



# Systemic Problems in Academia: The Positive Publication Bias and Solutions from a Human Factors Perspective

Stefan Gaillard<sup>1</sup>, Sean Devine<sup>2</sup>

## Keywords

Stefan Gaillard, Sean Devine

It will come as no surprise to readers of this journal that failure is systematically ignored in academia. The entire culture of academia revolves around a notion of linear progress made in leaps and bounds by great thinkers, rather than the more realistic image of science as an ongoing process of trial and error. For those already convinced that error deserves a place in scientific discussion, it is often tempting to treat the dismissal of failure as a moral failure on the part of researchers and publishers—if only we were more courageous to do the work that science demands of us, we would publish our failures all the time; if it was not for journals rejecting my work because it is not “innovative” enough, I would tell everyone of my important failures. This perspective, while understandable, misses the mark. Specifically, it neglects the *structure* of modern academia that perpetuates an ideal of “success-first” science on all fronts—a structure, like any structure, that severely limits researchers’ and journals’ ability to change it.

This point is described superbly in a recent editorial piece by public health researchers Sindall and Barrington (2020), who point out the many barriers to sharing and promoting scholarly understandings of failure in the public health interventionist field of Water, Sanita-

tion, and Hygiene (WASH), which they experienced personally as well.

In their in the Journal of Trial and Error’s (JOTE) first issue, the authors argue not only for the role of researchers themselves in pursuing submission and publication of reports of what went wrong, but also of funding partners in sharing information about projects that did not “turn out as planned,” and importantly, for the responsibility of journals in welcoming such sort of publications. Critically, they argue that the widespread fixation on “novelty” contributes to a “culture of success-ism” that is “highly unscientific”. In short, the outward-facing sheen of scientific success—propagated by scientists, funders, and stakeholders alike—challenges the very foundations scientific development is built on: transparency, collaboration, and trial and error.

Sindall and Barrington (2020) paint a compelling picture of an academic system that struggles to find a place for failure and the “multitude of sins” such a system necessarily perpetuates. Rightly so, they place the responsibility for change not only on researchers, but also funders and publication bodies.

But the buck doesn’t stop here. The same culture of success-ism that limits researchers limits individual publishers as well. The problem is diffuse and the blame cannot be laid on a single “bad actor”. Accordingly, we feel it pertinent to extend Sindall & Barrington’s argument by highlighting the structural barriers that re-

<sup>1</sup>Institute for Science in Society, Radboud University Nijmegen, the Netherlands & Center of Trial and Error, Utrecht, the Netherlands

<sup>2</sup>Department of Psychology, McGill University, Montreal, Canada

**Published**  
2022-03-02  
**Issued**  
March 3, 2022

**Correspondence**  
gaillard@centeroftrialanderror.com

**License** © ⓘ  
This article is licensed under the **Creative Commons Attribution 4.0 (CC-BY 4.0)** license, which allows you to copy, redistribute, transform and build upon the article and all its contents in any medium or format for any purpose, provided that appropriate credit is given.

© Gaillard & Devine 2022



strict the spread of failure in the sciences, even at the editorial level. To do so, we use systems theory to frame three limitations journals face when publishing failure and how these limitations cannot simply be solved by good agents engaging in good science.

### **| Failure and Systems**

In the field of Human Factors, and more specifically in Forensic Human Factors, it has long been recognized that failure usually occurs in the complex environment of a system. When accidents such as the Chernobyl nuclear disaster occur, it is easy to shift blame to a single individual or to a single team that was present when mistakes occurred. In response to these disasters, these employees are subsequently reprimanded and sometimes fired. A new person might fill their position and, in some cases, training protocols might be updated. However, these are often very ineffective measures. As previously indicated, mistakes are made in a complex environment. This means that mistakes can have occurred for any number of reasons, and often because of a combination of reasons. Accordingly, removing just one factor from the environment (i.e., the employee present during the disaster) is unlikely to significantly diminish the chances of similar mistakes happening in the future, because it was unlikely to be the determining factor (Holden, 2009).

