\mathbb{E}[\underline{C}(t_1)]$. The density function of T is given by $f_T(t) = \lambda(0)e^{-\lambda(0)t}$. If $T \geq t_1$ the credit accrued is zero, otherwise the expected credit is $\mathbb{E}[\underline{C}(t_1)|T]$. So

$$\begin{aligned}\mathbb{E}[\underline{C}(t_1)] &= \int_0^{t_1} \mathbb{E}[\underline{C}(t_1)|T = t] f_T(t) dt \\ &= \frac{\mu(c_g\lambda(0) + c_b\mu)(1 - e^{-\lambda(0)t_1}) - \lambda(0)^2(c_g + c_b)(1 - e^{-\mu t_1})}{\lambda(0)\mu(\mu - \lambda(0))} - c_b t_1.\end{aligned}$$

We are particularly interested in how $\mathbb{E}[\underline{C}(t_1)]$ varies with c_g:

$$\frac{d}{dc_g}\mathbb{E}[\underline{C}(t_1)] = \frac{\mu(1 - e^{-\lambda(0)t_1}) - \lambda(0)(1 - e^{-\mu t_1})}{\mu(\mu - \lambda(0))}.$$

So $\mathbb{E}[\underline{C}(t_1)]$ is also a linear function of c_g. To see that it is an increasing function of c_g it suffices to show that the derivative is positive. One way to see this is by interchanging the derivative and the integral (which is legitimate because the functions involved are differentiable and bounded):

$$\frac{d}{dc_g}\mathbb{E}[\underline{C}(t_1)] = \int_0^{t_1}\frac{d}{dc_g}\mathbb{E}[\underline{C}(t_1)|T = t]f_T(t)dt = \int_0^{t_1}\frac{1}{\mu}\left(1 - e^{-\mu(t_1-t)}\right)\lambda(0)e^{-\lambda(0)t}dt > 0$$

because the integrand is strictly positive for all $0 \leq t < t_1$.

How does $\mathbb{E}[\underline{C}(t_1)]$ relate to $\mathbb{E}[C(t_1)]$? Recall that $\underline{C}(t)$ only counts the credit associated with a single publication, whereas $C(t)$ tracks the credit for all publications combined. The probability distribution for the arrival time of each subsequent publication is (by design) quite complicated, as it depends on the number of previous publications, the number of previous exposures, and the precise arrival times of each of these. But conditional on its arrival time t, we know that the contribution each publication makes to the expected credit is equal to $\mathbb{E}[\underline{C}(t_1)|T = t]$. Moreover we know from the third observation above that if $\mathbb{E}[\underline{C}(t_1)|T = 0] > 0$ then $\mathbb{E}[\underline{C}(t_1)|T = t] > 0$ for all $t < t_1$. So it follows that if $\mathbb{E}[\underline{C}(t_1)|T = 0] > 0$ (which we can make sure is true by choosing c_g sufficiently high) then the contribution to the expected credit of each publication beyond the first is positive, and therefore $\mathbb{E}[C(t_1)] > \mathbb{E}[\underline{C}(t_1)]$.

Since $\mathbb{E}[\underline{C}(t_1)]$ increases linearly with c_g, we can make $\mathbb{E}[\underline{C}(t_1)]$ arbitrarily high by setting c_g sufficiently high. In particular, if $\lambda(0)$, t_1, c_g^*, c_b^*, and μ^* are fixed then $\mathbb{E}[C^*(t_1)]$ is thereby fixed as well. For arbitrary (fixed) values of $c_b^\dagger$ and $\mu^\dagger$ we can then choose $c_g^\dagger$ high enough such that $\mathbb{E}[\underline{C}^\dagger(t_1)] > \mathbb{E}[C^*(t_1)]$. If we also choose $c_g^\dagger$ high enough so the inequality from the previous paragraph obtains, we get the desired result regarding expectation:

$$\mathbb{E}[C^\dagger(t_1)] > \mathbb{E}[\underline{C}^\dagger(t_1)] > \mathbb{E}[C^*(t_1)].$$

