1   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 $\rightarrow$ 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)
- $\rightarrow$ 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 $\rightarrow$ implications for policy and meta-science

Registered Reports are an article format designed to make publication decisions independent from the reported research results (Chambers, 2013). In this format, the review process is split into 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

28 commits to publishing the eventual report, regardless of the direction of the results. Only

29 after in-principle acceptance has been issued do authors move on to data collection and

30 analysis and eventually complete the manuscript. At Stage 2, the final report (now including

31 the results) is subjected to a second round of peer review, but this time only to ensure that

32 the study was carried out as planned, that the data pass any pre-specified quality checks,

33 and that authors' conclusions are justified by the evidence.

34   Through this process, Registered Reports address publication bias as well as so-called

35 'questionable research practices' (QRPs). These two problems are considered important

36 contributors to psychology's replication crisis (Ferguson & Heene, 2012; Wagenmakers,

37 Wetzels, Borsboom, van der Maas, & Kievit, 2012) and to research waste in the biomedical

38 sciences (Chalmers & Glasziou, 2009) because they skew the available evidence for scientific

39 claims, causing overconfidence and higher rates of false-positive inferences. Publication bias

40 can result from editors and reviewers disproportionately rejecting submissions with negative

41 results ('reviewer bias,' Atkinson, Furlong, & Wampold, 1982; Greenwald, 1975; Mahoney,

42 1977) or from researchers failing to submit negative results for publication ('file-drawering,'

43 Franco, Malhotra, & Simonovits, 2014; Rosenthal, 1979). In Registered Reports, the

44 in-principle acceptance issued at Stage 1 reduces both of these issues: Editors and reviewers

45 cannot reject the Stage-2 report based on the direction of the results, which also reduces the

46 incentives for authors to file-drawer the study in case of negative results. QRPs are practices

47 that exploit undisclosed flexibility in data collection and analysis, for example when

48 analysing different justifiable combinations of variables, subsamples, and decision criteria,

49 and only reporting the ones with favourable results, or by presenting *post hoc* inferences as

50 having been predicted *a priori* (Agnoli, Wicherts, Veldkamp, Albiero, & Cubelli, 2017;

51 Fiedler & Schwarz, 2016; Fraser, Parker, Nakagawa, Barnett, & Fidler, 2018; John,

52 Loewenstein, & Prelec, 2012; Simmons, Nelson, & Simonsohn, 2011). Registered Reports

53 minimise the risk of QRPs via the two-stage review process, in which the Stage-1 protocol

54 acts as a preregistration and reviewers' task during Stage-2 review is to flag any undisclosed

55 deviations from it.

56 Registered Reports were first launched in 2013 at the journal *Cortex* (Chambers, 2013)

57 and are now offered by over 300 journals, predominantly in the behavioural sciences and life

58 sciences (see cos.io/rr). Nearly 600 Registered Reports had been published by 2021, with

59 uptake growing exponentially (Chambers & Tzavella, 2021). In Chapter 2, we presented

60 initial evidence that published Registered Reports have a substantially lower rate of positive

61 results than regular articles in psychology (44% vs 96%, Scheel, Schijen, & Lakens, 2021). A

62 study comparing Registered Reports with matched controls additionally found that

63 Registered Reports have higher median sample sizes and, in blind reviews, are judged to be

64 more rigorous in methodology and analysis and of higher overall quality (Soderberg et al.,

65 2021), which suggests that the increased amount of negative results is not an artifact of lower

66 power.[1]

67 These findings seem to support the effectiveness of Registered Reports in reducing

68 publication bias and QRPs. However, one remaining alternative explanation for the lower

69 rate of positive results is that the hypotheses tested in the Registered Reports literature are

70 less often true (have a lower prior probability) than those in the standard literature. This

71 idea is not implausible: If researchers expect that negative results are difficult to publish in

72 standard reports but pose no problem in Registered Reports, they might selectively choose

73 the Registered Report route when studying hypotheses that they think will yield negative

74 results. If researchers additionally perceive the standard route as less costly (e.g., more

75 habitual, more flexible, faster, requiring lower sample sizes, etc.), standard reports would

76 plausibly remain the preferred option for hypotheses that researchers are more certain are

77 true and will yield publishable results.

---

[1] We cannot rule out that the effect sizes studied in Registered Reports are smaller than in standard reports, in which case the larger samples may not translate into higher statistical power. However, it seems highly unlikely that such a difference would be large enough to explain the entire difference in the positive result rate (52 percentage points) reported in Chapter 2 (see Fig. 3 in Chapter 2).

Could such an effect alone explain why Registered Reports have fewer positive results? In Chapter 2, we showed that this is highly unlikely: Assuming that Registered Reports and standard reports have the same statistical power, the hypotheses tested in Registered Reports would need to be roughly half as likely to be true (see Fig. 3 in Scheel et al., 2021). The same analysis provides indirect evidence for bias in the standard literature, as the extremely high rate of positive results in standard reports (96%) is simply incompatible with plausible estimates of prior probability and power (see also Ingre & Nilsonne, 2018). Combined with more direct evidence of publication bias and QRPs in the standard literature (Agnoli et al., 2017; Franco et al., 2014; Franco, Malhotra, & Simonovits, 2016; John et al., 2012) as well as the face-valid mechanisms of Registered Reports to reduce the influence of these factors, our previous findings (Scheel et al., 2021) thus support the effectiveness of Registered Reports. However, we cannot rule out that the prior probability of tested hypotheses plays an additional role, which means that the 52 percentage points difference in positive results cannot be interpreted as a direct measure of the amount of bias in the standard literature.

Author incentives for Registered Reports

The above highlights the importance of understanding of when, where, and by whom Registered Reports are most likely to be used. First, knowing which factors influence researchers' choice between Registered Reports and the standard publication route is crucial for interpreting meta-scientific studies that compare the two formats, especially if these factors also affect study results (such as the prior probability of tested hypotheses). Second, such knowledge can help identify research areas in which Registered Reports are unlikely to gain traction by themselves and anticipate the need for further intervention (e.g., via policy) when there is a demand for unbiased results. In this chapter, we present a simulation model to examine in which situations Registered Reports can be expected to be particularly popular or unpopular. We focus on a key feature of Registered Reports: the results-independent publication guarantee as an incentive for authors.

104   Registered Reports are designed to serve the research community and other consumers

105 of the scientific literature by protecting against publication bias and QRPs. A key selling

106 point, however, is that they are thought to 'serve the interests of individual scientists' (p. 12,

107 Chambers & Tzavella, 2021) at the same time. The underlying argument is that because

108 scientists a) need to amass journal publications (which still are a central currency for hiring

109 and promotion decisions) and b) face shortfalls in publication output when their studies

110 yield negative results (which are more difficult to publish in the standard literature due to

111 publication bias), a publication guarantee before data collection should be highly valuable.

