Incentives for Registered Reports from a risk sensitivity perspective

Incentives for Registered Reports from a risk sensitivity perspective

New Intro:

- What are RRs
- Why: pub bias & QRPs
- Chapter 2: RRs indeed associated with lower rate of positive results
- Problem:
  - RRs may be used strategically for low priors
  - assuming that they work, uptake not as high as we'd like, and not in all fields → could be just basic diffusion of innovation process, but could also be because there are obstacles (e.g., in certain research areas, at certain career stages)
- → what are the incentives for/against RRs? Here, we'll look at this with a computational model
- RRs marketed as aligned with existing incentives: 'safe' choice for researchers
- But if that's true and they're a safe choice, we *wouldn't* expect them to always be preferred
- Risk-sensitivity theory
  - Intro to RST with example
  - Brief explanation of relationship with utility theory and prospect theory
  - Appliation to RR problem
- Goals of the chapter: apply RST to find out when & where RRs are expected to be particularly popular vs unpopular → implications for policy and meta-science

In the storybook version of science, researchers are driven by pure curiosity, conduct empirical studies to learn about the natural world, and impartially record the results. In reality, researchers are motivated and constrained by a wide range of psychological, social, political, and structural factors — and not all results are equally informative, interesting, newsworthy, or beneficial to their authors' career. Some combination of these factors likely explains the observation that results do not seem to be recorded impartially in many scientific disciplines. Specifically, negative results of statistical hypothesis tests are published less frequently than positive results (Fanelli, 2010)

Registered Reports are an article format designed to combat publication bias by moving the peer-review process to the planning stage of a study, thus separating the publication decision from the study results (Chambers, 2013). In these *Registered Reports*, the review process is split in two stages: At Stage 1, reviewers evaluate a pre-study protocol containing the research questions, hypotheses, methods, and planned analyses of a proposed study. In case of a positive decision, the journal issues an 'in-principle acceptance' and commits to publishing the eventual report regardless of the direction of the results. Once authors have collected and analysed the data and written up the results, the final report is submitted to a second stage of peer review, but this time only to ensure that the study was carried out as planned, that the data pass any pre-specified quality checks, and that authors' conclusions are justified by the evidence.

1. RRs: what they are (brief)

2. In contrast to other reforms, RRs try to align selfish incentives for authors with what's good for science

3. RRs to the rescue 2.1 RRs don't select based on results 2.2 Brief description of how they work 2.3 RRs reduce publication bias & QRPs –> explain pub bias & QRPs in more detail, cite evidence

4. RRs have become more popular

5. In the standard system, studies are evaluated for publication after they're completed

6. This introduces a risk of selecting based on results

7. Result-based selection bias is bad
8. RRs select before results are known to avoid this problem
9. Explain how RRs work
10.

Publishing a scientific article typically means writing up a report of a completed research project and submitting it to a

In the standard model of scientific publication, the peer review process

peer reviewers and journal editors evaluate reports of completed research projects and decide whether to publish them.

Peer-reviewed journal articles continue to be the standard model of scientific publishing.

In the standard model of scientific publishing, a researcher writes up a report of scientific work they have completed and submits it to a journal, where peer reviewers provide criticism, suggest improvements, evaluate the quality of the research, and help the journal editor decide whether or not to publish the manuscript.

Under-reporting of negative results skews the available evidence for scientific claims and can lead to overconfidence and an increased rate of false-positive inferences. Evidence for such publication bias — negative results getting published at a lower rate than positive results — has been been found in several disciplines Annie Franco, Malhotra, & Simonovits (2016) and is seen as an important contributor to poor replicability of published studies in biomedical and psychological research Chalmers & Glasziou (2009). In 2013, the journal *Cortex* pioneered a new article format designed to combat publication bias by moving the peer-review process to the planning stage of a study, thus separating the publication decision from the study results (Chambers, 2013). In these *Registered Reports*, the review process is split in two stages: At Stage 1, reviewers evaluate a pre-study protocol containing the research questions, hypotheses, methods, and planned analyses of a proposed study. In case of a positive decision, the journal issues an 'in-principle acceptance' and commits to publishing the eventual report regardless of the direction of the results. Once authors have collected and analysed the data and written up the results, the final report is submitted to a second stage of peer review, but this time only to ensure that the study was carried out as planned, that the data pass any pre-specified quality checks, and that authors' conclusions are justified by the evidence.

Through this process, Registered Reports address publication bias as well as so-called 'questionable research practices' (QRPs), two problems that are considered important contributors to psychology's replication crisis Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit (2012) and to research waste in the biomedical sciences (Chalmers & Glasziou, 2009). Publication bias can result from editors and reviewers disproportionately rejecting submissions with negative results [reviewer bias; Greenwald (1975),Mahoney (1977),Atkinson, Furlong, & Wampold (1982)] or researchers failing to submit negative results for publication[file-drawering; Rosenthal (1979),A. Franco, Malhotra, & Simonovits (2014)]. In Registered Reports, the virtual publication guarantee issued at Stage-1 reduces both of these issues by ensuring that editors and reviewers cannot reject the Stage-2 report based on the results[ˆ] and thus reducing the incentive for authors to file-drawer the study in case of negative results.

QRPs denote practices which exploit undisclosed flexibility in data collection and analysis, for example by running the analyses on different justifiable combinations of variables, exclusion and decision criteria, and only reporting the ones with favourable results, or by presenting post-hoc inferences as having been predicted a priori John, Loewenstein, & Prelec (2012). Registered Reports minimise the risk of QRPs

Publication bias and QRPs skew the available evidence for scientific claims and can lead to overconfidence and an increased rate of false-positive inferences. Second, the risk of QRPs is minimised by the two-stage review process, in which the Stage-1 protocol acts as a preregistration and reviewers' task during Stage-2 review is to flag any undisclosed deviations from it.

The format was first launched in 2013 at the journal *Cortex* (Chambers, 2013) and is now offered by over 300 journals, predominantly in the social and life sciences (see cos.io/rr). By 2021, nearly 600 Registered Reports had been published (Chambers & Tzavella, 2021). Initial evidence shows that published Registered Reports have a substantially lower rate of positive results than regular articles in psychology (44% versus 96%, Scheel, Schijen, & Lakens, 2021) and psychology, neuroscience, and the biomedical sciences (Allen & Mehler, 2019), and are judged to be of higher quality (Soderberg et al., 2021).

Assuming that researcher practices s' publication practices are at least in part strategic responses to incentives, changing the incentives should

`To do:`

- General introduction describing the problem of publication bias and questionable research practices (QRPs)
  - Publication bias: positive results are more likely to get published than negative results; can happen due to reviewer bias (Greenwald, 1975; Atkinson et al., 1982; evidence: Mahoney, 1977) or file-drawering (Rosenthal, 1979; evidence: A. Franco et al., 2014; Annie Franco et al., 2016)
  - QRPs: exploiting undisclosed flexibility in data collection and analysis to obtain more desirable results (Simmons et al., 2011; Agnoli et al., 2017; Fiedler & Schwarz, 2016; Fraser et al., 2018; evidence: John et al., 2012). QRPs inflate the error rates of statistical tests, typically the false-positive rate

  - consequence of publication bias and QRPs: the literature in psychology is excessively (> 90%) positive (Fanelli, 2010; Scheel et al., 2021; Theodore D. Sterling, 1959; Theodor D. Sterling, Rosenbaum, & Weinkam, 1995) and unreliable (Wacholder, Chanock, Garcia-Closas, El ghormli, & Rothman, 2004; evidence: Nosek et al., 2022)
- Reform efforts to address the problem:
  - preregistration to reduce QRPs (Lakens, 2019; Nosek, Ebersole, DeHaven, & Mellor, 2018; Wagenmakers et al., 2012)
  - various journals of negative results to reduce publication bias (these never seem to be successful though and always shut down after a while; add examples/references)
  - **Registered Reports** to reduce QRPs and publication bias at the same time (most powerful reform proposal to date)

By moving the publication decision to a time point before results are known, Registered Reports provide a powerful protection against publication bias (publication is results-independent by design) and remove one important incentive for authors to use QRPs.

Being a powerful bias-prevention tool that is increasingly popular, it is important to develop a better understanding of when, where, and by whom Registered Reports are most likely to be used. First, such knowledge can help identify research areas in which the format is unlikely to gain traction by itself and anticipate the need for further intervention (e.g., via policy) when there is a demand for unbiased results. Second, understanding when researchers' choice between Registered Reports and the standard publication route is likely to be influenced by factors that also influence the eventual results (e.g., the prior probability of the tested hypotheses) is important for meta-scientific studies that compare published studies in both formats and must take such confounds into account (e.g., Scheel et al., 2021). Such confounds could also lead to a situation in which Registered Reports become associated with certain types of results (e.g., negative results) and devalued if these results are deemed less interesting or important by the research community, making the format unsustainable in the long run. `To do: add note that this is likely what happened to several 'journals of negative results' that shut down due to lack of interest.` The goal of this chapter is to shed light on these questions by studying the potential impact of a key feature of Registered Reports: The results-independent publication guarantee as an incentive for authors.

Registered Reports as a low-risk publication option

Registered Reports serve the scientific community and other consumers of the scientific literature by protecting against publication bias and QRPs. But they are also designed to 'serve the interests of individual scientists' (p. 12, Chambers & Tzavella, 2021) by providing a publication guarantee irrespective of the study results. As such, Registered Reports make use of existing incentive structures in academia and do not rely on changes in norms or policy (in contrast to other reforms such as preregistration).

Peer-reviewed publications are a central currency for the careers of academic researchers, both in terms of publication quantity and publication impact (R. Müller, 2014; van Dalen & Henkens, 2012). In the standard publication model, researchers face uncertainty about whether and where they will be able to publish the results of their study. Translated into currency terms, the career benefit a researcher receives for conducting a study can vary extremely — from near zero when the resulting manuscript is rejected by all consulted journals (or when the author file-drawers the study because the chances of success do not justify the cost of repeated submissions and revisions) to an extremely high, perhaps career-making amount when a manuscript is published in a very high-impact journal like *Nature* or *Science*. In other words, success in the standard system is highly variable and highly volatile since it hinges on the one factor that is supposed to be outside of researchers' control — the study results. This unfortunate combination can be excessively stressful for researchers (especially junior scientists without secure positions) and tempt them to hype, spin, or even fabricate their results.

