

Incentives for a safe publication option from a risk sensitivity perspective

Anne M. Scheel<sup>1</sup>, Leo Tiokhin<sup>1</sup>, & Daniël Lakens<sup>1</sup>

<sup>1</sup> Eindhoven University of Technology

#### Author Note

Correspondence concerning this article should be addressed to Anne M. Scheel, Den Dolech 1, Atlas 9.417, 5600 MB, Eindhoven, The Netherlands. E-mail: a.m.scheel@tue.nl

Abstract

xxx Abstract xxx

*Keywords:* Publication bias, Registered Reports, hypothesis testing

### Incentives for a safe publication option from a risk sensitivity perspective

- publication bias & QRPs
- Registered Reports
  - what they are
  - how they protect against bias
  - evidence
- however, little data is available about the demographics of RR authors and characteristics of the research questions studied in RRs.

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 studied hypotheses) is important for meta-scientific studies that compare published studies in both formats and must take such confounds into account (e.g., Scheel, Schijen, & Lakens, 2021). As a final, perhaps less urgent motivation for such research, ... RISK OF DEVALUATION LIKE JOURNALS OF NEGATIVE RESULTS The goal of this chapter is to shed light on these questions by studying the potential impact of a key feature of Registered Reports: results-independent publication guarantee as an incentive for authors.

### **Registered Reports as a low-risk publication option**

In addition to protecting against publication bias and QRPs, Registered Reports are 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 () or data sharing ()).

Amount and impact of peer-reviewed publications is a central currency for academic researchers [add some evidence]. 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 payoff a researcher receives for conducting a study can vary extremely: from near zero when the resulting manuscript is rejected by all peer-reviewed journals the author submits to — or when the author file-drawers the study because the chances of successful publication appear too low to justify the costs of submission and peer review — to an extremely high, perhaps career-making amount when a manuscript is published in a high-impact journal like Nature or Science.

The basic argument is that partly because of publication bias, publication success in the normal system is highly volatile and dependent on the one factor that is supposed to be outside of researchers' control — the results —, putting especially early-career researchers under excessive stress and tempting them to hype, spin, or even fabricate their data. Because publication is results-independent in Registered Reports and authors receive an in-principle acceptance before investing in data collection or analysis, the format is a relatively safe, stress-free alternative. 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*, where ‘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). When organisms are sensitive not only to the mean outcomes of different options but also to their variance, they are called ‘risk sensitive’.

- Humans are risk-sensitive, therefore researchers should be too Like virtually all other organisms, humans are sensitive to risk but do not universally try to minimise (risk aversion) or maximise (risk proneness) it.

A large body of literature shows that humans

- RST is an explanatory theory for these situations Risk-Sensitivity Theory, a normative theory developed in behavioural ecology to explain foraging behaviour of animals Risk-Sensitivity Theory was originally developed to determine the optimal food-acquisition strategy for animals 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 been successfully applied to humans as well as several non-human animal species (Kacelnik & Bateson, 1996, 1997; Mishra, 2014; [add more refs]).

- What does the argument entail?
  - RRs are ‘safer’ than non-RRs  $\rightarrow$  less variance, less risk
  - (safer = better)
  - However, the argument that RRs are always more attractive for authors only holds if a) RRs are always worth at least as much as a normal publication or b) authors are always risk averse
- Enter Risk Sensitivity Theory
  - background
  - basics
  - comparison to utility theory and prospect theory
- RRs through an RST lens

## Goals

- Understand under which circumstances RRs are unlikely to be viable
- Understand how authors’ strategic risk-sensitive decision-making may affect what is studied in RRs/the difference

This chapter explores 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, which would make their results difficult to compare to the normal literature. The following section outlines central concepts of Risk-Sensitivity Theory, relates them to characteristics of academic careers, and describes a simulation model in which their effects on researchers’ risk-sensitive publication decisions are examined.

## Conceptual application of Risk-Sensitivity Theory to researchers' publication decisions

The following section describes general factors that affect the role of risk for individual's fitness and connects these factors to relevant elements of academic careers.