Fields such as the aviation industry have long recognized that there are better alternatives to simply replacing or retraining staff following a negative incident. For instance, at the beginning of the proliferation of commercial flights, it became clear that cockpits were often too complicated for pilots to understand, which lead to mistakes being made, and consequently accidents occurred. Instead of firing the pilots who made mistakes and/or making small changes to training protocols, the aviation industry started making changes to the cockpits themselves. In other words, the industry transformed the environment, not the agents.

By increasing the simplicity of the cockpit, pilots were less taxed and could invest more cognitive resources towards flying the plane. An example of such an environmental change is the adaptation of cockpit alarms. In older planes, cockpits would have the same audio-

visual cues for every alarm. Thus, the pilot got the same feedback from the system when the fuel was slightly low as when an important engine was malfunctioning. Pilots were then mentally taxed on two levels: they were constantly confronted with distracting alarms, some of which might not need their immediate attention; and when attending to these alarms, they had to invest great effort in distinguishing the alarm and assessing its importance. As a result, pilots learned to ignore most alarms. A research article published in 1996 summarized problems such as these at the time as follows: "The proliferation of alarms has led to the increased likelihood of false or multiple alarm events and the consequent inability of the operator to determine the underlying cause of the alarm(s)" (Gilson et al., 1996, pp. 12). Under such a system, a situation such as the following might occur: During a flight, one of the important engines malfunctions and needs immediate attention. At the same time, a number of other small, but non-urgent, malfunctions could occur. Because the very attentive, but overly taxed pilot cannot properly differentiate between the different alarms, they end up paying attention to the wrong malfunction. A short while later, the plane crashes. As recently as 2019, a plane crash occurred due to a poorly designed alarm system (Levin, 2019).

Faced with the issue of the inattentive pilot, an airline company is presented with two solutions. On the one hand, the company could fire the pilot, hire a new one, and train them for a longer period of time, with a more experienced instructor, and with more expensive training equipment. Besides the fact that this would be very expensive, it would also be unlikely to solve the problem. Humans' cognitive abilities are limited. This is because, like all human systems, their underlying biological systems operate on limited resources. Similarly, in the realms of cognition, some cognitive scientists believe you can train people for things like working memory, sustained attention, selective attention, and other cognitive skills, but all of these will have a natural ceiling—a point beyond which it is almost impossible to progress (Kwok et al., 2011; Traut et al., 2021).

In short, simply replacing or retraining bad agents doesn't work on its own; we cannot reduce airplane crashes by just getting new and better pilots. As in aviation, the same can be

said for scientific publication. The difference, however, as Sindall and Barrington point out, is that we are already in a nosedive when it comes to recognizing and publishing failure in science.

### Obstacles to Failure

As hinted at in the last section, we do not believe the issues Sindall and Barrington (2020) bring up can be solved by simply creating better, more failure-friendly, journals. Rather, we argue that these issues are *inherent* in the *structure* of modern scientific practice. By “structural”, we mean that science as it is currently understood is defined by success; failure is *not scientific* under the current paradigm. Even when scientific failures are recognized, they are framed within the larger context of long-run success—“failing now to succeed later”—rather than as an integral part of the scientific process in its own right. In this regard, “structural” is quite literal, such that to displace the structure of success that builds up the modern scientific apparatus would mean challenging its integrity to the point of collapse. But where one structure collapses, another is built. The question is whether we need to be so radical. Can journals not simply reform their practices to allow for more diverse results to be published? Do we risk throwing the baby out with the bathwater?

In this section, we raise the possibility that editorial reform may not be enough, or at least, that it may be very difficult. To do so, we discuss three issues that journals, who wish to publish scientific failure in the current landscape, face. Throughout, we use our experience at JOTE as a touchstone and example for some of these issues.

