

# Modeling science trustworthiness under publish or perish pressure

David Robert Grimes<sup>1,2</sup>, Chris T. Bauch<sup>3</sup>, John P.A. Ioannidis<sup>4</sup>

<sup>1</sup> Queen's University Belfast, UK, BT7 1NN, UK

<sup>2</sup> University of Oxford, Old Road Campus, Oxford, OX3 7DQ, UK

<sup>3</sup> University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada

<sup>4</sup> Meta-Research Innovation Center at Stanford (METRICS) / Departments of Medicine, Health Research and Policy, Biomedical Data Science, and Statistics, Stanford University, SPRC, MSOB X306, 1265 Welch Rd, Stanford, CA 94305, USA

d.r.grimes@qub.ac.uk

\* E-mail: d.r.grimes@qub.ac.uk

## Abstract

The scientific endeavor pivots on the accurate reporting of experimental and theoretical findings, and consequently scientific publication is immensely important. As the number of active scientists continues to increase, there is concern that rewarding scientists chiefly on publication creates a perverse incentive where careless and fraudulent research can thrive. This is compounded by the predisposition of top-tier journals towards novel or positive findings rather than negative results or investigations that merely confirm a null hypothesis, despite their intrinsic value, potentially compounding a reproducibility crisis in several fields. This is a serious problem for both science and public trust in scientific findings. To date, there has been comparatively little mathematical modeling on the factors that influence science trustworthiness, despite the importance of quantifying the problem. In this work, we present a simple phenomenological model with cohorts of diligent, careless and unethical scientists with funding allocated based on published outputs. The results of this analysis suggest that trustworthiness of published science in a given field is strongly influenced by the false positive rate and the pressures from journals for positive results, and that decreasing available funding has negative consequences for the resulting trustworthiness. We also examine strategies to combat propagation of irreproducible science, including increasing fraud detection and awarding diligence, discussing the implications of these findings.

## Introduction

In academia, the phrase "publish or perish" is more than a pithy witticism - it reflects the reality that researchers are under immense pressure to continuously produce outputs, with career advancement dependent upon them [1,2]. Academic publications are deemed a proxy for scientific productivity and ability, and with an increasing number of scientists competing for funding, the previous decades have seen an explosion in the rate of scientific publishing [3]. Yet whilst output has increased dramatically, increasing publication volume does not imply that the average trustworthiness of publications has improved. A

previous paper by Ioannidis [4] has outlined the reasons why many published research findings are false, and the dubious use of P-values for significance in research findings has of late been widely discussed [5–9]. Across much of experimental science from psychology [10] to biomedical science [11–13] and cancer research [14], there is concern over an apparent reproducibility crisis.

Despite their vital importance in conveying accurate science, top-tier journals possess a limited number of publication slots and are thus overwhelmingly weighed towards publishing only novel or significant results. Despite the fact that null results and replications are important scientific contributions, the reality is that journals do not much care for these findings. Researchers are not rewarded for submitting these findings nor for correcting the scientific record, as high profile examples attest [15,16]. This pressure to produce positive results may function as a perverse incentive. Edwards and Roy [17] argue that such incentives encourage a cascade of questionable findings and false positives. Heightened pressure on academics has created an environment where "*Work must be rushed out to minimize the danger of being scooped*" [18]. The range of questionable behavior itself is wide [19]. Classic 'fraud' (falsification, fabrication, and plagiarism (FFP) [20]) may be far less important than more subtle questionable research practices.

So how common are such practices? A study of National Institute of Health (NIH) funded early and mid-career scientists ( $n = 3247$ ) found that within the previous three years, 0.3% admitted to falsification of data, 6% to a failure to present conflicting evidence and a worrying 15.5% to changing of study design, methodology or results in response to funder pressure [21]. An overview by Fanelli [22] has shown that questionable research practices are as common as 75%, while fraud per se occurs only in 1-3% of scientists. These findings are alarming, yet quantification of these perverse incentives is vital if we are to understand the potential extent of the underlying problem, and formulate strategies to address it. This is an underdeveloped area, but one which is slowly growing - recent works by Smaldino and McElreath [23,24] have employed elegant dynamic models to demonstrate that even when there is no attempts at fraud or untoward research practices, selection based solely on published output tends to produce poorer methods and higher false discovery rates, a phenomenon they term "*the natural selection of bad science*".

Suboptimal science and fraud can take myriad forms which renders it difficult to detect [25]. For the purposes of this article, we define fraud as an explicit 'intention to deceive' [26]. A more recent investigation [22] put the weighed mean percentage of scientists committing research fraud as high as 1.97%, with over a third admitting to questionable research practices. The same investigation found that about 14.12 % of scientists reported observing fraudulent research behavior in colleagues. Another study [27] found that 5% of responding authors claimed to be personally aware of fabricated or misrepresented data in a trial they had participated in. A study of bio-statisticians [28] found that over half of respondents reported being aware of research misconduct.

A 2012 [29] analysis found that FFP offenses rather than honest error accounted for 67.4% of retracted publications, with the rate of retraction due to fraud increasing ten-

fold since 1975. An important question is whether scientists who are unethical (fraudulent) or sloppy (careless) may thrive and even outperform diligent scientists in a system driven by publish or perish pressure. Since it is impossible to identify all unethical and careless scientists, one can perform mathematical modeling of science under different assumptions and find out how these scientists fare and what the implications are for the overall trustworthiness of science.

To better understand the impact of publish or perish on scientific research, and to garner insight into what drives the trustworthiness of published science is of paramount importance if we are to counter-act any detrimental impacts of such practices. In this work, we present a simple but instructive model of scientific publishing trustworthiness under the assumption that researchers are rewarded for their published output, taking account of field-specific differences and the proportion of resources allocated with funding cycle. The factors that influence resultant trustworthiness are quantified and discussed, as well as implications for improving the trustworthiness of scientific publishing.

