

1 **Why does preregistration increase the persuasiveness of evidence? A Bayesian**  
2 **rationalization**

3 Aaron Peikert<sup>1,2,3</sup>, Maximilian S. Ernst<sup>1, 4</sup>, and & Andreas M. Brandmaier<sup>1, 3, 5</sup>

4 <sup>1</sup> Center for Lifespan Psychology

5 Max Planck Institute for Human Development

6 Berlin

7 Germany

8 <sup>2</sup> Department of Imaging Neuroscience

9 University College London

10 London

11 UK

12 <sup>3</sup> Max Planck UCL Centre for Computational Psychiatry and Ageing Research

13 Berlin

14 Germany

15 <sup>4</sup> Max Planck School of Cognition

16 Leipzig

17 Germany

18 <sup>5</sup> Department of Psychology

19 MSB Medical School Berlin

20 Berlin

21 Germany

22 The materials for this article are available on [GitHub](#) (Peikert & Brandmaier, 2023a). This  
23 version was created from git commit 3482d85. The manuscript is available as [preprint](#)  
24 (Peikert & Brandmaier, 2023b).

25 Submitted to *Meta-Psychology*. Participate in open peer review by sending an email to  
26 [open.peer.reviewer@gmail.com](mailto:open.peer.reviewer@gmail.com). The full editorial process of all articles under review at  
27 Meta-Psychology can be found following this link:

28 <https://tinyurl.com/mp-submissions>

29 You will find this preprint by searching for the first author's name.

### 30 Author Note

31  
32 The authors made the following contributions. Aaron Peikert: Conceptualization,  
33 Writing—Original Draft Preparation, Writing—Review & Editing, Methodology, Formal  
34 analysis, Software, Visualization, Project administration; Maximilian S. Ernst:  
35 Writing—Review & Editing, Formal analysis, Validation; Andreas M. Brandmaier:  
36 Writing—Review & Editing, Supervisions.

37 Correspondence concerning this article should be addressed to Aaron Peikert,  
38 Center for Lifespan Psychology, Max Planck Institute for Human Development, Lentzeallee  
39 94, 14195 Berlin, Germany. E-mail: [peikert@mpib-berlin.mpg.de](mailto:peikert@mpib-berlin.mpg.de)

**Abstract**

40

41 The replication crisis has led many researchers to preregister their hypotheses and data  
42 analysis plans before collecting data. A widely held view is that preregistration is supposed  
43 to limit the extent to which data may influence the hypotheses to be tested. Only if data  
44 have no influence an analysis is considered confirmatory. Consequently, many researchers  
45 believe that preregistration is only applicable in confirmatory paradigms. In practice,  
46 researchers may struggle to preregister their hypotheses because of vague theories that  
47 necessitate data-dependent decisions (aka exploration). We argue that preregistration  
48 benefits any study on the continuum between confirmatory and exploratory research. To  
49 that end, we formalize a general objective of preregistration and demonstrate that  
50 exploratory studies also benefit from preregistration. Drawing on Bayesian philosophy of  
51 science, we argue that preregistration should primarily aim to reduce uncertainty about the  
52 inferential procedure used to derive results. This approach provides a principled  
53 justification of preregistration, separating the procedure from the goal of ensuring strictly  
54 confirmatory research. We acknowledge that knowing the extent to which a study is  
55 exploratory is central, but certainty about the inferential procedure is a prerequisite for  
56 persuasive evidence. Finally, we discuss the implications of these insights for the practice of  
57 preregistration.

58 *Keywords:* preregistration; confirmation; exploration; hypothesis testing; Bayesian;

59 Open Science

60 Word count: 8390

61 **Why does preregistration increase the persuasiveness of evidence? A Bayesian**  
62 **rationalization**

63 The scientific community has long pondered the vital distinction between  
64 exploration and confirmation, discovery and justification, hypothesis generation and  
65 hypothesis testing, or prediction and postdiction (Hoyningen-Huene, 2006; Nosek et al.,  
66 2018; Shmueli, 2010; Tukey, 1980). Despite the different names, it is fundamentally the  
67 same dichotomy that is at stake here. There is a broad consensus that both approaches are  
68 necessary for science to progress; exploration, to make new discoveries and confirmation, to  
69 expose these discoveries to potential falsification, and assess empirical support for the  
70 theory. However, mistaking exploratory findings for empirically confirmed results is  
71 dangerous. It inflates the likelihood of believing that there is evidence supporting a given  
72 hypothesis, even if it is false. A variety of problems, such as researchers' degrees of freedom  
73 together with researchers' hindsight bias or naive p-hacking have led to such mistakes  
74 becoming commonplace yet unnoticed for a long time. Recognizing them has led to a crisis  
75 of confidence in the empirical sciences (Ioannidis, 2005), and psychology in particular  
76 (Open Science Collaboration, 2015). As a response to the crisis, evermore researchers  
77 preregister their hypotheses and their data collection and analysis plans in advance of their  
78 studies (Nosek et al., 2018). They do so to stress the predictive nature of their registered  
79 statistical analyses, often with the hopes of obtaining a label that marks the study as  
80 "confirmatory". Indeed, rigorous application of preregistration prevents researchers from  
81 reporting a set of results produced by an arduous process of trial and error as a simple  
82 confirmatory story (Wagenmakers et al., 2012) while keeping low false-positive rates. This  
83 promise of a clear distinction between confirmation and exploration has obvious appeal to  
84 many who have already accepted the practice. Still, the majority of empirical researchers  
85 do not routinely preregister their studies. One reason may be that some do not find that  
86 the theoretical advantages outweigh the practical hurdles, such as specifying every aspect of  
87 a theory and the corresponding analysis in advance. We believe that we can reach a greater

88 acceptance of preregistration by explicating a more general objective of preregistration that  
89 benefits all kinds of studies, even those that allow data-dependent decisions.

90 One goal of preregistration that has received widespread attention is to clearly  
91 distinguish confirmatory from exploratory research (Bakker et al., 2020; Mellor & Nosek,  
92 2018; Nosek et al., 2018; Simmons et al., 2021; Wagenmakers et al., 2012). In such a  
93 narrative, preregistration is justified by a confirmatory research agenda. However, two  
94 problems become apparent under closer inspection. First, many researchers do not  
95 subscribe to a purely confirmatory research agenda (Baumeister, 2016; Brandmaier et al.,  
96 2013; Finkel et al., 2017; Tukey, 1972). Second, there is no strict mapping of the categories  
97 preregistered vs. non-preregistered onto the categories confirmatory vs. exploratory  
98 research.

99 Obviously, researchers can conduct confirmatory research without preregistration —  
100 though it might be difficult to convince other researchers of the confirmatory nature of  
101 their research, that is, that they were free of cognitive biases, made no data-dependent  
102 decisions, and so forth. The opposite, that is, preregistered but not strictly confirmatory  
103 studies, are also becoming more commonplace (Chan et al., 2004; Dwan et al., 2008; Silagy  
104 et al., 2002).

105 This is the result of researchers applying one of two strategies to evade the  
106 self-imposed restrictions of preregistrations: writing a loose preregistration to begin with  
107 (Stefan & Schönbrodt, 2023) or deviating from the preregistration afterward. The latter is  
108 a frequent occurrence and, perhaps more worryingly, often remains undisclosed (Akker et  
109 al., 2023; Claesen et al., 2021). Both strategies may be used for sensible scientific reasons  
110 or with the self-serving intent of generating desirable results. Thus, insisting on equating  
111 preregistration and confirmation has led to the criticism that, all things considered,  
112 preregistration is actually harmful and neither sufficient nor necessary for doing good  
113 science (Pham & Oh, 2021; Szollosi et al., 2020).