**NEEDS REPHRASING B/C PREVIOUS PARAGRAPH WAS CHANGED** In this context, Risk-Sensitivity Theory's focus on reproductive fitness as the central outcome may be seen as problematic. 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 their survival and the propagation of their traits in an *academic* sense. 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 sole proximal motivation for researchers' behaviour in practice, just as evolutionary theory does not imply that reproductive success is the sole proximal motivation for every person's everyday decisions. **NEEDS WORK:** However, being a product of evolution, human decision making has been shown to be risk sensitive in many domains. significant competition for tenured positions in many academic disciplines inevitably creates a selection pressure for behaviours that further researchers' career success (Smaldino & McElreath, 2016).

***Non-linear fitness functions.*** The first and perhaps most important 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 (fitness). 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  $b_+$ , 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.

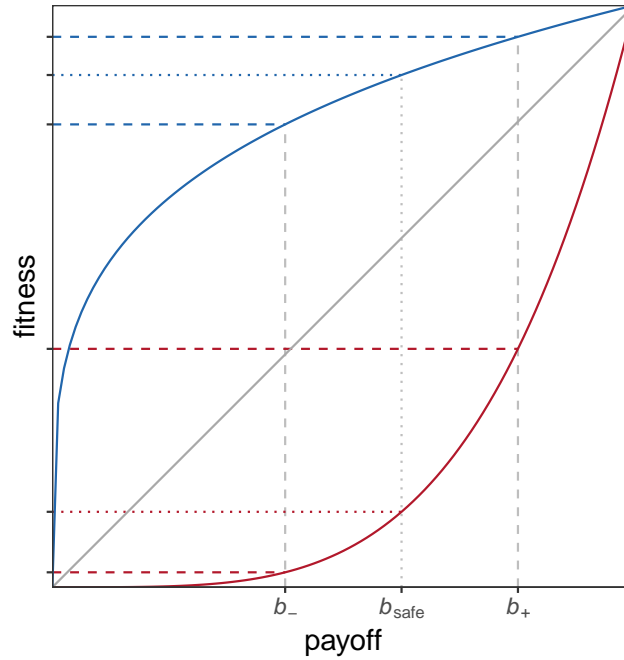


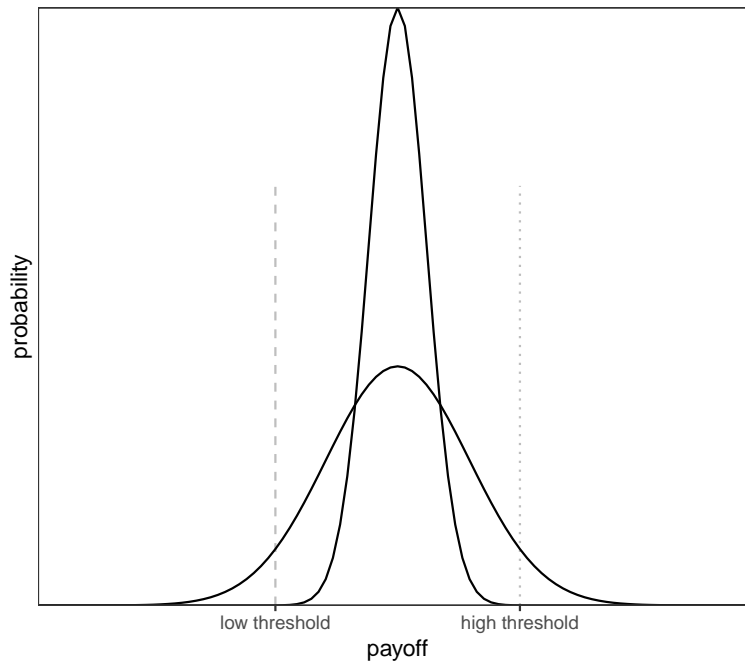
Figure 1. Consequences of non-linear fitness functions. Payoffs  $b_-$ ,  $b_{safe}$ , and  $b_+$  are converted into fitness with a diminishing (blue), linear (grey), or increasing (red) returns function.

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.