## Model Outline

### Basic model and assumptions

To construct a simple model of publication rewards, we define the total amount of available funding for research as  $R(t)$ . Per unit of funding in a given field, there is a global discovery rate of  $D_R$ , which includes a proportion  $p_T$  of positive / significant results, a proportion  $p_F$  of false positives, and a proportion  $n$  of null results. Null results in principle can include both true negatives and false negatives, but given the bias toward positive results we will not discriminate between these two in this investigation. The relative proportion of positives and nulls will be inherently field-specific - certain disciplines will be more prone to false positives, whilst others tend to yield less ambiguous results. Since the quantities are proportions we have that

$$p_T + p_F + n = 1. \quad (1)$$

In certain fields, the false positive rate may be high, and so diligent researchers take measures to falsify positive results and test their results multiple times. Even when research groups are very diligent, they may reasonably submit a fraction  $\epsilon$  of their false positives. Researchers exist on a spectrum, but for simplicity we may broadly sub-divide this spectrum into three distinct classes.

1. Diligent cohort - This group take pains to replicate experiments and do not dishonestly manipulate results. Their false positive submission fraction is  $\epsilon$ , thus as low as reasonably possible. They account for a fraction  $f_D$  of the initial total, and parameters relating to them have subscript  $D$ .
2. Careless cohort - This group do not falsify results, but are much less careful at eliminating spurious positive results. They may also have questionable practices that lead them to false positives. As a result, they have a false positive submission

rate of  $c\epsilon$  where  $c > 1$ . They account for a fraction  $f_C$  of the initial total, and parameters relating to them have subscript  $C$ .

3. Unethical cohort - This group appear broadly similar to the diligent group, but with one crucial difference in that may occasionally manipulate data or knowingly submit dubious results at a rate of  $\delta$  beyond global discovery rate. For convenience, instead of defining a higher value of  $D_R$  in this group to account for the higher “discovery” rate, we retain the same parameter value of  $D_R$  for the unethical cohort but allow  $p_T + p_F + n + \delta > 1$ , so that their realized “discovery” rate is higher than the other groups. They account for a fraction  $f_U$  of the initial total, and parameters relating to them have subscript  $U$ .

The funding held by the diligent cohort at a given time is  $x(t)$ , with  $y(t)$  held by the careless cohort and  $z(t)$  by the unethical cohort, so that

$$x(t) + y(t) + z(t) = R(t). \quad (2)$$

With these assumptions, we can model the theoretical impact of a paradigm where researchers are rewarded with funding and success in direct relation to their publication output. As outlined in the introduction, there is huge pressure on scientists to submit positive or ‘novel’ findings, whilst findings confirming the null hypothesis are frequently side-lined. Under such a selection pressure, all researchers will aim to submit their significant positive results for publication. The respective rates of submission per unit funding for the diligent, careless and unethical cohorts are accordingly

$$\begin{pmatrix} S_{D+} \\ S_{C+} \\ S_{U+} \end{pmatrix} = D_R \begin{pmatrix} p_T + \epsilon p_F \\ p_T + c\epsilon p_F \\ p_T + \epsilon p_F + \delta \end{pmatrix} \quad (3)$$

The rate at which null results are submitted is less clear - in general, there is a significant bias in publication towards significant results. As a consequence, negative findings are often shunned by high impact journals, and scientists are disinclined to submit them, meaning that potentially important null results may not ever see the light of publication, the so-called ‘file drawer’ problem. We assume that each cohort submit only a fraction of their null results in the proportions  $\beta_D, \beta_C, \beta_U$  such that

$$\begin{pmatrix} S_{D-} \\ S_{C-} \\ S_{U-} \end{pmatrix} = D_R n \begin{pmatrix} \beta_D \\ \beta_C \\ \beta_U \end{pmatrix}. \quad (4)$$

Equations 1-4 comprise the researcher specific parameters, and we must further quantify the journal specific elements also. Competition for space in field-specific top-tier journals is fierce, and we denote the combined carrying-capacity of these field-specific top-tier journals as  $J(t)$ . These journals exhibit a clear bias towards positive results, with a positive-publication weighing fraction of published articles,  $B$ , describing significant results. Thus, we can quantify the probability that a positive result ( $\nu_P(t)$ ) or a negative

result  $(\nu_N(t))$  is published. These probabilities are given by

$$\begin{pmatrix} \nu_P(t) \\ \nu_N(t) \end{pmatrix} = \begin{pmatrix} \frac{JB}{x(t)S_{D+}+y(t)S_{C+}+z(t)S_{U+}} \\ \frac{J(1-B)}{x(t)S_{D-}+y(t)S_{C-}+z(t)S_{U-}} \end{pmatrix}. \quad (5)$$

From this, we can then yield an expression for the publication rate per unit of funding for the diligent, careless and unethical cohorts, which are respectively

$$\begin{pmatrix} L_D(t) \\ L_C(t) \\ L_U(t) \end{pmatrix} = \nu_P(t) \begin{pmatrix} S_{D+} \\ S_{C+} \\ S_{U+} \end{pmatrix} + \nu_N(t) \begin{pmatrix} S_{D-} \\ S_{C-} \\ S_{U-} \end{pmatrix}. \quad (6)$$

The average rate of publications per unit of funding per unit time is thus

$$A(t) = \frac{J}{x(t) + y(t) + z(t)}. \quad (7)$$

If researchers are rewarded with funding based solely on their published output, we can quantify the impact of this with time by employing a recursive series solution at discrete time-steps, corresponding to funding cycles. If funding is allocated to each cohort is based upon their output at the beginning of the previous funding cycling, and we assume total funding remains constant ( $\frac{dR}{dt} = 0$ ) then the funding available for each cohort at each successive time step is

$$\begin{pmatrix} x(t+1) \\ y(t+1) \\ z(t+1) \end{pmatrix} = \begin{pmatrix} \frac{L_D(t)}{A(t)}x(t) \\ \frac{L_C(t)}{A(t)}y(t) \\ \frac{L_U(t)}{A(t)}z(t) \end{pmatrix}. \quad (8)$$

## Variable funding resources

We also consider the fact that the total amount of funding may not remain constant, so we may model the impact of changing funding scenarios. For simplicity, we assume it changes at some constant rate  $G$ , which can be negative (for diminishing funding, the likes of which might occur with a decrease in NIH or EU funding budgets), zero (for constant funding, as in equation 8) or positive (increasing funding). New funding is allocated at random in proportions reflecting the typical make-up of new researchers, and accordingly the refined equations are

$$\begin{pmatrix} x(t+1) \\ y(t+1) \\ z(t+1) \end{pmatrix} = \begin{pmatrix} \frac{L_D(t)}{A(t)}x(t) + f_D G \\ \frac{L_C(t)}{A(t)}y(t) + f_C G \\ \frac{L_U(t)}{A(t)}z(t) + f_U G \end{pmatrix}. \quad (9)$$