Compared to this, Registered Reports are a relatively safe, stress-free alternative because authors receive a results-independent publication guarantee before investing in data collection or analysis. As Registered-Reports inventor Chris Chambers put it in a recent talk (September 2021):

> And the second main benefit, the one that really is the main big one, the big draw, is that as a researcher you can get your paper accepted before you even start your research and regardless of how the results turn out in the end. So no more playing the *p*-value lottery, gambling on certain results going a certain way, otherwise you won't have your PhD or you won't get your next fellowship or your next grant — takes all of that pointless, and I think quite foolish, gambling out of the equation completely. (from minute 17:27)

But would researchers ever choose the gamble over the safe publication? Unless the net benefit of a Registered Report is always at least as valuable as the best possible outcome that could be achieved through the standard publication route, the answer is 'probably yes'. Authors deciding between Registered Reports and the standard publication route face the choice between a payoff with low variability (a relatively safe publication in the journal the Stage-1 protocol was submitted to) and a payoff with high variability (anywhere between no publication and a high-impact publication, or even several publications if the project yields enough 'fodder'). Situations like these are commonly termed *decision-making under risk*. 'Risk' is defined as 'unpredictable variation in the outcome of a behavior, with consequences for an organism's fitness or utility' (Winterhalder, Lu, & Tucker, 1999, p. 302). Organisms are *risk sensitive* when they are not only sensitive to the mean outcomes of different behavioural options but also to their variance.

Framing authors' choice between Registered Reports and standard publications as risk-averse versus risk-prone behaviour allows us to examine the problem with Risk-Sensitivity Theory, a normative theory developed in behavioural ecology to explain the foraging behaviour of animals. Risk-Sensitivity Theory was designed to determine the optimal food-acquisition strategy for an animal faced with a choice between a relatively safe (low-variance) food source and a risky (high-variance) source that sometimes yields large payoffs and sometimes small payoffs (or none at all). Despite this initial narrow scope, Risk-Sensitivity Theory has proven itself as a powerful framework for explaining risk-sensitive behaviour in a wide range of situations and species, including humans (Alex Kacelnik & Bateson, 1996; A. Kacelnik & Bateson, 1997; Mishra, 2014).

`To do:`

- Explain that RST is superior to utility theory and can incorporate prospect theory (Mishra, 2014)
- Better explain the evolutionary angle and why it matters

Goals of the chapter

In this chapter, a simulation model is used to explore how properties of academic careers and academic incentive structures that are relevant to risk sensitivity may affect the strategies of researchers choosing between Registered Reports and the standard publication format. The research goal is to understand in which circumstances Registered Reports should be particularly attractive, particularly unattractive, or particularly prone to highly selective use. The results of this analysis may help anticipate where the format is unlikely to take foot without additional changes to norms, incentives, or policy, and flag situations in which the results of published Registered Reports may be particularly difficult to compare to the normal literature. The following sections outline central concepts of Risk-Sensitivity Theory, relate them to characteristics of academic careers, and describe an evolutionary simulation model in which their effects on researchers' risk-sensitive publication decisions are examined.

Conceptual application of Risk-Sensitivity Theory to publication decisions

This section describes general factors that affect the role of risk for individual's fitness and connects these factors to relevant elements of academic careers. In this context, Risk-Sensitivity Theory's focus on reproductive fitness as the central outcome may be seen as misguided. But although researchers do not forage, grow, reproduce, and die in the *biological* sense (except in their role as human beings in general, of course), they undoubtedly are concerned with factors that influence 1) their survival and 2) the propagation of their traits in an *academic* sense. Even if we were to assume that researchers are not consciously trying to maximise their 'academic fitness', a competitive job market will by definition select for individuals whose behaviour increases their fitness. In applying Risk-Sensitivity Theory to researchers' publishing behaviour, we will therefore use a general notion of career success as the central outcome variable in place of reproductive fitness. This decision does not imply that career success is the only or the proximal motivation for researchers' behaviour in practice, just as evolutionary theory does not imply that reproductive success is the only or the proximal motivation for human behaviour in everyday life. `To do:`

- Refer back to a (not yet existing) section above to say that human decision making is a product of evolution
- In addition, narrow bottlenecks between early-career and tenured positions in many academic disciplines inevitably create a selection pressure for behaviours that further researchers' career success (Smaldino & McElreath, 2016).

Non-linear fitness functions. The first and perhaps most ubiquitous factor leading individuals to be risk sensitive are non-linear relationships between the outcomes of an individual's behaviour (e.g., harvested food items, publications) and its reproductive success (A. Kacelnik & Bateson, 1997). Consider two options, O_{safe} and O_{risky}. O_{safe} always gives the same payoff b_{safe}, whereas O_{risky} gives either a low payoff b_- or a high payoff b_+,

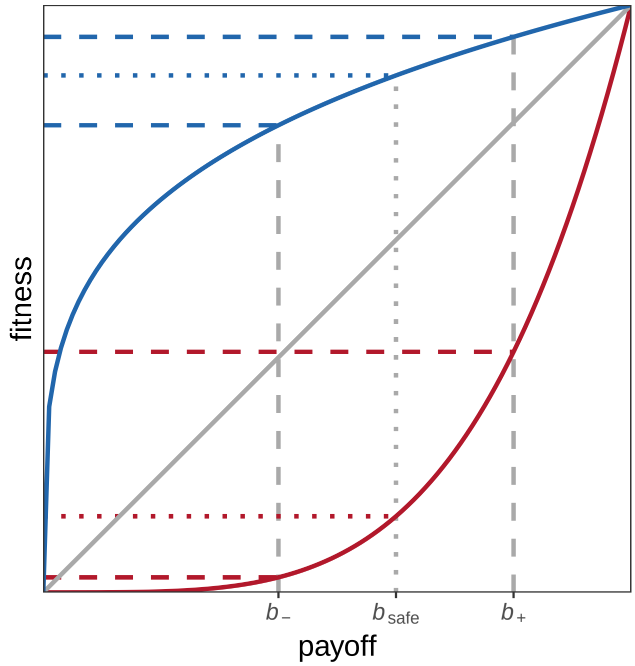

Figure 1. Consequences of non-linear fitness functions. Payoffs b_-, b_{safe}, and b_+ are converted into fitness with a diminishing (blue), linear (grey), or increasing (red) returns function.

each with probability $\frac{1}{2}$. When $b_{safe} = \frac{(b_- + b_+)}{2}$, O_{safe} and O_{risky} have the same expected payoff. However, we would only expect an individual to be indifferent between the two options if the consequences of their payoffs for the individual's fitness are linear. When the function relating payoffs to utility is instead convex or concave (yielding increasing or diminishing returns, respectively), the expected utility of O_{safe} and O_{risky} will differ and shift the individual's preference towards risk proneness or risk aversion. An illustration of this example is shown in Figure 1: While the payoffs b_-, b_{safe}, and b_+ are equidistant on the x-axis, b_{safe} is associated with greater fitness than the average of b_- and b_+ when the fitness function is concave, and with less fitness when the fitness function is convex. In other words, O_{safe} has greater expected fitness than O_{risky} when returns are diminishing, and O_{risky} has greater expected fitness than O_{safe} when returns are increasing.

Non-linear relationships are arguably the norm in the natural world and linear relationships the exception. This plausibly holds for academia as well, where the effect of publication success on researchers' career success might change over time: For early-career researchers, small increases in the number or impact of publications may have an accelerated effect on career success, whereas established professors may care little about any one additional publication on their record.

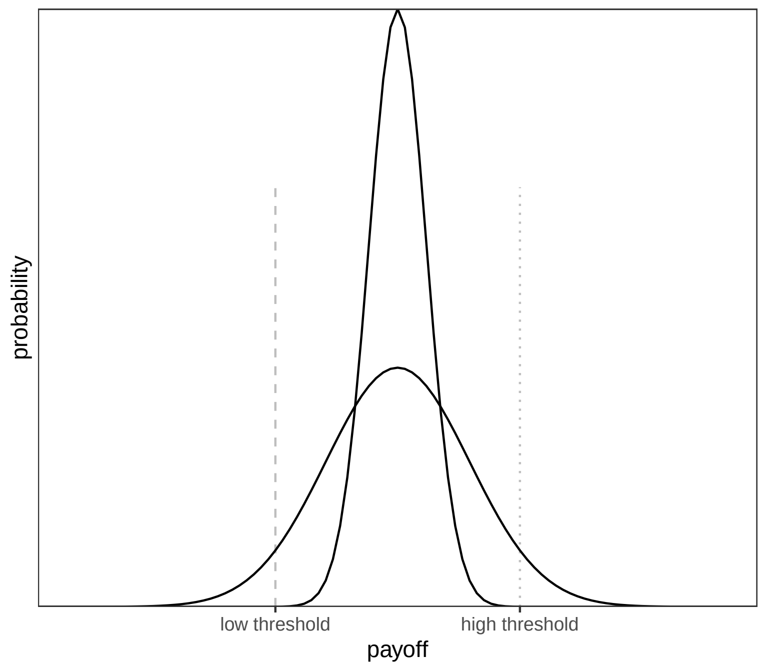

Figure 2. Survival thresholds. When fitness drops to zero below the low threshold (dashed line), individuals should be risk-averse because the outcomes of the low-risk option (narrow curve) are guaranteed to lie above the threshold and the outcomes of the high-risk option (wide curve) have a non-negligible risk of falling below the threshold. When fitness drops to zero below the high threshold (dotted line), individuals should be risk-prone because only the high-risk option provides a chance of passing the threshold.

Survival thresholds and competition. A second important factor for risk-sensitive behaviour are thresholds for survival and reproduction (Hurly, 2003; Winterhalder et al., 1999). Survival thresholds are cutoff points below which an individual's fitness drops to zero, for example due to starvation. Risk-Sensitivity Theory predicts that an individual will be risk averse when the resources provided by a low-variance option are sufficient to meet the threshold and risk-prone when they are not (Mishra, 2014). For example, a hummingbird that needs to acquire a certain amount of calories to survive the night will prefer a low-risk food source if this option's expected payoff is above the threshold, but avoid the low-risk source if only a higher-risk source provides a chance of survival. One such situation is depicted in Figure 2.

Although comparable cutoff points in academic careers may have somewhat less severe consequences, they certainly exist: Amount and impact of a researcher's publications are common and often explicit criteria in decisions that are central to the individual's career, such as whether they will be awarded a PhD, whether they will receive grant funding, whether they will be offered a tenure-track position, or whether they will be granted tenure. In some of these situations, the cutoff points are absolute and thus resemble survival thresholds in the biological sense, for example PhD-programme regulations that determine a minimal amount of peer-reviewed publications for a candidate to be awarded with a PhD, or tenure contracts that specify minimal publication targets. In other situations, the cutoff points are relative and depend on the number of eligible candidates, for example when grant funding is awarded to the 10 highest-ranked research proposals or a job is offered to the best candidate from a pool of applicants. In cases like these, one individual's success diminishes the chances of another — they represent *competition*. In the following, survival thresholds and competition will be treated as separate concepts to examine their differential effects on researchers' publication behaviour.