***Survival thresholds and competition.*** A second important factor for risk-sensitive behaviour are thresholds for survival and reproduction. Survival and reproduction 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 averse when they are not. 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 the 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.



*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.

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 the 10 highest-ranked research proposals receive funding or the best candidate out of a pool of applicants is offered a job. In cases like these, one individual's success diminishes the chances of other individuals—in other words, 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 factor with relevance to risk-sensitive behaviour considered in this chapter 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 each decision, choosing a risky option might yield an outcome that translates to zero fitness, meaning death or an ultimate failure to reproduce. When one risky decision might cost an individual's life or offspring, average fitness is best described by the geometric mean instead of the arithmetic mean. 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. 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 (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 system, decision events before fitness is evaluated ('per generation') could be seen as the time and resources a researcher has available for producing publications before a relevant decision like those mentioned in the previous section (award of a PhD or grant, job application, tenure decision) is made. Regarding time, the number of publications a PhD student or early-career researcher may be limited by **NEEDS WORK**

Each of the factors described above — non-linear fitness functions, survival thresholds, competition, and the number of decision events prior to 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, I examine the individual and interactive effects of these factors in a simulation model.

## Model description

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  $n = 200$  researchers. Each researcher has a fixed publication strategy  $s$ , called the submission threshold. In each round of the research phase, researchers randomly pick a hypothesis to test in a study. Hypotheses are true with prior probability  $p$ , which is uniformly distributed between 0 and 1. 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 test the hypothesis in a Registered Report. When  $p \geq s$ , the researcher chooses to gamble and test the hypothesis in a normal 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 proceeds to testing 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 conduct a Registered Report, their study gets published regardless of the result and the researcher receives a payoff  $b_{RR}$ . However, if the researcher chose to publish their results as a standard report, they face rampant publication bias: Only positive results are publishable as standard reports and yield a payoff  $b_{SR+}$ , whereas negative results are rejected or file-drawered and only yield a minimal payoff  $b_{SR-}$ . 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, 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) that diminishes the net reward.
2. These same reasons mean that 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 and reject negative results in standard report submissions, or simply because Registered Reports do not need to be resubmitted or require more extensive revisions in case of a negative result.
3. On average, standard reports with positive results are more valuable than Registered Reports ( $b_{RR} < b_{SR+}$ ), for example because most extremely high-impact journals (e.g.,

*Science* and *Nature*) do not (yet) offer Registered Reports, because not registering one's study *a priori* makes it easier to spin the results into an attention-grabbing story, 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 fact that most researchers do not seem to share this view. Once this changes, the present investigation may happily become redundant.

This 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 are evaluated by translating their accumulated publication payoffs  $b_1 + b_2 + \dots + b_m$  into fitness using exponent  $\epsilon$ , which reflects the shape of the fitness function ( $\epsilon = 1$  yields a linear function,  $0 < \epsilon < 1$  yields a concave function with diminishing returns,  $\epsilon > 1$  yields a convex function with increasing returns, see Figure 1):

$$fitness = \left( \sum_{i=1}^m b_i \right)^\epsilon \quad (1)$$

However, two situations may cause a researcher's fitness to fall to zero even when their accumulated payoffs are nonzero. First, the sum of their payoffs may fall below an absolute survival threshold  $\delta$ , 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  $\gamma$ , which reflects the intensity of

competition for scarce resources such as research grants or positions.  $\gamma$  is the proportion of the most productive researchers that 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  $(1 - \gamma)$  portion of researchers with the lowest sum of payoffs receives zero fitness and cannot reproduce.<sup>1</sup> For example,  $\gamma = 0.1$  means that only those researchers with accumulated payoffs in the top 10% of the population can reproduce, and the remaining 90% receives zero fitness.

**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  $w \sim R(\mu = 0, \sigma = 0.01)$ . This evolutionary dynamic of researchers passing on their traits to other researchers depending on their career success can be seen as reflecting mentorship and explicit teaching, such as when established professors advise their students to use the same strategies, or simply a generic social learning process in which successful researchers are more likely to be imitated by others.