114 We argue that such criticism is not directed against preregistration itself but against  
115 a justification through a confirmatory research agenda (Wagenmakers et al., 2012). When  
116 researchers criticize preregistration as being too inflexible to fit their research question,  
117 they often simply acknowledge that their research goals are not strictly confirmatory.  
118 Forcing researchers into adopting a strictly confirmatory research agenda does not only  
119 imply changing *how* they investigate a phenomenon but also *what* research questions they  
120 pose. However reasonable such a move is, changing the core beliefs of a large community is  
121 much harder than convincing them that a method is well justified. We, therefore, attempt  
122 to disentangle the *methodological* goals of preregistration from the *ideological* goals of  
123 confirmatory science. It might well be the case that psychology needs more confirmatory  
124 studies to progress as a science. However, independently of such a goal, preregistration can  
125 be useful for any kind of study on the continuum between strictly confirmatory and fully  
126 exploratory.

127 To form such an objective for preregistration, we first introduce some tools of  
128 Bayesian philosophy of science and map the exploration/confirmation distinction onto a  
129 dimensional quantity we call “theoretical risk” (a term borrowed from Meehl, 1978, but  
130 formalized as the probability of proving a hypothesis wrong if it does not hold).

131 We are interested in why preregistrations should change researchers’ evaluation of  
132 evidence. Applying a Bayesian framework allows us to investigate our research question  
133 most straightforwardly because it directly deals with what we ought to believe, given the  
134 evidence presented. Specifically, it allows us to model changes in subjective degrees of  
135 belief due to preregistration or, more simply, “persuasion”. Please note that our decision to  
136 adopt a Bayesian philosophy of science does not make assumptions about the statistical  
137 methods researchers use. In fact, this conceptualization is intentionally as minimal as  
138 possible to be compatible with a wide range of philosophies of science and statistical  
139 methods researchers might subscribe to. One feature of the Bayesian framework, is the

140 strong emphasis on subjective yet rational judgement. Therefore, we assume that  
141 researchers will differ significantly in how they value evidence but that by making  
142 assumptions about the general process, we can make general statements that apply to all  
143 these subjective evaluations. However, we should note that Popperians would be appalled  
144 that we are content with positive inductive inferences (but we regard “failing to disprove”  
145 as too limited), and Neopopperians would flinch that we assign probabilities to beliefs (we  
146 are fond of calculating things). While the latter move is not strictly necessary it allows us  
147 to connect the more abstract considerations more closely with what researchers believe.

148         Now, we outline two possible perspectives on the utility of preregistration. The first  
149 one corresponds to the traditional application of preregistration to research paradigms that  
150 focus on confirmation by maximizing the theoretical risk or, equivalently, by limiting type-I  
151 error (when dichotomous decisions about theories are an inferential goal). We argue that  
152 this view on the utility of preregistration can be interpreted as maximizing theoretical risk,  
153 which otherwise may be reduced by researchers’ degrees of freedom, p-hacking, and suchlike.  
154 The second interpretation is our main contribution: We argue that contrary to the classic  
155 view, the objective of preregistration is *not* the maximization of theoretical risk but rather  
156 the minimization of uncertainty about the theoretical risk. This interpretation leads to a  
157 broad applicability of preregistration to both exploratory and confirmatory studies.

158         To arrive at this interpretation, we rely on three arguments. The first is that  
159 theoretical risk is vital for judging evidential support for theories. The second argument is  
160 that the theoretical risk for a given study is generally uncertain. The third and last  
161 argument is that this uncertainty is reduced by applying preregistration. We conclude that  
162 because preregistration decreases uncertainty about the theoretical risk, which in turn  
163 increases the amount of knowledge we gain from a particular study, preregistration is  
164 potentially useful for any kind of study, no matter where it falls on the  
165 exploratory-confirmatory continuum.

### Epistemic value and the Bayesian rationale

166

167 Let us start by defining what we call expected epistemic value. If researchers plan  
168 to conduct a study, they usually hope that it will change their assessment of some theory's  
169 verisimilitude (Niiniluoto, 1998). In other words, they hope to learn something from  
170 conducting the study. The amount of knowledge researchers gain from a particular study  
171 concerning the verisimilitude of a specific theory is what we call epistemic value.  
172 Researchers cannot know what exactly they will learn from a study before they run it.  
173 However, they can develop an expectation that helps them decide about the specifics of a  
174 planned study. This expectation is what we term expected epistemic value. To make our  
175 three arguments, we must assume three things about what an ideal estimation process  
176 entails and how it relates to what studies (preregistered vs not preregistered) to conduct.

177

1. Researchers judge the evidence for or against a hypothesis rationally.

178

2. They expect other researchers to apply a similar rational process.

179

3. Researchers try to maximize the expected epistemic value for other researchers.

180

181 The assumption of rationality can be connected to Bayesian reasoning and leads to  
182 our adoption of the framework. Our rationale is as follows. Researchers who decide to  
183 conduct a certain study are actually choosing a study to bet on. They have to “place the  
184 bet” by conducting the study by investing resources and stand to gain epistemic value with  
185 some probability. This conceptualization of choosing a study as a betting problem allows  
186 us to apply a “Dutch book” argument (Christensen, 1991). This argument states that any  
187 better must follow the axioms of probability to avoid being “irrational,” i.e., accepting bets  
188 that lead to sure losses. Fully developing a Dutch book argument for this problem requires  
189 careful consideration of what kind of studies to include as possible bets, defining a  
190 conversion rate from the stakes to the reward, and modeling what liberties researchers have  
191 in what studies to conduct. Without deliberating these concepts further, we find it  
persuasive that researchers should not violate the axioms of probability if they have some



192 expectation about what they stand to gain with some likelihood from conducting a study.  
193 The axioms of probability are sufficient to derive the Bayes formula, on which we will  
194 heavily rely for our further arguments. The argument is not sufficient, however, to warrant  
195 conceptualizing the kind of epistemic value we reason about in terms of posterior  
196 probability; that remains a leap of faith. However, the argument applies to any reward  
197 function that satisfies the “statistical relevancy condition” (Fetzer, 1974; Salmon, 1970),  
198 that is, evidence only increases epistemic value for a theory if the evidence is more likely to  
199 be observed under the theory than under the alternative. In particular, “diagnosticity”  
200 (Fiedler, 2017; Oberauer & Lewandowsky, 2019), a concept highlighted in recent  
201 psychological literature, seems to adhere to the statistical relevancy condition.

### 202 **Epistemic value and theoretical risk**

203 Our first argument is that theoretical risk is crucial for judging evidential support  
204 for theories. Put simply, risky predictions create persuasive evidence if they turn out to be  
205 correct. This point is crucial because we attribute much of the appeal of a confirmatory  
206 research agenda to this notion.