Number of decision events before evaluation. A final risk-relevant factor considered here is the number of decision events taking place before an individual's fitness is evaluated. When a risky option is chosen repeatedly, the average of the accumulating payoffs gets closer and closer to the long-run expected payoff. This means that the danger of loosing out completely by only acquiring the lowest possible payoff of the risky option diminishes, making the risky option relatively more attractive. However, this relationship only holds for repeated decision events *before* an individual's fitness is evaluated. When fitness is evaluated after a single decision event, a risky option is more likely to yield an extreme outcome that translates to zero fitness (i.e., death or an ultimate failure to reproduce).

In situations like this, when a single risky decision might cost an individual's life or offspring, average fitness is best described by the geometric mean instead of the arithmetic mean (Haaland, Wright, & Ratikainen, 2019). The geometric mean is more sensitive to variance because it is multiplicative, capturing the fact that one failure to reproduce can end a genetic lineage. This circumstance has been shown to produce bet-hedging: Risk-averse strategies may be more adaptive across many generations even when more risk-prone strategies produce better outcomes in any one generation, simply because the latter are also more likely to lead to extinction by sheer bad luck (Haaland et al., 2019). While average fitness across generations is best represented with the geometric mean, average fitness *within* a generation is better captured by the arithmetic mean, reflecting the additive accumulation of payoffs from decision events before fitness is evaluated. Therefore, as the number of decision events per generation (i.e., before fitness is evaluated) increases, the variance-sensitive geometric mean of acquired payoffs becomes relatively less important and the less variance-sensitive arithmetic mean becomes more important. Consequently, individuals' behaviour should switch from relative risk-aversion to relative risk-proneness.

In the academic world, decision events before fitness is evaluated ('per generation') could seen as the time and resources a researcher has available for producing publications before a relevant selection event like those mentioned in the previous section (award of a PhD or grant, job application, tenure decision) is made. This parameter likely varies with career stage: A PhD student usually has three to four years to achieve the required publication output, a postdoc may work on a short-term contract of two years or even one year (after which their CV must be strong enough for the next application), and an assistant professor may have around seven years for receiving tenure. In addition, career progress often comes with greater research funds and, most importantly, the supervision of students and junior researchers whose efforts boost the supervisors' output (R. Müller, 2014). As a second, orthogonal aspect, the amount of publishable research that can be achieved before a selection event may vary between research areas. In some fields, data collection is fast and cheap, for example when experiments consist of short online questionnaires that are disseminated to large participant pools such as Amazon MTurk. In other fields, data collection is very expensive and slow, for example in clinical fMRI studies on specific patient groups. Irrespective of career stage, researchers in fields with fast and cheap data may thus be able to complete many more research cycles per time unit than researchers who use more expensive data.

Each of the risk-relevant factors described above — non-linear fitness functions, survival thresholds, competition, and number of decision events before evaluation — likely impacts researchers' decision strategies, including their choices between low-risk and high-risk publication options. To better understand when a low-risk option like Registered Reports should be particularly attractive or unattractive, the individual and interactive effects of these factors are examined in a simulation model.

Simulation model

We develop an evolutionary agent-based model which simulates a population of researchers who test hypotheses, (attempt to) publish the results either as Registered Reports or as standard reports, accumulate the payoffs for successful publications, and pass their publication strategies on to the next generation of researchers.

Research phase. Consider a population of $n = 500$ researchers. Each researcher has a fixed publication strategy s, the so-called submission threshold. In each round of the research phase, researchers randomly choose a hypothesis to test in a study. Hypotheses are true with prior probability p, which is uniformly distributed between 0 and 1 and known to the researcher. Before testing their chosen hypothesis, a researcher compares the prior p of their hypothesis with their submission threshold s. When $p < s$, the researcher chooses to play it safe and conduct a Registered Report to test the hypothesis. When $p \geq s$, the researcher chooses to gamble and test the hypothesis in a regular study which is then submitted as a standard report.

For simplicity, we assume that p is an ideal objective prior and that researchers' hypothesis tests are free from additional sources of error. Thus, when a researcher tests hypothesis i, they obtain a positive result with probability p_i and a negative result with probability $1 - p_i$. If the researcher chose to submit a Registered Report, their study is published regardless of the result and the researcher receives a payoff b_{RR}. However, if the researcher chose to submit a standard report, they face rampant publication bias: Only positive results are publishable as standard reports and yield a payoff $b_{SR+} = 1$, whereas negative results are rejected or file-drawered and yield no payoff, $b_{SR-} = 0$. For all variations of the model tested here, we assume that $b_{SR-} < b_{RR} < b_{SR+}$. This assumption reflects the following considerations:

1. Due to publication bias in the standard-report literature, negative results are less valuable than positive results ($b_{SR-} < b_{SR+}$), for example because they do not lead to a publication at all, because only very low-impact journals are willing to publish them, or because getting them published requires a lot of extra effort (e.g., via frequent resubmissions following rejection or substantial revisions demanded by reviewers) which diminishes the net reward.
2. For these same reasons, Registered Reports are on average more valuable than standard reports with negative results ($b_{SR-} < b_{RR}$), for example because Registered Reports are offered by journals that may display publication bias for standard reports (rejecting standard report submissions with negative results), or simply because Registered Reports need to be resubmitted less often or require less extensive revisions.
3. On average, standard reports with positive results are more valuable than Registered Reports ($b_{RR} < b_{SR+}$), for example because many high-impact journals do not (yet) offer Registered Reports, because not registering one's study *a priori* makes it easier to spin the results to appear more impactful and thus increases the chances to be published in a high-impact journal, or because Registered Reports may require more effort due to their stricter quality criteria, lowering the net reward. While proponents of Registered Reports may argue that the format has such tremendous advantages that authors' resulting career benefits are superior to any alternative, this chapter is

predicated on the assumption that most researchers currently do not share this view. Once this changes, the present investigation may happily become redundant.

This entire research cycle — choosing a hypothesis, choosing a publication route by comparing its prior p to one's submission threshold s, testing the hypothesis, and receiving payoff b_{RR} for a Registered Report or b_{SR-} or b_{SR+} for a positive and negative standard report, respectively — is repeated m times.

Evaluation phase. At the end of the research phase, researchers' accumulated publication payoffs $b_1 + b_2 + ... + b_m$ are translated into fitness. Fitness is calculated with a function characterised by exponent ϵ, which determines the shape of the function. $\epsilon = 1$ yields a linear function, $0 < \epsilon < 1$ yields a concave function with diminishing returns, and $\epsilon > 1$ yields a convex function with increasing returns (see Figure 1):

$$fitness = (\sum_{i=1}^{m} b_i)^{\epsilon} \tag{1}$$

However, two situations may cause a researcher's fitness to fall to zero even when their accumulated payoffs are non-zero. First, the sum of their payoffs may fall below an absolute survival threshold δ, for example when a researcher fails to meet an agreed publication target by the time their 'tenure clock' runs out. Thus, when $\sum_{i=1}^{m} b_i < \delta$, $fitness = 0$. Second, the sum of their payoffs may fall below a relative threshold γ, which reflects the intensity of competition (e.g., for scarce research grants or positions). γ is the proportion of researchers who are considered for reproduction. When $\gamma = 1$, all researchers in the population are considered for reproduction and their fitness is calculated according to Eq. 1. When $\gamma < 1$, the $(1 - \gamma) * 500$ least successful researchers receive zero fitness and cannot reproduce.[1] For

[1] In the simulation, γ is applied *after* fitness has been calculated, not before. This change has purely technical reasons and leads to the same result as applying γ to accumulated payoffs and then calculating fitness because all fitness functions are monotonic increasing and fitness functions do not vary within a population.

example, $\gamma = 0.1$ means that only those researchers with accumulated payoffs in the top 10% of the population can reproduce, and the fitness of the remaining 90% is set to zero.

Table 1
Parameter definitions and values

Parameter	Definition	Value [range]
n	population size	500
g	number of generations	250
p	prior probability of hypotheses	uniform [0–1]
b_{SR-}	payoff for negative standard report	0
b_{SR+}	payoff for positive standard report	1
b_{RR}	payoff for Registered Report	[.1, .2, . . . , .9]
ϵ	fitness function exponent	[0.2, 1, 5]
m	research cycles per generation	[1, 2, 4, 8, 16, 32]
δ	survival threshold below which fitness = 0, expressed as proportion of m	[0, .25, .5, .75]
γ	proportion of most successful researchers selected for reproduction (competition)	[1, .9, .5, .1, .05, .01]

Reproduction phase. Finally, the researchers in the current population retire and a new (non-overlapping) generation of researchers is created. A researcher in the new generation inherits their publication strategy (submission threshold) s from a researcher in the previous generation with the probability of the previous researcher's fitness (i.e., the new generation's submission thresholds are sampled with replacement from the previous generation, probability-weighted by fitness). The new generation's submission thresholds are inherited with a small amount of random noise, such that $s_{new} = s_{old} + w$, with $w \sim N(\mu = 0, \sigma = 0.01)$. Authors of similar evolutionary agent-based models have described

That is, applying the fitness function does not affect the rank order of researchers in the population.

such heredity transmission as reflecting mentorship and teaching (e.g., when established professors advise mentees to copy their strategies) or simply a generic social learning process in which successful researchers are more likely to be imitated by others (Smaldino & McElreath, 2016). Although this interpretation may be useful, the main purpose of this aspect of the model is purely technical and not specifically intended to reflect reality — it simply provides the machinery for determining which publication strategies are optimal in the various situations we are investigating.

Outcome variable s. We study how the evolution of researchers' submission thresholds s is affected by the payoff for Registered Reports b_{RR} (relative to the payoffs for standard reports, which are fixed at $b_{SR-} = 0$ and $b_{SR+} = 1$), by the shape of the fitness function determined by exponent ϵ, by the number of research cycles per generation m, by survival threshold δ, and by competition γ (see Table 1 for an overview of the model parameters and their values considered in the simulation). A researcher's submission threshold s is a *strategy*, not an absolute decision — it determines *how* the choice between Registered Reports and standard reports is made, not which format is chosen. As such, s indicates the amount of risk a researcher is willing to take. Very low values of s reflect risk proneness: The researcher prefers to gamble and chooses the standard publication route for almost all hypotheses they encounter, using the Registered Report route only for hypotheses that are virtually guaranteed to be false (and yield negative results). Very high values of s reflect risk aversion: The researcher is unwilling to risk a negative result in a standard report and studies almost all hypotheses they encounter in the Registered Report format, reserving the standard publication route for hypotheses that are virtually guaranteed to be true (and yield positive results).