112 As Registered-Reports advocate Chris Chambers once put it (September 2021):

113   And the second main benefit, the one that really is the main big one, the big

114   draw, is that as a researcher you can get your paper accepted before you even

115   start your research and regardless of how the results turn out in the end. So no

116   more playing the *p*-value lottery, gambling on certain results going a certain way,

117   otherwise you won't have your PhD or you won't get your next fellowship or your

118   next grant — takes all of that pointless, and I think quite foolish, gambling out of

119   the equation $(\dots)^2$

120   However, although it is objectively true that Registered Reports provide more certainty

121 about eventual publication success early in a project, this certainty may not always be

122 preferred over the 'gamble'

123   being a 'safe' alternative to the 'gamble' of the standard publication route

124   But if the standard publications are indeed a gamble and Registered Reports a safe

125 alternative, does it follow that Registered Reports

126   Peer-reviewed publications are a central currency for the careers of academic

─────

2 https://youtu.be/FiVI3cwVMZI?list=PLChfyH8TVDGmYENpXUDPaeeq2SLh8q9dt&t=1047, from
minute 17:27

127 researchers, both in terms of publication quantity and publication impact (R. Müller, 2014;

128 van Dalen & Henkens, 2012). In the standard publication model, researchers face uncertainty

129 about whether and where they will be able to publish the results of their study. Translated

130 into currency terms, the career benefit a researcher receives for conducting a study can vary

131 extremely — from near zero when the resulting manuscript is rejected by all consulted

132 journals (or when the author file-drawers the study because the chances of success do not

133 justify the cost of repeated submissions and revisions) to an extremely high, perhaps

134 career-making amount when a manuscript is published in a very high-impact journal like

135 *Nature* or *Science*. In other words, success in the standard system is highly variable and

136 highly volatile since it hinges on the one factor that is supposed to be outside of researchers'

137 control — the study results. This unfortunate combination can be excessively stressful for

138 researchers (especially junior scientists without secure positions) and tempt them to hype,

139 spin, or even fabricate their results.

140 Compared to this, Registered Reports are a relatively safe, stress-free alternative

141 because authors receive a results-independent publication guarantee before investing in data

142 collection or analysis. As Registered-Reports inventor Chris Chambers put it in a recent talk

143 (September 2021):

144 And the second main benefit, the one that really is the main big one, the big

145 draw, is that as a researcher you can get your paper accepted before you even

146 start your research and regardless of how the results turn out in the end. So no

147 more playing the *p*-value lottery, gambling on certain results going a certain way,

148 otherwise you won't have your PhD or you won't get your next fellowship or your

149 next grant — takes all of that pointless, and I think quite foolish, gambling out of

150 the equation completely. (from minute 17:27)

151 But would researchers ever choose the gamble over the safe publication? Unless the net

152 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 (Kacelnik & Bateson, 1996; 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, we use a simulation model 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 past behaviour increased their prospects. Such competition can create bottlenecks between early-career and tenured positions in many academic disciplines, which inevitably induce a selection pressure for career-promoting behaviours (Smaldino & McElreath, 2016).

In applying Risk-Sensitivity Theory to researchers' publishing behaviour, we will

203 therefore use a general notion of career success as the central outcome variable in place of

204 reproductive fitness. This decision does not imply that career success is the only or the

205 proximal motivation for researchers' behaviour in practice, just as evolutionary theory does

206 not imply that reproductive success is the only or the proximal motivation for human

207 behaviour in everyday life. However, we do assume that selection for career-promoting

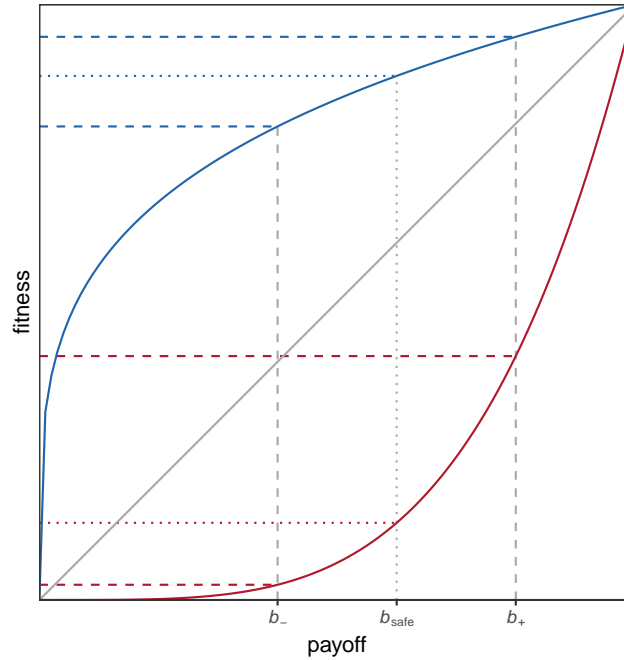208 behaviours has a noticeable impact on research practice.

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.

209 ***Non-linear fitness functions.***   The first and perhaps most ubiquitous factor

210 leading individuals to be risk sensitive are non-linear relationships between the outcomes of

211 an individual's behaviour (e.g., harvested food items, publications) and its reproductive

212 success (Kacelnik & Bateson, 1997). Consider two options, O_{safe} and O_{risky}. O_{safe} always

213 gives the same payoff b_{safe}, whereas O_{risky} gives either a low payoff b_- or a high payoff b_+,

214 each with probability $\frac{1}{2}$. When $b_{safe} = \frac{(b_- + b_+)}{2}$, O_{safe} and O_{risky} have the same expected

215 payoff. However, we would only expect an individual to be indifferent between the two

216 options if the consequences of their payoffs for the individual's fitness are linear. When the

²¹⁷ function relating payoffs to fitness is instead convex or concave (yielding increasing or

²¹⁸ diminishing returns, respectively), the expected fitness of O_{safe} and O_{risky} will differ and

²¹⁹ shift the individual's preference towards risk proneness or risk aversion. An illustration of

²²⁰ this example is shown in Figure 1: While the payoffs b_-, b_{safe}, and b_+ are equidistant on the

²²¹ x-axis, b_{safe} is associated with greater fitness than the average of b_- and b_+ when the

²²² function is concave, and with lower fitness when the function is convex. In other words,

²²³ O_{safe} has greater expected fitness than O_{risky} when returns are diminishing, and O_{risky} has

²²⁴ greater expected fitness than O_{safe} when returns are increasing.

²²⁵   Non-linear relationships are arguably the norm in the natural world and linear

²²⁶ relationships the exception. This plausibly holds for academia as well, where the effect of

²²⁷ publication success on researchers' career success might change over time: For early-career

²²⁸ researchers, small increases in the number or impact of publications may have an accelerated

²²⁹ effect on career success, whereas established professors may care little about any one

²³⁰ additional publication to their record.

²³¹   ***Survival thresholds and competition.*** A second important factor for

²³² risk-sensitive behaviour are thresholds for survival and reproduction (Hurly, 2003;

²³³ Winterhalder et al., 1999). Survival thresholds are cutoff points below which an individual's

²³⁴ fitness drops to zero, for example due to starvation. Risk-Sensitivity Theory predicts that an

²³⁵ individual will be risk averse when the resources provided by a low-variance option are

²³⁶ sufficient to meet the threshold and risk-prone when they are not (Mishra, 2014). For

²³⁷ example, a hummingbird that needs to acquire a certain amount of calories to survive the

²³⁸ night will prefer a low-risk food source if this option's expected payoff is above the threshold,

²³⁹ but avoid the low-risk source if only a higher-risk source provides a chance of survival. One

²⁴⁰ such situation is depicted in Figure 2.