207 Let us make some simplifying assumptions and define our notation. To keep the  
208 notation simple, we restrict ourselves to evidence of a binary nature (either it was observed  
209 or not). We denote the probability of a hypothesis before observing evidence as  $P(H)$  and  
210 its complement as  $P(\neg H) = 1 - P(H)$ . The probability of observing evidence under some  
211 hypothesis is  $P(E|H)$ . We can calculate the probability of the hypothesis after observing  
212 the evidence with the help of the Bayes formula:

$$P(H|E) = \frac{P(H)P(E|H)}{P(E)} \quad (1)$$

213 The posterior probability  $P(H|E)$  is of great relevance since it is often used directly  
214 or indirectly as a measure of confirmation of a hypothesis. In the tradition of Carnap, in its

215 direct use, it is called *confirmation as firmness*; in its relation to the a priori probability  
216  $P(H)$ , it is called *increase in firmness* (Carnap, 1950, preface to the 1962 edition). We  
217 concentrate on the posterior probability because of its simplicity but take it only as one  
218 example of a possible measure. In reality, researchers surely differ in what function they  
219 apply to judge evidence and it is often most fruitful to compare more than two competing  
220 hypotheses. The goal is therefore to reason about the space of possible measures  
221 researchers might apply. However, since any measure fulfilling the statistical relevancy  
222 condition increases monotonically with an increase in posterior probability  $P(H|E)$ , we  
223 might well take it to illustrate our reasoning.

224 In short, we want to increase posterior probability  $P(H|E)$ . Increases in posterior  
225 probability  $P(H|E)$  are associated with increased epistemic value, of which we want to  
226 maximize the expectation. So how can we increase posterior probability? The Bayes  
227 formula yields three components that influence confirmation, namely  $P(H)$ ,  $P(E|H)$  and  
228  $P(E)$ . The first option leads us to the unsurprising conclusion that higher a priori  
229 probability  $P(H)$  leads to higher posterior probability  $P(H|E)$ . If a hypothesis is more  
230 probable to begin with, observing evidence in its favor will result in a hypothesis that is  
231 more strongly confirmed, all else being equal. However, the prior probability of a  
232 hypothesis is nothing our study design can change. The second option is equally  
233 reasonable; that is, an increase in  $P(E|H)$  leads to a higher posterior probability  $P(H|E)$ .  
234  $P(E|H)$  is the probability of obtaining evidence for a hypothesis when it holds. We call  
235 this probability of detecting evidence, given that the hypothesis holds “detectability.”  
236 Consequently, researchers should ensure that their study design allows them to find  
237 evidence for their hypothesis, in case it is true. When applied strictly within the bounds of  
238 null hypothesis testing, detectability is equivalent to power (or the complement of type-II  
239 error rate). However, while detectability is of great importance for study design, it is not  
240 directly relevant to what a preregistration is communicating to other researchers. We later  
241 discuss how issues of detectability must be considered in a preregistration. Thus,  $P(E)$

242 remains to be considered. Since  $P(E)$  is the denominator, decreasing it can increase the  
 243 posterior probability. In other words, high risk, high reward.

244 If we equate riskiness with a low probability of obtaining evidence (when the  
 245 hypothesis is false), the Bayesian rationale perfectly aligns with the observation that risky  
 246 predictions lead to persuasive evidence. This tension between high risk leading to high gain  
 247 is central to our consideration of preregistration. A high-risk, high-gain strategy is bound  
 248 to result in many losses that are eventually absorbed by the high gains. Sustaining many  
 249 “failed” studies is not exactly aligned with the incentive structure under which many, if not  
 250 most, researchers operate. Consequently, researchers are incentivized to appear to take  
 251 more risks than they actually do, which misleads their readers to give their claims more  
 252 credence than they deserve. It is at this juncture that the practice and mispractice of  
 253 preregistration comes into play. We argue that the main function of preregistration is to  
 254 enable proper judgment of the riskiness of a study.

255 To better understand how preregistrations can achieve that, let us take a closer look  
 256 at the factors contributing to  $P(E)$ . Using the law of total probability, we can split  $P(E)$   
 257 into two terms:

$$P(E) = P(H)P(E|H) + P(\neg H)P(E|\neg H) \quad (2)$$

258 We have already noted that there is not much to be done about prior probability  
 259 ( $P(H)$ , and hence its counter probability  $P(\neg H)$ ), and that it is common sense to increase  
 260 detectability  $P(E|H)$ . The real lever to pull is therefore  $P(E|\neg H)$ . This probability tells  
 261 us how likely it is that we find evidence in favor of the theory when in fact, the theory is  
 262 not true. Its counter probability  $P(\neg E|\neg H) = 1 - P(E|\neg H)$  is what we call “theoretical  
 263 risk”, because it is the risk a theory takes on in predicting the occurrence of particular  
 264 evidence in its favor. We borrow the term from Meehl (1978), though he has not assigned

265 it to the probability  $P(\neg E|\neg H)$ . Kukla (1990) argued that the core arguments in Meehl  
266 (1990) can be reconstructed in a purely Bayesian framework. However, while he did not  
267 mention  $P(\neg E|\neg H)$  he suggested that Meehl (1978) used the term “very strange  
268 coincidence” for a small  $P(E|\neg H)$  which would imply, that  $P(\neg E|\neg H)$  can be related to or  
269 even equated to theoretical risk.

270 Let us note some interesting properties of theoretical risk  $P(\neg E|\neg H)$ . First,  
271 increasing theoretical risk leads to higher posterior probability  $P(H|E)$ , our objective.  
272 Second, if the theoretical risk is smaller than detectability  $P(E|H)$  it follows that the  
273 posterior probability must decrease when observing the evidence. If detectability exceeds  
274 theoretical risk, the evidence is less likely under the theory than it is when the theory does  
275 not hold (the inverse of statistical relevancy). Third, if the theoretical risk equals zero, then  
276 posterior probability is at best equal to prior probability but only if detectability is perfect  
277 ( $P(H|E) = 1$ ). In other words, observing a sure fact does not lend credence to a hypothesis.

278 The last statement sounds like a truism but is directly related to Popper’s seminal  
279 criterion of demarcation. He stated that if it is impossible to prove that a hypothesis is false  
280 ( $P(\neg E|\neg H) = 0$ , theoretical risk is zero), it cannot be considered a scientific hypothesis  
281 (Popper, 2002, p. 18). We note these relations to underline that the Bayesian rationale we  
282 apply here is able to reconstruct many commonly held views on riskiness and epistemic  
283 value (but we of course differ from Popper on the central role of induction in science).

284 Both theoretical risk  $P(\neg E|\neg H)$  and detectability  $P(E|H)$  aggregate countless  
285 influences; otherwise, they could not model the process of evidential support for theories.  
286 To illustrate the concepts we have introduced here, consider the following example of a  
287 single theory and three experiments that may test it. The experiments were created to  
288 illustrate how they may differ in their theoretical risk and detectability. Suppose the  
289 primary theory is about the cognitive phenomenon of “insight.” For the purpose of  
290 illustration, we define it, with quite some hand-waving, as a cognitive abstraction that

291 allows agents to consistently solve a well-defined class of problems. We present the  
292 hypothesis that the following problem belongs to such a class of insight problems:

293 Use five matches (IIII) to form the number eight.

294 We propose three experiments that differ in theoretical risk and detectability. All  
295 experiments take a sample of ten psychology students. We present the students with the  
296 problem for a brief span of time. After that, the three experiments differ as follows:

- 297 1. The experimenter gives a hint that the problem is easy to solve when using Roman  
298 numerals; if all students come up with the solution, she records it as evidence for the  
299 hypothesis.
- 300 2. The experimenter shows the solution “VIII” and explains it; if all students come up  
301 with the solution, she records it as evidence for the hypothesis.
- 302 3. The experimenter does nothing; if all students come up with the solution, she records  
303 it as evidence for the hypothesis.