**Outcome variables.** We study how the evolution of researchers’ submission thresholds  $s$  is affected by the payoff parameters  $b_{RR}$ ,  $b_{SR-}$ , and  $b_{SR+}$ , by the shape of the fitness function determined by exponent  $\epsilon$ , by the absolute survival threshold  $\delta$ , by competition  $\gamma$ , and by the number of research cycles per generation  $m$  (see Table 1 for an overview of the model parameters and their values considered in the simulation). It is important to keep in mind that a researcher’s submission threshold  $s$  is a *strategy*, not an

---

<sup>1</sup> The computer code of the simulation applies  $\gamma$  after fitness has been calculated according to the accumulated payoffs. This change has purely technical reasons and yields the same result as calculating fitness after  $\gamma$  has been applied to accumulated payoffs, since all fitness functions are monotonic increasing.

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 is willing to gamble and chooses the standard publication route for almost all hypotheses they encounter, using the Registered Reports 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 Reports format, reserving the standard publication route only for hypotheses that are virtually guaranteed to be true (and yield positive results).

The evolved values of  $s$  over many generations indicate the optimal strategy for a given set of parameter values.

## Results

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, the parameter 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, and thus the results are only meaningful in relation to each other.

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. All variations of the simulation model reported here were run with a population size of  $n = 500$  researchers over 250 generations and payoffs for negative and positive results in standard reports fixed at  $b_{SR-} = 0$  and  $b_{SR+} = 1$ , respectively. Figure 3 shows the effect of varying the payoffs for

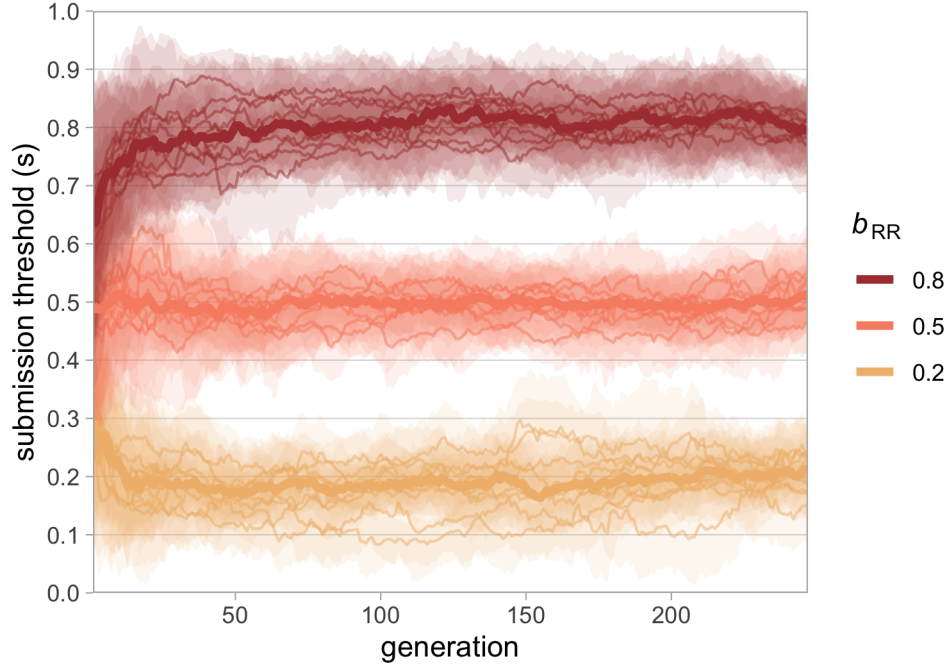


Figure 3. Evolution of submission threshold  $s$  with 3 different payoffs for Registered Reports ( $b_{RR}$ ). Simulations are based on a population of  $n = 500$  researchers over 250 generations, with payoffs for standard reports fixed at 0 for negative results ( $b_{SR-} = 0$ ) and 1 for positive 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.