## Research fraud detection

For unethical researchers, we can look at a slightly more complicated scenario where dubious publications have a probability of detection leading to denial of funding,  $\eta$ . We further assume this penalization only applies to dubious results which were published rather than just submitted. If this consideration is taken into account, then we modify the last part of equation 9 to reflect this so that

$$z(t+1) = \left( \frac{L_U(t)}{A(t)} - D_R \eta \delta \nu_p(t) \right) z(t) + f_U G \quad (10)$$

## Rewarding diligence

The diligent cohort have intrinsically lower submission rates than other groups, and consequently are more likely to suffer under a publish or perish regime, despite the importance of their reproducible work. To counter-act this, it has been suggested that rewarding diligence might counteract this trend [30,31]. We might envision a situation where scientific works are audited for reproducibility, with groups who keep their reproducibility high and error rates below a certain unavoidable threshold (given by  $D_R \nu_P \epsilon$ ) garnering a reward of  $R_W$ . This in practice could only be achieved by the diligent cohort, and in the most simple case, their funding resources are given by

$$x(t+1) = \frac{L_D(t)}{A(t)} x(t) + f_D G + R_W \quad (11)$$

## Counter-acting publication bias

It is also possible to envision a situation where journals don't give any preference to positive results over null results. In this case, we would expect researchers to submit all their results so that  $\beta_D = \beta_C = \beta_U = 1$ . In this case,  $\nu_P$  and  $\nu_N$  are replaced by a single function of time  $\nu$ , given by

$$\nu(t) = \frac{J}{x(t)(S_{D+} + S_{D-}) + y(t)(S_{C+} + S_{C-}) + z(t)(S_{U+} + S_{U-})} \quad (12)$$

## Trustworthiness of published science

Finally, we define a metric for the trustworthiness of published science, defined as the proportion of reproducible results,  $T(t)$ . This is given by

$$T(t) = 1 - \frac{\nu_p D_R (x(\epsilon p_F) + y(c \epsilon p_F) + z(\epsilon p_F + \delta))}{J} \quad (13)$$

where the time arguments of  $x, y, z$  and  $\nu_p$  have been excluded for clarity.

## Parameter estimation and assumptions

Details on the parameter estimation and assumptions and the range of values considered for each parameter in the model appear in supplement 1 and supplementary table 1.

## Results

### Impact of the field-specific false positive rate

Figure 1 shows the change in funding proportions with time for each cohort in a field with a low rate of false positives ( $\frac{p_F}{p_T+p_F} = 0.25$ ) and a field with a high rate of false positives ( $\frac{p_F}{p_T+p_F} = 0.8$ ). What is immediately evident is that in that in fields where false positives comprise the bulk of positive results, the trustworthiness of published science suffers markedly, and careless and unethical cohorts are disproportionately rewarded at the expense of diligent researchers. Simulation results suggests that the trustworthiness of published science in any given field is strongly dependent on the false positive rate in that field under a publish or perish paradigm.

### Impact of funding growth rate

As depicted in figure 2. The increasing of available funds has the net effect of reducing publication pressure by bring down the average number of publications expected per unit funding, provided journal capacity stays roughly constant, reducing the likelihood of dubious publications being selected. Conversely, reducing funding increases the publication pressure and results in increases selection of suspect works and a fall in scientific reproducibility.

### Impact of increased fraud detection

Figure 3 depicts the impact of aggressive fraud detection and punishment. Increased fraud detection seems to improve science trustworthiness, but  $\eta$  has to be very high in practice to have a substantial impact on the proportion of funding allocated to unethical cohorts. Negating growth, the funding allocation to this group would only be expected to decrease with time if

$$\eta > \frac{1}{D_R \delta \nu_P} \left( \frac{L_U}{A} - 1 \right). \quad (14)$$

In practice this is quite high, and for the values in supplementary table 1, a value of  $\eta > 0.7688$  would be required to fully diminish funding to this cohort in time.