304 We argue that experiment 1 has high theoretical risk  $P(\neg E_1|\neg H)$  and high  
305 detectability  $P(E_1|H)$ . If “insight” has nothing to do with solving the problem ( $\neg H$ ), then  
306 presenting the insight that Roman numerals can be used should not lead to all students  
307 solving the problem ( $\neg E_1$ ); the experiment, therefore, has high theoretical risk  
308  $P(\neg E_1|\neg H)$ . Conversely, if insight is required to solve the problem ( $H$ ), then it is likely to  
309 help all students to solve the problem ( $E_1$ ), the experiment, therefore, has high  
310 detectability  $P(E_1|H)$ . The second experiment, on the other hand, has low theoretical risk  
311  $P(\neg E_2|\neg H)$ . Even if “insight” has nothing to do with solving the problem ( $\neg H$ ), there are  
312 other plausible reasons for observing the evidence ( $E_2$ ), because the students could simply  
313 copy the solution without having any insight. With regard to detectability, experiments 1  
314 and 2 differ in no obvious way. Experiment 3, however, also has low detectability. It is  
315 unlikely that all students will come up with the correct solution in a short time ( $E_3$ ), even

316 if insight is required ( $H$ ); experiment 3 therefore has low detectability  $P(E_3|H)$ . The  
317 theoretical risk, however, is also low in absolute terms, but high compared to the  
318 detectability (statistical relevancy condition is satisfied). In the unlikely event that all 10  
319 students place their matches to form the Roman numeral VIII ( $E_3$ ), it is probably due to  
320 insight ( $H$ ) and not by chance  $P(\neg E_3|\neg H)$ . Of course, in practice, we would allow the  
321 evidence to be probabilistic, e.g., relax the requirement of “all students” to nine out of ten  
322 students, more than eight, and so forth.

323 As mentioned earlier, we restrict ourselves to binary evidence, to keep the  
324 mathematical notation as simple as possible. We discuss the relation between statistical  
325 methods and theoretical risk in the [Statistical Methods](#) section.

### 326 Preregistration as a means to increase theoretical risk?

327 Having discussed that increasing the theoretical risk will increase the epistemic  
328 value, it is intuitive to task preregistration with maximizing theoretical risk, i.e., a  
329 confirmatory research agenda. Indeed, limiting the type-I error rate is commonly stated as  
330 *the* central goal of preregistration (Nosek et al., 2018; Oberauer, 2019; Rubin, 2020). We  
331 argue that while such a conclusion is plausible, we must first consider at least two  
332 constraints that place an upper bound on the theoretical risk.

333 First, the theory itself limits theoretical risk: Some theories simply do not make  
334 risky predictions, and preregistration will not change that. Consider the case of a  
335 researcher contemplating the relation between two sets of variables. Suppose each set is  
336 separately well studied, and strong theories tell the researcher how the variables within the  
337 set relate. However, our imaginary researcher now considers the relation between these two  
338 sets. For lack of a better theory, they assume that some relation between any variables of  
339 the two sets exists. This is not a risky prediction to make in psychology (Orben & Lakens,  
340 2020). However, we would consider it a success if the researcher would use the evidence  
341 from this rather exploratory study to develop a more precise (and therefore risky) theory,



**368 Statistical methods**

369 One widely known factor is the contribution of statistical methods to theoretical  
370 risk. Theoretical risk  $P(\neg E|\neg H)$  is deeply connected with statistical methods, because it is  
371 related to the type-I error rate in statistical hypothesis testing  $P(E|\neg H)$  by  
372  $P(\neg E|\neg H) = 1 - P(E|\neg H)$ , if you consider the overly simplistic case where the research  
373 hypothesis is equal to the statistical alternative-hypothesis because then the null-hypothesis  
374 is  $\neg H$ . Because many researchers are familiar with the type-I error rate, it can be helpful  
375 to remember this connection to theoretical risk. Researchers who choose a smaller type-I  
376 error rate can be more sure of their results, if significant, because the theoretical risk is  
377 higher. However, this connection should not be overinterpreted for two reasons. First,  
378 according to most interpretations of null hypothesis testing, the absence of a significant  
379 result should not generally be interpreted as evidence against the hypothesis (Mayo, 2018,  
380 p. 5.3). Second, the research hypothesis rarely equals the statistical alternative hypothesis  
381 (most research hypothesis are more specific than “any value except zero”). In fact, it is  
382 entirely possible to assume the null hypothesis as a research hypothesis, as is commonly  
383 done in e.g., structural equation modelling, where the roles of detectability, theoretical risk  
384 and type-I/II error rate switch. We argue that theoretical risk (and hence its complement,  
385  $P(E|\neg H)$ ) also encompasses factors outside the statistical realm, most notably the study  
386 design and broader analytical strategies. Type-I error rate is the property of a statistical  
387 test under some assumptions, whereas theoretical risk is a researchers’ belief. One may  
388 take such theoretical properties as a first starting point to form a substantive belief but  
389 surely researchers ought to take other factors into consideration. For example, if a  
390 researcher believes that there might be confounding variables at play for the relation  
391 between two variables, this should decrease theoretical risk; after all they might find an  
392 association purely on account of the confounders (Fiedler, 2017).

393 Statistical methods stand out among these factors because we have a large and



394 well-understood toolbox for assessing and controlling their contribution to theoretical risk.  
395 Examples of our ability to exert this control are the choice of type-I error rate, adjustments  
396 for multiple testing, the use of corrected fit measures (i.e., adjusted  $R^2$ ), information  
397 criteria, or cross-validation in machine learning. These tools help us account for biases in  
398 statistical methods that influence theoretical risk (and hence,  $P(E|\neg H)$ ).

399         The point is that the contribution of statistical methods to theoretical risk can be  
400 formally assessed. For many statistical models it can be analytically computed under some  
401 assumptions. For those models or assumptions where this is impossible, one can employ  
402 Monte Carlo simulation to estimate the contribution to theoretical risk. The precision with  
403 which statisticians can discuss contributions to theoretical risk has lured the community  
404 concerned with research methods into ignoring other factors that are much more uncertain.  
405 We cannot hope to resolve this uncertainty; but we have to be aware of its implications.  
406 These are presented in the following.