Registered Reports when the fitness function is linear ( $\epsilon = 1$ ), with no survival threshold ( $\delta = 0$ ) or competition ( $\gamma = 1$ ), and one research cycle per generation ( $m = 1$ ). The overall pattern of results is unsurprising—the higher the payoff for Registered Reports, the more popular they become. When  $b_{RR}$  is low, Registered Reports are unpopular and only used for the least probable hypotheses; when  $b_{RR}$  is high, Registered Reports are very popular and only hypotheses with extremely high priors are studied in standard reports.

In this very simple case illustrated here, evolved submission thresholds 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



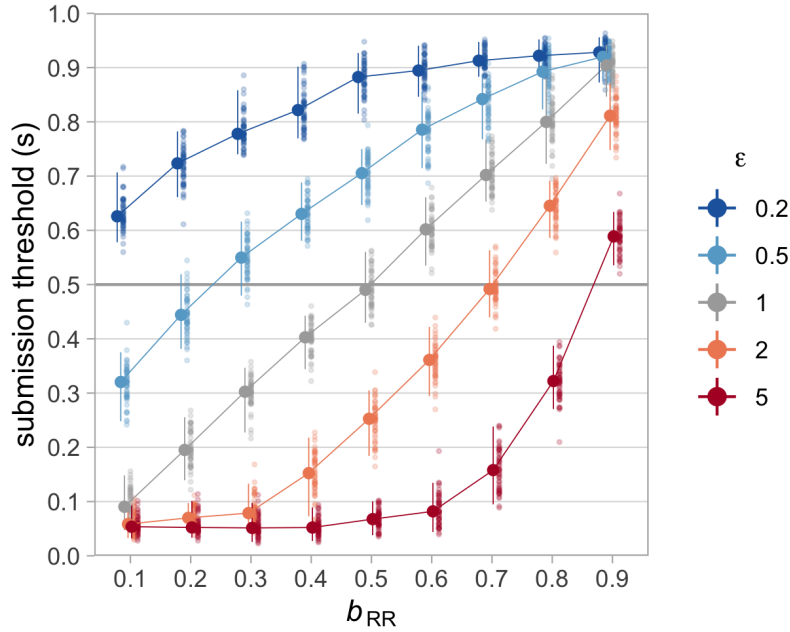
[0–1] of hypothesis priors and the payoff structure  $b_{SR-} = 0$  and  $b_{SR+} = 1$ , which means that the expected payoff of a standard report is always equal to the prior of the tested hypothesis:

$$E[b_{SR}] = pb_{SR+} + (1 - p)b_{SR-} = p * 1 + (1 - p) * 0 = p \quad (2)$$

For example, testing a hypothesis with  $p = 0.2$  in a standard report would yield the expected payoff  $0.2 * 1 + 0.8 * 0 = 0.2$ . The optimal strategy is to submit a Registered Report whenever the expected payoff of a standard report is lower than the payoff for a Registered Report,  $E[b_{SR}] < b_{RR}$ , and thus whenever  $p < b_{RR}$ . The strategy is optimal because it 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  $E[b_{SR}]$  for all hypotheses with  $p < 0.5$  but lower than  $E[b_{SR}]$  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.

### Non-linear fitness functions

With this basic understanding of the payoff structure in hand, we can take a look at what happens when payoffs have non-linear consequences for researchers' fitness. Figure 4 contrasts the effects of two diminishing fitness functions ( $\epsilon = 0.2$  and  $\epsilon = 0.5$ , shown in blue shades) and two increasing fitness functions ( $\epsilon = 2$  and  $\epsilon = 5$ , shown in red shades) with a linear function ( $\epsilon = 1$ , grey line) for different payoffs for Registered Reports. The grey line for  $\epsilon = 1$  represents the already familiar situation from Figure 3 above: When the fitness



*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  $\epsilon$ ), 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.

