1   Incentives for Registered Reports from a risk sensitivity perspective

Incentives for Registered Reports from a risk sensitivity perspective

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

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

1. RRs: what they are (brief)

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

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

27   more detail, cite evidence

28   4. RRs have become more popular

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

30   6. This introduces a risk of selecting based on results

31   7. Result-based selection bias is bad

32   8. RRs select before results are known to avoid this problem

33   9. Explain how RRs work

34   10.

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

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

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

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

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

46   Under-reporting of negative results skews the available evidence for scientific claims
47 and can lead to overconfidence and an increased rate of false-positive inferences. Evidence

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

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

⁷⁵ QRPs denote practices which exploit undisclosed flexibility in data collection and

⁷⁶ analysis, for example by running the analyses on different justifiable combinations of

⁷⁷ variables, exclusion and decision criteria, and only reporting the ones with favourable results,

⁷⁸ or by presenting post-hoc inferences as having been predicted a priori John, Loewenstein, &

⁷⁹ Prelec (2012). Registered Reports minimise the risk of QRPs

⁸⁰ Publication bias and QRPs skew the available evidence for scientific claims and can

⁸¹ lead to overconfidence and an increased rate of false-positive inferences. Second, the risk of

⁸² QRPs is minimised by the two-stage review process, in which the Stage-1 protocol acts as a

⁸³ preregistration and reviewers' task during Stage-2 review is to flag any undisclosed

⁸⁴ deviations from it.

⁸⁵ The format was first launched in 2013 at the journal *Cortex* (Chambers, 2013) and is

⁸⁶ now offered by over 300 journals, predominantly in the social and life sciences (see cos.io/rr).

⁸⁷ By 2021, nearly 600 Registered Reports had been published (Chambers & Tzavella, 2021).

⁸⁸ Initial evidence shows that published Registered Reports have a substantially lower rate of

⁸⁹ positive results than regular articles in psychology (44% versus 96%, Scheel, Schijen, &

⁹⁰ Lakens, 2021) and psychology, neuroscience, and the biomedical sciences (Allen & Mehler,

⁹¹ 2019), and are judged to be of higher quality (Soderberg et al., 2021).

⁹² Assuming that researcher practices s' publication practices are at least in part strategic

⁹³ responses to incentives, changing the incentives should

⁹⁴ To do:

⁹⁵ • General introduction describing the problem of publication bias and questionable

⁹⁶ research practices (QRPs)

⁹⁷ – Publication bias: positive results are more likely to get published than negative

⁹⁸ results; can happen due to reviewer bias (Greenwald, 1975; Atkinson et al., 1982;

⁹⁹ evidence: Mahoney, 1977) or file-drawering (Rosenthal, 1979; evidence: A. Franco

et al., 2014; Annie Franco et al., 2016)

– QRPs: exploiting undisclosed flexibility in data collection and analysis to obtain more desirable results (Simmons et al., 2011; Agnoli et al., 2017; Fiedler & Schwarz, 2016; Fraser et al., 2018; evidence: John et al., 2012). QRPs inflate the error rates of statistical tests, typically the false-positive rate

– consequence of publication bias and QRPs: the literature in psychology is excessively ($> 90\%$) positive (Fanelli, 2010; Scheel et al., 2021; Theodore D. Sterling, 1959; Theodor D. Sterling, Rosenbaum, & Weinkam, 1995) and unreliable (Wacholder, Chanock, Garcia-Closas, El ghormli, & Rothman, 2004; evidence: Nosek et al., 2022)

- Reform efforts to address the problem:

– preregistration to reduce QRPs (Lakens, 2019; Nosek, Ebersole, DeHaven, & Mellor, 2018; Wagenmakers et al., 2012)

– various journals of negative results to reduce publication bias (these never seem to be successful though and always shut down after a while; add examples/references)

– **Registered Reports** to reduce QRPs and publication bias at the same time (most powerful reform proposal to date)

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

Being a powerful bias-prevention tool that is increasingly popular, it is important to develop a better understanding of when, where, and by whom Registered Reports are most likely to be used. First, such knowledge can help identify research areas in which the format is unlikely to gain traction by itself and anticipate the need for further intervention (e.g., via policy) when there is a demand for unbiased results. Second, understanding when

researchers' choice between Registered Reports and the standard publication route is likely to be influenced by factors that also influence the eventual results (e.g., the prior probability of the tested hypotheses) is important for meta-scientific studies that compare published studies in both formats and must take such confounds into account (e.g., Scheel et al., 2021). Such confounds could also lead to a situation in which Registered Reports become associated with certain types of results (e.g., negative results) and devalued if these results are deemed less interesting or important by the research community, making the format unsustainable in the long run. `To do: add note that this is likely what happened to several "journals of negative results" that shut down due to lack of interest.` The goal of this chapter is to shed light on these questions by studying the potential impact of a key feature of Registered Reports: The results-independent publication guarantee as an incentive for authors.

Registered Reports as a low-risk publication option

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

Peer-reviewed publications are a central currency for academic researchers, both in terms of publication quantity and publication impact (R. Müller, 2014; van Dalen & Henkens, 2012). In the standard publication model, researchers face uncertainty about whether and where they will be able to publish the results of their study. Translated into currency terms, the payoff a researcher receives for conducting a study can vary extremely: from near zero when the resulting manuscript is rejected by all consulted journals (or when the author file-drawers the study because the chances of success seem too low to justify the

152 costs of repeated submissions and revisions) to an extremely high, perhaps career-making

153 amount when a manuscript is published in a very high-impact journal like *Nature* or *Science*.

154 In other words, success in the standard system is highly variable and highly volatile since it

155 hinges on the one factor that is supposed to be outside of researchers' control — the study

156 results. This unfortunate combination can be excessively stressful for researchers (especially

157 junior scientists without secure positions) and tempt them to hype, spin, or even fabricate

158 their results.

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

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

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

162 (September 2021):

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

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

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

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

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

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

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

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

171 benefit of a Registered Report is always at least as valuable as the best possible outcome

172 that could be achieved through the standard publication route, the answer is "probably yes".

173 Authors deciding between Registered Reports and the standard publication route face the

174 choice between a payoff with low variability (a relatively safe publication in the journal the

175 Stage-1 protocol was submitted to) and a payoff with high variability (anywhere between no

176 publication and a high-impact publication, or even several publications if the project yields

177 enough "fodder"). Situations like these are commonly termed *decision-making under risk*.

178 "Risk" is defined as "unpredictable variation in the outcome of a behavior, with consequences
179 for an organism's fitness or utility" (Winterhalder, Lu, & Tucker, 1999, p. 302). Organisms
180 are *risk sensitive* when they are not only sensitive to the mean outcomes of different
181 behavioural options but also to their variance.

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

192 To do:

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

196 **Goals of the chapter**

197 In this chapter, a simulation model is used to explore how properties of academic
198 careers and academic incentive structures that are relevant to risk sensitivity may affect the
199 strategies of researchers choosing between Registered Reports and the standard publication
200 format. The research goal is to understand in which circumstances Registered Reports should
201 be particularly attractive, particularly unattractive, or particularly prone to highly selective
202 use. The results of this analysis may help anticipate where the format is unlikely to take foot

²⁰³ without additional changes to norms, incentives, or policy, and flag situations in which the

²⁰⁴ results of published Registered Reports may be particularly difficult to compare to the

²⁰⁵ normal literature. The following sections outline central concepts of Risk-Sensitivity Theory,

²⁰⁶ relate them to characteristics of academic careers, and describe an evolutionary simulation

²⁰⁷ model in which their effects on researchers' risk-sensitive publication decisions are examined.