Moving on to the second part of the theorem, we now consider the probability of exceeding a credit threshold $\theta > 0$, which requires a more detailed analysis. Let N (a random variable) be the total number of publications that occurs in the time interval $[0, t_1]$. The event $N = 0$ occurs just in case $T > t_1$, and it entails that $C(t_1) = 0$. The probability of this event is

$$\Pr(C(t_1) = 0) \geq \Pr(N = 0) = \Pr(T > t_1) = e^{-\lambda(0)t_1}.$$

Note that this probability does not depend on c_g. Assuming (as the statement of the theorem does) that $\lambda(0) > 0$ and $t_1 > 0$ are fixed, this establishes an upper bound on the probability of interest:

$$\Pr(C(t_1) \geq \theta) \leq \Pr(N \geq 1) = \Pr(T < t_1) = 1 - e^{-\lambda(0)t_1}.$$

In light of this, the following two claims are equivalent:

$$\lim_{c_g \to \infty} \Pr(C(t_1) \geq \theta) = \Pr(N \geq 1) \quad \text{and} \quad \lim_{c_g \to \infty} \Pr(C(t_1) \geq \theta | N \geq 1) = 1.$$

The bulk of this proof aims to establish the latter (which will allow us to infer the former). So we restrict attention to cases where $N \geq 1$. We begin by claiming that if $N \geq 1$, then we can choose c_g high enough that (with arbitrarily high probability) N is 'large'.

In fact, the claim we will prove is stronger. Let S denote the sum over all publications of the time interval from publication until t_1 (ignoring whether they are exposed). That is, if $T_0, T_1, \ldots, T_{N-1}$ are the publication times (where $T_0 = T$), then $S = \Sigma_{i=0}^{N-1}(t_1 - T_i)$. We show that with an appropriate choice of c_g, we can make S arbitrarily large with arbitrarily high probability. That is, not only are there arbitrarily many publications, but they do not all occur so close to t_1 that the combined time they are in existence remains small.

Let $\varepsilon > 0$. We want to show for some arbitrary (large) constant k that

$$\Pr(S \geq k | N \geq 1) = \frac{\Pr(S \geq k, N \geq 1)}{\Pr(N \geq 1)} = \frac{\Pr(S \geq k, T < t_1)}{\Pr(T < t_1)} > 1 - \varepsilon.$$

Let

$$t^* = \min\left\{\frac{1}{\lambda(0)} \log\left(1 + \frac{\varepsilon}{3}(e^{\lambda(0)t_1} - 1)\right), \frac{1}{\mu} \log\left(\frac{1}{1 - \varepsilon/3}\right)\right\} > 0.$$

Then

$$\begin{aligned} \Pr(S \geq k | N \geq 1) &\geq \frac{\Pr(S \geq k, T < t_1 - t^*)}{\Pr(T < t_1)} \\ &= \Pr(S \geq k | T < t_1 - t^*) \frac{1 - e^{-\lambda(0)(t_1 - t^*)}}{1 - e^{-\lambda(0)t_1}} \\ &\geq \left(1 - \frac{\varepsilon}{3}\right) \Pr(S \geq k | T < t_1 - t^*). \end{aligned}$$

Let N^* denote the number of publications that occur in the time interval $[T, T + (1/2)t^*]$. Conditional on $T < t_1 - t^*$, we know that $T + t^*/2 < t_1 - t^*/2$, that is, we have N^* publications occurring before $t_1 - t^*/2$, so $S \geq N^* \cdot t^*/2$. Thus if we let $n = 2k/t^*$ we have

$$\Pr(S \geq k \big| T < t_1 - t^*) \geq \Pr(N^* \geq n \big| T < t_1 - t^*).$$

Let G_2 denote the event that the first two publications fail to be exposed during the first $t^*/2$ time units after their publication (assuming these publications occur at all). Note that