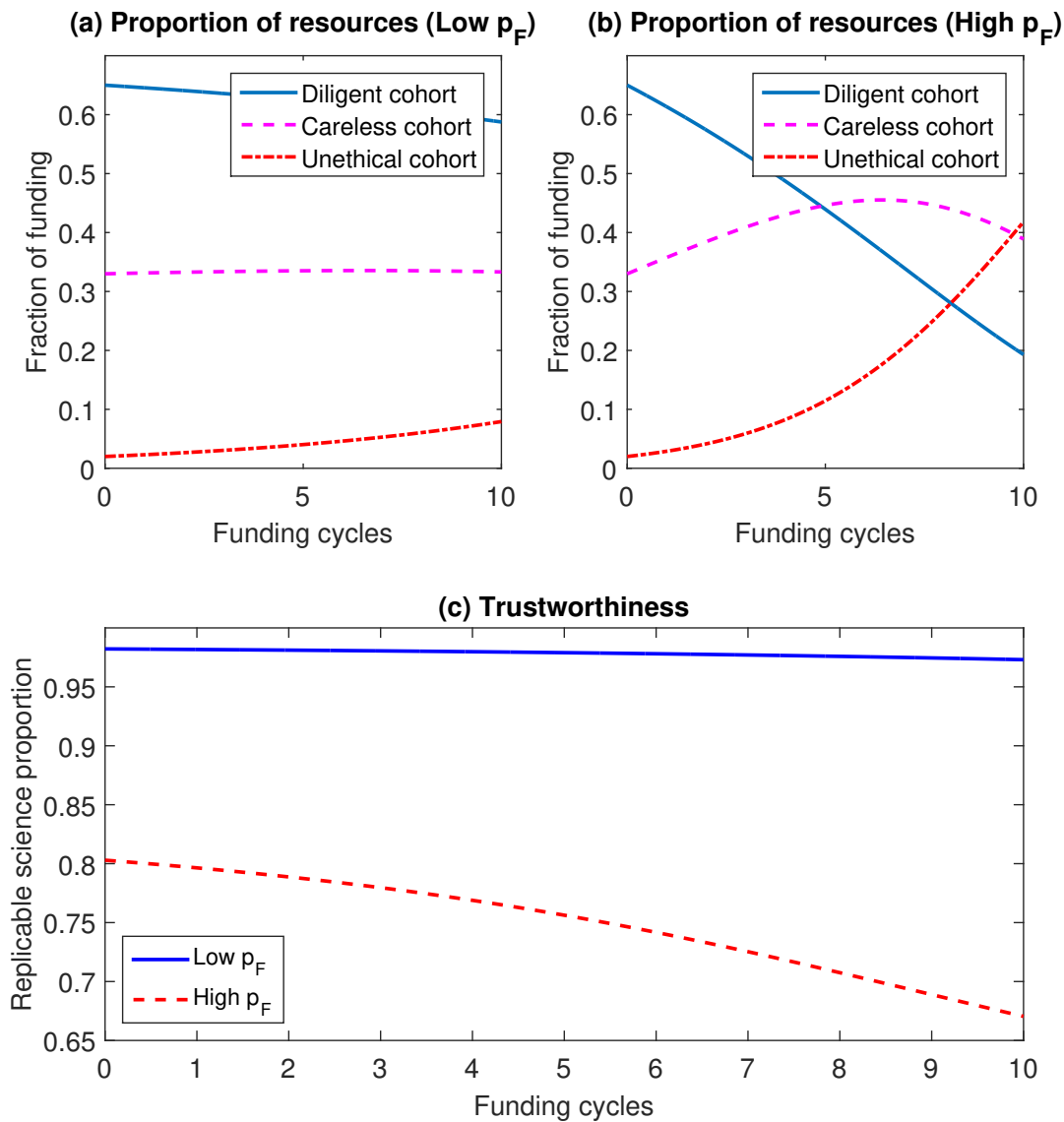


Figure 1: The impact of field-specific false positive rate on resources allocated and science trustworthiness. (a) depicts the projected funding allocations in a field where false positives are relatively rare ( $p_T = 0.32, p_F = 0.08$ ). By contrast, (b) shows the impact on resources consumed when false positives are the norm ( $p_T = 0.08, p_F = 0.32$ ). In (c), the trustworthiness (proportion of reproducible science) for both scenarios are depicted, indicating false positive rate highly drives the trustworthiness of scientific publication in a given field.

## Impact of rewarding diligence

By inspection, it is straight-forward to show that for the amount of funding held by the diligent cohort to stay the same or increase, then the condition on  $R_W$  is

$$R_W \geq \left(1 - \frac{L_D(t)}{A(t)}\right) x(t) - f_D G \quad (15)$$



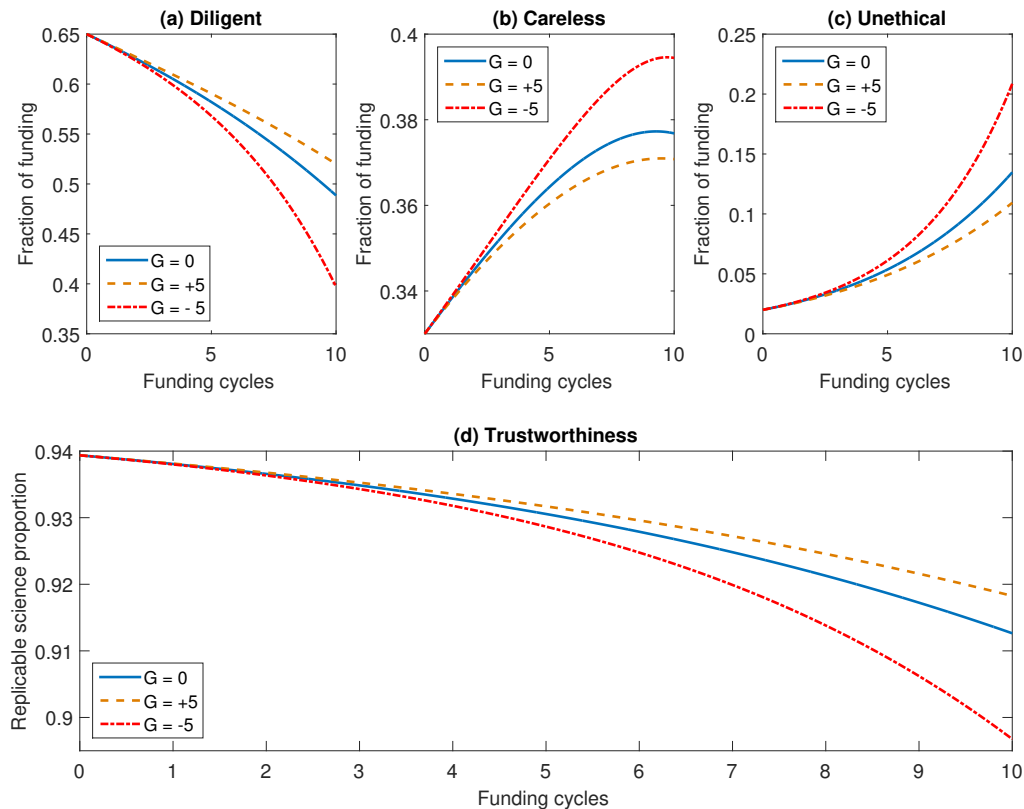


Figure 2: The impact of funding increases and decreases on funding allocation and science trustworthiness. Figures (a) to (c) depict the absolute proportion of funding resources allocated to diligent, careless and unethical cohorts when funding changes at rates of 0, 5 and -5 per cycle respectively.

though in practice for most situations,  $R_W$  will have to be much greater than this minimum value. For the example depicted in figure 4, a large reward for diligence ( $R_W = 10$ ) substantially increases the funds awarded to the diligent cohort. However, reproducibility still falls slowly if the unethical cohort are not removed. It is possible to both reward diligence and punish fraud, which can improve trustworthiness, illustrated in figure 4.