²⁰⁸ **Conceptual application of Risk-Sensitivity Theory to publication decisions**

²⁰⁹   This section describes general factors that affect the role of risk for individual's fitness

²¹⁰ and connects these factors to relevant elements of academic careers. In this context,

²¹¹ Risk-Sensitivity Theory's focus on reproductive fitness as the central outcome may be seen

²¹² as misguided. But although researchers do not forage, grow, reproduce, and die in the

²¹³ *biological* sense (except in their role as human beings in general, of course), they undoubtedly

²¹⁴ are concerned with factors that influence 1) their survival and 2) the propagation of their

²¹⁵ traits in an *academic* sense. Even if we were to assume that researchers are not consciously

²¹⁶ trying to maximise their 'academic fitness', a competitive job market will *per definitionem*

²¹⁷ select for individuals whose behaviour increases their fitness. In applying Risk-Sensitivity

²¹⁸ Theory to researchers' publishing behaviour, we will therefore use a general notion of career

²¹⁹ success as the central outcome variable in place of reproductive fitness. This decision does

²²⁰ not imply that career success is the only or the proximal motivation for researchers'

²²¹ behaviour in practice, just as evolutionary theory does not imply that reproductive success is

²²² the only or the proximal motivation for human behaviour in everyday life. `To do:`

²²³ • Refer back to a (not yet existing) section above to say that human decision making is a

²²⁴   product of evolution

²²⁵ • In addition, narrow bottlenecks between early-career and tenured positions in many

²²⁶   academic disciplines inevitably create a selection pressure for behaviours that further

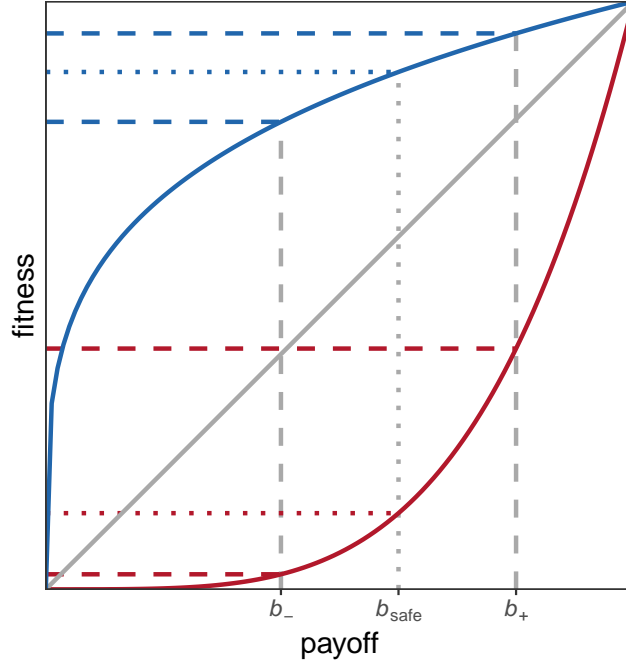²²⁷   researchers' career success (Smaldino & McElreath, 2016).

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.

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

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

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

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

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

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

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

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

236   function relating payoffs to utility is instead convex or concave (yielding increasing or

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

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

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

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

241   function is concave, and with less fitness when the fitness function is convex. In other words,

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

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

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

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

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

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

248 effect on career success, whereas established professors may care little about any one
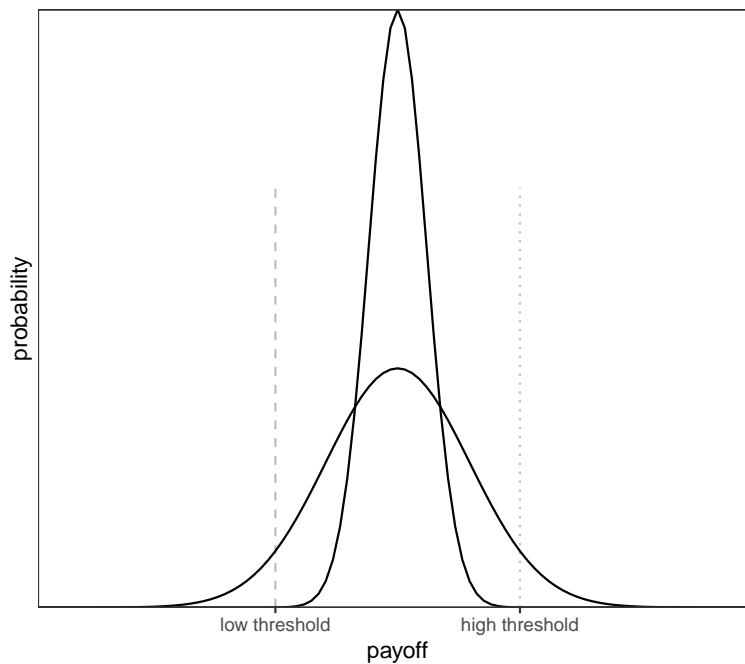
249 additional publication on their record.

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

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

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

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

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

²⁵⁴ individual will be risk averse when the resources provided by a low-variance option are

²⁵⁵ sufficient to meet the threshold and risk-prone when they are not (Mishra, 2014). For

²⁵⁶ example, a hummingbird that needs to acquire a certain amount of calories to survive the

²⁵⁷ night will prefer a low-risk food source if this option's expected payoff is above the threshold,

²⁵⁸ but avoid the low-risk source if only a higher-risk source provides a chance of survival. One

²⁵⁹ such situation is depicted in Figure 2.

²⁶⁰ Although comparable cutoff points in academic careers may have somewhat less severe

²⁶¹ consequences, they certainly exist: Amount and impact of a researcher's publications are

²⁶² common and often explicit criteria in decisions that are central to the individual's career,

²⁶³ such as whether they will be awarded a PhD, whether they will receive grant funding,

²⁶⁴ whether they will be offered a tenure-track position, or whether they will be granted tenure.

²⁶⁵ In some of these situations, the cutoff points are absolute and thus resemble survival

²⁶⁶ thresholds in the biological sense, for example PhD-programme regulations that determine a

²⁶⁷ minimal amount of peer-reviewed publications for a candidate to be awarded with a PhD, or

²⁶⁸ tenure contracts that specify minimal publication targets. In other situations, the cutoff

²⁶⁹ points are relative and depend on the number of eligible candidates, for example when grant

²⁷⁰ funding is awarded to the 10 highest-ranked research proposals or a job is offered to the best

²⁷¹ candidate from a pool of applicants. In cases like these, one individual's success diminishes

²⁷² the chances of another — they represent *competition*. In the following, survival thresholds

²⁷³ and competition will be treated as separate concepts to examine their differential effects on

²⁷⁴ researchers' publication behaviour.