function is linear, the optimal strategy is  $s_{optimal} = b_{RR}$ , making Registered Reports relatively popular when they are worth more than 0.5 and relatively unpopular when they are worth less than 0.5. Non-linear fitness functions change this picture. When additional payoffs yield diminishing returns ( $\epsilon < 1$ ), Registered Reports become more attractive even when they are worth less than half of published (positive) standard reports. This is because concave functions ‘shrink’ the difference between moderate and high payoffs relative to the difference between low and moderate payoffs (as illustrated in Figure 1). 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 — such that senior researchers’ returns on career success per publication (or per increment of publication impact) are diminishing and those of early-career researchers are increasing — this pattern suggests that Registered Reports should be more attractive for senior researchers and a tough sell for early-career researchers. This observation is interesting because it seems at odds with preliminary evidence suggesting that Registered Reports may be more likely to have early-career researchers as first authors than standard reports (Chambers & Tzavella, 2021). One explanation for such data (if robust) could be that the effect of concave versus convex fitness functions is swamped out by factors unrelated to risk sensitivity (e.g., 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.

### Number of decision events before evaluation

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 only a single 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, Wright, & Ratikainen, 2019). However, Figure 5 shows that the effect of increasing numbers of research cycles per generation ( $m$ ) interacts with the shape of the fitness function: Moving up on the y-axis of each panel, we see that submission thresholds are decreasing (indicating risk proneness) only in the top panel ( $\epsilon = 0.2$ ) but stay constant in the middle panel ( $\epsilon = 1$ ) and even *increase* in the bottom panel ( $\epsilon = 5$ ). Why does  $m$  appear to have opposite effects for diminishing and increasing fitness functions? To understand this pattern, it helps to first consider only the

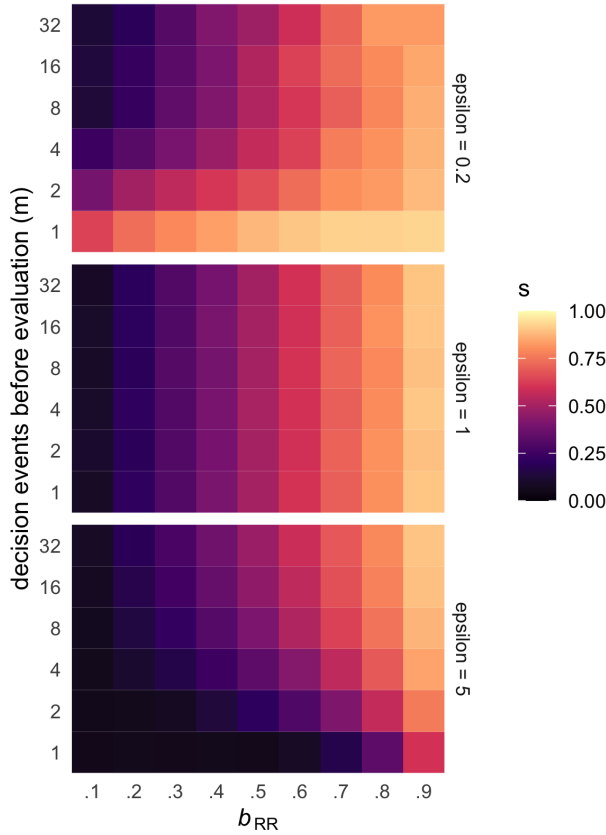


Figure 5. Effect of research cycles per generation on evolved submission thresholds. Shown are median evolved submission thresholds ( $s$ ) after 250 generations in 50 runs (tile colour represents the median of 50 run medians) for varying numbers of research cycles per generation ( $m$ , y-axis), different values of  $b_{RR}$  (x-axis), and different fitness functions (characterised by exponent  $\epsilon$ ).

bottom row of each panel, where  $m = 1$ . These three rows contain the same data 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 dynamic 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 use an illustrating example, consider two researchers with extreme submission strategies: Emma conducts only Registered Reports ( $s_{Emma} = 1$ ) and Gordon conducts only

standard reports ( $s_{Gordon} = 0$ ). The payoff for Registered Reports is fixed at  $b_{RR} = 0.5$ .

After one research cycle, Emma receives a payoff of 0.5 and Gordon receives either 0 or 1.