$$\Pr(G_2) \geq e^{-\mu t^*} \geq 1 - \frac{\varepsilon}{3}.$$

Putting this together, so far we have

$$
\begin{aligned}
\Pr(S \geq k | N \geq 1) &\geq \left(1 - \frac{\varepsilon}{3}\right) \Pr(S \geq k | T < t_1 - t^*) \\
&\geq \left(1 - \frac{\varepsilon}{3}\right) \Pr(N^* \geq n | T < t_1 - t^*) \\
&\geq \left(1 - \frac{\varepsilon}{3}\right) \Pr(N^* \geq n, G_2 | T < t_1 - t^*) \\
&\geq \left(1 - \frac{\varepsilon}{3}\right)^2 \Pr(N^* \geq n, | G_2, T < t_1 - t^*).
\end{aligned}
$$

We now consider the probability in the last line above, which is the probability of at least n publications in the time interval $[T, T + (1/2)t^*]$ conditional on the first two publications not being exposed during that same time interval. Choose

$$
c_g = \max\left\{ n \cdot c_b, e^{(16n + 24/\varepsilon)/(t^*)^2} \right\}.
$$

For the purpose of determining the probability that $N^* \geq n$, we may assume that $c(t) \geq c_g$ for all $t \in [T, T + (1/2)t^*]$. This is because (a) after the first publication at time T, $p_g = 1$ and $p_b = 0$ so $c_g p_g - c_b p_b = c_g$, (b) because of the condition G_2, the second publication occurs before the first exposure, (c) after the second publication but before the n-th publication, because of condition G_2, $p_g \geq 2$ and $p_b \leq n - 2$ and hence $c_g p_g - c_b p_b \geq c_g + nc_b - (n - 2)c_b \geq c_g$, and ($d$) anything that happens after the n-th publication (including any exposures) cannot reduce the total number of publications that already occurred, and so is irrelevant to the probability that $N^* \geq n$.

It follows from equation 3 that the rate of publication is

$$
\lambda(t) \geq \lambda(0) + \log(c_g + 1)(t - T)
$$

for all $t \in [T, T + (1/2)t^*]$. This means that (at least as long as $N^* < n$) the 'average intensity' during the time interval is

$$
\Lambda\left(T, T + \frac{1}{2}t^*\right) = \int_T^{T+t^*/2} \lambda(t)dt \geq \frac{1}{2}\lambda(0)t^* + \frac{1}{8}\log(c_g + 1)(t^*)^2.
$$

Let Y be a Poisson-distributed random variable with rate parameter $\bar{\Lambda} = \lambda(0)t^*/2 + \log(c_g + 1)(t^*)^2/8$. Then $\Pr(N^* \geq n, | G_2, T < t_1 - t^*) \geq \Pr(Y \geq n)$ (see Tijms [2003], theorem 1.3.1). Note that

$$
\bar{\Lambda} \geq \frac{1}{8}\log(c_g)(t^*)^2 \geq 2n + \frac{3}{\varepsilon}.
$$

Since Y follows a Poisson distribution, $\mathbb{E}[Y] = \mathrm{Var}(Y) = \bar{\Lambda}$. Since $\mathbb{E}[Y] = \bar{\Lambda} > n$, we can apply Cantelli's inequality (a one-sided version of the Chebyshev inequality) to get

$$\Pr(Y \geq n) = \Pr(Y - \mathbb{E}[Y] \geq n - \mathbb{E}[Y]) \geq 1 - \frac{\mathrm{Var}(Y)}{\mathrm{Var}(Y) + (n - \mathbb{E}[Y])^2}.$$

Thus

$$\Pr(Y \geq n) \geq 1 - \frac{1}{\bar{\Lambda} + 1 - 2n + n^2/\bar{\Lambda}} \geq 1 - \frac{1}{\bar{\Lambda} - 2n} \geq 1 - \frac{\varepsilon}{3}.$$