## Impact of the positive publication weighing

To simulate how published science fares under the rather artificial fixation of top-tier journals with positive novel results, figure 5 depicts how funding is allocated and science trustworthiness changes with varying values for  $B$ . In this simulation,  $\beta_D = \beta_C = \beta_D = 0.5$  when publications were  $B$  dependent. For  $B$ -independence, null results were as likely to be published so all were submitted and thus  $\beta_D = \beta_C = \beta_D = 1$ . Higher values of  $B$  lead to perverse rewarding of false positives and fraudulent results at the expense of diligence science. Best outcome for science trustworthiness was observed when journals were simulated as completely agnostic to findings.

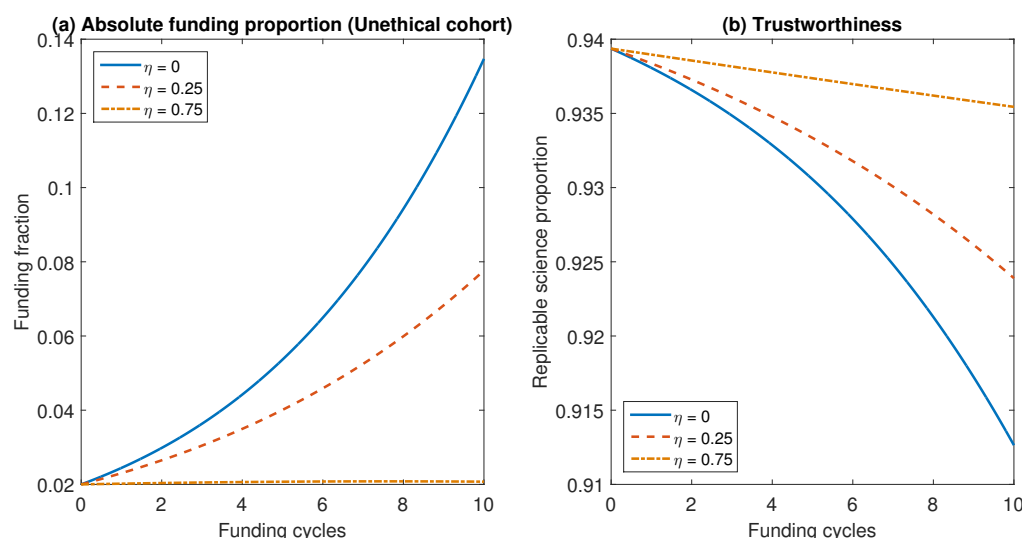


Figure 3: The impact of strict fraud detection / penalization. (a) Increasing the rate at which fraud is detected limits the amount of resources garnered by unethical cohorts, but (b) high values of  $\eta$  are required to markedly improve the trustworthiness of published science.

## Discussion

The model presented is a simplification of a complex eco-system, but gives some insight into what factors shape scientific trustworthiness. A full discussion of assumptions, further findings and avenues for future work are included in supplement 1. The model suggests that a fixation in top-tier journals significant or positive findings tends to drive trustworthiness of published science down, and is more likely to select for false positives and fraudulent results. In our simulations, best outcome was obtained by simply paying no heed to whether a result was significant or not. This is akin to the model used by many emerging open access peer-reviewed journals such as PLOS One, who have a policy of accepting any work provided it is scientifically rigorous. Our simulation suggests this model of publishing should improve science trustworthiness, and it is encouraging that many other publishers are taking this approach too, including Royal Society Open Science and Nature Scientific Reports. As of 2017, Scientific Reports has surpassed PLOS One as the world's biggest mega-journal [32].

While this is encouraging in one respect, there is still a perception that such journals are for 'trivial' or unimportant results, and that positive or important results should still go to a few journals with extreme competition for space. Empirical evaluations show that small studies published in top-impact journals have markedly exaggerated results on average compared with similar studies on the same questions published in journals of lesser impact factor [33]. This suggests that the pressure to publish in these flagship journals may still be very real, despite the option of publishing in less competitive journals. The analysis here suggests that science trustworthiness is affected too by changes in funding resources, and that when an increase of funding improves the over-all trustworthiness of science, as depicted in figure 2. Conversely when this is diminished, the increased