When fitness is calculated after this one round with  $\epsilon = 0.2$ , Emma's fitness is

$b_{Emma}^\epsilon = \frac{1}{2}^{\frac{1}{5}} = 0.87$ , and Gordon's fitness is either  $b_{Gordon-}^\epsilon = 0^{\frac{1}{5}} = 0$  or  $b_{Gordon+}^\epsilon = 1^{\frac{1}{5}} = 1$ .

In a population of Emmas and Gordons, lucky Gordons who got a positive result have a narrow fitness advantage over all Emmas (1 versus 0.87), while unlucky Gordons lose to all Emmas by a wide margin (0 versus 0.87). Since there are twice as many Emmas as lucky Gordons, the Emma strategy is quite successful.

Now consider the same scenario with 4 research cycles before fitness is calculated.

Emmas receive the same payoff in every round and accumulate  $\frac{1}{2} * 4 = 2$ . Lucky Gordons (who win every time) accumulate a total payoff of  $1 * 4 = 4$ , while unlucky Gordons (who lose every time) again receive 0 total payoff. Now, however, the probabilistic outcomes over 4 rounds lead to more versions of Gordon, including average Gordons (who win half of the time and lose half of the time) who accumulate the same total payoff as Emmas ( $\frac{1}{2} * 0 + \frac{1}{2} * 1$ ) \* 4 = 2. This translates into fitness values of 0 for unlucky Gordons,  $2^{\frac{1}{5}} = 1.15$  for Emmas and average Gordons, and  $4^{\frac{1}{5}} = 1.32$  for lucky Gordons. The Emma strategy still yields an enormous advantage compared to unlucky Gordons and only a small disadvantage compared to lucky Gordons. But this time, there are fewer Gordons who are less successful than Emmas because Emmas now share their place with average Gordons, meaning that the relative fitness advantage of the Emma strategy decreases. As the number of research cycles per generation grows, the law of large numbers dictates that more Gordons achieve average total payoffs and fewer Gordons achieve extreme total payoffs (winning 32 times in a row is much less probable than winning 4 times in a row), which reduces the width of the Gordon distribution until it approximates the Emma distribution.

When the fitness function is increasing ( $\epsilon = 5$ ), the overall effect of increasing values of  $m$  is identical, with the only difference that Emmas are initially disadvantaged (because their

fitness distance to the lucky half of Gordons is much greater than to the unlucky Gordons). With larger  $m$ , more and more Gordons receive average total payoffs and share Emma's disadvantaged position (decreasing Emma's relative disadvantage), until the Gordon distribution is again virtually equal to the Emma distribution. These results show that rather than causing absolute risk aversion, increasing values of  $m$  simply swamp out the effect of  $\epsilon$  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.

## Survival thresholds

## Competition

A reference to Figure 6.