But now we are done:

$$\begin{aligned}\Pr(S \geq k|N \geq 1) &\geq \left(1 - \frac{\varepsilon}{3}\right)^2 \Pr(N^* \geq n, |G_2, T < t_1 - t^*) \\ &\geq \left(1 - \frac{\varepsilon}{3}\right)^2 \Pr(Y \geq n) \\ &\geq \left(1 - \frac{\varepsilon}{3}\right)^3 > 1 - \varepsilon.\end{aligned}$$

Returning to the big picture, recall that we are aiming to show that

$$\lim_{c_g \to \infty} \Pr(C(t_1) \geq \theta | N \geq 1) = 1.$$

For each publication $i = 0, 1, \ldots, N - 1$, consider the amount of time between publication and the cutoff time t_1, which is $t_1 - T_i$. Suppose we form groups of publications such that for each group b, the sum of these amounts of time is at least t_1 and at most $2t_1$:

$$t_1 \leq \sum_{i \in b} (t_1 - T_i) \leq 2t_1.$$

Because $t_1 - T_i \leq t_1$ for each i, we can be sure that we can form such groups without this sum exceeding $2t_1$ for any of them. Moreover, since we have just shown that $S = \Sigma_{i=0}^{N-1}(t_1 - T_i)$ becomes large for large values of c_g, we can guarantee that the number of such groups M is large.

We now place the following lower bound on the amount of credit accrued by the combined publications in each group: if at least one of the publications in the group is exposed at any time, we assume all the publications in the group are exposed the entire time. Thus, the credit associated with each group in which at least one exposure occurs is at least $-2c_b t_1$, and the credit associated with groups with no exposures is at least $c_g t_1$. The probability of at least one exposure in a given group is at most $q = 1 - e^{-2\mu t_1}$. Conversely, the probability of no exposures is at least $1 - q$.

Let Z be a random variable following a binomial distribution with parameters M (the number of trials) and $1 - q$ (the success probability). Then Z gives a conservative estimate of the number of groups in which no exposures occur, in the sense that the probability of no exposures in at least m groups is bounded below by the probability that $Z \geq m$. As explained above, we associate $c_g t_1$ credit with 'successes' (groups with no exposures) and $-2c_b t_1$ with 'failures'. It follows that

$$\begin{aligned}\Pr(C(t_1) \geq \theta | N \geq 1) &\geq \Pr(c_g t_1 Z - 2c_b t_1 (M - Z) \geq \theta | N \geq 1)\\ &= \Pr\left(Z \geq \frac{2c_b M + \theta/t_1}{c_g + 2c_b} | N \geq 1\right)\\ &= \sum_{m=0}^{\infty} \Pr\left(Z \geq \frac{2c_b m + \theta/t_1}{c_g + 2c_b} | M = m\right) \Pr(M = m | N \geq 1).\end{aligned}$$

Let $\varepsilon > 0$. Choose c_g larger than $(2c_b + \theta/t_1)/(1-q)^2$ but also large enough such that

$$\Pr\left(S \geq \frac{4t_1}{q(1-q)\varepsilon} | N \geq 1\right) \geq 1 - \frac{\varepsilon}{2}$$

(which we have previously shown is possible). Since Z is binomial, we know its mean is $\mathbb{E}[Z|M=m] = (1-q)m$ and its variance $\mathrm{Var}(Z|M=m) = q(1-q)m$. From $c_g \geq (2c_b + \theta/t_1)/(1-q)^2$ it follows that for all $m \geq 1$,

$$c_g(1-q)^2 m \geq \left(2c_b + \frac{\theta}{t_1}\right) m \geq 2c_b(1-(1-q)^2)m + \frac{\theta}{t_1}.$$