²⁷⁵ ***Number of decision events before evaluation.*** A final risk-relevant factor

²⁷⁶ considered here is the number of decision events taking place before an individual's fitness is

²⁷⁷ evaluated. When a risky option is chosen repeatedly, the average of the accumulating payoffs

²⁷⁸ gets closer and closer to the long-run expected payoff. This means that the danger of loosing

²⁷⁹ out completely by only acquiring the lowest possible payoff of the risky option diminishes,

²⁸⁰ making the risky option relatively more attractive. However, this relationship only holds for

repeated decision events *before* an individual's fitness is evaluated. When fitness is evaluated after a single decision event, a risky option is more likely to yield an extreme outcome that translates to zero fitness (i.e., death or an ultimate failure to reproduce).

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

In the academic world, decision events before fitness is evaluated ("per generation") could seen as the time and resources a researcher has available for producing publications before a relevant selection event like those mentioned in the previous section (award of a PhD or grant, job application, tenure decision) is made. This parameter likely varies with career stage: A PhD student usually has three to four years to achieve the required publication output, a postdoc may work on a short-term contract of two years or even one year (after which their CV must be strong enough for the next application), and an assistant professor may have around seven years for receiving tenure. In addition, career progress

307   often comes with greater research funds and, most importantly, the supervision of students

308   and junior researchers whose efforts boost the supervisors' output (R. Müller, 2014). As a

309   second, orthogonal aspect, the amount of publishable research that can be achieved before a

310   selection event may vary between research areas. In some fields, data collection is fast and

311   cheap, for example when experiments consist of short online questionnaires that are

312   disseminated to large participant pools such as Amazon MTurk. In other fields, data

313   collection is very expensive and slow, for example in clinical fMRI studies on specific patient

314   groups. Irrespective of career stage, researchers in fields with fast and cheap data may thus

315   be able to complete many more research cycles per time unit than researchers who use more

316   expensive data.

317   Each of the risk-relevant factors described above — non-linear fitness functions, survival

318   thresholds, competition, and number of decision events before evaluation — likely impacts

319   researchers' decision strategies, including their choices between low-risk and high-risk

320   publication options. To better understand when a low-risk option like Registered Reports

321   should be particularly attractive or unattractive, the individual and interactive effects of

322   these factors are examined in a simulation model.

Simulation model

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

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

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

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

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

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

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

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

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

333 their hypothesis with their submission threshold s. When $p < s$, the researcher chooses to

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

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

336 submitted as a standard report.

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

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

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

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

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

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

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

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

345 of the model tested here, we assume that $b_{SR-} < b_{RR} < b_{SR+}$. This assumption reflects the

346 following considerations:

347 1. Due to publication bias in the standard-report literature, negative results are less

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

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

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

351 resubmissions following rejection or substantial revisions demanded by reviewers)

352 which diminishes the net reward.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

369 comparing its prior p to one's submission threshold s, testing the hypothesis, and receiving

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

371 report, respectively — is repeated m times.

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

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

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

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

376 $\epsilon > 1$ yields a convex function with increasing returns (see Figure 1):

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

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

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

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

380 by the time their "tenure clock" runs out. Thus, when $\sum_{i=1}^{m} b_i < \delta$, $fitness = 0$. Second, the

381 sum of their payoffs may fall below a relative threshold γ, which reflects the intensity of

382 competition (e.g., for scarce research grants or positions). γ is the proportion of researchers

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

384 considered for reproduction and their fitness is calculated according to Eq. 1. When $\gamma < 1$,

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

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

387 of the population can reproduce, and the fitness of the remaining 90% is set to zero.

Table 1
Parameter definitions and values

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

---

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

388   ***Reproduction phase.***   Finally, the researchers in the current population retire and

389   a new (non-overlapping) generation of researchers is created. A researcher in the new

390   generation inherits their publication strategy (submission threshold) s from a researcher in

391   the previous generation with the probability of the previous researcher's fitness (i.e., the new

392   generation's submission thresholds are sampled with replacement from the previous

393   generation, probability-weighted by fitness). The new generation's submission thresholds are

394   inherited with a small amount of random noise, such that $s_{new} = s_{old} + w$, with

395    $w \sim N(\mu = 0, \sigma = 0.01)$. Authors of similar evolutionary agent-based models have described

396   such heredity transmission as reflecting mentorship and teaching (e.g., when established

397   professors advise mentees to copy their strategies) or simply a generic social learning process

398   in which successful researchers are more likely to be imitated by others (Smaldino &

399   McElreath, 2016). Although this interpretation may be useful, the main purpose of this

400   aspect of the model is purely technical and not specifically intended to reflect reality — it

401   simply provides the machinery for determining which publication strategies are optimal in

402   the various situations we are investigating.

403   **Outcome variable s.**   We study how the evolution of researchers' submission

404   thresholds s is affected by the payoff for Registered Reports b_{RR} (relative to the payoffs for

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

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

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

408   parameters and their values considered in the simulation). A researcher's submission

409   threshold s is a *strategy*, not an absolute decision — it determines *how* the choice between

410   Registered Reports and standard reports is made, not which format is chosen. As such, s

411   indicates the amount of risk a researcher is willing to take. Very low values of s reflect risk

412   proneness: The researcher prefers to gamble and chooses the standard publication route for

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

414   that are virtually guaranteed to be false (and yield negative results). Very high values of s

reflect risk aversion: The researcher is unwilling to risk a negative result in a standard report and studies almost all hypotheses they encounter in the Registered Report format, reserving the standard publication route for hypotheses that are virtually guaranteed to be true (and yield positive results).

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

Simulation 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, their values do not represent real-world units. The goal of this analysis is to understand the relative effects of the model parameters in a simplified artificial system, which means that the results are only meaningful in relation to each other.

The first generation of researchers in each simulation run is initialised with randomly distributed submission thresholds s (drawn from a uniform distribution [0–1]), which are then allowed to evolve over the subsequent generations. Figure 3 shows the effect of varying the payoffs for Registered Reports when the fitness function is linear ($\epsilon = 1$), with no survival threshold ($\delta = 0$) 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
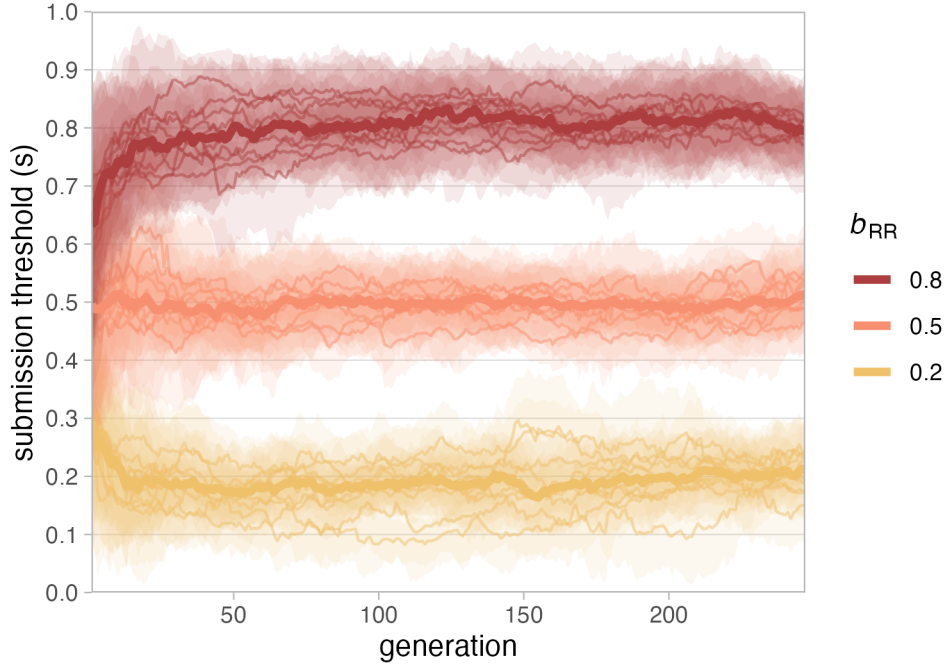
Figure 3. Evolution of submission threshold s with 3 different payoffs for Registered Reports (b_{RR}). Simulations are based on a population of $n = 500$ researchers over 250 generations, with payoffs for standard reports fixed at 0 for negative results ($b_{SR-} = 0$) and 1 for positive results ($b_{SR+} = 1$), 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.