Simulation approach. We use the evolutionary mechanism of this agent-based model as a means for identifying optimal behaviour under different conditions. But this goal can also be achieved in other ways. One non-evolutionary alternative is to calculate expected fitness (i.e., the long-run average) for a wide range of s and determine which strategy maximises expected fitness in each condition. A drawback of this approach is that it does not account for population dynamics and therefore cannot easily simulate the effects of competition. Because of this limitation, our study is based on the evolutionary model. However, we use the expected-fitness model to validate all analyses except those involving competition and show that both models produce virtually identical results (see Appendix).

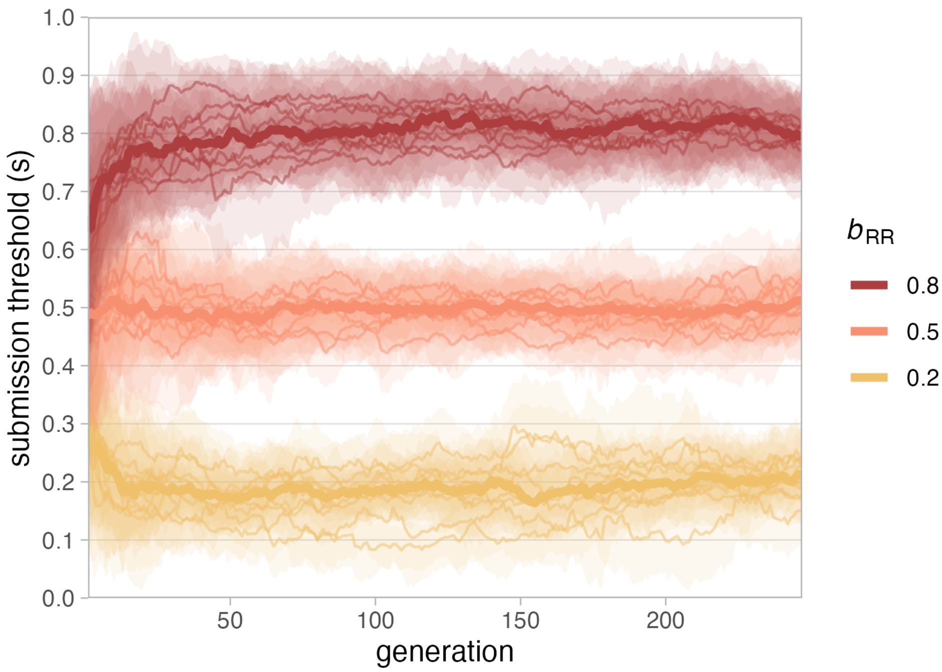

Figure 3. Evolution of submission threshold s with 3 different payoffs for Registered Reports (b_{RR}). Simulations are based on a population of $n = 500$ researchers over 250 generations, with payoffs for standard reports fixed at 0 for negative results ($b_{SR-} = 0$) and 1 for positive results ($b_{SR+} = 1$), a linear fitness function $\epsilon = 1$, one research cycle per generation ($m = 1$), no survival threshold ($\delta = 0$) and no competition ($\gamma = 1$). Each condition was run 10 times. Thin lines represent the median submission threshold of the population in each run, shaded areas represent the inter-quartile range of submission thresholds in the population in each run, thick lines represent the median of run medians per condition.

Simulation results

The results of the simulation models will be presented in order of increasing model complexity. We start by explaining the very simple scenarios shown in Figure 3 and Figure 4. These conditions are identical to situations discussed above and the results should thus be unsurprising. However, while they may seem trivial to some, we hope that these explanations will help unfamiliar readers understand the functioning of our model and the less intuitive results presented later.

When interpreting the results below, it is important to keep in mind that the analysed parameter values are inherently arbitrary. Although the model parameters are chosen to capture important characteristics of real-world concepts, their values do not represent real-world units. The goal of this analysis is to understand the relative effects of the model parameters in a simplified, artificial system, which means that the results are only meaningful in relation to each other.

Single research cycle per generation, linear fitness function

The first generation of researchers in each simulation run is initialised with randomly distributed submission thresholds s (drawn from a uniform distribution [0–1]), which are then allowed to evolve over the subsequent generations. Figure 3 shows the effect of varying the payoffs for Registered Reports when the fitness function is linear ($\epsilon = 1$), with no survival threshold ($\delta = 0$), no competition ($\gamma = 1$), and one research cycle per generation ($m = 1$). In this very simple scenario, evolved submission thresholds (s) approximate the payoff for Registered Reports in each condition, indicating that the optimal submission threshold is always equal to b_{RR} ($s_{optimal} = 0.2$ when $b_{RR} = 0.2$, $s_{optimal} = 0.5$ when $b_{RR} = 0.5$, $s_{optimal} = 0.8$ when $b_{RR} = 0.8$). The reason behind this is the uniform distribution [0–1] of hypothesis priors, the payoff structure $b_{SR-} = 0$ and $b_{SR+} = 1$, and the linear fitness function ($\epsilon = 1$ means that fitness equals payoff). In this constellation, the expected fitness obtained from a standard report is always equal to the prior of the tested hypothesis:

$$E[f_{SR}] = (p * b_{SR+} + (1-p) * b_{SR-})^1 = p * 1 + (1-p) * 0 = p \tag{2}$$

For example, testing a hypothesis with $p = 0.2$ in a standard report would yield the expected fitness $E[f_{SR}] = (0.2 * 1 + 0.8 * 0)^1 = 0.2$. The optimal strategy is to submit a Registered Report whenever the expected fitness provided by a standard report is lower than the fitness provided by a Registered Report, $E[f_{SR}] < b_{RR}$, and thus whenever $p < b_{RR}$. This ensures that researchers always get the best of both worlds, minimising shortfalls when priors are (too) low and maximising winning chances when priors are (sufficiently) high. For example, $b_{RR} = 0.5$ is larger than $E[f_{SR}]$ for all hypotheses with $p < 0.5$ but lower than $E[f_{SR}]$ for all hypotheses with $p > 0.5$. In this situation, researchers who submit Registered Reports whenever $p < 0.5$ and standard reports whenever $p > 0.5$ protect themselves against losing a bad bet by instead taking the fixed payoff $b_{RR} = 0.5$, but always play a good bet and thus maximise their chances of winning $b_{SR+} = 1$. Every alternative is inferior in the long run because researchers with $s > b_{RR}$ lose out on increased chances of publishing a standard report and researchers with $s < b_{RR}$ take unnecessary risks and go empty-handed too often.

Allowing for non-linear fitness functions

In the real world, the career benefits researchers receive from publications are rarely, if ever, linear. In early career, we may assume a convex fitness function, with each addition to the short publication record of a young researcher yielding increasing returns for their prospects on the job market and their ability to obtain grant funding. A notable exception may be PhD students who plan to leave academia after obtaining their degree, and for whom the career returns of publications exceeding the PhD requirements are thus strongly decreasing (concave fitness function). Researchers who stay in academia may experience that the career returns for each additional publication begin to decrease as their publication record grows, meaning that advanced career stages may also be characterised by a concave fitness function.

Figure 4 contrasts the effects of two concave fitness functions ($\epsilon = 0.2$ and $\epsilon = 0.5$, shown in blue shades) and two convex fitness functions ($\epsilon = 2$ and $\epsilon = 5$, shown in red

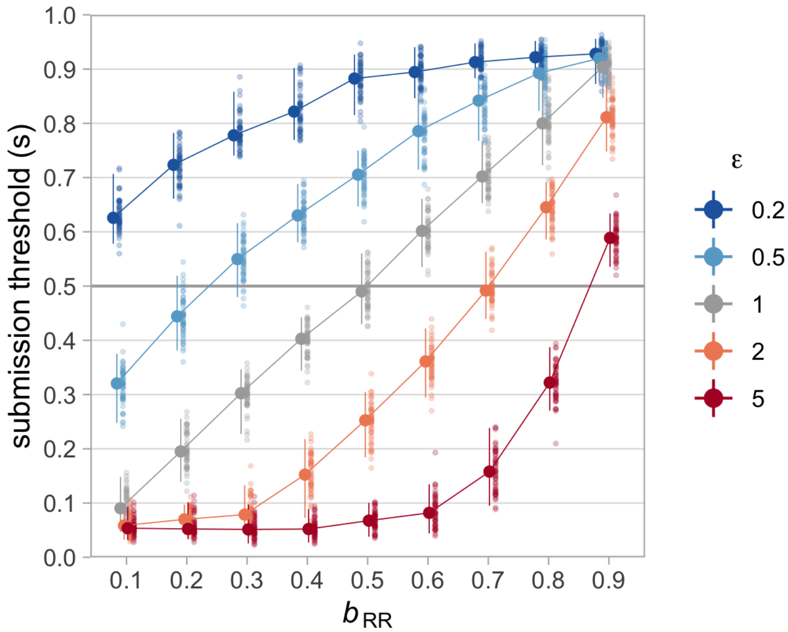

Figure 4. Effect of fitness functions on evolved submission thresholds. Shown are median submission thresholds in the final (250^{th}) generations of 50 runs for different values of b_{RR} (x-axis) and different fitness functions (characterised by exponent ϵ), with one research cycle per generation ($m = 1$), no survival threshold ($\delta = 0$) and no competition ($\gamma = 1$). Fitness functions with $\epsilon = 0.2$ and $\epsilon = 0.5$ (blue lines) are concave with diminishing returns, functions with $\epsilon = 2$ and $\epsilon = 5$ (red lines) are convex with increasing returns, and the function with $\epsilon = 1$ (grey line) is linear. Small dots represent median s of the final generation in each run, large dots represent the median of these 50 run medians per condition. Error bars represent the 95% capture probability around the median of medians.

shades) with a linear function ($\epsilon = 1$, grey line) for different payoffs for Registered Reports, in the same simple scenario with only one research cycle per generation. The grey line for $\epsilon = 1$ represents the already familiar situation from Figure 3 above: When the fitness function is linear, the optimal strategy is $s_{optimal} = b_{RR}$. Non-linear fitness functions deviate from this pattern exactly as expected based on Figure 1. When additional payoffs yield diminishing returns ($\epsilon < 1$), Registered Reports become more attractive even when they are worth less than the expected payoff for standard reports. As explained above, this is because concave functions 'shrink' the difference between moderate and high payoffs relative to the difference between low and moderate payoffs. Conversely, when additional payoffs yield increasing returns ($\epsilon > 1$), Registered Reports are unattractive unless their payoffs are almost as large as those for published standard reports because convex functions increase the difference between moderate and high payoffs relative to low versus moderate payoffs.