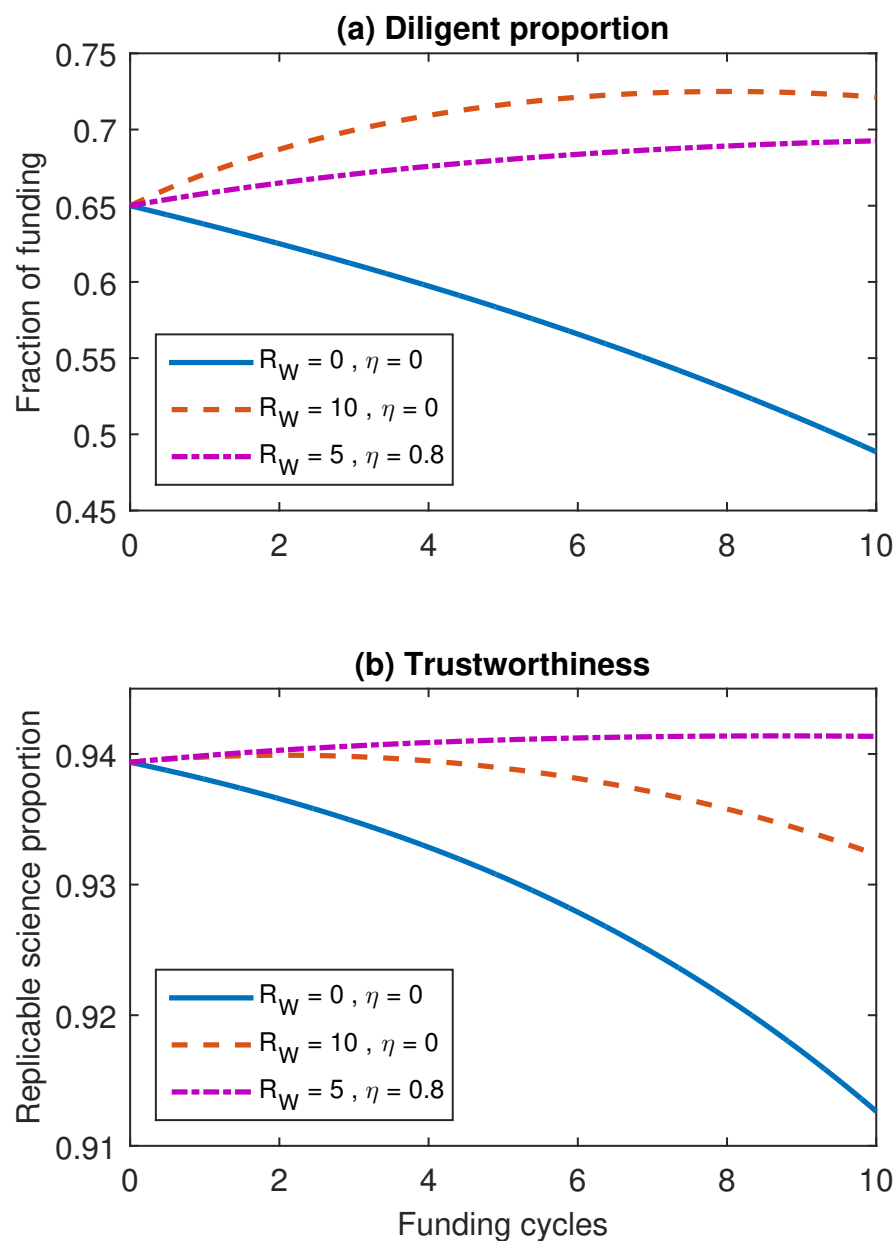


Figure 4: The impact of rewarding researcher for diligence. This improves the proportion of funding allocated to diligent researchers, but to improve science trustworthiness still requires non-zero values of  $\eta$  under this schema.

competition on scientists appears to create conditions when false positives and dubious results are more likely to be selected for and rewarded. This is a natural consequence of the model.

One curious result persistently seen in the model was that diligent researchers are unfairly affected by careless or unethical conduct, with avoidable false positives or unethical publications garnering disproportionate reward at their expense. Simply increasing fraud

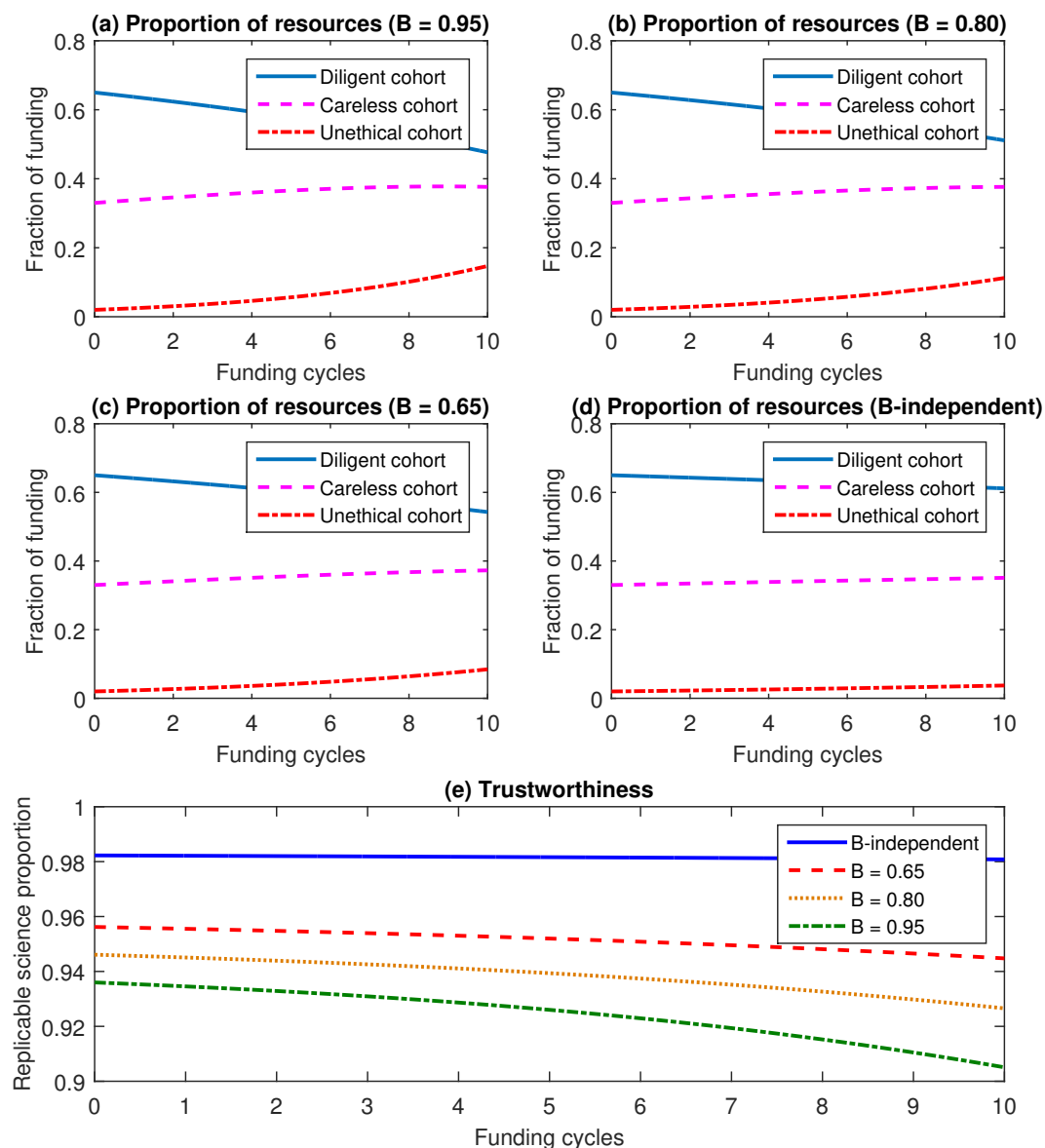


Figure 5: The impact of positive publication weighing on the trustworthiness of published science. (a), (b) and (c) show funding allocation 95%, 80%, 65% of published results are positive respectively, whilst (d) depicts the situation when publications are completely agnostic. Science trustworthiness of all these scenarios is shown in figure (e), suggesting best trustworthiness obtained when journals were completely agnostic to whether a result was positive or null.

detection doesn't do much to stop this, as careless researchers benefit from the gap in the market, out-producing their diligent colleagues, as seen in figure 3. This appears to be an unfortunate and seemingly unavoidable consequence of a 'publish or perish' system. However, in good scientific environments carelessness would be sooner or later detected

and potentially penalized. We can estimate how much of a penalty for carelessness or reward for diligence we need so as to inverse the worsening trends that we observe, by manipulating equations similar to the manner outlined for unethical conduct. However, this approach risks being ruthlessly punitive, punishing honest mistakes with the same severity reserved for the most egregious abuses of scientific trust.