²⁴¹   Although comparable cutoff points in academic careers may have somewhat less severe

²⁴² consequences, they certainly exist: Amount and impact of a researcher's publications are
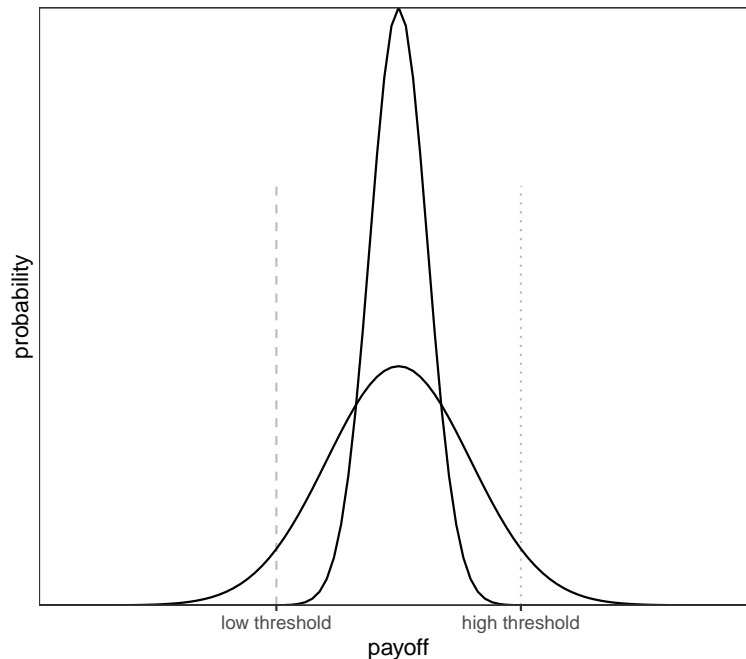
Figure 2. Survival thresholds. When fitness drops to zero below the low threshold (dashed line), individuals should be risk-averse because the outcomes of the low-risk option (narrow distribution) are guaranteed to lie above the threshold and the outcomes of the high-risk option (wide distribution) have 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.

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 number 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

²⁵⁵ researchers' publication behaviour.

²⁵⁶ ***Number of decision events before evaluation.*** A final risk-relevant factor

²⁵⁷ considered here is the number of decision events taking place before an individual's fitness is

²⁵⁸ evaluated. When a risky option is chosen repeatedly, the average of the accumulating payoffs

²⁵⁹ gets closer and closer to the long-run expected payoff. This means that the danger of loosing

²⁶⁰ out completely by only acquiring the lowest possible payoff of the risky option diminishes,

²⁶¹ making the risky option relatively more attractive. However, this relationship only holds for

²⁶² repeated decision events *before* an individual's fitness is evaluated. When fitness is evaluated

²⁶³ after a single decision event, a risky option is more likely to yield an extreme outcome that

²⁶⁴ translates to zero fitness (i.e., death or an ultimate failure to reproduce).

²⁶⁵ In situations like this, when a single risky decision might cost an individual's life or

²⁶⁶ offspring, average fitness is best described by the geometric mean instead of the arithmetic

²⁶⁷ mean (Haaland, Wright, & Ratikainen, 2019). The geometric mean is more sensitive to

²⁶⁸ variance because it is multiplicative, capturing the fact that one failure to reproduce can end

²⁶⁹ a genetic lineage. This circumstance has been shown to produce bet-hedging: Risk-averse

²⁷⁰ strategies may be more adaptive across many generations even when more risk-prone

²⁷¹ strategies produce better outcomes in any one generation, simply because the latter are also

²⁷² more likely to lead to extinction by sheer bad luck (Haaland et al., 2019). While average

²⁷³ fitness across generations is best represented with the geometric mean, average fitness *within*

²⁷⁴ a generation is better captured by the arithmetic mean, reflecting the additive accumulation

²⁷⁵ of payoffs from decision events before fitness is evaluated. Therefore, as the number of

²⁷⁶ decision events per generation (i.e., before fitness is evaluated) increases, the

²⁷⁷ variance-sensitive geometric mean of acquired payoffs becomes relatively less important and

²⁷⁸ the less variance-sensitive arithmetic mean becomes more important. Consequently, an

²⁷⁹ individual's behaviour should switch from relative risk-aversion to relative risk-proneness.

²⁸⁰ For the purpose of the present study, 'decision events' refer to researchers' decisions of

whether to conduct a Registered Report or pursue the standard publication route. Because Registered Reports must be submitted before data collection, such decisions occur whenever researchers start a new empirical project that they later may want to publish.[3] The number of decision events before evaluation thus reflects the number of empirical projects that a researcher can conduct before their publication record is considered for hiring, promotion, or grant funding decisions. We will call this parameter 'empirical pace'.

Key factors influencing empirical pace are the time and resources required to conduct a study and the time and resources researchers have available. Empirical pace may thus differ between 1) research areas that vary in speed and/or cost of data collection (e.g., a field relying on online questionnaires *vs* a field relying on fMRI studies), 2) research labs that vary in funding and manpower, and 3) career stages, for example because career progress often comes with increased funding and the supervision of junior researchers whose efforts boost the supervisors' output (R. Müller, 2014), or because junior researchers often have short-term contracts that limit the available time for producing research output before their CVs are evaluated for the next application.

Each of the risk-relevant factors described above — non-linear fitness functions, survival thresholds, competition, and empirical pace — 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, we examine the individual and interactive effects of these factors in a simulation model.

---

[3] At the current moment, most researchers likely never consciously consider Registered Reports as a publication option. However, the fact that they *could* nonetheless renders their pursuit of standard publications a choice, albeit an implicit one.

Simulation model

302   We develop an evolutionary agent-based model which simulates a population of

303   researchers who test hypotheses, (attempt to) publish the results either as Registered

304   Reports or as standard reports, accumulate the payoffs for successful publications, and pass

305   their publication strategies on to the next generation of researchers.

306   ***Research phase.***   Consider a population of $n = 500$ researchers. Each researcher

307   has a fixed publication strategy s, the so-called submission threshold. In each round of the

308   research phase, researchers randomly choose a hypothesis to test in a study. Hypotheses are

309   true with prior probability p, which is uniformly distributed between 0 and 1 and known to

310   the researcher. Before testing their chosen hypothesis, a researcher compares the prior p of

311   their hypothesis with their publication strategy s. When $p < s$, the researcher chooses to

312   play it safe and conduct a Registered Report to test the hypothesis. When $p \geq s$, the

313   researcher chooses to gamble and test the hypothesis in a regular study which is then

314   submitted as a standard report.

315   For simplicity, we assume that p is an ideal objective prior and that researchers'

316   hypothesis tests are free from additional sources of error. Thus, when a researcher tests

317   hypothesis i, they obtain a positive result with probability p_i and a negative result with

318   probability $1 - p_i$. If the researcher chose to submit a Registered Report, their study is

319   published regardless of the result and the researcher receives a payoff b_{RR}. However, if the

320   researcher chose to submit a standard report, they face rampant publication bias: Only

321   positive results are publishable as standard reports and yield a payoff $b_{SR+} = 1$, whereas

322   negative results are rejected or file-drawered and yield no payoff, $b_{SR-} = 0$. For all variations

323   of the model tested here, we assume that the payoff for a Registered Report falls between

324   these bounds, such that $b_{SR-} < b_{RR} < b_{SR+}$. This assumption reflects the following

325   considerations:

326   1. Due to publication bias in the standard literature, negative results are less valuable

328   than positive results ($b_{SR-} < b_{SR+}$), for example because they do not lead to a

329   publication at all, because only very low-impact journals are willing to publish them,

330   or because getting them published requires a lot of extra effort (e.g., via frequent

331   resubmissions following rejection or substantial revisions demanded by reviewers),

332   which diminishes the net reward.

333   2. For these same reasons, Registered Reports are on average more valuable than

334   standard reports with negative results ($b_{SR-} < b_{RR}$), for example because Registered

335   Reports are offered by journals that may display publication bias for standard reports

336   (rejecting standard report submissions with negative results), or simply because

337   Registered Reports need to be resubmitted less often or require less extensive revisions.

338   3. On average, standard reports with positive results are more valuable than Registered

339   Reports ($b_{RR} < b_{SR+}$), for example because many high-impact journals do not (yet)

340   offer Registered Reports, because not registering one's study *a priori* makes it easier to

341   spin the results to appear more impactful and thus increases the chances to be

342   published in a high-impact journal, or because Registered Reports may require more

343   effort due to their stricter quality criteria, lowering the net reward. While proponents

344   of Registered Reports may argue that the format has such tremendous advantages that

345   authors' resulting career benefits are superior to any alternative, this chapter is

346   predicated on the assumption that most researchers currently do not share this view.