When different fitness functions are taken to reflect different career stages, this pattern suggests that Registered Reports should be more attractive for senior researchers and a tough sell for early-career researchers. Interestingly, preliminary empirical evidence suggests the opposite: Registered Reports appear to be more likely to have early-career researchers as first authors than standard reports (77% vs 67% in the journal *Cortex*, Chambers & Tzavella, 2021). One explanation for this counterintuitive result could be that Registered Reports are disproportionally used by early-career researchers who intend to leave academia and thus have a concave fitness function. Alternatively, it is possible that factors or dynamics not considered in this simulation swamp out the effects of concave vs convex fitness functions, such as younger researchers being more likely to adopt new methods. However, as we will see below, the effects of different fitness functions are not always as straightforward as in the simple case illustrated in Figure 4 but produce different results in interaction with other risk-related factors.

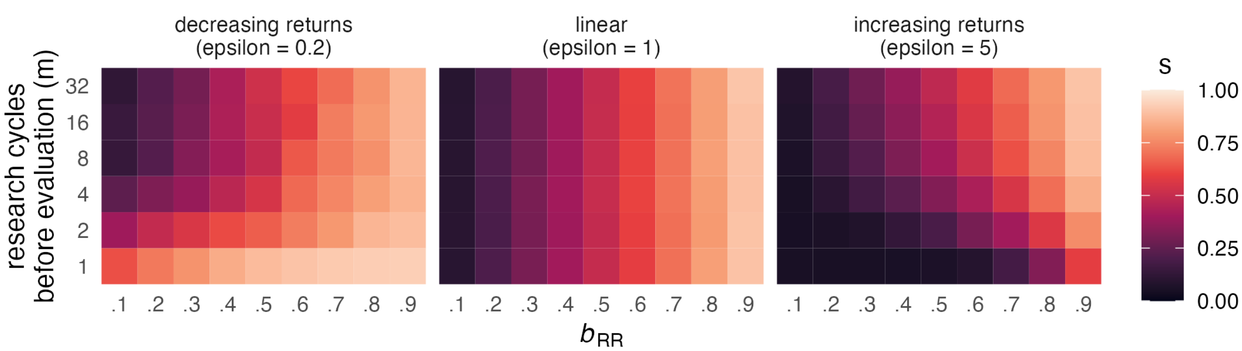

Figure 5. Effect of research cycles per generation on evolved submission thresholds. Shown are median evolved submission thresholds (s) after 250 generations in 50 runs (tile colour represents the median of 50 run medians) depending on the number of research cycles per generation (m, y-axis), different values of b_{RR} (x-axis), and different fitness functions (characterised by exponent ϵ) with no survival threshold ($\delta = 0$) and no competition ($\gamma = 1$).

Varying the number of research cycles per generation

The analyses discussed so far focused on the simple case of one research cycle (or decision event) per generation, meaning that researchers' fitness was calculated based on the payoff from one single study. As discussed above, increasing numbers of decision events prior to evaluation may make individuals more risk-prone because single negative outcomes are less catastrophic for reproduction (Haaland et al., 2019). However, Figure 5 shows that this is not universally true — rather, the effect of increasing numbers of research cycles per generation (m) depends on the shape of the fitness function. Moving up on the y-axis of each panel, we see that submission thresholds are decreasing (indicating greater risk proneness) only when the fitness function is concave ($\epsilon = 0.2$, left panel) but stay constant when it is linear ($\epsilon = 1$, middle panel) and even *increase* when it is convex ($\epsilon = 5$, right panel).

Why does m appear to have opposite effects for concave and convex fitness functions? To understand this pattern, it helps to first consider only the bottom row of each panel, where $m = 1$. These three rows contain the same results as the top, middle, and bottom curves in Figure 4 and show risk aversion when $\epsilon = 0.2$ (i.e., Registered Reports are attractive even when they yield a low payoff), risk proneness when $\epsilon = 5$ (Registered Reports are unattractive even when they yield a high payoff), and a linear strategy $s_{optimal} = b_{RR}$ when $\epsilon = 1$. From this starting point, the two panels with non-linear fitness functions start to approximate the linear case as m increases. This pattern reflects the idea that fitness is better captured by the geometric mean when m is low, and better captured by the arithmetic mean when m is high (Haaland et al., 2019).

To better understand this dynamic, let's consider two researchers with extreme submission strategies: Regina Register conducts only Registered Reports ($s_{Regina} = 1$), Darren Daring conducts only standard reports ($s_{Darren} = 0$). The payoff for Registered Reports is fixed at $b_{RR} = 0.5$. After one research cycle, Regina receives a payoff of 0.5 and Darren receives either 0 or 1 (with 50/50 odds). If fitness is calculated after this one round with $\epsilon = 0.2$ (concave function, yielding diminishing returns), Regina's fitness is $f_{Regina} = \frac{1}{2}^{\frac{1}{5}} = 0.87$, and Darren's fitness is either $f_{Darren-} = 0^{\frac{1}{5}} = 0$ or $f_{Darren+} = 1^{\frac{1}{5}} = 1$. In a population of 100 Reginas and 100 Darrens, there will be roughly 50 lucky Darrens who get a positive result and 50 Darrens who get a negative result. Lucky Darrens have a narrow fitness advantage over all Reginas (1 versus 0.87), while unlucky Darrens lose to all Reginas by a wide margin (0 versus 0.87). Since there are twice as many Reginas as lucky Darrens, the Regina strategy is relatively more successful.

Let's now consider the same scenario with $m = 4$ research cycles per generation. Reginas receive the same payoff in every round and accumulate $b_{total} = \frac{1}{2} * 4 = 2$. Lucky Darrens (who win every time) accumulate $b_{total} = 1 * 4 = 4$, while unlucky Darrens (who lose every time) again receive 0 total payoff. Now, however, the probabilistic outcomes over 4 rounds lead to three additional versions of Darren: moderately lucky (winning 3/4 times), average (2/4, receiving the same total payoff as Reginas), and moderately unlucky (1/4). Translating payoffs into fitness, the Regina strategy ($f_{Regina} = 2^{\frac{1}{5}} = 1.15$) still yields an enormous advantage compared to unlucky Darrens ($f_{Darren_{unlucky}} = 0$) and only a small disadvantage compared to lucky Darrens ($f_{Darren_{lucky}} = 4^{\frac{1}{5}} = 1.32$). But this time, there are fewer Darrens who are less successful than Reginas because Reginas now share their place with average Darrens — the relative fitness advantage of the Regina strategy decreases.

As the rate of research cycles per generation grows, the law of large numbers dictates that more and more Darrens achieve average total payoffs, while fewer and fewer Darrens achieve extreme total payoffs (winning 32 times in a row is much less probable than winning 4 times in a row). This reduces the width of the Darren distribution until it approximates the Regina distribution. When the fitness function is convex ($\epsilon = 5$, yielding increasing returns), the overall effect of increasing values of m is the same, with the only difference that Reginas are initially disadvantaged (because their fitness distance to the lucky half of Darrens is much greater than than to the unlucky Darrens). With larger m, more and more

Darrens receive average total payoffs and share Regina's disadvantaged position (decreasing Regina's relative disadvantage), until the Darren distribution is again virtually equal to the Regina distribution. Rather than causing absolute risk aversion, increasing values of m thus counter the effect of ϵ and reduce the effects of all fitness functions to the linear case. Consequently, the top rows ($m = 32$) of the top and bottom panels in Figure 5 resemble the stable pattern across all m shown in the middle panel.

Translated into terms of academic careers, this less intuitive pattern indicates that being able to complete empirical studies at a higher rate — e.g., when working in a field where data collection is fast and cheap or when having more resources for data collection available — may cancel out the effects of different career stages. This could partly explain why Registered Reports appear to be less popular among senior researchers (Chambers & Tzavella, 2021) than we would expect based on the effects of different fitness functions alone: Although additional publications likely yield diminishing returns in later career stages (concave fitness function), academic seniority often comes with resources that boost research output per time (e.g., more lab members). As a consequence, established professors may be relatively indifferent to Registered Reports. Junior researchers, for whom additional publications have increasing returns on career success, may be especially reluctant to use Registered Reports when they have very limited time or resources to produce publications before an important selection event, such as on short-term postdoc contracts (R. Müller & de Rijcke, 2017).

Absolute survival thresholds

The survival thresholds in our model represent absolute publication targets that researchers must meet in order to progress in their career. The clearest examples for such thresholds are PhD regulations and tenure agreements. To be awarded with a PhD, many institutions and faculties require candidates to have a certain number of their thesis chapters published in peer-reviewed journals. Similarly, tenure agreements may include publication