#### 407 **Sources of uncertainty**

408         As we have noted, it is possible to quantify how statistical models affect the  
409 theoretical risk based on mathematical considerations and simulation. However, other  
410 factors in the broader context of a study are much harder to quantify. If one chooses to  
411 focus only on the contribution of statistical methods to theoretical risk, one is bound to  
412 overestimate it. Take, for example, a t-test of mean differences in two samples. Under ideal  
413 circumstances (assumption of independence, normality of residuals, equal variance), it  
414 stays true to its type-I error rate. However, researchers may do many very reasonable  
415 things in the broader context of the study that affect theoretical risk: They might exclude  
416 outliers, choose to drop an item before computing a sum score, broaden their definition of  
417 the population to be sampled, translate their questionnaires into a different language,  
418 impute missing values, switch between different estimators of the pooled variance, or any  
419 number of other things. All of these decisions carry a small risk that they will increase the

420 likelihood of obtaining evidence despite the underlying research hypothesis being false.  
 421 Even if the t-test itself perfectly maintains its type I error rate, these factors influence  
 422  $P(E|\neg H)$ . While, in theory, these factors may leave  $P(E|\neg H)$  unaffected or even decrease  
 423 it, we argue that this is not the case in practice. Whether researchers want to or not, they  
 424 continuously process information about how the study is going, except under strict  
 425 blinding. While one can hope that processing this information does not affect their  
 426 decision-making either way, this cannot be ascertained. Therefore, we conclude that  
 427 statistical properties only guarantee a lower bound for theoretical risk. The only thing we  
 428 can conclude with some certainty is that theoretical risk is not higher than what the  
 429 statistical model guarantees without knowledge about the other factors at play.

### 430 **The effects of uncertainty**

431 Before we ask how preregistration influences this uncertainty, we must consider the  
 432 implications of being uncertain about the theoretical risk. Within the Bayesian framework,  
 433 this is both straightforward and insightful. Let us assume a researcher is reading a study  
 434 from another lab and tries to decide whether and how much the presented results confirm  
 435 the hypothesis. As the researcher did not conduct the study (and the study is not  
 436 preregistered), they can not be certain about the various factors influencing theoretical risk  
 437 (researcher degrees of freedom). We therefore express this uncertainty about the theoretical  
 438 risk as a probability distribution  $Q$  of  $P(E|\neg H)$  (remember that  $P(E|\neg H)$  is related to  
 439 theoretical risk by  $P(E|\neg H) = 1 - P(\neg E|\neg H)$ , so it does not matter whether we consider  
 440 the distribution of theoretical risk or  $P(E|\neg H)$ ). To get the expected value of  $P(H|E)$   
 441 that follows from the researchers' uncertainty about the theoretical risk, we can compute  
 442 the expectation using Bayes theorem:

$$\mathbb{E}_Q[P(H|E)] = \mathbb{E}_Q \left[ \frac{P(H)P(E|H)}{P(H)P(E|H) + P(\neg H)P(E|\neg H)} \right] \quad (3)$$

443 Of course, the assigned probabilities and the distribution  $Q$  vary from study to

444 study and researcher to researcher (and even the measure of confirmation), but we can  
 445 illustrate the effect of uncertainty with an example. Assuming  $P(E|H) = 0.8$  (relative of  
 446 the typically strived for power of 80%). Let us further assume that the tested hypothesis is  
 447 considered unlikely to be true by the research community before the study is conducted  
 448 ( $P(H) = 0.1$ ) and assign a uniform distribution for  $P(E|\neg H) \sim U([1 - \tau, 1])$  where  $\tau$  is set  
 449 to  $1 - \alpha$ , reflecting our assumption that this term gives an upper bound for theoretical risk  
 450  $P(\neg E|\neg H)$ . We chose this uniform distribution as it is the maximum entropy distribution  
 451 with support  $[1 - \tau, 1]$  and hence conforms to our Bayesian framework (Giffin & Caticha,  
 452 2007).

With this, we derive the expected value of  $P(H|E)$  as

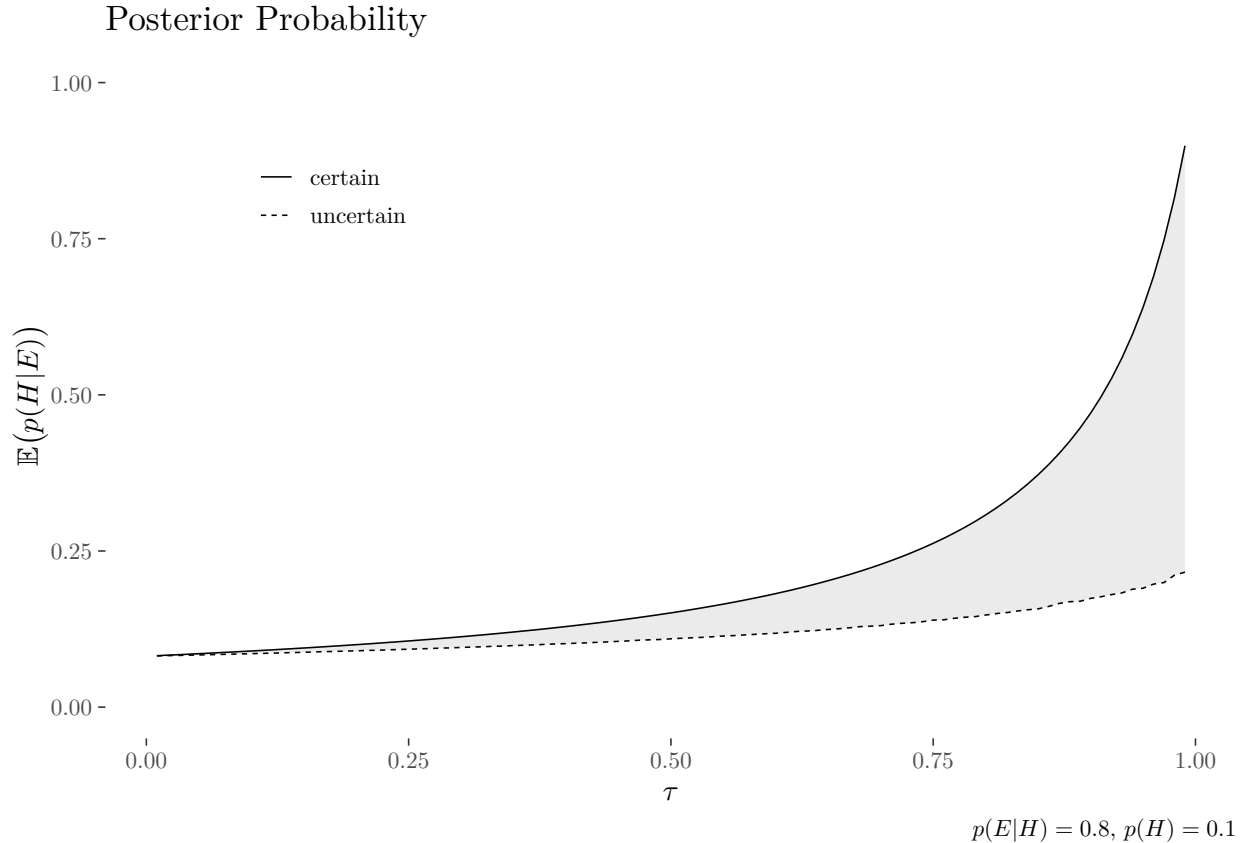
$$\mathbb{E}_Q[P(H|E)] = \mathbb{E}_Q \left[ \frac{P(H)P(E|H)}{P(H)P(E|H) + P(\neg H)P(E|\neg H)} \right] \quad (4)$$

$$= \int_{[1-\tau, 1]} \tau^{-1} \frac{P(H)P(E|H)}{P(H)P(E|H) + P(\neg H)P(E|\neg H)} dP(E|\neg H) \quad (5)$$

$$= \frac{P(H)P(E|H)}{P(\neg H)\tau} \ln \left( \frac{P(H)P(E|H) + P(\neg H)}{P(H)P(E|H) + P(\neg H)(1 - \tau)} \right) \quad (6)$$