Hence, for all $m \geq 1$, $(c_g + 2c_b)(1-q)^2 m \geq 2c_b m + \theta/t_1$, which entails $(2c_b m + \theta/t_1)/(c_g + 2c_b) \leq (1-q)^2 m = (1-q)m - q(1-q)m$. It follows that $(2c_b m + \theta/t_1)/(c_g + 2c_b) - (1-q)m \leq -q(1-q)m < 0$. So we can again use Cantelli's inequality to get

$$\begin{aligned}\Pr\left(Z \geq \frac{2c_b m + \theta/t_1}{c_g + 2c_b} | M = m\right) &\geq 1 - \frac{q(1-q)m}{q(1-q)m + ((2c_b m + \theta/t_1)/(c_g + 2c_b) - (1-q)m)^2}\\ &\geq 1 - \frac{q(1-q)m}{q(1-q)m + (-q(1-q)m)^2}\\ &= 1 - \frac{1}{q(1-q)m + 1}.\end{aligned}$$

From the above we get that whenever $m \geq 2/(q(1-q)\varepsilon)$,

$$\Pr\left(Z \geq \frac{2c_b m + \theta/t_1}{c_g + 2c_b} | M = m\right) \geq 1 - \frac{\varepsilon}{2}.$$

Thus

$$\begin{aligned}\Pr(C(t_1) \geq \theta | N \geq 1) &\geq \sum_{m=\lceil 2/(q(1-q)\varepsilon) \rceil}^{\infty} \Pr\left(Z \geq \frac{2c_b m + \theta/t_1}{c_g + 2c_b} | M = m\right) \Pr(M = m | N \geq 1)\\ &\geq \left(1 - \frac{\varepsilon}{2}\right) \Pr\left(M \geq \frac{2}{q(1-q)\varepsilon} | N \geq 1\right)\\ &\geq \left(1 - \frac{\varepsilon}{2}\right) \Pr\left(S \geq \frac{4t_1}{q(1-q)\varepsilon} | N \geq 1\right)\\ &\geq \left(1 - \frac{\varepsilon}{2}\right)^2 > 1 - \varepsilon.\end{aligned}$$

This establishes the result that

$$\lim_{c_g \to \infty} \Pr(C(t_1) \geq \theta | N \geq 1) = 1,$$

and hence

$$\lim_{c_g \to \infty} \Pr(C(t_1) > \theta) = \Pr(N \geq 1) = 1 - e^{-\lambda(0)t_1}.$$

The desired claim follows straightforwardly from this. In particular, if $\lambda(0)$, t_1, c_g^*, c_b^*, and μ^* are fixed then $\Pr(C^*(t_1) > \theta)$ is thereby fixed as well. Moreover, we know that $\Pr(C^*(t_1) > \theta) < 1 - e^{-\lambda(0)t_1}$ since $\theta > 0$ and $\Pr(C^*(t_1) = 0) = e^{\lambda(0)t_1}$ and $\Pr(0 < C^*(t_1) < \theta) > 0$. Thus $\Pr(C^*(t_1) > \theta)$ is less than the limiting value by some definite amount. For arbitrary fixed values of $c_b^\dagger$ and $\mu^\dagger$, the limiting result just established guarantees that we can choose $c_g^\dagger$ large enough such that $\Pr(C^\dagger(t_1) > \theta)$ is closer to $1 - e^{-\lambda(0)t_1}$ than $\Pr(C^*(t_1) > \theta)$. Thus, in particular,

$$\Pr(C^\dagger(t_1) > \theta) > \Pr(C^*(t_1) > \theta). \qquad \square$$

Acknowledgements

Thanks to Liam Bright, Kevin Zollman, two anonymous referees, and one of the editors for valuable comments and discussion. This work was supported by the Dutch Research Council (NWO) through the NWO Talent Programme (grant no. 016.Veni.195.141).

Department of Philosophy,
Logic and Scientific Method
London School of Economics
and
Political Science
London, UK
r.heesen@lse.ac.uk

References

Abraham, P. [2000]: 'Duplicate and Salami Publications', *Journal of Postgraduate Medicine*, **46**, pp. 67–69.

Boyer, T. [2014]: 'Is a Bird in the Hand Worth Two in the Bush? Or, Whether Scientists Should Publish Intermediate Results', *Synthese*, **191**, pp. 17–35.