### Funding

It is JOTE’s mission to promote and normalize the process of failure in science. However, we recognize that actually doing so comes at high costs for new and established journals. As Sindall & Barrington highlight, scientific funding is a scarce resource. This is true not only for scientists themselves, but for the journals that publish their work; and especially true for journals that are sympathetic to publishing scientific failures. Funding bodies want to support a

compendium of amazing, world-changing findings, not a collection of brick-building errors.

As a result of this scarcity, journals may turn to article processing charges (APCs) as a way to maintain their services. This business-model is effective when the journal can provide services that financially-secure journals cannot (open access, equity, less emphasis on positive results), but becomes intractable when additional social costs are added on top of the economic ones. These social costs refer both to those experienced by the journal—loss of reputation (more on this below), difficulty finding reviewers and submissions, accusations of predation—and by the submitters, who bear the brunt of the cost and, for social reasons (e.g., working at a lesser-known institution, being from a non-Western background, etc.) may be restricted in their available funds. This issue is exacerbated for journals that encourage researchers to publish their failures, which, because of the “culture of success-ism” in academia, is already perceived as a socially costly endeavor. As a result, publishers that aim to counter this culture start off in a precarious position, limited in terms of traditional funding and at a disadvantage in asking would-be authors—who themselves experience the social and professional costs of publishing null results—for support.

Therefore, journals willing to publish and promote failure in science are left scrounging for funds, often relying on the goodwill of the editorial team to keep them afloat. Alternatively, journals can support the publicizing of failure through the publicizing of success, spinning the failure as an altruistic side-benefit of an otherwise traditional, positive-result-publishing, outlet for scientific advancement. While sometimes the only option, this “failure as an afterthought” mindset often contributes to the very culture of success-ism that the publication of failure in science aims to combat.

### Prestige

Above and beyond financial stability, journals comfortable with the idea of publishing failure face an uphill battle with their own reputation. While science makes room for failure, a scientific method cannot be said to exist without people enacting it. This so-called scientific

method is typically associated with academia, and while Sindall and Barrington are right to encourage people to be more “scientific” by sharing failure, the problem is an academic system that does not. Thus, many would-be authors are disincentivized from submitting their work to failure-friendly journals for fear that publicizing their null, negative, or unclear results would harm their reputation, decrease their chances to acquire funding in the future, and reflect poorly upon them as scientists. This reluctance to submit is a unique challenge that journals who publish failure face. Even positive-results-publishing low impact-factor journals have the benefit—if it can be called one—of researchers submitting their work there after they have exhausted more well-established outlets. In contrast, publishers that want to promote failure find themselves in the awkward position of having to convince researchers to publish their work *at all* (i.e., as opposed to scrapping the whole project), not to mention in their journal, for their peers to see.

Increasing the prestige of failure would contribute to an increased willingness on the part of academics to send in their failures. Examples of initiatives and changes that can increase the prestige of failure within academia include brilliant failure awards and changes to the priority rule. Brilliant failure awards, like other awards, signal what the awarder considers to be important and/or valuable. The priority rule refers to the tendency within science to award scientists who are the first to make a discovery, usually by giving them credit—for example, in the form of naming a theory or subject matter after them. Implicitly, the priority rule indicates that only the final success is worthy of reward, although this success often hasn’t been possible without the failures of others along the way. Changing these incentives can motivate more researchers to publish their failure.

At the same time, those researchers that do already want to publish about failure often encounter unexpected hurdles in the publication process. If the journal is subscription-based, the readership of an already niche topic—failure within some field of science—is reduced further, rendering the whole exercise much less appealing to the open-minded researcher who wants their failure to be read and understood. Conversely, if the journal is open-access, it is likely to rely on APCs, requiring researchers

that send in an article to a journal to pay a fee upon submission. As discussed in the previous section (3.1), with the increasing trend towards open science, APCs are sometimes seen as necessary by open access journals to cover costs. However, these charges selectively disadvantage researchers at less financially prosperous institutions. Often funding to cover high article processing charges is simply not available. This, in turn, prevents researchers in less financially secure situations from sharing their failures.