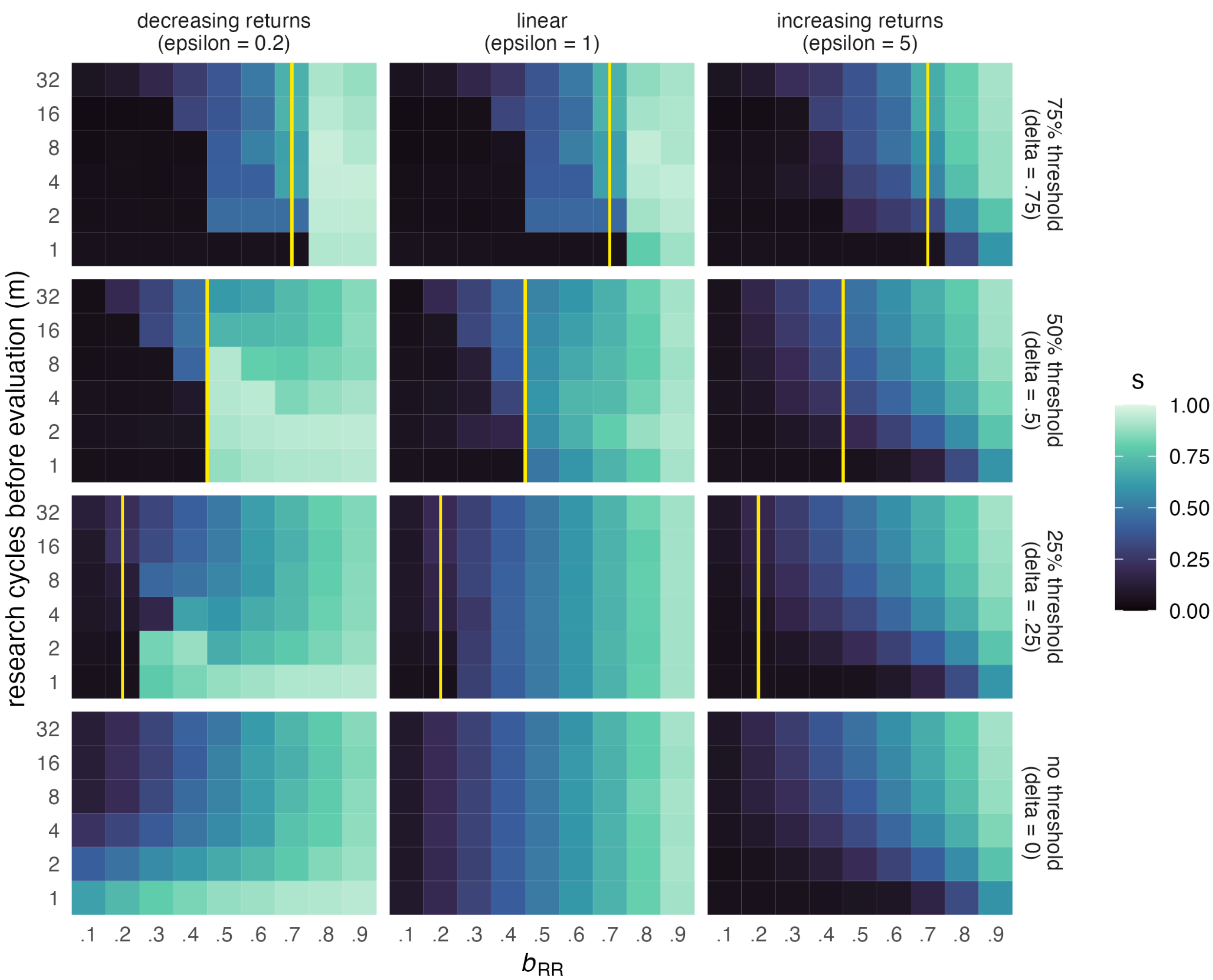

Figure 6. Effect of survival thresholds on evolved submission thresholds. Shown are median submission thresholds (s) after 250 generations in 50 runs (tile colour represents the median of 50 run medians) depending on survival thresholds (δ, shown as vertical yellow line), fitness functions (characterised by exponent ϵ), numbers of research cycles per generation (m), and values of b_{RR}, in the absence of competition ($\gamma = 1$). Survival thresholds are set as proportions of m, i.e., as a percentage of the maximum possible payoff in each condition. To reproduce, researchers must accumulate a total payoff exceeding $\delta * m$.

targets in the form of a minimum number of peer-reviewed publications within a certain time, sometimes also specifying minimal journal ranks (Liner & Sewell, 2009). How demanding such requirements are (e.g., what proportion of thesis chapters must be published) then determines whether they represent low, medium, or high survival thresholds.

We investigate the effects of survival thresholds representing 25%, 50%, and 75% of the maximum possible payoff researchers can achieve in one generation. When $\delta > b_{RR}$, Registered Reports alone are not sufficient to reach the survival threshold (b_{RR} values to the left of the yellow line in Figure 6). For example, at $m = 4$, a survival threshold of 75% ($\delta = .75$) means that researchers must gain at least 3 points to be able to reproduce. When $b_{RR} = .7$, submitting four Registered Reports will only amount to 2.8 points in total, just short of meeting the threshold. On the other hand, when $b_{RR} = .8$ (i.e., just above δ), four Registered Reports would yield 3.2 points and thus ensure reproduction. Choosing the standard route some of the time can increase fitness even further, but also increases the risk of not meeting the survival threshold. As a consequence, one may intuitively expect Registered Reports to be popular whenever $\delta \leq b_{RR}$ and unpopular whenever $\delta > b_{RR}$.

Figure 6 shows that this is true in many, but not all conditions. First, we can see that survival thresholds have their biggest effect when the number of research cycles per generation is low — at high values of m, publication strategies are virtually unaffected in all conditions. Second, survival thresholds have a stronger effect when the fitness function is linear ($\epsilon = 1$) or concave ($\epsilon = 0.2$). In these two conditions, they produce very similar patterns: The Registered Report route is almost never chosen when b_{RR} is too low to meet the survival threshold (particularly at $\delta = .25$ and $\delta = .5$; less so at $\delta = .75$), and this effect tapers off as the number of research cycles increases. Compared to baseline, the change is particularly striking for the concave fitness function ($\epsilon = 0.2$, left column in Fig. 6), where RRs are normally preferred at low m. When the survival threshold is high ($\delta = .75$) or the fitness function is concave, we can also see that Registered Reports become *more* popular when they are worth just enough to pass the survival threshold. For the convex fitness function ($\epsilon = 5$) on the other hand, survival thresholds of 25% and 50% seem to have no effect at all. Only a high threshold of 75% makes RRs even less popular when they have low value ($b_{RR} \leq 0.4$), especially when the number of research cycles is low.

What does this mean in practice? In our model, fitness (according to the three different fitness functions) is calculated after the survival threshold has been met. This is meant to mimic publication requirements that are expressed in raw numbers rather than some form of value. Importantly, it also means that our simulation shows which strategies during a PhD or on the tenure track lead to maximal fitness *after* researchers have successfully obtained their PhD or have been granted tenure. With this in mind, it becomes easier to understand the meaning of the different fitness functions. As discussed above, PhD candidates plausibly receive increasing returns for additional publications (convex fitness function), unless they intend not to stay in academia, in which case returns are strongly decreasing (concave fitness function). For researchers on the tenure track, the fitness function after achieving tenure is also likely concave, assuming a) that achieving tenure is one of the most important career goals for many (making further progress less important) and b) that such individuals have already built up substantial publication records, to which any single addition makes less and less of a difference. However, exceptions from this scenario may well exist, for example in situations where even tenured researchers are under great pressure to obtain grant funding.

Translated to real-world scenarios, our results thus suggest the following implications: First, survival thresholds are almost irrelevant when researchers can complete large numbers of studies before they are evaluated (reflecting characteristics of the research field, available resources, or length of the evaluation period). Second, researchers with a convex fitness function — such as PhD candidates who are pursuing an academic career — are only affected by high survival thresholds, which lead them to choose Registered Reports even less often than normal when their value is low. Third, researchers with a concave fitness function — such as tenure candidates or PhD students who aim for careers outside of academia — are highly sensitive to the value of Registered Reports: They virtually never conduct Registered Reports when their value is too low for meeting the survival threshold, but strongly prefer them when their value is sufficient (especially when empirical pace is low and/or the survival threshold is high).

Competition

Competition occurs whenever the demand for academic positions or grant funding exceeds the supply. Figure 7 shows that competition generally leads to an aversion of Registered Reports, as can be seen by the darkening of the plots when moving up from the bottom row of panels. The only exception to this rule is very low competition: When the top 90% are allowed to reproduce (and only the bottom 10% are rejected, $\gamma = .9$), Registered Reports become more popular than they are in the absence of competition. This effect is strongest for the concave fitness function ($\epsilon = 0.2$), where it holds for almost all values of b_{RR} at very low numbers of m and for high values of b_{RR} at high numbers of m. When the fitness function is linear or convex, Registered Reports are chosen more often only when both b_{RR} and m are high. At higher levels of competition ($\gamma > .5$), the differences between the fitness functions disappear. In all three cases, Registered Reports are essentially wiped out for low numbers of research cycles (m), and this effect increases with competition (the higher the competition, the higher m must be for Registered Reports to still be viable). Intense competition also negatively affects Registered Reports at high numbers of m, but here the baseline pattern (a linear increase of RR popularity with b_{RR}) remains intact.

Looking at the first three rows of panels in Figure 7 (1%, 5%, and 10% competition), the extreme effect of competition at low m appears to decrease slightly when competition is highest ($\gamma = .01$), indicated by the dark bar at the bottom of each panel becoming slightly lighter. This paradoxial result is not due to Registered Reports being more lucrative in those conditions. Rather, competition is so extreme that the natural selection in our model starts

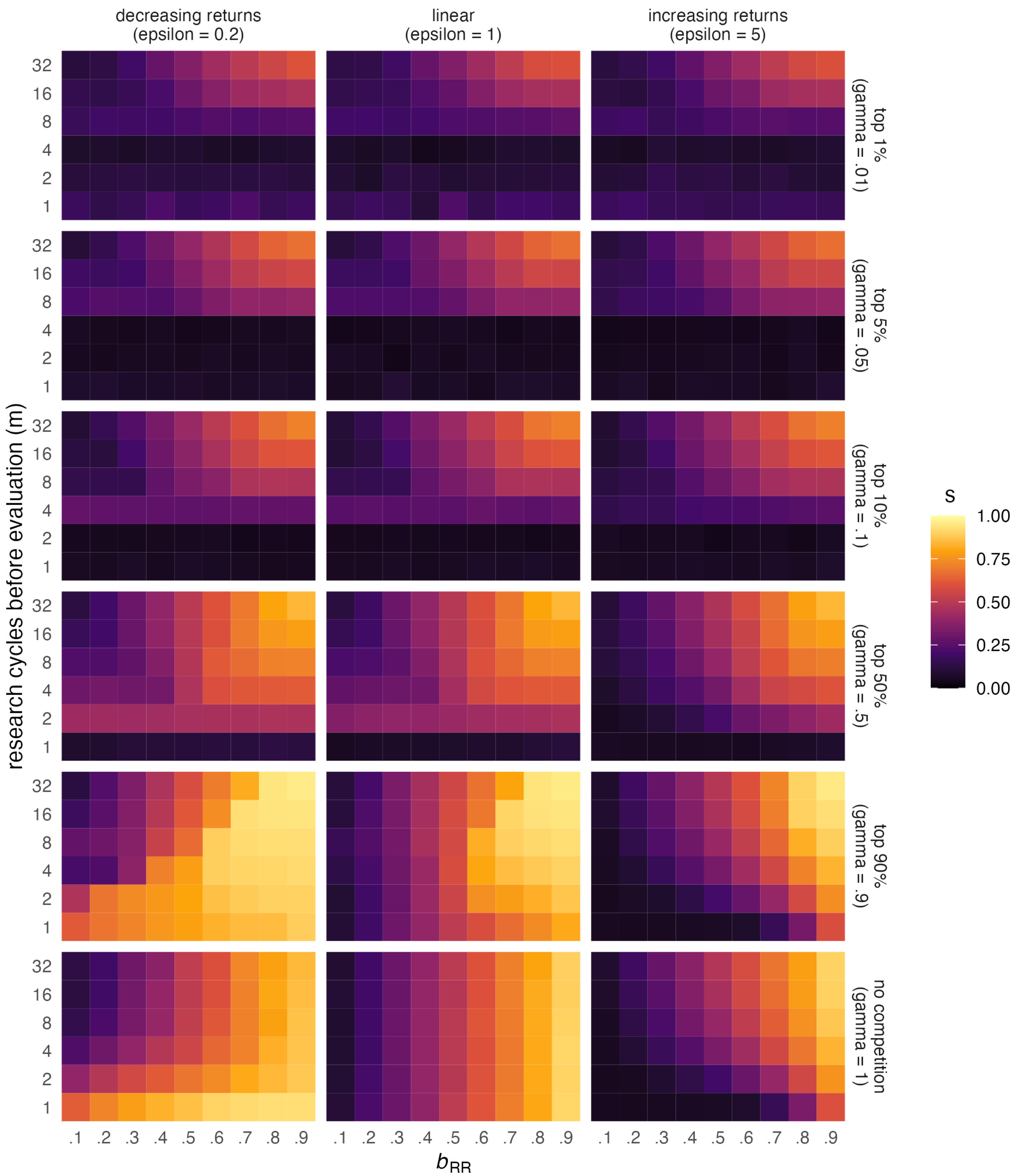

Figure 7. Effect of competition on evolved submission thresholds. Shown are median evolved submission thresholds (s) after 250 generations in 50 runs (tile colour represents the median of 50 run medians) depending on the intensity of competition (γ, y-axis), numbers of research cycles per generation (m), different values of b_{RR} (x-axis), and different fitness functions (characterised by exponent ϵ) with no survival threshold ($\delta = 0$). To reproduce, researchers must accumulate a total payoff in the top γ proportion of the population.