Bright, L. K. [2017]: 'On Fraud', *Philosophical Studies*, **174**, pp. 291–310.

Brischoux, F. and Angelier, F. [2015]: 'Academia's Never-Ending Selection for Productivity', *Scientometrics*, **103**, pp. 333–36.

Bruner, J. P. [2013]: 'Policing Epistemic Communities', *Episteme*, **10**, pp. 403–16.

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., et al. [2018]: 'Evaluating the Replicability of Social Science Experiments in *Nature* and *Science* between 2010 and 2015', *Nature Human Behaviour*, **2**, pp. 637–44.

Clauset, A., Arbesman, S. and Larremore, D. B. [2015]: 'Systematic Inequality and Hierarchy in Faculty Hiring Networks', *Science Advances*, **1**, available at <https://doi.org/10.1126/sciadv.1400005>.

Cole, J. R. and Cole, S. [1973]: *Social Stratification in Science*, Chicago, IL: University of Chicago Press.

Dasgupta, P. and David, P. A. [1994]: 'Toward a New Economics of Science', *Research Policy*, **23**, pp. 487–521.

de Cruz, H. [2018]: 'Prestige Bias: An Obstacle to a Just Academic Philosophy', *Ergo*, **5**, pp. 259–87.

DiPrete, T. A. and Eirich, G. M. [2006]: 'Cumulative Advantage as a Mechanism for Inequality: A Review of Theoretical and Empirical Developments', *Annual Review of Sociology*, **32**, pp. 271–97.

Fanelli, D. [2010]: 'Do Pressures to Publish Increase Scientists' Bias? An Empirical Support from US States Data', *PLOS ONE*, **5**, available at <https://doi.org/10.1371/journal.pone.0010271>.

Flanagin, A., Carey, L. A., Fontanarosa, P. B., Phillips, S. G., Pace, B. P., Lundberg, G. D. and Rennie, D. [1998]: 'Prevalence of Articles with Honorary Authors and Ghost Authors in Peer-Reviewed Medical Journals', *Journal of the American Medical Association*, **280**, pp. 222–24.

Frey, D. and Šešelja, D. [2018]: 'What Is the Epistemic Function of Highly Idealized Agent-Based Models of Scientific Inquiry?', *Philosophy of the Social Sciences*, **48**, pp. 407–33.

Heesen, R. [2017a]: 'Academic Superstars: Competent or Lucky?', *Synthese*, **194**, pp. 4499–518.

Heesen, R. [2017b]: 'Communism and the Incentive to Share in Science', *Philosophy of Science*, **84**, pp. 698–716.

Heesen, R. [2018]: 'Why the Reward Structure of Science Makes Reproducibility Problems Inevitable', *Journal of Philosophy*, **115**, pp. 661–74.

Heesen, R. [2019]: 'The Credit Incentive to Be a Maverick', *Studies in History and Philosophy of Science Part A*, **76**, pp. 5–12.

Heesen, R. [2021]: 'R Code for Cumulative Advantage and the Incentive to Commit Fraud in Science', OSF, available at <https://doi.org/10.17605/OSF.IO/AWR8X>.

Heesen, R., Bright, L. K. and Zucker, A. [2019]: 'Vindicating Methodological Triangulation', *Synthese*, **196**, pp. 3067–81.

Huebner, B. and Bright, L. K. [2020]: 'Collective Responsibility and Fraud in Scientific Communities', in S. Bazargan-Forward and D. Tollefsen (*eds*), *The Routledge Handbook of Collective Responsibility*, New York: Routledge, pp. 358–72.

Hvide, H. K. [2002]: 'Tournament Rewards and Risk Taking', *Journal of Labor Economics*, **20**, pp. 877–98.

John, L. K., Loewenstein, G. and Prelec, D. [2012]: 'Measuring the Prevalence of Questionable Research Practices with Incentives for Truth Telling', *Psychological Science*, **23**, pp. 524–32.