441 Reports, the more popular they are. When b_{RR} is low, Registered Reports are unpopular

442 and only used for the least probable hypotheses; when b_{RR} is high, Registered Reports are

443 very popular and only hypotheses with extremely high priors are studied in standard reports.

444   In this very simple case illustrated here, evolved submission thresholds approximate

445 the payoff for Registered Reports in each condition, indicating that the optimal submission

446 threshold is always equal to b_{RR} ($s_{optimal} = 0.2$ when $b_{RR} = 0.2$, $s_{optimal} = 0.5$ when

447 $b_{RR} = 0.5$, $s_{optimal} = 0.8$ when $b_{RR} = 0.8$). The reason behind this is the uniform distribution

448 [0–1] of hypothesis priors and the payoff structure $b_{SR-} = 0$ and $b_{SR+} = 1$. In this

449 constellation, the expected payoff of a standard report is always equal to the prior of the

450 tested hypothesis:

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

451   For example, testing a hypothesis with $p = 0.2$ in a standard report would yield the

452   expected payoff $0.2*1 + 0.8*0 = 0.2$. The optimal strategy is to submit a Registered Report

453   whenever the expected payoff of a standard report is lower than the payoff for a Registered

454   Report, $E[b_{SR}] < b_{RR}$, and thus whenever $p < b_{RR}$. The strategy is optimal because it

455   ensures that researchers always get the best of both worlds, minimising shortfalls when

456   priors are (too) low and maximising winning chances when priors are (sufficiently) high. For

457   example, $b_{RR} = 0.5$ is larger than $E[b_{SR}]$ for all hypotheses with $p < 0.5$ but lower than

458    $E[b_{SR}]$ for all hypotheses with $p > 0.5$. In this situation, researchers who submit Registered

459   Reports whenever $p < 0.5$ and standard reports whenever $p > 0.5$ protect themselves against

460   losing a bad bet by instead taking the fixed payoff $b_{RR} = 0.5$, but always play a good bet and

461   thus maximise their chances of winning $b_{SR+} = 1$. Every alternative is inferior in the long

462   run because researchers with $s > b_{RR}$ lose out on increased chances of publishing a standard

463   report and researchers with $s < b_{RR}$ take unnecessary risks and go empty-handed too often.

464   **Non-linear fitness functions**

465   With this basic understanding of the payoff structure in hand, we can take a look at

466   what happens when payoffs have non-linear consequences for researchers' fitness. Figure 4

467   contrasts the effects of two diminishing fitness functions ($\epsilon = 0.2$ and $\epsilon = 0.5$, shown in blue

468   shades) and two increasing fitness functions ($\epsilon = 2$ and $\epsilon = 5$, shown in red shades) with a

469   linear function ($\epsilon = 1$, grey line) for different payoffs for Registered Reports. The grey line

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

471   function is linear, the optimal strategy is $s_{optimal} = b_{RR}$, making Registered Reports

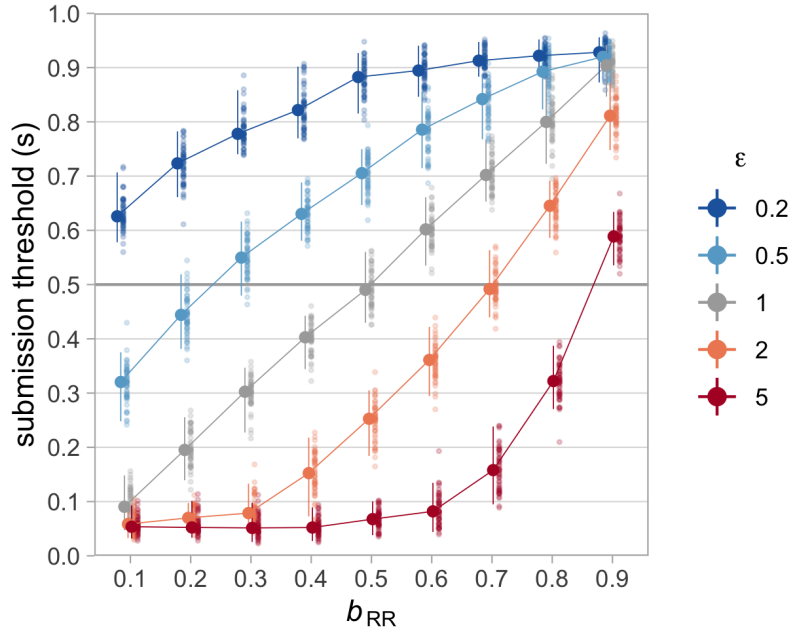472   relatively popular when they are worth more than 0.5 and relatively unpopular when they

Figure 4. Effect of fitness functions on evolved submission thresholds. Shown are median submission thresholds 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.

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
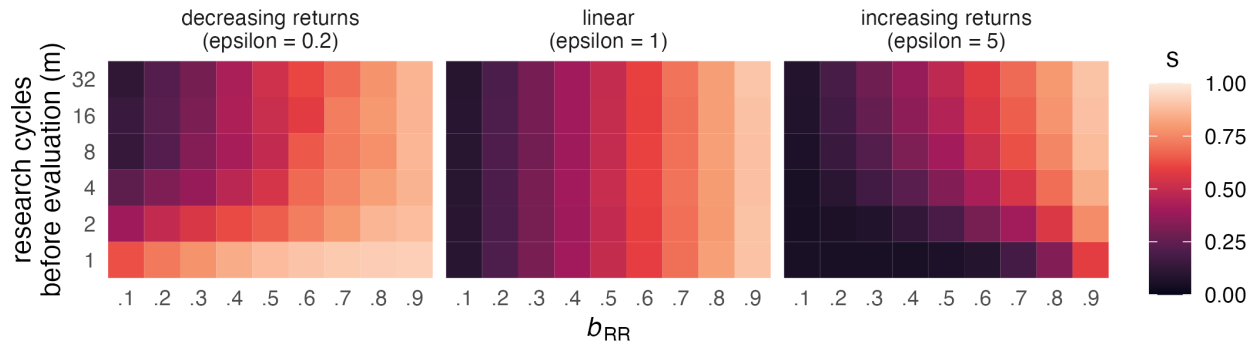
factors.