Sindall and Barrington (2020) provide several examples from humanitarian aid projects which went wrong. Two relevant cases in South Africa are mentioned, where failures led to a waste of money and physical injuries. Publishing about failure could prevent this from happening again in the future, but only if the relevant people are able to access these publications. In countries with a lower quality of life, the relative difference that can be made to improving quality of life by preventing waste of money and physical injuries is larger.

It is not only at the submission phase, however, that the culture of success-ism rears its head. Peer reviewers may also be reluctant to contribute their expertise to results that do not reveal some great discovery about their area of study. On the one hand, this is reasonable because experts are busy and unsolicited requests for reviews are often perceived as a nuisance. On the other hand, reviewers’ increased willingness to respond to larger-scale positive-publishing journals over smaller-scale negative-publishing ones reflects again the equating of experimental success with prestige—a relationship that undergirds the philosophy of success-ism that journals like JOTE seek to overturn.

Thus, in addition to battling for financial stability, publishers that seek to promote failure in science struggle to meet publication quotas or rally enthusiasm for publishing in their journal. The culture of success-ism in academia does double duty in this regard, draining failure-friendly journals’ wallets and potential contributors.

### Lack of Training

Despite the challenges presented so far in this article, it is clear that not all academics or fun-

ders share the same unwavering commitment to success-ism (otherwise we wouldn't be able to publish this editorial). Still, when journals that do promote failure and trial and error become operational, they face the stark realization that neither researchers nor editors nor reviewers are properly equipped to analyze and interpret failure or trial and error.

Academic training in the sciences, whether it be in experimental design or statistical inference, focuses little on the positive interpretation of negative results. That is, scientists today are unequipped to constructively answer the question "what went wrong?" While more and more spaces are opening up to discuss failure in science, such as the WASH Failures Team created by Sindall and Barrington and others, there remains a lack of knowledge about how to critically evaluate failure in the sciences. This is reflected at all levels of the editorial process, from the submissions themselves, wherein reviewers often struggle to quantify or explain the scientific failure in question, to the peer reviews, wherein peer reviewers are used to critiquing a piece based on its positive contribution to the field rather than its negative influence on well-held ideas or techniques.

This lack of knowledge about how to critically evaluate failure extends beyond the sciences to many other types of work such as so-called high-risk industries (Bevilacqua & Ciapica, 2018). This lack of knowledge is not due to a lack of literature on the topic. In philosophy, the process of what we have called "trial and error" has been analyzed in a wide range of contexts. In fields like evolutionary epistemology, agnotology, and within the thought experiments of intuition pumps, much attention is paid to the role of trial and error and failure in the process of gaining knowledge. Additionally, a significant part of the field of Human Factors is dedicated to the analysis and prevention of failure—taking into consideration processes of trial and error. The problem doesn't seem to be what knowledge is available, but rather what knowledge is known by the relevant actors and put into practice.

In one sense, the novelty of interpreting failures in academic practice leads to exciting new horizons where experts are pushed outside their comfort zone and must engage in live, critical, evaluation of data. In another sense, however, this new frontier beckons for

structure—structure that can only be achieved through the long-term, symbiotic, practice of conducting and interpreting failed research and the normalization of failure in science.

In the meantime however, the onus of structuring failure falls on the editorial team of any given failure-friendly journal—a difficult and daunting task that may further discourage the widespread publication of scientific failure. Although training can help, it is insufficient on its own; just like with pilots, there is only so much we can cognitively expect from academics.

### I Destined to Fail?

Sindall and Barrington (2020) highlight the need for "more structured ways to report on failures" in science. We could not agree more. However, there are very real, structural, limitations from a publisher's perspective that makes this goal very difficult to achieve. Thus, we argue that it is not only that "[j]ournals *want* to publish novel research and failure is too often not seen as novel" (italics added), but that oftentimes doing otherwise is self-destructive.