347   Once this changes, the present investigation may happily become redundant.

348   This entire research cycle — choosing a hypothesis, choosing a publication route by

349 comparing its prior p to one's publication strategy s, testing the hypothesis, and receiving

350 payoff b_{RR} for a Registered Report or b_{SR-} or b_{SR+} for a positive and negative standard

351 report, respectively — is repeated m times.

352   ***Evaluation phase.***   At the end of the research phase, researchers' accumulated

353 publication payoffs $b_1 + b_2 + ... + b_m$ are translated into fitness f. Fitness is calculated with a

354 function characterised by exponent ϵ, which determines the shape of the function. $\epsilon = 1$

355 yields a linear function, $0 < \epsilon < 1$ yields a concave function with diminishing returns, and

356 $\epsilon > 1$ yields a convex function with increasing returns (see Figure 1):

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

357 However, two situations may cause a researcher's fitness to fall to zero even when their

358 accumulated payoffs are non-zero. First, the sum of their payoffs may fall below an absolute

359 survival threshold δ, for example when a researcher fails to meet an agreed publication target

360 by the time their 'tenure clock' runs out. Thus, when $\sum_{i=1}^{m} b_i < \delta$, $f = 0$. Second, the sum of

361 their payoffs may fall below a relative threshold γ, which reflects the intensity of competition

362 (e.g., for scarce research grants or positions). γ is the proportion of researchers who are

363 considered for reproduction. When $\gamma = 1$, all researchers in the population are considered for

364 reproduction and their fitness is calculated according to Eq. 1. When $\gamma < 1$, the

365 $(1 - \gamma) * 500$ least successful researchers receive zero fitness and cannot reproduce.[4] For

366 example, $\gamma = 0.1$ means that only those researchers with accumulated payoffs in the top 10%

367 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]

---

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

Parameter	Definition	Value [range]
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 ('empirical pace')	[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 s from a researcher in the previous generation with the probability of the previous researcher's fitness (i.e., the new generation's publication strategies are sampled with replacement from the previous generation, probability-weighted by fitness). The new generation's publication strategies 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 such hereditary 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' publication strategies s is affected by the payoff for Registered Reports b_{RR} (relative to the payoffs for

384 standard reports, which are fixed at $b_{SR-} = 0$ and $b_{SR+} = 1$), by the shape of the fitness

385 function determined by exponent ϵ, by the number of research cycles per generation m, by

386 survival threshold δ, and by competition γ (see Table 1 for an overview of the model

387 parameters and their values considered in the simulation). It is important to keep in mind

388 that a researcher's publication strategy s is not an absolute decision: It determines *how* the

389 choice between Registered Reports and standard reports is made, not which format is chosen.

390 As such, s indicates the amount of risk a researcher is willing to take. Very low values of s

391 reflect risk proneness: The researcher prefers to gamble and chooses the standard publication

392 route for almost all hypotheses they encounter, using the Registered Report route only for

393 hypotheses that are virtually guaranteed to be false (and yield negative results). Very high

394 values of s reflect risk aversion: The researcher is unwilling to risk a negative result in a

395 standard report and studies almost all hypotheses they encounter in the Registered Report

396 format, reserving the standard publication route for hypotheses that are virtually guaranteed

397 to be true (and yield positive results).

398 **Simulation approach.** We use the evolutionary mechanism of this agent-based

399 model as a means for identifying optimal behaviour under different conditions. But this goal

400 can also be achieved in other ways. One non-evolutionary alternative is to calculate expected

401 fitness (i.e., the long-run average) for a wide range of s and determine which strategy

402 maximises it in each condition. A drawback of this approach is that it does not account for

403 population dynamics and therefore cannot easily simulate the effects of competition. Because

404 of this limitation, our study is based on the evolutionary model. However, we validate all

405 analyses except those involving competition on the expected-fitness model and show that

406 both models produce virtually identical results (see Appendix).

Simulation results

408 The results of the simulation models will be presented in order of increasing model

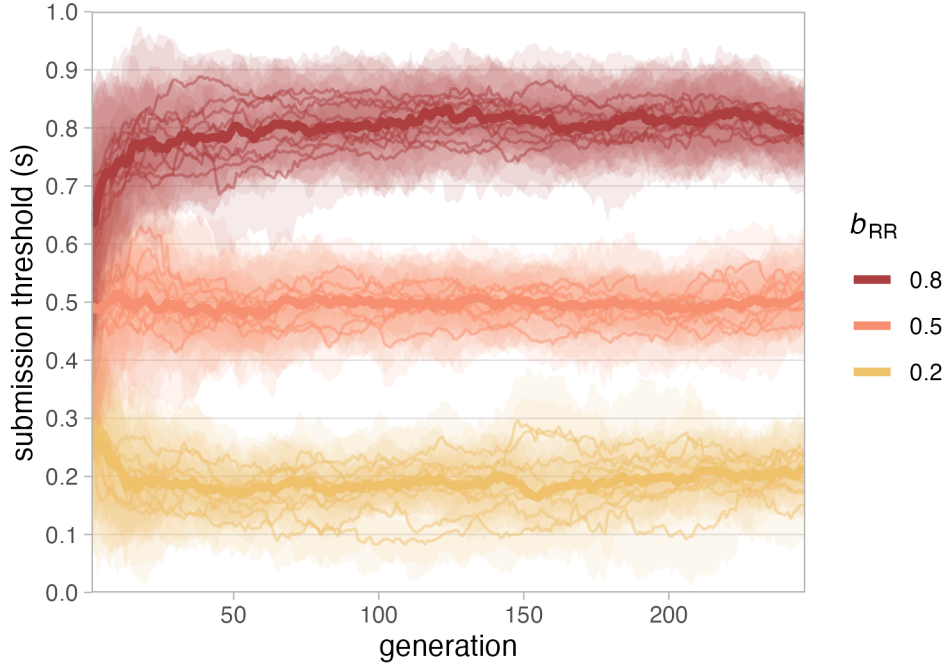409 complexity. We start by explaining the very simple scenarios shown in Figure 3 and Figure 4.

Figure 3. Evolution of publication strategy 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 publication strategy of the population in each run, shaded areas represent the inter-quartile range of publication strategies in the population in each run, and thick lines represent the median of run medians per condition.

410   These scenarios are identical to situations discussed above and the results should thus be

411   unsurprising. However, while they may seem trivial to some, we hope that these

412   explanations will help unfamiliar readers understand the basic functioning of our model as

413   well as the less intuitive results presented later.