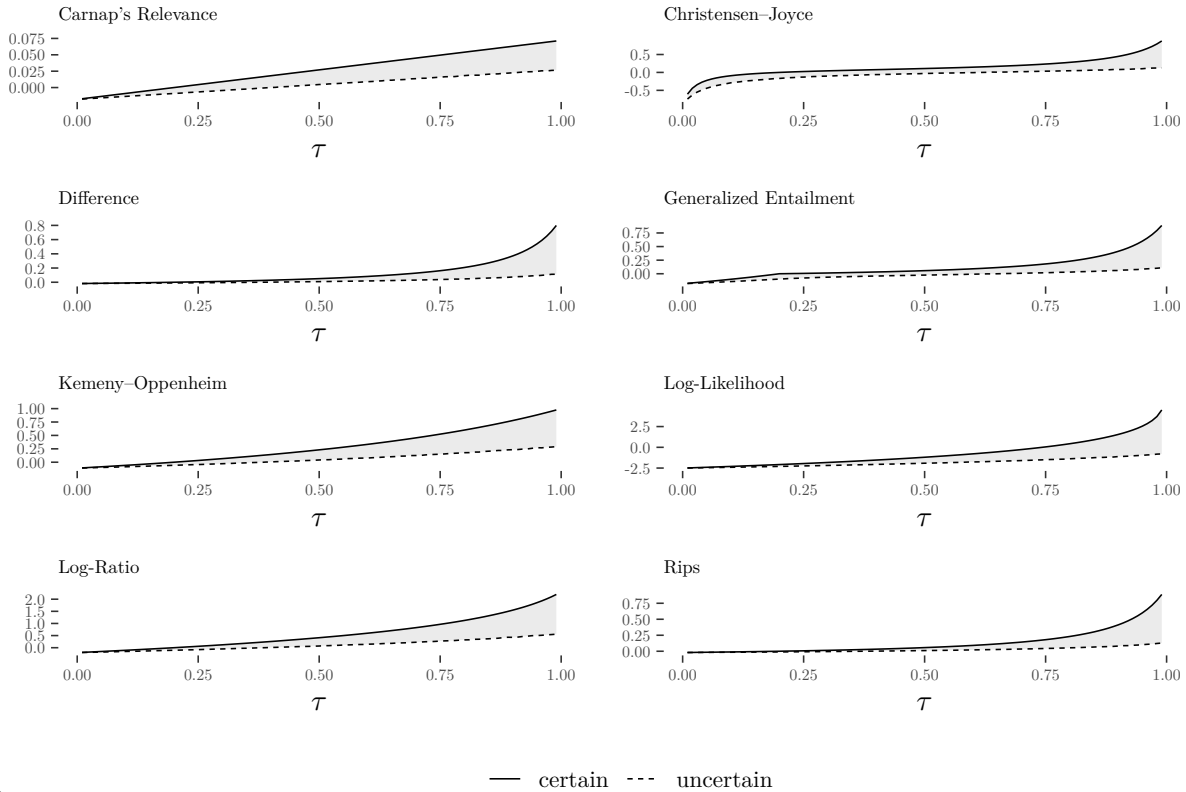
453 Figure 1 shows exemplary the effect of theoretical risk (x-axis) on the posterior  
 454 probability (y-axis) being certain (solid line) or uncertain (dashed line) about the  
 455 theoretical risk of a study. Our expectation of the gained epistemic value varies  
 456 considerably depending on how uncertain we are about the theoretical risk a study took on.  
 457 Mathematically, uncertainty about theoretical risk is expressed through the variance (or  
 458 rather entropy) of the distribution. The increase in uncertainty (expressed as more entropic  
 459 distributions) leads to a decreased expected epistemic value.

460 The argument for a confirmatory research agenda is that by increasing theoretical  
 461 risk we increase expected epistemic value, i.e., moving to the right on the x-axis in Figure 1  
 462 increases posterior probability (on the y-axis). However, if a hypothesis in a certain study

**Figure 1**

*Posterior probability (confirmation as firmness) as a function of theoretical risk  $\tau$ , where  $\tau$  is either certain (solid line) or maximally uncertain (dotted line).*

463 has low theoretical risk, there is not much researchers can do about it. However, studies do  
 464 not only differ by how high the theoretical risk is but also by how certain the recipient is  
 465 about the theoretical risk. A study that has a very high theoretical risk (e.g., 1.00% chance  
 466 that if the hypothesis is wrong, evidence in its favor will be observed,) but has also  
 467 maximum uncertainty will result in a posterior probability of 22%, while the same study  
 468 with maximum certainty will result in 90% posterior probability. The other factors  
 469 (detectability, prior beliefs, measure of epistemic value) and, therefore, the extent of the  
 470 benefit varies, of course, with the specifics of the study. Crucially, even studies with some  
 471 exploratory aspects benefit from preregistration, e.g., in this scenario with a  $\tau = 0.80$  (false  
 472 positive rate of 0.20) moving from uncertain to certain increases the posterior from 0.15 to

**Figure 2**

*Several measures for confirmation as an increase in firmness as a function of  $\tau$ , where  $\tau$  is either certain (solid line) or maximally uncertain (dotted line). Measures taken from Sprenger and Hartmann (2019), Table 1.3, p. 51.*

473 0.31. We find it helpful to calculate an example because of the nonlinear nature of the  
 474 evidence functions.

#### 475 Preregistration as a means to decrease uncertainty about the theoretical risk

476 We hope to have persuaded the reader to accept two arguments: First, the  
 477 theoretical risk is important for judging evidential support for theories. Second, the  
 478 theoretical risk is inherently uncertain, and the degree of uncertainty diminishes the  
 479 persuasiveness of the gathered evidence. The third and last argument is that  
 480 preregistrations reduce this uncertainty. Following the last argument, a preregistered study  
 481 is represented by the solid line (certainty about theoretical risk), and a study that was not  
 482 preregistered is more similar to the dashed line (maximally uncertain about theoretical

483 risk) in Figure 1 and Figure 2.

484 Let us recall our three assumptions:

- 485 1. Researchers judge the evidence for or against a hypothesis rationally.
- 486 2. They expect other researchers to apply a similar rational process.
- 487 3. Researchers try to maximize the expected epistemic value for other researchers.

488 The point we make with these assumptions is that researchers aim to persuade  
489 other researchers, for example, the readers of their articles. Not only the original authors  
490 are concerned with the process of weighing evidence for or against a theory but really the  
491 whole scientific community the study authors hope to persuade. Unfortunately, readers of a  
492 scientific article (or, more generally, any consumer of a research product) will likely lack  
493 insight into the various factors that influence theoretical risk. While the authors  
494 themselves may have a clear picture of what they did and how it might have influenced the  
495 theoretical risk they took, their readers have much greater uncertainty about these factors.  
496 In particular, they never know which relevant factors the authors of a given article failed to  
497 disclose, be it intentionally or not. From the perspective of the ultimate skeptic, they may  
498 claim maximum uncertainty.

499 Communicating clearly how authors of a scientific report collected their data and  
500 consequently analyzed it to arrive at the evidence they present is crucial for judging the  
501 theoretical risk they took. Preregistrations are ideal for communicating just that because  
502 any description after the fact is prone to be incomplete. For instance, the authors could  
503 have opted for selective reporting, that is, they decided to exclude a number of analytic  
504 strategies they tried out. That is not to say that every study that was not-preregistered  
505 was subjected to practices of questionable research practices. The point is that we cannot  
506 exclude it with certainty. This uncertainty is drastically reduced if the researchers have  
507 described what they intended to do beforehand and then report that they did exactly that.