That being said, JOTE is in a fortunate position where we have (and continue to) overcome many of the challenges discussed in this editorial. We do so by leaning into our unorthodoxy, embracing failure wholeheartedly and interdisciplinarily, and tirelessly working to draw informative conclusions from inconclusive results. Nevertheless, the question must be asked: "In the current scientific climate, is this model sustainable?" To that, we must answer with a clear "no". The evidence is clear that failure is required for a healthy science to flourish. The fact that failure has not even yet budded in the sciences highlights the incompatibility of the current model of scientific success and the promotion of scientific failure.

Rather than aim for a reformed system that accommodates failure, our goal is to help usher in a new system wherein failure is a necessary component in scientific progress and discourse. In other words, JOTE's goal is to become obsolete. That is, one day, we hope that our efforts contribute to the formation of a scientific mainstream that recognizes the integral role that failure plays in the scientific process and actively normalizes its publication.

But how to achieve this goal? This piece, alongside Sindall and Barrington's, has empha-



sized the many obstacles that publishers and researchers alike face in the publication and promotion of failure. While it may seem like there is no way out of success-first science, we offer some suggestions here for changes in academia and publishing. In the aviation industry, a nosedived airplane leads to multiple thorough investigations into the circumstances that lead to the problem. This is then followed up with an integral solution, taking into account different factors and their relation to each other. Likewise, in academia and publishing, multiple studies have analyzed the conditions which gave rise to the problem of positive publication bias. Now is the time to start with an integral solution, where instead of tackling positive publication bias from just one angle, researchers, editors, publishers, and managers all work together to solve the problem. In an attempt to facilitate this process, JOTE is increasing its collaborations - for example, by setting up a brilliant failure award at Utrecht University together with Utrecht Young Academy (and hopefully elsewhere), and by providing lunch lectures about failure and uncertainty at University Medical Centers, which we are starting to do with the help of The New Utrecht School. To reflect this shift from mostly publishing to a broader focus, we will launch the Center of Trial and Error—a place where all our activities come together. The long-term goal is to create an academic culture where failure is an accepted part of the scientific process.

Therefore, while we are flattered by Sindall's and Barrington's (2020) hope that the *Journal of Trial and Error* won't be the last "outlet for failure", we certainly hope that we will be the last of our kind. We look forward to the day where founding a journal based only on failure is regarded as just as inane as founding one based solely on success.

## References

- Gilson, R., Deaton, J., & Mouloua, M. (1996). Coping with complex alarms: Sophisticated aircraft cockpit alarm systems demand a shift in training strategies. *Ergonomics in Design: The Quarterly of Human Factors Applications*, 4(4), 12–18. <https://doi.org/10.1177/106480469600400404>
- Holden, R. (2009). People or systems? to blame is human. the fix is to engineer. *Professional safety*, 54(12), 34–41.
- Kwok, T. C., Chau, W. W., Yuen, K. S., Wong, A. Y., Li, J. C., Shiu, R. Y., & Ho, F. K. (2011). Who would benefit from memory training? a pilot study examining the ceiling effect of concurrent cognitive stimulation. *Clinical Interventions in Aging*, 6, 83–88. <https://doi.org/10.2147/CIA.S16802>
- Levin, A. (2019). Boeing failed to predict that slew of 737 max warning alarms would confuse pilots, investigators say.
- Sindall, R., & Barrington, D. (2020). Fail fast, fail forward, fail openly: The need to share failures in development. *Journal of Trial & Error*, 1(1), 6–8. <https://doi.org/10.36850/ed2>
- Traut, H. J., Guild, R. M., & Munakata, Y. (2021). Why does cognitive training yield inconsistent benefits? a meta-analysis of individual differences in baseline cognitive abilities and training outcomes. *Frontiers in psychology*, 12, 1–20. <https://doi.org/10.3389/fpsyg.2021.662139>