414   When interpreting the results below, one should bear in mind that the analysed

415   parameter values are inherently arbitrary. Although the model parameters are intended to

416   capture important characteristics of real-world concepts, their values do not represent

417   real-world units. The goal of this analysis is to understand the relative effects of the model

418   parameters in a simplified, artificial system, which means that the results are only

419   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 publication strategies 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 publication strategies (s) approximate the payoff for Registered Reports in each condition, indicating that the optimal publication strategy 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:

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

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.
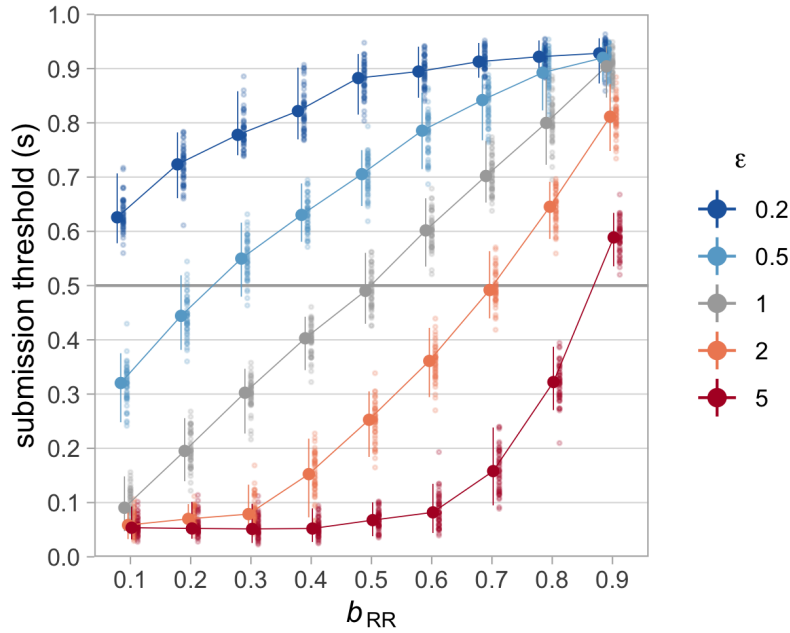

Figure 4. Effect of fitness functions on evolved publication strategies. Shown are median publication strategies in the final (250th) generations of 50 runs for different values of b_{RR} (x-axis) and different fitness functions (characterised by exponent ϵ), with one research cycle per generation ($m = 1$), 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.

Allowing for non-linear fitness functions

Arguably, the career benefits researchers receive from publications in the real world 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

452 for whom the career returns of publications exceeding the PhD requirements are thus

453 strongly decreasing (concave fitness function). Researchers who stay in academia may

454 experience that the career returns for each additional publication begin to decrease as their

455 publication record grows, meaning that advanced career stages may also be characterised by

456 a concave fitness function.

457   Figure 4 contrasts the effects of two concave fitness functions ($\epsilon = 0.2$ and $\epsilon = 0.5$,

458 shown in blue shades) and two convex fitness functions ($\epsilon = 2$ and $\epsilon = 5$, shown in red

459 shades) with a linear function ($\epsilon = 1$, grey line) for different payoffs for Registered Reports,

460 in the same simple scenario with only one research cycle per generation. The grey line for

461 $\epsilon = 1$ represents the already familiar situation from Figure 3 above: When the fitness

462 function is linear, the optimal strategy is $s_{optimal} = b_{RR}$. Non-linear fitness functions deviate

463 from this pattern exactly as expected based on Figure 1. When additional payoffs yield

464 diminishing returns ($\epsilon < 1$), Registered Reports become more attractive even when they are

465 worth less than the expected payoff for standard reports. As explained above, this is because

466 concave functions 'shrink' the difference between moderate and high payoffs relative to the

467 difference between low and moderate payoffs. Conversely, when additional payoffs yield

468 increasing returns ($\epsilon > 1$), Registered Reports are unattractive unless their payoffs are

469 almost as large as those for published standard reports because convex functions increase the

470 difference between moderate and high payoffs relative to low versus moderate payoffs.

471   When different fitness functions are taken to reflect different career stages, this pattern

472 suggests that Registered Reports should be more attractive for senior researchers and a

473 tough sell for early-career researchers. Interestingly, preliminary empirical evidence suggests

474 the opposite: Registered Reports appear to be more likely to have early-career researchers as

475 first authors than standard reports (77% vs 67% in the journal *Cortex*, Chambers & Tzavella,

476 2021). One explanation for this counterintuitive result could be that Registered Reports are

477 disproportionally used by early-career researchers who intend to leave academia and thus

have a concave fitness function. Alternatively, factors or dynamics not considered in this

simulation may 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
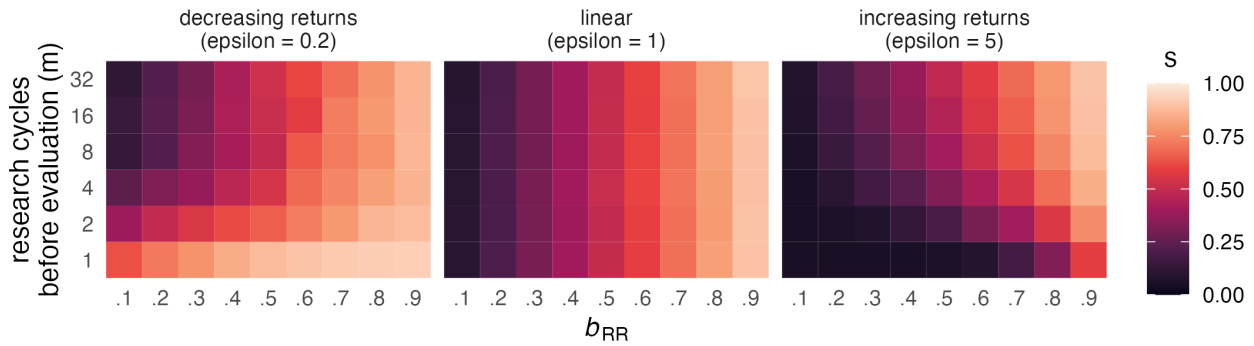
factors.

Figure 5. Effect of research cycles per generation on evolved publication strategies. Shown are median evolved publication strategies (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 presented 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 s decreases (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

494   panel) and even *increases* when it is convex ($\epsilon = 5$, right panel).

495   Why does m appear to have opposite effects for concave and convex fitness functions?

496   As a starting point, it helps to first consider only the bottom row of each panel, where

497   $m = 1$. These three rows contain the same results as the top, middle, and bottom curves in

498   Figure 4 and show risk aversion when $\epsilon = 0.2$ (i.e., Registered Reports are attractive even

499   when they yield a low payoff), risk proneness when $\epsilon = 5$ (Registered Reports are

500   unattractive even when they yield a high payoff), and a linear strategy $s_{optimal} = b_{RR}$ when

501   $\epsilon = 1$. From this starting point, the two panels with non-linear fitness functions start to

502   approximate the linear case as m increases. This pattern reflects the idea that fitness is

503   better captured by the geometric mean when m is low, and better captured by the

504   arithmetic mean when m is high (Haaland et al., 2019).

505   To better understand this dynamic, let's consider two researchers with extreme

506   submission strategies: Regina Register conducts only Registered Reports ($s_{Regina} = 1$),

507   Darren Daring conducts only standard reports ($s_{Darren} = 0$). The payoff for Registered

508   Reports is fixed at $b_{RR} = 0.5$. After one research cycle, Regina receives a payoff of 0.5 and

509   Darren receives either 0 or 1 (with 50/50 odds). If fitness is calculated after this one round

510   with $\epsilon = 0.2$ (concave function, yielding diminishing returns), Regina's fitness is

511   $f_{Regina} = \frac{1}{2}^{\frac{1}{5}} = 0.87$, and Darren's fitness is either $f_{Darren-} = 0^{\frac{1}{5}} = 0$ or $f_{Darren+} = 1^{\frac{1}{5}} = 1$.

512   In a population of 100 Reginas and 100 Darrens, there will be roughly 50 lucky Darrens who

513   get a positive result and 50 Darrens who get a negative result. Lucky Darrens have a narrow

514   fitness advantage over all Reginas (1 versus 0.87), while unlucky Darrens lose to all Reginas

515   by a wide margin (0 versus 0.87). Since there are twice as many Reginas as lucky Darrens,

516   the Regina strategy is relatively more successful.