508 In that case, readers can be certain they received a complete account of the situation.  
509 They still might be uncertain about the actual theoretical risk the authors took, but to a  
510 much smaller extent than if the study would not have been preregistered.

511 The remaining sources of uncertainty might be unfamiliarity with statistical  
512 methods or experimental paradigms used, the probability of an implementation error in the  
513 statistical analyses, a bug in the software used for analyses, etc. To further reduce the  
514 uncertainty about theoretical risk, researchers must therefore publish code and ideally data.  
515 After all, computational reproducibility is only possible if the data analytic procedure was  
516 communicated clearly enough to allow others to retrace the computational steps (Peikert &  
517 Brandmaier, 2021).

518 In any case, a well-written preregistration should aim to reduce the uncertainty  
519 about the theoretical risk and hence increase the persuasiveness of evidence. Therefore, a  
520 study that perfectly adhered to its preregistration will resemble the solid line in Figure 1/2.  
521 Crucially, perfect means here that the theoretical risk can be judged with low uncertainty,  
522 not that the theoretical risk is necessarily high.

### 523 **Hacking, harking, and other harms**

524 The importance of distinguishing between low and highly uncertain theoretical risk  
525 becomes perhaps clearer if we consider a few hypothetical cases for illustration.

- 526 1. We know with absolute certainty that researchers will revert to p-hacking to create  
527 evidence that is favorable for the theory.
- 528 2. A hypothesis was picked to explain reported results after the fact (HARKing, Kerr,  
529 1998).
- 530 3. We cannot exclude the possibility of p-hacking having led to the reported results.
- 531 4. Reported results were obtained by planned exploration.
- 532 5. Reported results were obtained by unplanned exploration.

533 In case 1, there is no theoretical risk ( $P(\neg E|\neg H) = 0$ ). If we know that the results  
534 will be engineered to support the hypothesis no matter what, there is no reason to collect  
535 data. A prime example of this case is the  $p_{\text{pointless}}$  metric (Hussey, 2021). Case 2 has a  
536 similar problem. After all, the hypothesis that it had to happen the way it did happen is  
537 irrefutable. In fact, both cases should be problematic to anyone who subscribes to the  
538 statistical relevancy condition because if we choose the hypothesis in accordance with the  
539 data or vice versa, without restrictions, they are not related anymore (i.e., observing the  
540 data does not tell us anything about the hypothesis and the other way around). Case 3 is  
541 different since here the theoretical risk is not necessarily low but simply uncertain (and  
542 perhaps best represented by the dotted line in Figure 1/2). In case 4, the theoretical risk is  
543 neither zero (unless the researcher plans to do run variations of analyses until a favourable  
544 outcome is obtained, then we have a particular instance case of 1) nor high (as this is the  
545 nature of exploratory approaches). However, we can take advantage of computational  
546 reproducibility, use statistical properties, simulation or resampling methods, together with  
547 scientific reasoning, to get a reasonably certain evaluation of the theoretical risk and hence  
548 are in a somewhat favourable position (i.e., close to the solid line in n Figure 1/2). This  
549 favorable position leads us to recommend preregistration of exploratory studies. Case 5  
550 shares the neither zero nor high theoretical risk of case 4 but has additional uncertainty  
551 about how much exploration was going on (how hard did researcher try to come up with  
552 favourable results). Its low *and uncertain* theoretical risk make it difficult to produce  
553 compelling evidence.

554

## Discussion

555 To summarize, we showed that both higher theoretical risk and lower uncertainty  
556 about theoretical risk lead to higher expected epistemic value across a variety of measures.  
557 The former result that increasing theoretical risk leads to higher expected epistemic value  
558 reconstructs the appeal and central goal of preregistration of confirmatory research  
559 agendas. However, theoretical risk is something researchers have only limited control over.



560 For example, theories are often vague and ill-defined, resources are limited, and increasing  
561 theoretical risk usually decreases detectability of a hypothesized effect (a special instance of  
562 this trade-off is the well-known tension between type-I error and statistical power). While  
563 we believe that preregistration is always beneficial, it might be counterproductive to pursue  
564 high theoretical risk if the research context is inappropriate for strictly confirmatory  
565 research. Specifically, appropriateness here entails the development of precise theories and  
566 the availability of necessary resources (often, large enough sample size, but also see  
567 Brandmaier et al. (2015)) to adequately balance detectability against theoretical risk.

568 In terms of preparing the conditions for confirmatory research, preregistration may  
569 at most help to invest some time into developing more specific, hence riskier, implications  
570 of a theory. But for a confirmatory science, it will not be enough to preregister all studies.  
571 This undertaking requires action from the whole research community (Lishner, 2015).  
572 Incentive structures must be created to evaluate not the outcomes of a study but the rigor  
573 with which it was conducted (Cagan, 2013; Schönbrodt et al., 2022). Journal editors could  
574 encourage theoretical developments that allow for precise predictions that will be tested by  
575 other researchers and be willing to accept registered reports (Fried, 2020a, 2020b; van  
576 Rooij & Baggio, 2021, 2020). Funding agencies should demand an explicit statement about  
577 theoretical risk in relation to detectability and must be willing to provide the necessary  
578 resources to reach adequate levels of both (Koole & Lakens, 2012).

579 Theoretical risk may conceptually be related to the framework of “severity” (Mayo,  
580 2018; Mayo & Spanos, 2011). However, there are crucial differences between the two. First,  
581 our perspective on theoretical risk is not primarily concerned with avoiding inductive  
582 reasoning but with subjective changes of belief. This is important because, while severity is  
583 calculable, it remains unclear how severity should be valued, e.g. if an increase in severity  
584 from .80 to .81 should be as impressive as from .99 to .999. Second, severity considerations  
585 are mainly after the fact. Severity, a measure with which we can rule out alternative

586 explanations, can only be calculated after evidence was observed. However, there also are  
587 communalities, like the strong emphasis on counterfactual consideration (imagining the  
588 hypothesis was false), and there are even proposals to reconcile Bayesian and severity  
589 considerations (van Dongen et al., 2023).

590 Our latter result, on the importance of preregistration for minimizing uncertainty,  
591 has two important implications. The first is, that even if all imaginable actions regarding  
592 promoting higher theoretical risk are taken, confirmatory research should be preregistered.  
593 Otherwise, the uncertainty about the theoretical risk will diminish the advantage of  
594 confirmatory research. Second, even under less-than-ideal circumstances for confirmatory  
595 research, preregistration is beneficial. Preregistering exploratory studies increases the  
596 expected epistemic value by virtue of reducing uncertainty about theoretical risk.  
597 Nevertheless, exploratory studies will have a lower expected epistemic value than a more  
598 confirmatory study if both are preregistered and have equal detectability.

599 Focusing on uncertainty reduction also explains two common practices of  
600 preregistration that do not align with a confirmatory research agenda. First, researchers  
601 seldomly predict precise numerical outcomes, instead they use preregistrations to describe  
602 the process that generates the results. Precise predictions would have very high theoretical  
603 risk (they are likely incorrect if the theory is wrong). A statistical procedure may have high  
604 or low theoretical risk depending on the specifics of the model used. Specifying the process,  
605 therefore, is in line with the rationale we propose here, but is less reasonable when the goal  
606 of preregistration is supposed to be a strictly confirmatory research agenda.