While a penalty for carelessness has intuitive appeal, distinguishing between honest and careless errors is fraught with difficulty. As has been argued elsewhere [30,31], rewarding diligence is perhaps a better way to ensure researchers do not suffer for good conduct. A simple model of this is shown in figure 4, and indeed this suggests rewarding diligence improves the proportion of funding allocated to diligent groups. However, it requires some penalty for bad conduct to keep unethical cohorts from benefiting at the expense of others. In practice this level of detection appears to have to be relatively high, which of course would require considerable resources to achieve. It should be noted too that the false positive rate of a field has a significant impact on science trustworthiness, as illustrated in figure 1. A high type II error rate provides ample cover for an unethical researcher to cheat without overt fear of detection [22,26], perhaps explaining the elevated prevalence of dubious practice in biomedical science [22] in particular.

Future work with more sophisticated models could explore how best to implement these and other possible interventions designed to improve science trustworthiness. For instance, trustworthiness as a function of positive publication bias ( $B$ ) and fraud detection rate ( $\eta$ ) could be computed and optimization approaches could be applied to determine the optimal combination of  $B$  and  $\eta$  to improve science trustworthiness. These parameters can be somewhat influenced by large academic societies, government agencies, or independent foundations for instance, who could fund efforts to detect fraud in published work and support research concerning null results.

It's also important to note that the model results pivot explicitly on the assumption that scientists are forced to operate under a 'publish-or-perish' regime, and rewarded solely on output. Thus there is another way to improve the trustworthiness of published science - while publications are indeed one measure of productivity, they are not necessarily the sole measure. While a much harder aspect to gauge, trustworthiness is more fundamentally important. For their part, scientific journals should realize that issues such as replication and null findings are equally vital to good science as eye-catching 'new' results. This is slowly beginning to be recognized, with some groups coming to the forefront of championing reproducible research methods [34]. The consequences detailed in this manuscript only arise when publishing quantity is the dominant measure of an academic's worth, but in reality this should only be one consideration amongst many. The model suggests that if publishing is the sole criteria under which academics are judged, then dubious conduct can thrive.

We accordingly need to address alternative ways to assess researchers, and to encourage judicious diligence over dubious publishing. The model outlined here is far from complete, but yields some insights into the factors that shape the trustworthiness of published science. There is already evidence that pressure to publish is driving researcher burn-

out and cynicism in published research [35], negatively affecting both research and the researchers themselves [36, 37]. Other studies have not found a clear association of some productivity incentives with bias [38], but these incentives may be confounded in that sometimes they coexist with other features and research practices that tend to increase also quality of research, rather than just quantity of publications. Crucially, bogus findings risk undermining public confidence in science. Amongst notable examples [39–41], the fraudulent *Lancet* MMR-Autism paper [42] is especially infamous, remaining a cornerstone of anti-vaccine narratives [43].

Publishing is not intrinsically flawed, and conversely complete, unbiased publication is essential for scientific progress. We need a better understanding of the factors driving publication and productivity-related behaviors. This is key not only to appreciating the exceptional pressures wrought upon researchers by a strict publish-or-perish imposition, but to improving science itself. This would not only benefit those working in the field, but is crucial if public trust in science is to be maintained.

## Acknowledgements

DRG would like to thank Dr. Alex G. Fletcher and Dr David Basanta for their useful feedback and thoughtful imbibing. DRG would also like to thank Prof. David Colquhoun FRS for his thoughtful discussion and advice.

## References

- [1] D. Fanelli, “Do pressures to publish increase scientists’ bias? An empirical support from US states data,” *PLoS ONE*, vol. 5, no. 4, 2010.
- [2] U. S. Neill, “Publish or perish, but at what cost?,” *Journal of Clinical Investigation*, vol. 118, pp. 2368–2368, jul 2008.
- [3] L. Bornmann and R. Mutz, “Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references,” *Journal of the Association for Information Science and Technology*, vol. 66, pp. 2215–2222, nov 2015.
- [4] J. P. A. Ioannidis, “Why most published research findings are false,” *PLoS Medicine*, vol. 2, no. 8, pp. 0696–0701, 2005.
- [5] M. Krawczyk, “The search for significance: A few peculiarities in the distribution of p values in experimental psychology literature,” *PLoS ONE*, vol. 10, no. 6, pp. 1–19, 2015.
- [6] D. Chavalarias, J. D. Wallach, A. H. T. Li, and J. P. A. Ioannidis, “Evolution of reporting P-values in the biomedical literature, 1990-2015,” *Jama*, vol. 315, no. 11, pp. 1141–1148, 2016.
- [7] L. G. Halsey, D. Curran-Everett, S. L. Vowler, and G. B. Drummond, “The fickle P value generates irreproducible results,” *Nature Methods*, vol. 12, no. 3, pp. 179–185, 2015.