operating more on chance than on individuals' traits. Essentially, only individuals with the maximum possible payoff (publishing only standard reports with positive results) are able to reproduce. Most likely to receive this maximum payoff are individuals who investigate hypotheses with high prior probabilities. In our model, this is not a trait that can be passed on, but determined by random chance. Among individuals who experience this kind of luck, the variance of publication strategy s should be high: A hypothesis with prior $p = .95$ will be submitted as a standard report and likely yield a positive result (and thus the maximum payoff) regardless of whether the researcher's submission threshold is as low as $s = .1$ or has high as $s = .9$. The higher average s at low m under extreme competition thus reflects relaxed selection pressure on s. This is also evident by the shades of the dark bar at the bottom of the panels for $\gamma = .01$ Figure 7 fluctuating randomly for each level of m rather than showing a specific pattern. A clearer illustration of the effect can be found in Figure XXX in the appendix, which shows large increases in the variance of evolved submission thresholds in these conditions. At higher m, selection on s stays intact simply because much fewer individuals will be very lucky 4, 8, 16, or 32 times in a row than once or twice in a row, and the submission strategy thus remains an important factor.

This effect of relaxed selection is not an arbitrary feature of our model, but commonly encountered in natural populations (Snyder, Ellner, & Hooker, 2021). In many species, luck can have an outsized impact on survival and reproduction, rendering the effects of individual traits relatively less important. Luck does not eliminate natural selection[2], but it can significantly slow it. XXX TRY LONGER SIM RUNS & REPORT HERE XXX The phenomenon is related to one form of survivorship bias: Looking at 'survivors' of a highly selective process, one may erroneously infer that specific observable traits or behaviours of such individuals were the cause of their success when they were actually merely coincidental.

[2] This is also apparent in Figure XXX: Although the variance of evolved s increases dramatically with high competition, it never spans the entire range of s.

In the academic world, researchers compete for tenured positions and grants. The level of competition may vary between research areas, countries, institutions, grant programmes, and so on. Our findings suggest that intense competition may be a significant threat for the viability of Registered Reports, regardless of career stage. This effect is particularly extreme in fields with low empirical pace, where submission strategies that involve any amount of Registered Reports are only viable when competition is so high that success requires extraordinary luck. In contrast, very low but non-zero levels of competition increase the popularity of Registered Reports, especially when their value is high, when the fitness function is concave (e.g., in later career stages), and when researchers can complete studies at a fast rate (high empirical pace).

Discussion

In the artificial world of the model presented here, the standard publication route is a coin toss — the probability of obtaining a publishable result is 50%[3], which translates to an expected payoff of 0.5 points per study. If Registered Reports are a safe alternative to this gamble and guarantee publication in every case, one might think that payoff-maximising researchers would prefer them whenever they are worth more than 0.5 points and avoid them whenever they are worth less. This intuition, however, rests on the assumption that the career benefits researchers receive from publications are linear.[4] We argue that this assumption is violated in many, if not all, real-world situations. Here, we investigated the impact of four factors that likely shape real-world situations: convex vs concave fitness functions (additional publications yielding either increasing or decreasing career benefits, reflecting early vs later career stages), empirical pace (reflecting differences in speed and cost

[3] This is the case because we modelled the prior probability of tested hypotheses as being uniformly distributed from 0 to 1 and as being identical to the probability of obtaining a positive (i.e., publishable) result.

[4] Linearity is violated when the fitness function is concave or convex ($\epsilon \neq |1|$), but also in the presence of survival thresholds or competition, because these effectively introduce a step-change in the fitness function (low but non-zero payoffs yield zero career benefits).

of data collection), survival thresholds (reflecting absolute publication targets researchers must meet in a given time), and competition for jobs or grants. Our results show that in isolation or combined, many of these factors would lead researchers with career-maximising strategies to avoid Registered Reports — even when Registered Reports are worth more than the expected payoff from standard reports.

To understand the results, it is useful to take the middle panel of Figure 5 ($\epsilon = 1$) as a baseline. In this panel, publication payoffs translate into linear career benefits (the fitness curve is linear and there is no survival threshold and no competition), and the outcome is highly intuitive: Researchers prefer Registered Reports whenever they are worth more than 0.5 points, their preference is exactly proportional to b_{RR}, and it is not affected by empirical pace. Compared to this baseline, Registered Reports are *less* popular when a) additional publications yield increasing returns (e.g., in early career) and empirical pace is low, b) when researchers face a survival threshold that cannot be met with Registered Reports alone, especially when publications yield decreasing returns (e.g., in advanced career stages) and empirical pace is low, and c) when there is substantial competition. Competition has the most extreme effect and can cause a complete avoidance of Registered Reports when empirical pace is low. Conversely, Registered Reports are *more* popular than at baseline when a) additional publications yield decreasing returns and empirical pace is low, b) Registered Reports are worth just enough to reach a survival threshold and publications yield decreasing returns, especially when empirical pace is low, and c) when there is very low but non-zero competition, especially when publications yield decreasing returns or empirical pace is high.

Looking at the interactions of the different factors, three observations stand out. First, high empirical pace attenuates the effects of all other factors — at the highest pace we considered ($m = 32$), outcomes are identical to baseline in almost all conditions. The only exception to this rule is high competition, but although Registered Reports are relatively less attractive in this condition, the basic pattern is preserved and they remain viable when their value is high. Second, the effect of survival thresholds strongly depends on the shape of the fitness function, suggesting that publication targets may have the strongest impact in advanced career stages. Third, the opposite is true for high competition, which cancels out the effects of different fitness functions and thus appears to have virtually the same impact across career stages.

Implications

- Fields with low pace/labs with low resources are most susceptible to other factors
- Tenure track: value of RRs extremely important
- Grants: strategy to only sift out the worst application and raffle among the rest would favour RR-heavy strategy
- competition: relate to competition for priority & potential interaction with up-front cost of RRs

`To do:`

- Implications of results
  - cautious mapping of model factors to real-world situations
  - potential implications for meta-science
  - potential implications for policy

Limitations

- Narrow focus on one specific (and highly stylised) difference between Registered Reports and standard reports; there are many others. Model ignores a myriad other factors that influences who chooses Registered Reports for which studies when

- Concept of publication bias as filtering positive results of hypothesis tests (and the respective connection to hypothesis priors such that high priors –> better) is cartoonish and not entirely accurate for the simple reason that positive results of trivial (or otherwise boring) hypotheses are usually not highly valued (also, this approach only focuses on hypothesis testing, which is widely used in psychology but by far not the only means of doing science). A more valid solution may be the concept of publication bias as favouring belief-shifting results presented by Gross & Bergstrom (2021). Adapting the model presented here to capture this concept of bias could be an interesting future direction. However, the present version of the model also allows a conservative interpretation in which the prior probability of hypotheses simply reflects authors' predictions of the eventual publication value of different research questions. This interpretation is still concordant with Registered Reports and standard reports differing in risk, because the publication value of standard reports certainly depends more strongly on the study results than the publication value of Registered Reports (even if not in the simplistic sense of positive hypothesis tests having higher value).

- Fitness concept: one caveat is that

- RRs may actually *slow* the empirical pace, introducing an interaction that our model doesn't take into account

- Fitness curves: more senior researchers may also take the needs of their early-career mentees into account

Future directions

Ability-based risk taking. The model presented in this chapter only considers the effects of situational factors on individuals' risk sensitivity. However, risk sensitivity can also be influenced by individual differences, such that individuals with traits or abilities that increase their expected payoff from a risky option (e.g., traits that increase their winning chances or the payoff when winning or that buffer the impact of losses) should be more risk-prone (Barclay, Mishra, & Sparks, 2018). Such factors may be important to consider in the context of research and publication practices. For example, researchers who are better at choosing research questions that are likely to result in high-impact publications (e.g., through talent or experience) may find Registered Reports less attractive. As a more nefarious version of this idea, Registered Reports may be relatively unpopular among researchers who are more inclined to using questionable research practices (or even fraud) to obtain publishable or impactful results.

Registered Reports and post-publication peer review. The post-publication peer review platform *Peer Community In* (PCI) recently launched a new model of Registered Reports (PCI Registered Reports) in which authors are no longer tied to a specific journal. PCI offers authors the regular process of stage-1 and stage-2 review, the end result of a successful submission is 'only' a preprint with a so-called 'recommendation' from PCI. Authors can subsequently publish their manuscript in one of several journals who partnered with PCI and either rely on the PCI review process alone or offer a streamlined review process for PCI-recommended preprints, or they can submit to any other journal as if their manuscript were a standard report. This innovation gives Registered-Reports authors significantly more freedom to capitalise on the results of their study because a submission to PCI Registered Reports does not preclude the chance of a high-impact publication. PCI Registered Reports thus constitute a significant change to the relative incentives and risk structure of Registered Reports compared to standard reports that merits a closer investigation in the future.

Conclusion

Disclosures

Data, materials, and online resources. This manuscript was created using RStudio (1.2.5019, RStudio Team, 2019) and R (Version 4.2.1; R Core Team, 2019) and the R-packages *bookdown* (Version 0.34; Xie, 2016), *ggplot2* (Version 3.5.0; Wickham, 2016), *here* (Version 1.0.1; K. Müller, 2017), *knitr* (Version 1.46; Xie, 2015), *papaja* (Version 0.1.1.9001;

Aust & Barth, 2018), *rmarkdown* (Version 2.26; Xie, Allaire, & Grolemund, 2018), *stringr* (Version 1.5.1; Wickham, 2023), and *tinylabels* (Version 0.2.3; Barth, 2022).