517   Let's now consider the same scenario with $m = 4$ research cycles per generation.

518   Reginas receive the same payoff in every round and accumulate $b_{total} = \frac{1}{2} * 4 = 2$. Lucky

519   Darrens (who win every time) accumulate $b_{total} = 1 * 4 = 4$, while unlucky Darrens (who lose

520 every time) again receive 0 total payoff. Now, however, the probabilistic outcomes over 4

521 rounds lead to three additional versions of Darren: moderately lucky (winning 3/4 times),

522 average (2/4, receiving the same total payoff as Reginas), and moderately unlucky (1/4).

523 Translating payoffs into fitness, the Regina strategy ($f_{Regina} = 2^{\frac{1}{5}} = 1.15$) still yields an

524 enormous advantage compared to unlucky Darrens ($f_{Darren_{unlucky}} = 0$) and only a small

525 disadvantage compared to lucky Darrens ($f_{Darren_{lucky}} = 4^{\frac{1}{5}} = 1.32$). But this time, there are

526 fewer Darrens who are less successful than Reginas because Reginas now share their place

527 with average Darrens. The relative fitness advantage of the Regina strategy thus decreases.

528 As the rate of research cycles per generation grows, the law of large numbers dictates that

529 more and more Darrens achieve average total payoffs, while fewer and fewer Darrens achieve

530 extreme total payoffs (winning 32 times in a row is much less probable than winning 4 times

531 in a row). This reduces the width of the Darren distribution until it approximates the

532 Regina distribution—meaning that optimal publication strategies become identical to those

533 optimal for a linear fitness function.

534   When the fitness function is convex ($\epsilon = 5$, yielding increasing returns), the overall

535 effect of increasing values of m is the same, with the only difference that Reginas are initially

536 disadvantaged (because their fitness distance to the lucky half of Darrens is much greater

537 than than to the unlucky Darrens). With larger m, more and more Darrens receive average

538 total payoffs and share Regina's disadvantaged position (decreasing Regina's relative

539 disadvantage), until the Darren distribution is again virtually equal to the Regina

540 distribution. Rather than causing absolute risk aversion, increasing values of m thus counter

541 the effect of ϵ and reduce the effects of concave and convex fitness functions to the linear

542 case. Consequently, the top rows ($m = 32$) of the top and bottom panels in Figure 5

543 resemble the stable pattern across all m shown in the middle panel.

544   Translated into terms of academic careers, this less intuitive pattern indicates that

545 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. Regarding junior researchers (for whom

additional publications have increasing returns on career success), the results suggest that

they 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 (δ) 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

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). Such

requirements may represent low, medium, or high survival thresholds depending on how

demanding they are (e.g., the proportion of thesis chapters that must be published).

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%
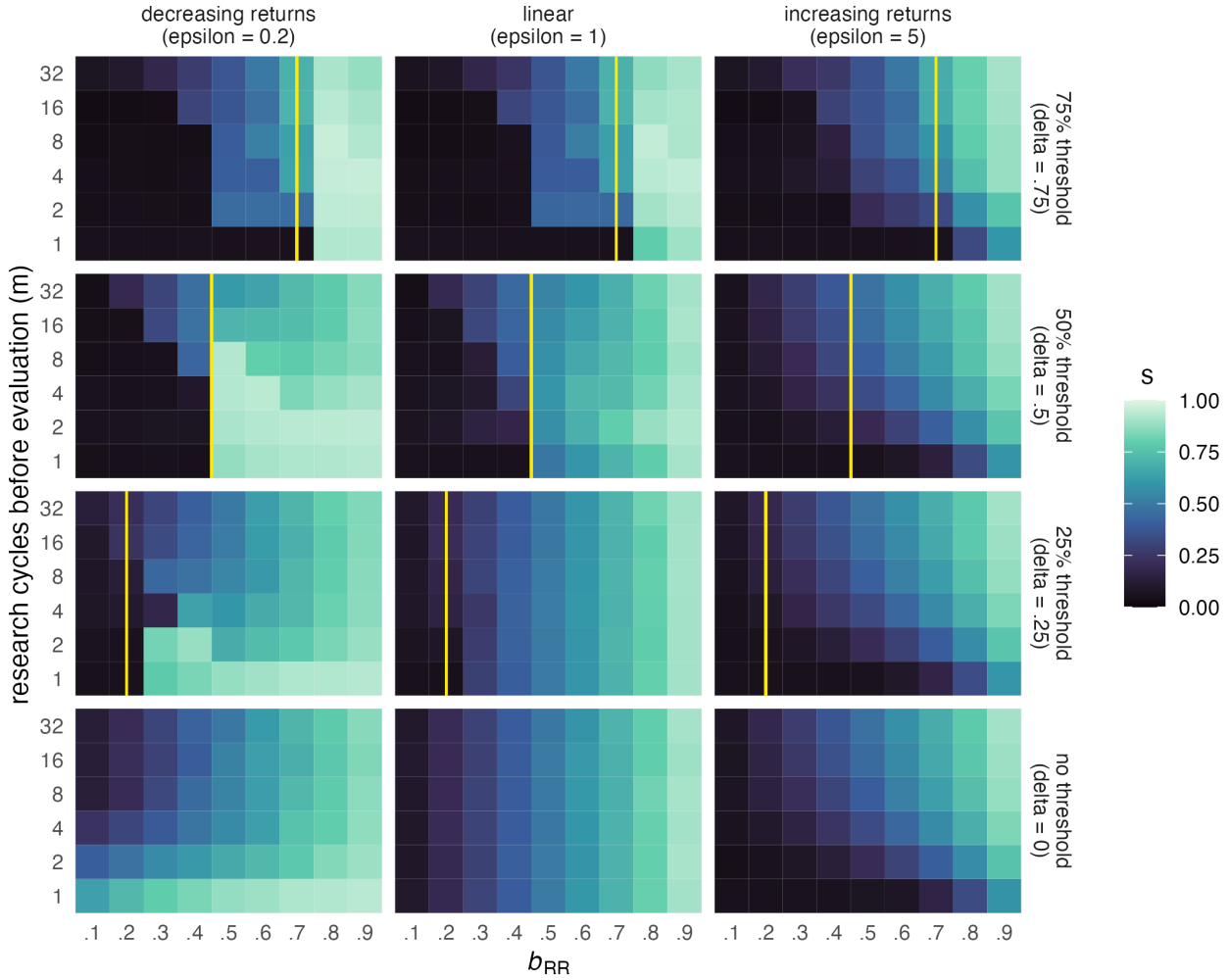
Figure 6. Effect of survival thresholds on evolved publication strategies. Shown are median publication strategies (s) after 250 generations in 50 runs (tile colour represents the median of 50 run medians) depending on survival thresholds (δ, shown as vertical yellow line), fitness functions (characterised 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 $\delta * m$.

572 ($\delta = .75$) means that researchers must gain at least 3 points to be able to reproduce. When

573 $b_{RR} = .7$, submitting four Registered Reports will only amount to 2.8 points in total, just

574 short of meeting the threshold. On the other hand, when $b_{RR} = .8$ (i.e., just above δ), four

575 Registered Reports would yield 3.2 points and thus ensure reproduction. Choosing the

576 standard route some of the time can increase fitness even further, but also increases the risk

577 of not meeting the survival threshold. As a consequence, one may intuitively expect

578 Registered Reports to be popular whenever $\delta \leq b_{RR}$ and unpopular whenever $\delta > b_{RR}$.

579 Figure 6 shows that this is true in many, but not all conditions. First, we can see that

580 survival thresholds have their biggest effect when the number of research cycles per

581 generation is low — at high values of m, publication strategies are virtually unaffected in all

582 conditions. Second, survival thresholds have a stronger effect when the fitness function is

583 linear ($\epsilon = 1$) or concave ($\epsilon = 0.2$). In these two conditions, they produce very similar

584 patterns: The Registered Report route is almost never chosen when b_{RR} is too low to meet

585 the survival threshold (particularly at $\delta = .25$ and $\delta = .5$; less so at $\delta = .75$), and this effect

586 tapers off as the number of research cycles increases. Compared to baseline, the change is

587 particularly striking for the concave fitness function ($\epsilon = 0.2$, left column in Fig. 6), where

588 RRs are normally preferred at low m. When the survival threshold is high ($\delta = .75$) or the

589 fitness function is concave, we can also see that Registered Reports become *more* popular

590 than baseline when they are worth just enough to pass the survival threshold. For the

591 convex fitness function ($\epsilon = 5$) on the other hand, survival thresholds of 25% and 50% seem

592 to have no effect at all. Only a high threshold of 75% makes RRs even less popular when

593 they have low value ($b_{RR} \leq 0.4$), especially when the number of research cycles is low.