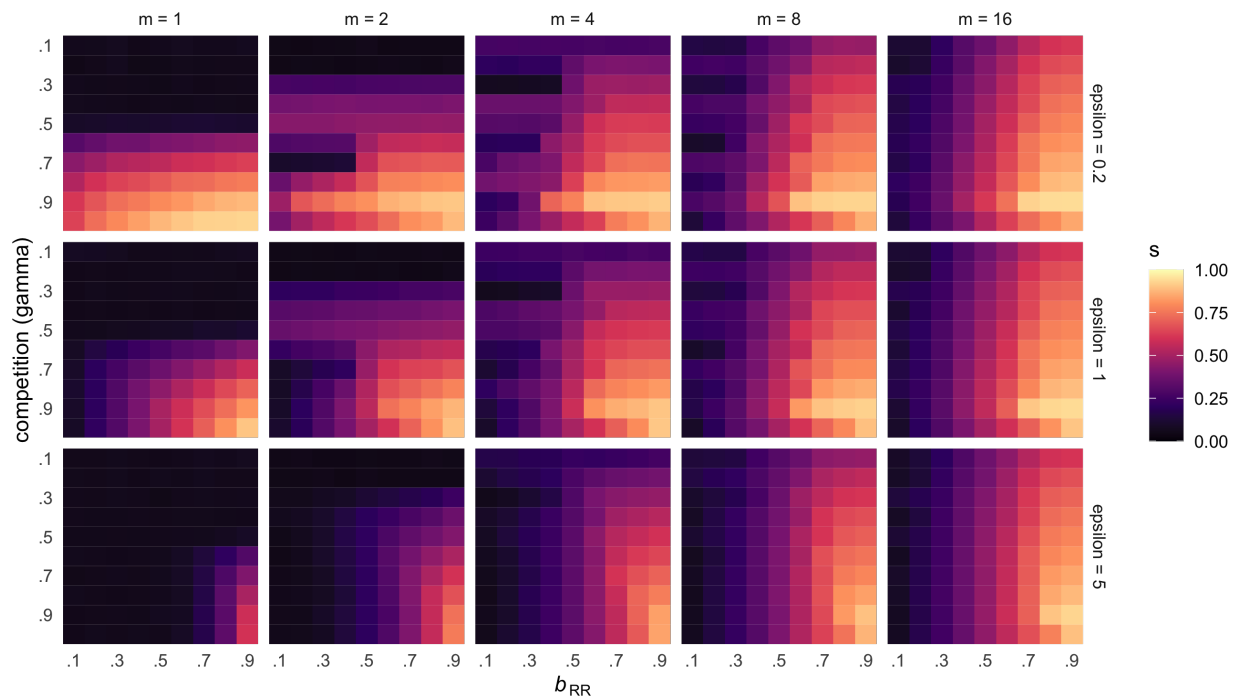


Figure 6. Figure note).

## Measures and coding procedure

Print settings from the script ‘utilitycurves\_plot.R’:

payoff\_RR is 0.60 and e\_low is 0.25.

## Results

## Discussion

## Limitations

## Future directions

- state-dependent risk taking (Barclay et al.): some researchers are more willing or able to tweak their results than others, making SRs more attractive(?)

## Conclusion

## Disclosures

**Data, materials, and online resources.** Data and code necessary to reproduce all analyses reported here, as well as the Appendix, the preregistration, and additional supplementary files, are available at <https://osf.io/dbhgr>. The manuscript, including figures and statistical analyses, the Appendix, and the codebook available in the supplement were created using RStudio (1.2.5019, RStudio Team, 2019) and R (Version 3.6.0; R Core Team, 2019) and the R-packages *bookdown* (Version 0.24; Xie, 2016), *ggplot2* (Version 3.3.5; Wickham, 2016), *here* (Version 1.0.1; Müller, 2017), *knitr* (Version 1.37; Xie, 2015), *papaja* (Version 0.1.0.9997; Aust & Barth, 2018), *rmarkdown* (Version 2.12; Xie, Allaire, & Golemund, 2018), *stringr* (Version 1.4.0; Wickham, 2019), and *tinylabels* (Version 0.2.3; Barth, 2022).

**Reporting.** We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study.

**Author Contributions.** Conceptualisation: A.S. & D.L.; data curation, formal analysis, and software: A.S. & M.R.M.J.S.; investigation, methodology, and validation: A.S., M.R.M.J.S., & D.L.; supervision: A.S. & D.L.; visualisation and writing—original draft: A.S.; writing—review and editing: A.S., M.R.M.J.S., & D.L.

**Conflicts of Interest.** The authors declare that they have no conflicts of interest with respect to the authorship or the publication of this article.

**Acknowledgements.** This work was funded by VIDI grant 452-17-013. We thank Chris Chambers, Emma Henderson, Leo Tiokhin, Stuart Ritchie, and Simine Vazire for valuable comments that helped improve this manuscript.

## References

- Aust, F., & Barth, M. (2018). *papaja: Create APA manuscripts with R Markdown*.
- Barth, M. (2022). *tinylabels: Lightweight variable labels*. Retrieved from <https://cran.r-project.org/package=tinylabels>
- 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>
- 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>
- Müller, K. (2017). *Here: A simpler way to find your files*.
- R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- 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>
- 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>
- Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York.
- Wickham, H. (2019). *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-sensitive 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* (Second). 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., & Grolemund, G. (2018). *R markdown: The definitive guide*. Boca Raton, Florida: Chapman and Hall/CRC.