Kerr, N. L. [1998]: 'HARKing: Hypothesizing after the Results Are Known', *Personality and Social Psychology Review*, **2**, pp. 196–217.

Kitcher, P. [1990]: 'The Division of Cognitive Labor', *Journal of Philosophy*, **87**, pp. 5–22.

Latour, B. and Woolgar, S. [1986]: *Laboratory Life: The Construction of Scientific Facts*, Princeton, NJ: Princeton University Press.

Lazear, E. P. and Rosen, S. [1981]: 'Rank-Order Tournaments as Optimum Labor Contracts', *Journal of Political Economy*, **89**, pp. 841–64.

Lee, C. J., Sugimoto, C. R., Zhang, G. and Cronin, B. [2013]: 'Bias in Peer Review', *Journal of the American Society for Information Science and Technology*, **64**, pp. 2–17.

Merton, R. K. [1957]: 'Priorities in Scientific Discovery: A Chapter in the Sociology of Science', *American Sociological Review*, **22**, pp. 635–59.

Merton, R. K. [1968]: 'The Matthew Effect in Science', *Science*, **159**, pp. 56–63.

Norris, J. R. [1998]: *Markov Chains*, Cambridge: Cambridge University Press.

Nosek, B. A. and Errington, T. M. [2017]: 'Reproducibility in Cancer Biology: Making Sense of Replications', *eLife*, **6**, available at <https://doi.org/10.7554/eLife.23383>.

O'Connor, C. [2019]: 'The Natural Selection of Conservative Science', *Studies in History and Philosophy of Science Part A*, **76**, pp. 24–29.

O'Connor, C. and Bruner, J. [2019]: 'Dynamics and Diversity in Epistemic Communities', *Erkenntnis*, **84**, pp. 101–19.

Open Science Collaboration [2015]: 'Estimating the Reproducibility of Psychological Science', *Science*, **349**, available at <science.sciencemag.org/content/349/6251/aac4716>.

R Core Team [2020]: *R: A Language and Environment for Statistical Computing*, Vienna: R Foundation for Statistical Computing, available at <R-project.org>.

Robson, A. J. [2002]: 'Evolution and Human Nature', *Journal of Economic Perspectives*, **16**, pp. 89–106.

Simmons, J. P., Nelson, L. D. and Simonsohn, U. [2011]: 'False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant', *Psychological Science*, **22**, pp. 1359–66.

Smaldino, P. E. and McElreath, R. [2016]: 'The Natural Selection of Bad Science', *Royal Society Open Science*, **3**, available at <rsos.royalsocietypublishing.org/content/3/9/160384>.

Smith, M. [2018]: 'When Does Evidence Suffice for Conviction?', *Mind*, **127**, pp. 1193–218.

Strevens, M. [2003]: 'The Role of the Priority Rule in Science', *Journal of Philosophy*, **100**, pp. 55–79.

Stroebe, W., Postmes, T. and Spears, R. [2012]: 'Scientific Misconduct and the Myth of Self-Correction in Science', *Perspectives on Psychological Science*, **7**, pp. 670–88.

Tijms, H. C. [2003]: *A First Course in Stochastic Models*, Chichester: Wiley.

Tomkins, A., Zhang, M. and Heavlin, W. D. [2017]: 'Reviewer Bias in Single- versus Double-Blind Peer Review', *Proceedings of the National Academy of Sciences USA*, **114**, pp. 12708–13.

Wray, K. B. [2017]: 'The Impact of Collaboration on the Epistemic Cultures of Science', in T. Boyer-Kassem, C. Mayo-Wilson and M. Weisberg (*eds*), *Scientific Collaboration and Collective Knowledge*, Oxford: Oxford University Press, pp. 117–34.

Zollman, K. J. S. [2018]: 'The Credit Economy and the Economic Rationality of Science', *Journal of Philosophy*, **115**, pp. 5–33.

Zollman, K. J. S. [unpublished]: 'The Scientific Ponzi Scheme', available at <philsci-archive.pitt.edu/16264>.