594 What does this mean in practice? In our model, fitness (according to the three

595 different fitness functions) is calculated after the survival threshold has been met. This is

596 meant to mimic publication requirements that are expressed in raw numbers. Importantly, it

597 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 relatively 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 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
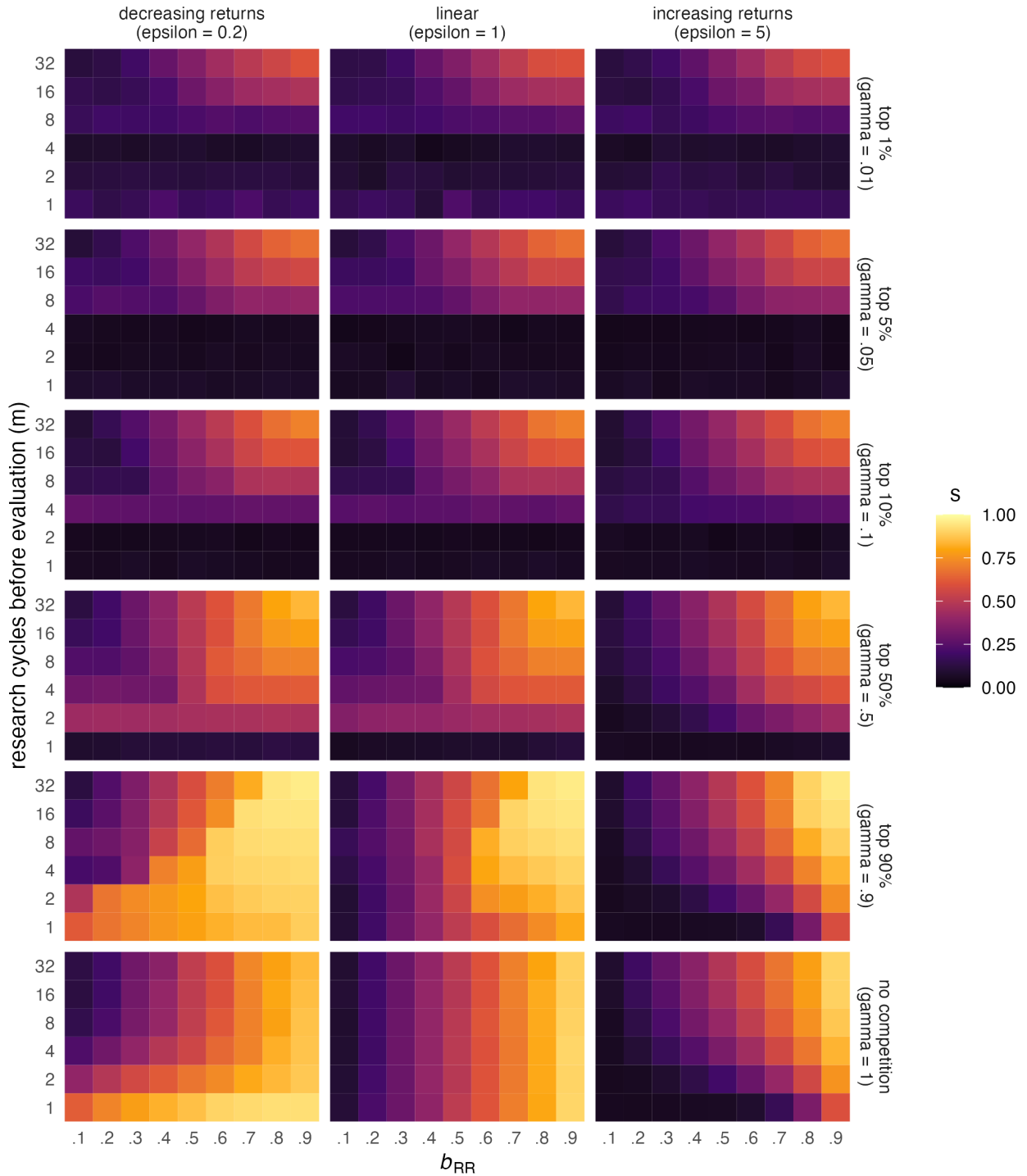
Figure 7. Effect of competition on evolved publication strategies. Shown are median evolved publication strategies (s) after 250 generations in 50 runs (tile colour represents the median of 50 run medians) depending on the intensity of competition (γ, y-axis), numbers of research cycles per generation (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.

624 Registered Reports, as can be seen by the darkening of the plots when moving up from the

625 bottom row of panels. The only exception to this rule is very low competition: When the top

626 90% are allowed to reproduce (and only the bottom 10% are rejected, $\gamma = .9$), Registered

627 Reports become more popular than they are in the absence of competition. This effect is

628 strongest for the concave fitness function ($\epsilon = 0.2$), where it holds for almost all values of

629 b_{RR} at very low numbers of m and for high values of b_{RR} at high numbers of m. When the

630 fitness function is linear or convex, Registered Reports are chosen more often only when both

631 b_{RR} and m are high. At higher levels of competition ($\gamma > .5$), the differences between the

632 fitness functions disappear. In all three cases, Registered Reports are essentially wiped out

633 for low numbers of research cycles (m), and this effect increases with competition (the higher

634 the competition, the higher m must be for Registered Reports to still be viable). Intense

635 competition also negatively affects Registered Reports at high numbers of m, but here the

636 general pattern of the baseline condition (a linear increase of Registered Reports popularity

637 with b_{RR}) remains intact.

638 Looking at the first three rows of panels in Figure 7 (1%, 5%, and 10% competition),

639 the extreme effect of competition at low m appears to decrease slightly when competition is

640 highest ($\gamma = .01$), indicated by the dark bar at the bottom of each panel becoming slightly

641 lighter. This paradoxial result is not due to Registered Reports being more lucrative in those

642 conditions. Rather, competition is so extreme that the natural selection in our model starts

643 operating more on chance than on individuals' traits. Essentially, only individuals with the

644 maximum possible payoff (publishing only standard reports with positive results) are able to

645 reproduce. Most likely to receive this maximum payoff are individuals who investigate

646 hypotheses with high prior probabilities. In our model, this is not a trait that can be passed

647 on, but determined by random chance. Among individuals who experience this kind of luck,

648 the variance of publication strategy s should be high: A hypothesis with prior $p = .95$ will be

649 submitted as a standard report and likely yield a positive result (and thus the maximum

650 payoff) regardless of whether the researcher's publication strategy 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 $\gamma = .01$ (Fig. 7), which fluctuate 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 publication strategies 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 publication 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[5], 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 those were actually merely coincidental.