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

Number of research cycles 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 the

payoff from one single study. As discussed above, increasing numbers of decision events prior

499   to evaluation may make individuals more risk-prone because single negative outcomes are

500   less catastrophic for reproduction (Haaland et al., 2019). However, Figure 5 shows that the

501   effect of increasing numbers of research cycles per generation (m) interacts with the shape of

502   the fitness function: Moving up on the y-axis of each panel, we see that submission

503   thresholds are decreasing (indicating risk proneness) only in the top panel ($\epsilon = 0.2$) but stay

504   constant in the middle panel ($\epsilon = 1$) and even *increase* in the bottom panel ($\epsilon = 5$). Why

505   does m appear to have opposite effects for diminishing and increasing fitness functions? To

506   understand this pattern, it helps to first consider only the bottom row of each panel, where

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

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

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

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

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

512   approximate the linear case as m increases. This dynamic reflects the idea that fitness is

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

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

515   To use an illustrating example, consider two researchers with extreme submission

516   strategies: Emma conducts only Registered Reports ($s_{Emma} = 1$) and Gordon conducts only

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

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

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

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

521   In a population of Emmas and Gordons, lucky Gordons who got a positive result have a

522   narrow fitness advantage over all Emmas (1 versus 0.87), while unlucky Gordons lose to all

523   Emmas by a wide margin (0 versus 0.87). Since there are twice as many Emmas as lucky

524   Gordons, the Emma strategy is quite successful.

525 Now consider the same scenario with 4 research cycles per generation. Emmas receive

526 the same payoff in every round and accumulate $\frac{1}{2} * 4 = 2$. Lucky Gordons (who win every

527 time) accumulate a total payoff of $1 * 4 = 4$, while unlucky Gordons (who lose every time)

528 again receive 0 total payoff. Now, however, the probabilistic outcomes over 4 rounds lead to

529 more versions of Gordon, including average Gordons (who win half of the time and lose half

530 of the time) who accumulate the same total payoff as Emmas $(\frac{1}{2} * 0 + \frac{1}{2} * 1) * 4 = 2$. This

531 translates into fitness values of 0 for unlucky Gordons, $2^{\frac{1}{5}} = 1.15$ for Emmas and average

532 Gordons, and $4^{\frac{1}{5}} = 1.32$ for lucky Gordons. The Emma strategy still yields an enormous

533 advantage compared to unlucky Gordons and only a small disadvantage compared to lucky

534 Gordons. But this time, there are fewer Gordons who are less successful than Emmas

535 because Emmas now share their place with average Gordons, meaning that the relative

536 fitness advantage of the Emma strategy decreases. As the number of research cycles per

537 generation grows, the law of large numbers dictates that more Gordons achieve average total

538 payoffs and fewer Gordons achieve extreme total payoffs (winning 32 times in a row is much

539 less probable than winning 4 times in a row), which reduces the width of the Gordon

540 distribution until it approximates the Emma distribution.

541 When the fitness function is increasing ($\epsilon = 5$), the overall effect of increasing values of

542 m is identical, with the only difference that Emmas are initially disadvantaged (because their

543 fitness distance to the lucky half of Gordons is much greater than than to the unlucky

544 Gordons). With larger m, more and more Gordons receive average total payoffs and share

545 Emma's disadvantaged position (decreasing Emma's relative disadvantage), until the Gordon

546 distribution is again virtually equal to the Emma distribution. These results show that

547 rather than causing absolute risk aversion, increasing values of m simply swamp out the

548 effect of ϵ and reduce the effects of all fitness functions to the linear case. Consequently, the

549 top rows ($m = 32$) of the top and bottom panels in Figure 5 resemble the stable pattern

550 across all m shown in the middle panel.

551 Translated into terms of academic careers, this effect may indicate that senior

552 researchers — for whom new publications likely have diminishing returns — are more risk

553 prone than we might expect when only considering the fitness function, because academic

554 seniority also brings resources that boost research output per time. As a consequence,

555 established professors may be relatively indifferent to Registered Reports. Junior researchers,

556 for whom additional publications may have increasing returns on career success, may be

557 reluctant to use Registered Reports when they have very limited time or resources to

558 produce publications before an important selection event, such as postdocs on very short

559 contracts (R. Müller & de Rijcke, 2017).

Survival thresholds

561 When $\delta > b_{RR}$, Registered Reports alone are not sufficient to reach the survival

562 threshold (b_{RR} values to the left of the yellow line in Figure 6). For example, at $m = 4$, a

563 survival threshold of 75% ($\delta = .75$) means that researchers must gain at least 3 points to be

564 able to reproduce. When $b_{RR} = .7$, submitting four Registered Reports will only amount to

565 2.8 points in total, just short of meeting the threshold. On the other hand, when $b_{RR} = .8$

566 (i.e., just above δ), four RRs would yield 3.2 and thus ensure reproduction. Choosing the

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

568 of not meeting the survival threshold. As a consequence, one would expect Registered

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

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

571 survival thresholds have their biggest effect when the number of research cycles is low — at

572 high values of m, publication strategies are virtually unaffected in all conditions. Second,

573 survival thresholds have a stronger effect when the fitness function is linear ($\epsilon = 1$) or

574 concave ($\epsilon = 0.2$). In these two conditions, they produce very similar patterns: The RR

575 route is almost never chosen when b_{RR} is too low to meet the survival threshold (particularly