- [8] N. Altman and M. Krzywinski, “Points of significance: P values and the search for significance,” *Nature Methods*, vol. 14, no. 1, pp. 3–4, 2016.
- [9] D. Colquhoun, “An investigation of the false discovery rate and the misinterpretation of P values,” *Royal Society Open Science*, pp. 1–15, 2014.
- [10] M. Baker, “Is there a reproducibility crisis?,” *Nature*, vol. 533, pp. 452–454, 2016.
- [11] R. Al-Shahi Salman, E. Beller, J. Kagan, E. Hemminki, R. S. Phillips, J. Savulescu, M. Macleod, J. Wisely, and I. Chalmers, “Increasing value and reducing waste in biomedical research regulation and management,” *The Lancet*, vol. 383, no. 9912, pp. 176–185, 2014.
- [12] C. G. Begley and J. P. A. Ioannidis, “Reproducibility in science: Improving the standard for basic and preclinical research,” *Circulation Research*, vol. 116, no. 1, pp. 116–126, 2015.
- [13] Nature, “Fostering reproducible fMRI research,” *Nature Neuroscience*, vol. 20, pp. 298–298, feb 2017.
- [14] C. G. Begley and L. M. Ellis, “Drug development: Raise standards for preclinical cancer research.,” *Nature*, vol. 483, no. 7391, pp. 531–3, 2012.
- [15] S. J. Ritchie, R. Wiseman, and C. C. French, “Failing the future: Three unsuccessful attempts to replicate bem’s ‘retroactive facilitation of recall’ effect,” *PLoS ONE*, vol. 7, no. 3, pp. 1–5, 2012.
- [16] M. Xpress, “Controversial study promoting psychic ability debunked,” 2012.
- [17] M. A. Edwards and S. Roy, “Academic Research in the 21st Century: Maintaining Scientific Integrity in a Climate of Perverse Incentives and Hypercompetition,” *Environmental Engineering Science*, vol. 34, no. 1, p. ees.2016.0223, 2016.
- [18] P. a. Lawrence, “The politics of publication.,” *Nature*, vol. 422, no. 6929, pp. 259–261, 2003.
- [19] A. Williamson, “COPE: Committee on Publication Ethics,” *Learned Publishing*, vol. 14, pp. 149–150, apr 2001.
- [20] R. de Vries, M. S. Anderson, and B. C. Martinson, “Normal Misbehavior: Scientists Talk about the Ethics of Research,” *Journal of Empirical Research on Human Research Ethics: An International Journal*, vol. 1, pp. 43–50, mar 2006.
- [21] B. C. Martinson, M. S. Anderson, and R. de Vries, “Scientists behaving badly,” *Nature*, vol. 435, no. 7043, pp. 737–738, 2005.
- [22] D. Fanelli, “How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data,” *PLoS ONE*, vol. 4, no. 5, 2009.
- [23] P. E. Smaldino and R. McElreath, “The natural selection of bad science,” *Royal Society Open Science*, vol. 3, p. 160384, sep 2016.



- [24] R. McElreath and P. E. Smaldino, “Replication, communication, and the population dynamics of scientific discovery,” *PLoS ONE*, vol. 10, no. 8, pp. 1–16, 2015.
- [25] J. Giles, “Special Report: Taking on the cheats,” *Nature*, vol. 435, no. 7040, pp. 258–259, 2005.
- [26] M. Franzen, S. Rödder, and P. Weingart, “Fraud: causes and culprits as perceived by science and the media,” *EMBO reports*, vol. 8, no. 1, pp. 3–7, 2007.
- [27] W. Gardner, C. W. Lidz, and K. C. Hartwig, “Authors’ reports about research integrity problems in clinical trials,” *Contemporary Clinical Trials*, vol. 26, no. 2, pp. 244–251, 2005.
- [28] J. Ranstam, M. Buyse, S. L. George, S. Evans, N. L. Geller, B. Scherrer, E. Lesaffre, G. Murray, L. Edler, J. L. Hutton, T. Colton, and P. Lachenbruch, “Fraud in medical research: an international survey of biostatisticians. ISCB Subcommittee on Fraud.,” *Controlled clinical trials*, vol. 21, no. 2000, pp. 415–427, 2000.
- [29] F. C. Fang, R. G. Steen, and A. Casadevall, “Misconduct accounts for the majority of retracted scientific publications,” *Proceedings of the National Academy of Sciences*, vol. 109, pp. 17028–17033, oct 2012.
- [30] J. P. Ioannidis, “How to make more published research true,” *PLoS Med*, vol. 11, no. 10, p. e1001747, 2014.
- [31] J. P. Ioannidis and M. J. Khoury, “Assessing value in biomedical research: the pqrst of appraisal and reward,” *Jama*, vol. 312, no. 5, pp. 483–484, 2014.
- [32] P. Davis, “Scientific reports overtakes plos one as largest megajournal,” 2017.
- [33] K. C. Siontis, E. Evangelou, and J. P. Ioannidis, “Magnitude of effects in clinical trials published in high-impact general medical journals,” *International journal of epidemiology*, p. dyr095, 2011.
- [34] M. R. Munafò, B. A. Nosek, D. V. M. Bishop, K. S. Button, C. D. Chambers, N. Percie, U. Simonsohn, and E.-j. Wagenmakers, “A manifesto for reproducible science,” *Nature Publishing Group*, vol. 1, no. January, pp. 1–9, 2017.
- [35] J. K. Tijdkink, A. C. M. Vergouwen, and Y. M. Smulders, “Publication pressure and burn out among dutch medical professors: A nationwide survey,” *PLOS ONE*, vol. 8, pp. 1–6, 09 2013.
- [36] S. Rawat and S. Meena, “Publish or perish: Where are we heading?,” *Journal of research in medical sciences: the official journal of Isfahan University of Medical Sciences*, vol. 19, no. 2, p. 87, 2014.
- [37] M. van Wesel, “Evaluation by citation: Trends in publication behavior, evaluation criteria, and the strive for high impact publications,” *Science and Engineering Ethics*, vol. 22, no. 1, pp. 199–225, 2016.

- [38] D. Fanelli, R. Costas, and J. P. Ioannidis, “Meta-assessment of bias in science,” *Proceedings of the National Academy of Sciences*, p. 201618569, 2017.
- [39] M. A. G. van der Heyden, T. van de Ven, and T. Opthof, “Fraud and misconduct in science: the stem cell seduction: Implications for the peer-review process,” *Netherlands Heart Journal*, vol. 17, no. 1, pp. 25–29, 2009.
- [40] J. Rasko and C. Power, “What pushes scientists to lie? the disturbing but familiar story of haruko obokata,” 2015.
- [41] G. Vogel, “Jan hendrik schn loses his ph.d.,” 2011.
- [42] “A timeline of the wakefield retraction,” *Nature Medicine*, vol. 16, pp. 248–248, mar 2010.
- [43] D. R. Grimes, “On the viability of conspiratorial beliefs,” *PLoS ONE*, vol. 11, no. 1, pp. 1–17, 2016.