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 when very few research cycles can be completed before an evaluation event (e.g., in fields with low empirical pace, in labs with few resources, or on short-term contracts): In such situations, publication strategies that involve any amount of Registered Reports are only viable when competition is so high that success requires extraordinary luck. In contrast, very

---

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

675 low but non-zero levels of competition increase the popularity of Registered Reports,

676 especially when their value is high, when the fitness function is concave (e.g., in later career

677 stages), and when researchers can complete many studies before being evaluated.

Discussion

679 In the artificial world of the model presented here, the standard publication route is a

680 coin toss — the probability of obtaining a publishable result is 50% on average[6], translating

681 to an expected payoff of 0.5 points per study. If Registered Reports are a safe alternative to

682 this gamble and guarantee publication in every case, one might think that payoff-maximising

683 researchers would prefer them whenever they are worth more than the expected payoff from

684 standard reports and avoid them whenever they are worth less. This intuition, however, rests

685 on the assumption that the career benefits researchers receive from publications are linear

686 and involve no step changes.[7] We argue that this assumption is violated in many, if not all,

687 real-world situations. Here, we investigated the impact of four factors that likely shape

688 real-world situations: convex vs concave fitness functions (additional publications yielding

689 either increasing or decreasing returns, reflecting early vs later career stages), empirical pace

690 (reflecting differences in speed and cost of data collection, available resources, or available

691 time), survival thresholds (reflecting absolute publication targets researchers must meet in a

692 given time), and competition for jobs or grants. Our results show that in isolation or

693 combined, many of these factors would lead researchers with career-maximising strategies to

694 avoid Registered Reports — even when Registered Reports are worth more than the expected

695 payoff from standard reports.

696 To summarise the results, it is useful to take the middle panel of Figure 5 ($\epsilon = 1$) as a

---

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

[7] 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).

697 baseline. In this panel, publication payoffs translate into linear career benefits (the fitness

698 curve is linear and there is no survival threshold and no competition), and the outcome is

699 highly intuitive: Researchers prefer Registered Reports whenever they are worth more than

700 0.5 points, their preference is exactly proportional to b_{RR}, and it is not affected by empirical

701 pace. Compared to this baseline, Registered Reports are *less* popular when a) additional

702 publications yield increasing returns (e.g., in early career) and empirical pace is low, b) when

703 researchers face a survival threshold that cannot be met with Registered Reports alone,

704 especially when publications yield decreasing returns (e.g., in advanced career stages) and

705 empirical pace is low, and c) when there is substantial competition. Competition has the

706 most extreme effect and can cause a complete avoidance of Registered Reports when

707 empirical pace is low. Conversely, Registered Reports are *more* popular than at baseline

708 when a) additional publications yield decreasing returns and empirical pace is low, b)

709 Registered Reports are worth just enough to reach a survival threshold and publications

710 yield decreasing returns, especially when empirical pace is low, and c) when there is very low

711 but non-zero competition, especially when publications yield decreasing returns or empirical

712 pace is high.

713   Looking at the interactions of the different factors, three observations stand out. First,

714 high empirical pace attenuates the effects of all other factors — at the highest pace we

715 considered ($m = 32$), outcomes are identical to baseline in almost all conditions. The only

716 exception to this rule is high competition, but although Registered Reports are relatively less

717 attractive in this condition, the basic pattern is preserved and they remain viable when their

718 value is high. Second, the effect of survival thresholds strongly depends on the shape of the

719 fitness function, suggesting that publication targets may have the strongest impact in

720 advanced career stages. Third, the opposite is true for high competition, which cancels out

721 the effects of different fitness functions and thus appears to have virtually the same impact

722 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

748   conservative interpretation in which the prior probability of hypotheses simply reflects

749   authors' predictions of the eventual publication value of different research questions.

750   This interpretation is still concordant with Registered Reports and standard reports

751   differing in risk, because the publication value of standard reports certainly depends

752   more strongly on the study results than the publication value of Registered Reports

753   (even if not in the simplistic sense of positive hypothesis tests having higher value).

754   • Fitness concept: one caveat is that

755   • RRs may actually *slow* the empirical pace, introducing an interaction that our model

756   doesn't take into account

757   • Fitness curves: more senior researchers may also take the needs of their early-career

758   mentees into account

Future directions

760   ***Ability-based risk taking.***   The model presented in this chapter only considers the

761   effects of situational factors on individuals' risk sensitivity. However, risk sensitivity can also

762   be influenced by individual differences, such that individuals with traits or abilities that

763   increase their expected payoff from a risky option (e.g., traits that increase their winning

764   chances or the payoff when winning or that buffer the impact of losses) should be more

765   risk-prone (Barclay, Mishra, & Sparks, 2018). Such factors may be important to consider in

766   the context of research and publication practices. For example, researchers who are better at

767   choosing research questions that are likely to result in high-impact publications (e.g.,

768   through talent or experience) may find Registered Reports less attractive. As a more

769   nefarious version of this idea, Registered Reports may be relatively unpopular among

770   researchers who are more inclined to using questionable research practices (or even fraud) to

771   obtain publishable or impactful results.

772   ***Registered Reports and post-publication peer review.***   The post-publication

773   peer review platform *Peer Community In* (PCI) recently launched a new model of Registered

774   Reports (PCI Registered Reports) in which authors are no longer tied to a specific journal.

775   PCI offers authors the regular process of stage-1 and stage-2 review, the end result of a

776   successful submission is 'only' a preprint with a so-called 'recommendation' from PCI.

777   Authors can subsequently publish their manuscript in one of several journals who partnered

778   with PCI and either rely on the PCI review process alone or offer a streamlined review

779   process for PCI-recommended preprints, or they can submit to any other journal as if their

780   manuscript were a standard report. This innovation gives Registered-Reports authors

781   significantly more freedom to capitalise on the results of their study because a submission to

782   PCI Registered Reports does not preclude the chance of a high-impact publication. PCI

783   Registered Reports thus constitute a significant change to the relative incentives and risk

784   structure of Registered Reports compared to standard reports that merits a closer

785   investigation in the future.

786   **Conclusion**

787   **Disclosures**

788   **Data, materials, and online resources.**   This manuscript was created using

789   RStudio (1.2.5019, RStudio Team, 2019) and R (Version 4.2.1; R Core Team, 2019) and the

790   R-packages *bookdown* (Version 0.34; Xie, 2016), *ggplot2* (Version 3.5.0; Wickham, 2016), *here*

791   (Version 1.0.1; K. Müller, 2017), *knitr* (Version 1.46; Xie, 2015), *papaja* (Version 0.1.1.9001;

792   Aust & Barth, 2018), *rmarkdown* (Version 2.26; Xie, Allaire, & Grolemund, 2018), *stringr*

793   (Version 1.5.1; Wickham, 2023), and *tinylabels* (Version 0.2.3; Barth, 2022).

References

795   Agnoli, F., Wicherts, J. M., Veldkamp, C. L. S., Albiero, P., & Cubelli, R. (2017).

796   Questionable research practices among italian research psychologists. *PLOS ONE*,

797   *12*(3), e0172792. https://doi.org/10.1371/journal.pone.0172792

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

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, A., 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

Ingre, M., & Nilsonne, G. (2018). Estimating statistical power, posterior probability and publication bias of psychological research using the observed replication rate. *Royal Society Open Science*, *5*(9), 181190. https://doi.org/10.1098/rsos.181190

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, A., & 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

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

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

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

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

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

907   Chapman and Hall/CRC.

908 Xie, Y. (2016). *Bookdown: Authoring books and technical documents with R markdown.*

909   Boca Raton, Florida: Chapman and Hall/CRC.

910 Xie, Y., Allaire, J. J., & Grolemund, G. (2018). *R markdown: The definitive guide.* Boca

911   Raton, Florida: Chapman and Hall/CRC.