576 at $\delta = .25$ and $\delta = .5$; less so at $\delta = .75$), and this effect tapers off as the number of research
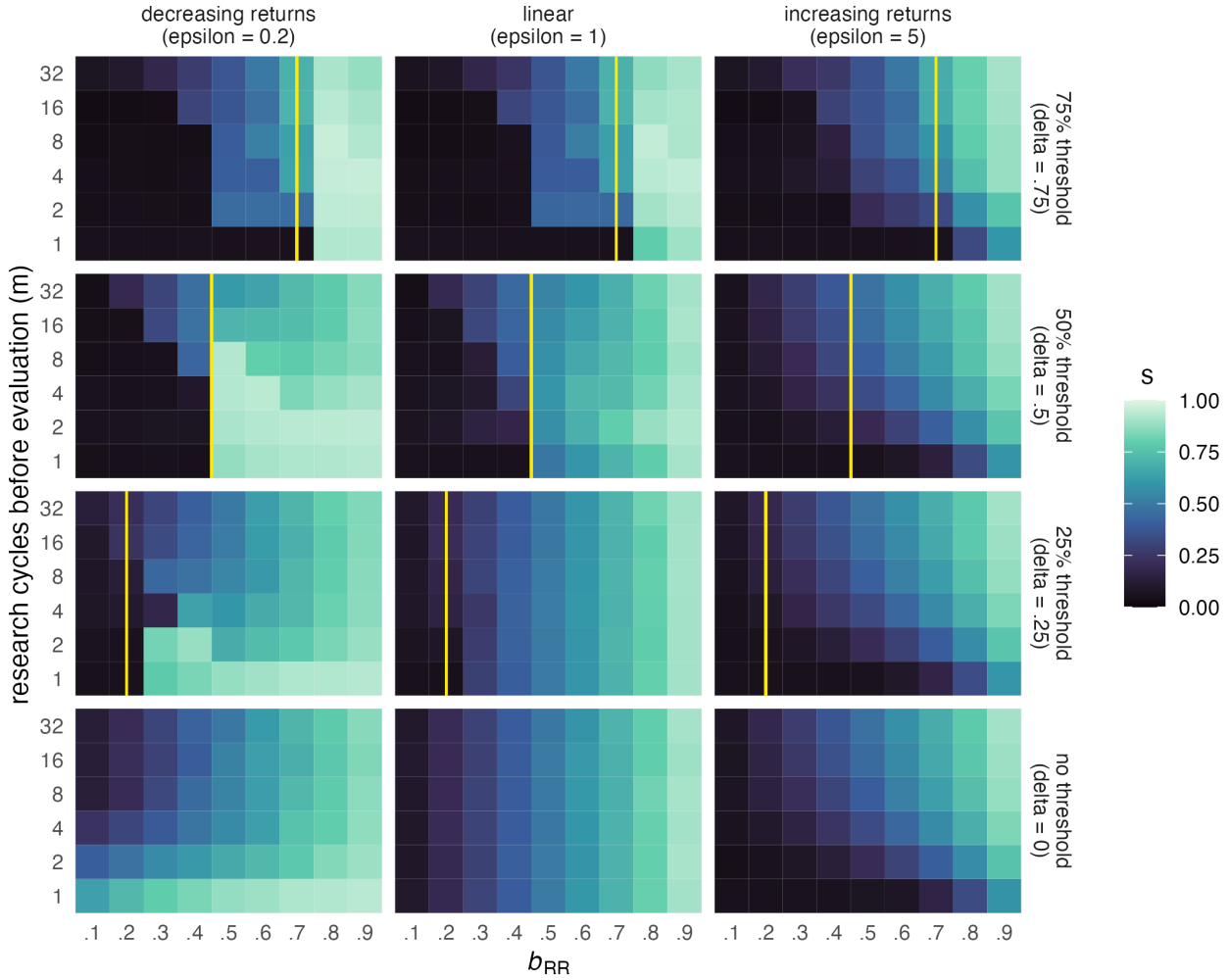
Figure 6. Effect of survival thresholds on evolved submission thresholds. Shown are median submission thresholds (s) after 250 generations in 50 runs (tile colour represents the median of 50 run medians) depending on survival thresholds (δ), fitness functions (characterised by exponent ϵ), numbers of 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 $\delta * m$.

577 cycles increases. Compared to baseline, the change is particularly striking for the concave

578 fitness function ($\epsilon = 0.2$, left column in Fig. 6), where RRs are normally preferred at low m.

579 When the survival threshold is high ($\delta = .75$) or the fitness function is concave, we can also

580 see that Registered Reports become *more* popular when they are worth just enough to pass

581 the survival threshold. For the convex fitness function ($\epsilon = 5$) on the other hand, survival

582 thresholds of 25% and 50% seem to have no effect at all. Only a high threshold of 75%

583 makes RRs even less popular when they have low value ($b_{RR} \leq 0.4$), especially when the

584 number of research cycles is low.

585   What does this mean in practice? The survival thresholds in our model represent

586 absolute publication targets that researchers must meet in order to progress in their career.

587 The clearest examples for such situations are PhD regulations and tenure agreements. To be

588 awarded with a PhD, many institutions and faculties require candidates to have a certain

589 number of their thesis chapters published in peer-reviewed journals. Similarly, tenure

590 agreements may include publication targets in the form of a minimum number of

591 peer-reviewed publications within a certain time, sometimes also specifying a minimum

592 journal rank (Liner & Sewell, 2009). The strictness of such requirements (e.g., the proportion

593 of thesis chapters that must be published) then determines whether they represent low,

594 medium, or high survival thresholds.

595   In our model, fitness (according to the three different fitness functions) is calculated

596 after the survival threshold has been met. This is meant to mimic publication requirements

597 that are expressed in raw numbers rather than some form of value. Importantly, it also

598 means that our simulation shows which strategies during a PhD or on the tenure track lead

599 to maximal fitness *after* researchers have successfully obtained their PhD or have been

600 granted tenure. With this in mind, it becomes easier to understand the meaning of the

601 different fitness functions. For PhD candidates, the fitness function is plausibly convex

602 ($\epsilon > 1$), as every additional publication to their (usually short) record may yield increasing

603   returns on the job market and when applying for grants. One notable exception are

604   candidates who do not intend to stay in academia, and for whose careers publications will

605   not be a meaningful currency — here, publications beyond the survival threshold would

606   instead yield strongly decreasing returns (concave fitness function, $\epsilon < 1$). For researchers on

607   the tenure track, the fitness function after achieving tenure may also be concave, assuming a)

608   that achieving tenure is one of the most important career goals for many (making further

609   progress less important) and b) that such individuals have already built up substantial

610   publication records, to which any single addition makes less and less of a difference. However,

611   exceptions from this scenario may well exist, for example in situations where even tenured

612   researchers are under great pressure to compete for grants.

613   Translated to real-world scenarios, our results thus suggest the following implications:

614   First, survival thresholds are almost irrelevant when the empirical pace of a research area

615   (the number of studies that can be completed in a given amount of time) is high. Second,

616   researchers with a convex fitness function — such as PhD candidates who are pursuing an

617   academic career — are only affected by high survival thresholds, which lead them to choose

618   Registered Reports even less often than normal when their value is low. Third, researchers

619   with a concave fitness function — such as tenure candidates or PhD students who pursue

620   careers outside of academia — are highly sensitive to the value of Registered Reports: They

621   virtually never conduct Registered Reports when their value is too low for meeting the

622   survival threshold, but strongly prefer them when their value is sufficient (especially when

623   empirical pace is low and/or the survival threshold is high).

Competition

625   Figure 7 shows that competition generally leads to an aversion of Registered Reports,

626   as can be seen by the darkening of the plots when moving up from the bottom row of panels.