References

Agnoli, F., Wicherts, J. M., Veldkamp, C. L. S., Albiero, P., & Cubelli, R. (2017). Questionable research practices among italian research psychologists. *PLOS ONE*, *12*(3), e0172792. https://doi.org/10.1371/journal.pone.0172792

Allen, C., & Mehler, D. M. A. (2019). Open science challenges, benefits and tips in early career and beyond. *PLOS Biology*, *17*(5), e3000246. https://doi.org/10.1371/journal.pbio.3000246

Atkinson, D. R., Furlong, M. J., & Wampold, B. E. (1982). Statistical significance, reviewer evaluations, and the scientific process: Is there a (statistically) significant relationship? *Journal of Counseling Psychology*, *29*(2), 189–194. https://doi.org/10.1037/0022-0167.29.2.189

Aust, F., & Barth, M. (2018). *papaja: Create APA manuscripts with R Markdown.*

Barclay, P., Mishra, S., & Sparks, A. M. (2018). State-dependent risk-taking. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1881), 20180180. https://doi.org/10.1098/rspb.2018.0180

Barth, M. (2022). *tinylabels: Lightweight variable labels.* Retrieved from https://cran.r-project.org/package=tinylabels

Chalmers, I., & Glasziou, P. (2009). Avoidable waste in the production and reporting of research evidence. *The Lancet*, *374*(9683), 86–89. https://doi.org/10.1016/S0140-6736(09)60329-9

Chambers, C. D. (2013). Registered reports: A new publishing initiative at Cortex. *Cortex*, *49*, 606–610. https://doi.org/10.1016/j.cortex.2012.12.016

Chambers, C. D., & Tzavella, L. (2021). The past, present and future of Registered Reports. *Nature Human Behaviour*, 1–14. https://doi.org/10.1038/s41562-021-01193-7

Csada, R. D., James, P. C., & Espie, R. H. M. (1996). The "File Drawer Problem" of

Non-Significant Results: Does It Apply to Biological Research? *Oikos*, *76*(3), 591–593. https://doi.org/10.2307/3546355

de Vries, Y. A., Roest, A. M., Jonge, P. de, Cuijpers, P., Munafò, M. R., & Bastiaansen, J. A. (2018). The cumulative effect of reporting and citation biases on the apparent efficacy of treatments: The case of depression. *Psychological Medicine*, *48*(15), 2453–2455. https://doi.org/10.1017/S0033291718001873

Dickersin, K., & Min, Y. I. (1993). Publication bias: The problem that won't go away. *Annals of the New York Academy of Sciences*, *703*, 135-146; discussion 146-148. https://doi.org/10.1111/j.1749-6632.1993.tb26343.x

Fanelli, D. (2010). "Positive" results increase down the hierarchy of the sciences. *PLoS ONE*, *5*(4), e10068. https://doi.org/10.1371/journal.pone.0010068

Ferguson, C. J., & Heene, M. (2012). A Vast Graveyard of Undead Theories: Publication Bias and Psychological Science's Aversion to the Null. *Perspectives on Psychological Science*, *7*(6), 555–561. https://doi.org/10.1177/1745691612459059

Fiedler, K., & Schwarz, N. (2016). Questionable Research Practices Revisited. *Social Psychological and Personality Science*, *7*(1), 45–52. https://doi.org/10.1177/1948550615612150

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, *345*(6203), 1502–1505. https://doi.org/10.1126/science.1255484

Franco, Annie, Malhotra, N., & Simonovits, G. (2016). Underreporting in Psychology Experiments: Evidence From a Study Registry. *Social Psychological and Personality Science*, *7*(1), 8–12. https://doi.org/10.1177/1948550615598377

Fraser, H., Parker, T., Nakagawa, S., Barnett, A., & Fidler, F. (2018). Questionable research practices in ecology and evolution. *PLOS ONE*, *13*(7), e0200303. https://doi.org/10.1371/journal.pone.0200303

Greenwald, A. G. (1975). Consequences of Prejudice Against the Null Hypothesis.

Psychological Bulletin, *82*(1), 1–20.

Gross, K., & Bergstrom, C. T. (2021). Why ex post peer review encourages high-risk research while ex ante review discourages it. *Proceedings of the National Academy of Sciences*, *118*(51). https://doi.org/10.1073/pnas.2111615118

Haaland, T. R., Wright, J., & Ratikainen, I. I. (2019). Bet-hedging across generations can affect the evolution of variance-sensitive strategies within generations. *Proceedings of the Royal Society B*. https://doi.org/10.1098/rspb.2019.2070

Hurly, A. T. (2003). The twin threshold model: Risk-intermediate foraging by rufous hummingbirds, Selasphorus rufus. *Animal Behaviour*, *66*(4), 751–761. https://doi.org/10.1006/anbe.2003.2278

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling. *Psychological Science*, *23*(5), 524–532. https://doi.org/10.1177/0956797611430953

Kacelnik, Alex, & Bateson, M. (1996). Risky Theories—The Effects of Variance on Foraging Decisions. *Integrative and Comparative Biology*, *36*(4), 402–434. https://doi.org/10.1093/icb/36.4.402

Kacelnik, A., & Bateson, M. (1997). Risk-sensitivity: Crossroads for theories of decision-making. *Trends in Cognitive Sciences*, *1*(8), 304–309. https://doi.org/10.1016/s1364-6613(97)01093-0

Lakens, D. (2019). The value of preregistration for psychological science: A conceptual analysis. *Japanese Psychological Review*, *62*(3), 221–230. https://doi.org/10.24602/sjpr.62.3_221

Liner, G. H., & Sewell, E. (2009). Research requirements for promotion and tenure at PhD granting departments of economics. *Applied Economics Letters*. https://doi.org/10.1080/13504850701221998

Mahoney, M. J. (1977). Publication Prejudices: An Experimental Study of Confirmatory Bias in the Peer Review System. *Cognitive Therapy and Research*, *1*(2), 161–175.

https://doi.org/10.1007/BF01173636

Mishra, S. (2014). Decision-Making Under Risk: Integrating Perspectives From Biology, Economics, and Psychology. *Personality and Social Psychology Review*, *18*(3), 280–307. https://doi.org/10.1177/1088868314530517

Müller, K. (2017). *Here: A simpler way to find your files.*

Müller, R. (2014). Postdoctoral Life Scientists and Supervision Work in the Contemporary University: A Case Study of Changes in the Cultural Norms of Science. *Minerva*, *52*(3), 329–349. https://doi.org/10.1007/s11024-014-9257-y

Müller, R., & de Rijcke, S. (2017). Thinking with indicators. Exploring the epistemic impacts of academic performance indicators in the life sciences. *Research Evaluation*, *26*(3), 157–168. https://doi.org/10.1093/reseval/rvx023

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, *115*(11), 2600–2606. https://doi.org/10.1073/pnas.1708274114

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., … Vazire, S. (2022). Replicability, Robustness, and Reproducibility in Psychological Science. *Annual Review of Psychology*, *73*(1), annurev-psych-020821-114157. https://doi.org/10.1146/annurev-psych-020821-114157

R Core Team. (2019). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

RStudio Team. (2019). *RStudio: Integrated development environment for r.* Boston, MA: RStudio, Inc.

Scheel, A. M., Schijen, M. R. M. J., & Lakens, D. (2021). An Excess of Positive Results: Comparing the Standard Psychology Literature With Registered Reports. *Advances in Methods and Practices in Psychological Science*, *4*(2), 251524592110074.

https://doi.org/10.1177/25152459211007467

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. *Royal Society Open Science*, *3*, 160384. https://doi.org/10.1098/rsos.160384

Snyder, R. E., Ellner, S. P., & Hooker, G. (2021). Time and Chance: Using Age Partitioning to Understand How Luck Drives Variation in Reproductive Success. *The American Naturalist*, *197*(4), E110–E128. https://doi.org/10.1086/712874

Soderberg, C. K., Errington, T. M., Schiavone, S. R., Bottesini, J., Thorn, F. S., Vazire, S., . . . Nosek, B. A. (2021). Initial evidence of research quality of registered reports compared with the standard publishing model. *Nature Human Behaviour*, *5*(8), 990–997. https://doi.org/10.1038/s41562-021-01142-4

Sterling, Theodore D. (1959). Publication Decisions and their Possible Effects on Inferences Drawn from Tests of Significance—or Vice Versa. *Journal of the American Statistical Association*, *54*(285), 30–34. https://doi.org/10.1080/01621459.1959.10501497

Sterling, Theodor D., Rosenbaum, W. L., & Weinkam, J. J. (1995). Publication Decisions Revisited: The Effect of the Outcome of Statistical Tests on the Decision to Publish and Vice Versa. *The American Statistician*, *49*(1), 108. https://doi.org/10.2307/2684823

van Dalen, H. P., & Henkens, K. (2012). Intended and unintended consequences of a publish-or-perish culture: A worldwide survey. *Journal of the American Society for Information Science and Technology*, *63*(7), 1282–1293. https://doi.org/10.1002/asi.22636

Wacholder, S., Chanock, S., Garcia-Closas, M., El ghormli, L., & Rothman, N. (2004).

Assessing the Probability That a Positive Report is False: An Approach for Molecular Epidemiology Studies. *JNCI Journal of the National Cancer Institute*, *96*(6), 434–442. https://doi.org/10.1093/jnci/djh075

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An Agenda for Purely Confirmatory Research. *Perspectives on Psychological Science*, *7*(6), 632–638. https://doi.org/10.1177/1745691612463078

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York.

Wickham, H. (2023). *Stringr: Simple, consistent wrappers for common string operations*. Retrieved from https://CRAN.R-project.org/package=stringr

Winterhalder, B., Lu, F., & Tucker, B. (1999). Risk-senstive adaptive tactics: Models and evidence from subsistence studies in biology and anthropology. *Journal of Archaeological Research*, *7*(4), 301–348. https://doi.org/10.1007/BF02446047

Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd ed.). Boca Raton, Florida: Chapman and Hall/CRC.

Xie, Y. (2016). *Bookdown: Authoring books and technical documents with R markdown*. Boca Raton, Florida: Chapman and Hall/CRC.

Xie, Y., Allaire, J. J., & Grolemund, G. (2018). *R markdown: The definitive guide*. Boca Raton, Florida: Chapman and Hall/CRC.