627   The only exception to this rule is very low competition: When the top 90% are allowed to

628   reproduce (and only the bottom 10% are rejected, $\gamma = .9$), Registered Reports become more
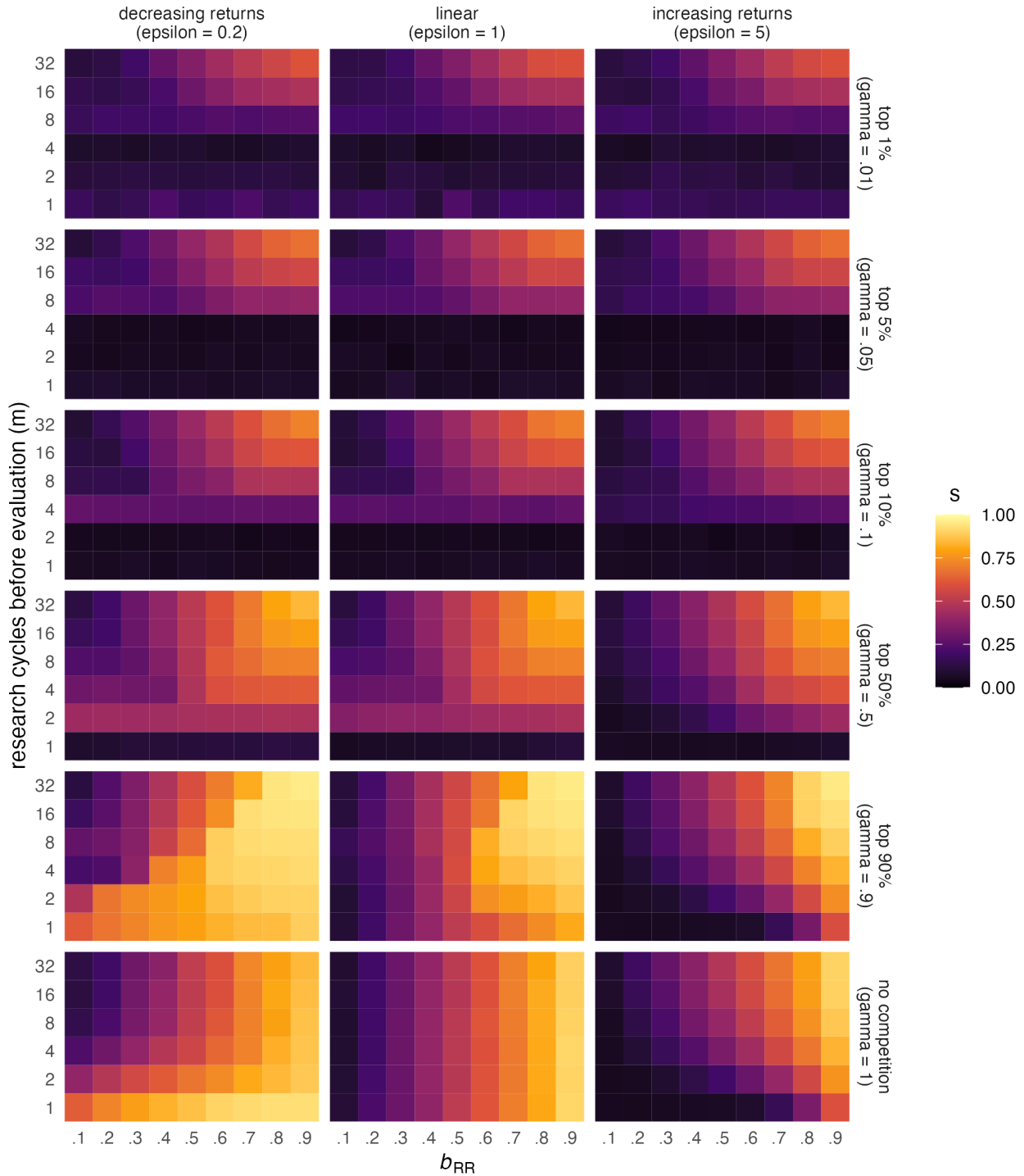
Figure 7. Effect of competition on evolved submission thresholds. Shown are median evolved submission thresholds (s) after 250 generations in 50 runs (tile colour represents the median of 50 run medians) depending on the intensity of competition (γ, y-axis), numbers of research cycles per generation (m), different values of b_{RR} (x-axis), and different fitness functions (characterised by exponent ϵ) with no survival threshold ($\delta = 0$). To reproduce, researchers must accumulate a total payoff in the top γ proportion of the population.

629 popular than they are in the absence of competition. This effect is strongest for the concave

630 fitness funcation ($\epsilon = 0.2$), where it holds for almost all values of b_{RR} at very low numbers of

631 m and for high values of b_{RR} at high numbers of m. When the fitness function is linear or

632 convex, Registered Reports are chosen more often only when both b_{RR} and m are high. At

633 higher levels of competition ($\gamma > .5$), the differences between the fitness functions disappear.

634 In all three cases, Registered Reports are essentially wiped out for low numbers of research

635 cycles (m), and this effect increases with competition (the higher the competition, the higher

636 m must be for Registered Reports to still be viable). At high numbers of m, more

637 competition also negatively affects Registered Reports, but here the baseline pattern (a

638 linear increase of RR popularity with b_{RR}) remains intact.

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

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

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

642 lighter. This paradoxial phenomenon is not due to Registered Reports being more lucrative

643 in those conditions. Rather, competition is so extreme that the natural selection in our model

644 starts operating more on chance than on individuals' traits. Essentially, only individuals

645 with the maximum possible payoff (publishing only standard reports with positive results)

646 are able to reproduce. Most likely to receive this maximum payoff are individuals who

647 investigate hypotheses with high prior probabilities. In our model, this is not a trait that can

648 be passed on, but determined by random chance. Among individuals who experience this

649 kind of luck, the variance of publication strategy s should be high: A hypothesis with prior

650 $p = .95$ will be submitted as a standard report and likely yield a positive result (and thus the

651 maximum payoff) regardless of whether the researcher's submission threshold is as low as

652 $s = .1$ or has high as $s = .9$. The higher average s at low m under extreme competition thus

653 reflects relaxed selection pressure on s. This is also evident by the shades of the dark bar at

654 the bottom of the panels for $\gamma = .01$ Figure 7 fluctuating randomly for each level of m rather

655 than showing a specific pattern. A clearer illustration of the effect can be found in Figure

XXX in the appendix, which shows the (large increases in) variance of evolved submission thresholds in addition to the medians. At higher m, selection on s stays intact simply because much fewer individuals will be very lucky 8, 16, or 32 times in a row than 4 or fewer times in a row, and the submission strategy thus remains a relevant 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 (Figure XXX shows that although the variance of evolved s increases dramatically with high competition, it never spans the entire range of s), but it can significantly slow it. XXX TRY LONGER SIM RUNS & REPORT HERE XXX The phenomenon is related to one form of survivorship bias: Looking at 'survivors' of a highly selective process, one may erroneously infer that specific observable traits or behaviours of such individuals were the cause of their success when they were actually merely coincidental.

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 high competition may be a significant threat for the viability of Registered Reports, regardless of career stage. This effect is particularly extreme in fields with low empirical pace, where submission strategies that involve any amount of Registered Reports are only viable when competition is so high that success requires extraordinary luck. In contrast, very low but non-zero levels of competition increase the popularity of Registered Reports, especially when their value is high, when the fitness function is concave (e.g., in later career stages), and in fields with high empirical pace.

Discussion

To do:

681 • Brief recap of results

682 • Implications of results

683 – cautious mapping of model factors to real-world situations

684 – potential implications for meta-science

685 – potential implications for policy

Limitations
686

687 • Narrow focus on one specific (and highly stylised) difference between Registered

688 Reports and standard reports; there are many others. Model ignores a myriad other

689 factors that influences who chooses Registered Reports for which studies when

690 • Concept of publication bias as filtering positive results of hypothesis tests (and the

691 respective connection to hypothesis priors such that high priors –> better) is

692 cartoonish and not entirely accurate for the simple reason that positive results of

693 trivial (or otherwise boring) hypotheses are usually not highly valued (also, this

694 approach only focuses on hypothesis testing, which is widely used in psychology but by

695 far not the only means of doing science). A more valid solution may be the concept of

696 publication bias as favouring belief-shifting results presented by Gross & Bergstrom

697 (2021). Adapting the model presented here to capture this concept of bias could be an

698 interesting future direction. However, the present version of the model also allows a

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

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

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

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

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

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

705 • Fitness concept: one caveat is that

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

707   doesn't take into account

Future directions

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

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

711   be influenced by individual differences, so that individuals with traits or abilities that

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

713   chances or the payoff when winning, or that buffer losses) should be more risk-prone

714   (Barclay, Mishra, & Sparks, 2018). Such factors may be important to consider in the context

715   of research and publication practices. For example, researchers who are better at choosing

716   research questions that are likely to result in high-impact publications (e.g., through talent

717   or experience) may find Registered Reports less attractive. As a more nefarious version of

718   this idea, Registered Reports may be relatively unpopular among researchers who are more

719   willing or able to use questionable research practices (or even fraud) to obtain publishable or

720   impactful results.

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

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

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

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

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

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

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

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

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

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

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

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

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

734 investigation in the future.

735 **Conclusion**

736 **Disclosures**

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

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

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

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

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

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

743   **References**

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

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

746   *12*(3), e0172792. https://doi.org/10.1371/journal.pone.0172792

747 Allen, C., & Mehler, D. M. A. (2019). Open science challenges, benefits and tips in early

748   career and beyond. *PLOS Biology*, *17*(5), e3000246.

749   https://doi.org/10.1371/journal.pbio.3000246

750 Atkinson, D. R., Furlong, M. J., & Wampold, B. E. (1982). Statistical significance,

751   reviewer evaluations, and the scientific process: Is there a (statistically) significant

752   relationship? *Journal of Counseling Psychology*, *29*(2), 189–194.

753   https://doi.org/10.1037/0022-0167.29.2.189

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

755 Barclay, P., Mishra, S., & Sparks, A. M. (2018). State-dependent risk-taking. *Proceedings*

756   *of the Royal Society B: Biological Sciences*, *285*(1881), 20180180.

757   https://doi.org/10.1098/rspb.2018.0180

Barth, M. (2022). *tinylabels: Lightweight variable labels.* Retrieved from

https://cran.r-project.org/package=tinylabels

Chalmers, I., & Glasziou, P. (2009). Avoidable waste in the production and reporting of

research evidence. *The Lancet, 374*(9683), 86–89.

https://doi.org/10.1016/S0140-6736(09)60329-9

Chambers, C. D. (2013). Registered reports: A new publishing initiative at Cortex.

Cortex, 49, 606–610. https://doi.org/10.1016/j.cortex.2012.12.016

Chambers, C. D., & Tzavella, L. (2021). The past, present and future of Registered

Reports. *Nature Human Behaviour,* 1–14. https://doi.org/10.1038/s41562-021-01193-7

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

Non-Significant Results: Does It Apply to Biological Research? *Oikos, 76*(3), 591–593.

https://doi.org/10.2307/3546355

de Vries, Y. A., Roest, A. M., Jonge, P. de, Cuijpers, P., Munafò, M. R., & Bastiaansen,

J. A. (2018). The cumulative effect of reporting and citation biases on the apparent

efficacy of treatments: The case of depression. *Psychological Medicine, 48*(15),

2453–2455. https://doi.org/10.1017/S0033291718001873

Dickersin, K., & Min, Y. I. (1993). Publication bias: The problem that won't go away.

Annals of the New York Academy of Sciences, 703, 135-146; discussion 146-148.

https://doi.org/10.1111/j.1749-6632.1993.tb26343.x

Fanelli, D. (2010). "Positive" results increase down the hierarchy of the sciences. *PLoS*

ONE, 5(4), e10068. https://doi.org/10.1371/journal.pone.0010068

Ferguson, C. J., & Heene, M. (2012). A Vast Graveyard of Undead Theories: Publication

Bias and Psychological Science's Aversion to the Null. *Perspectives on Psychological*

Science, 7(6), 555–561. https://doi.org/10.1177/1745691612459059

Fiedler, K., & Schwarz, N. (2016). Questionable Research Practices Revisited. *Social*

Psychological and Personality Science, 7(1), 45–52.

https://doi.org/10.1177/1948550615612150

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

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

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

Greenwald, A. G. (1975). Consequences of Prejudice Against the Null Hypothesis. *Psychological Bulletin*, *82*(1), 1–20.

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

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

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

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

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

Kacelnik, A., & Bateson, M. (1997). Risk-sensitivity: Crossroads for theories of

decision-making. *Trends in Cognitive Sciences*, *1*(8), 304–309.

https://doi.org/10.1016/s1364-6613(97)01093-0

Lakens, D. (2019). The value of preregistration for psychological science: A conceptual

analysis. *Japanese Psychological Review*, *62*(3), 221–230.

https://doi.org/10.24602/sjpr.62.3_221

Liner, G. H., & Sewell, E. (2009). Research requirements for promotion and tenure at

PhD granting departments of economics. *Applied Economics Letters*.

https://doi.org/10.1080/13504850701221998

Mahoney, M. J. (1977). Publication Prejudices: An Experimental Study of Confirmatory

Bias in the Peer Review System. *Cognitive Therapy and Research*, *1*(2), 161–175.

https://doi.org/10.1007/BF01173636

Mishra, S. (2014). Decision-Making Under Risk: Integrating Perspectives From Biology,

Economics, and Psychology. *Personality and Social Psychology Review*, *18*(3),

280–307. https://doi.org/10.1177/1088868314530517

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

Müller, R. (2014). Postdoctoral Life Scientists and Supervision Work in the

Contemporary University: A Case Study of Changes in the Cultural Norms of Science.

Minerva, *52*(3), 329–349. https://doi.org/10.1007/s11024-014-9257-y

Müller, R., & de Rijcke, S. (2017). Thinking with indicators. Exploring the epistemic

impacts of academic performance indicators in the life sciences. *Research Evaluation*,

26(3), 157–168. https://doi.org/10.1093/reseval/rvx023

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The

preregistration revolution. *Proceedings of the National Academy of Sciences*, *115*(11),

2600–2606. https://doi.org/10.1073/pnas.1708274114

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., . . .

Vazire, S. (2022). Replicability, Robustness, and Reproducibility in Psychological

Science. *Annual Review of Psychology*, *73*(1), annurev-psych-020821-114157.

https://doi.org/10.1146/annurev-psych-020821-114157

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

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

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

Scheel, A. M., Schijen, M. R. M. J., & Lakens, D. (2021). An Excess of Positive Results: Comparing the Standard Psychology Literature With Registered Reports. *Advances in Methods and Practices in Psychological Science, 4*(2), 251524592110074. https://doi.org/10.1177/25152459211007467

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

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

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

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

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

https://doi.org/10.1080/01621459.1959.10501497

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

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

Wacholder, S., Chanock, S., Garcia-Closas, M., El ghormli, L., & Rothman, N. (2004). Assessing the Probability That a Positive Report is False: An Approach for Molecular Epidemiology Studies. *JNCI Journal of the National Cancer Institute*, *96*(6), 434–442. https://doi.org/10.1093/jnci/djh075

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

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

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

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

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

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

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

894   Raton, Florida: Chapman and Hall/CRC.