607 Second, researchers often have to deviate from the preregistration and make  
608 data-dependent decisions after the preregistration. If the only goal of preregistration is to  
609 ensure confirmatory research, such changes are not justifiable. However, under our rational,  
610 some changes may be justified. Any change increases the uncertainty about the theoretical  
611 risk and may even decrease the theoretical risk. The changes still may be worthwhile if the

612 negative outcomes may be offset by an increase in detectability due to the change.  
613 Consider a preregistration that failed to specify how to handle missing values, and  
614 researchers subsequently encountering missing values. In such case, detectability becomes  
615 zero because the data cannot be analyzed without a post-hoc decision about how to handle  
616 the missing data. Any such decision would constitute a deviation from the preregistration,  
617 which is possible under our proposed objective. Note that a reader cannot rule out that the  
618 researchers leveraged the decision to decrease theoretical risk, i.e., picking among all  
619 options the one that delivers the most beneficial results for the theory (in the previous  
620 example, choosing between various options of handling missing values). Whatever decision  
621 they make, increased uncertainty about the theoretical risk is inevitable and the expected  
622 epistemic value is decreased compared to a world where they anticipated the need to deal  
623 with missing data. However, it is still justified to deviate. After all they have not  
624 anticipated the case and are left with a detectability of zero. Any decision will increase  
625 detectability to a non-zero value offsetting the increase in uncertainty. The researchers also  
626 may do their best to argue that the deviation was not motivated by increasing theoretical  
627 risk, thereby, decreasing the uncertainty. Ideally, there is a default decision that fits well  
628 with the theory or with the study design. Or, if there is no obvious candidate, the  
629 researchers could conduct a multiverse analysis of the available options to deal with  
630 missings to show the influence of the decision (Steen et al., 2016). In any case, deviations  
631 must be transparently reported and we applaud recent developments to standardize and  
632 normalize this process (Willroth & Atherton, 2023).

633 As explained above, reduction in uncertainty as the objective for preregistration  
634 does not only explain some existing practice, that does not align with confirmation as a  
635 goal, it also allows to form recommendations to improve the practice of preregistration.  
636 Importantly, we now have a theoretical measure to gauge the functionality of  
637 preregistrations, which can only help increase its utility. In particular, a preregistration  
638 should be specific about the procedure that is intended to generate evidence for a theory.

639 Such a procedure may accommodate a wide range of possible data, i.e., it may be  
640 exploratory. The theoretical risk, however low, must be communicated clearly. Parts of the  
641 process left unspecified imply uncertainty, which preregistration should reduce. However,  
642 specifying procedures that can be expected to fail will lead to deviation and, subsequently,  
643 to larger uncertainty.

644 Our emphasis on transparency aligns with other justifications of preregistration,  
645 especially those put forth by Lakens (2019)'s, although based on quite different  
646 philosophical foundations. Our goal is to contribute a rationale that more comprehensively  
647 captures the spectrum of exploration and confirmation in relation to preregistrations,  
648 post-hoc changes of preregistrations, and subjective evaluations of evidence. We find it  
649 difficult to content ourselves with vague terms like “control” or “transparency” if they  
650 ultimately remain unconnected to how much researchers believe in a theory. Within our  
651 framework, researchers have the ability to input their assumptions regarding the  
652 perspectives of other researchers and calculate the potential impact of their actions on their  
653 readership, whether these actions relate to study design, to the preregistration itself, or  
654 subsequent deviations from it. We put subjective evaluations at the center of our  
655 considerations; we deal explicitly with researchers who are proponents of some theory (they  
656 have higher priors for the theory being true), researchers who suspect confounding variables  
657 (they assume lower theoretical risk), or those who remain doubtful if everything relevant  
658 was reported (they have higher uncertainty about theoretical risk) or even those who place  
659 greater value on incongruent evidence than others (they differ in their confirmation  
660 function). We, therefore, hope to not only provide a rationale for preregistration for those  
661 who subscribe to a Bayesian philosophy of science but also a framework to navigate the  
662 complicated questions that arise in the practice of preregistration.

663 At the same time, approaching the evaluation of evidence using a Bayesian  
664 formalism is far from novel Fiedler (2017). To our knowledge, it was not yet applied to the

665 problem of preregistration. However, Oberauer and Lewandowsky (2019) made use of the  
666 formalism to model the relation between theory, hypothesis, and evidence. In the context  
667 of this conceptualization, they discussed the usefulness of preregistration, though without  
668 applying the formalism there. Most importantly, they are rather critical of the idea that  
669 preregistration has tangible benefits. Instead, they prefer multiverse analyses but contend  
670 that those could be preregistered if one fancies it. Their reasoning is based on two  
671 intuitions about what should *not* influence the evaluation of evidence: temporal order and  
672 the mental state of the originator. In our opinion, they disregard the temporal order a bit  
673 too hastily, as it is a long-standing issue in Bayesian philosophy of science known as the  
674 “problem of old evidence” (Chihara, 1987). However, we agree that not the temporal order  
675 is decisive but if the researchers incorporated the information into the hypothesis the  
676 evidence is supposed to confirm. For the other, we argue that the mental state of the  
677 originator does matter. Suppose there are  $k = 1, 2, \dots, K$  ways to analyze data, where each  
678  $k$  has a  $P(E_k|\neg H) > 0$ . If they intend to try each way after another but happen to be  
679 “lucky” on the first try and stop, should we then apply  $P(E|\neg H) = P(E_1|\neg H)$  or  
680  $P(E|\neg H) = P(E_1 \vee \dots \vee E_k|\neg H)$ ? We think the latter. However, this “Defeatist” intuition  
681 is not universally warranted and depends on what we take  $H$  to mean specifically (Kotzen,  
682 2013). Addressing, this problem might benefit from combining Oberauer and Lewandowsky  
683 (2019)’s idea of updating on two nested levels (theory-hypothesis layered on top of  
684 hypothesis-evidence) with our approach to modelling uncertainty.

685         Whatever the difference in evaluating preregistration as a tool, maybe conceptually  
686 more profound is that Oberauer and Lewandowsky (2019) conceptualizes  
687 “discovery-oriented research” differently than we do “exploratory”. They assume the same  
688 theoretical risk ( $P(\neg E|\neg H) = .05$ ) and detectability ( $P(E|H) = .8$ ) in their calculation  
689 example as we do but assign different prior probabilities, namely .06 for discovery versus .6  
690 for theory testing. Then, they conclude that discovery-oriented researcher requires a much  
691 lower type-I error rate to control false positive in light of the low prior probability. This

692 runs counter to our definition of exploratory research having low theoretical risk. Of course,  
693 we agree that low priors require more persuasive evidence; our disagreement, therefore, lies  
694 mainly in terminology. They imagine discovery-oriented researchers to conduct  
695 experiments where they have low expectations that they obtain positive evidence  
696 ( $.06 \cdot .8 + .94 \cdot .05 = 0.095$ ), but if they do, it raises the posterior significantly (from .06 to  
697 .51) In our view, researchers who set out to explore a data set often find “something” (due  
698 to low  $P(\neg E|\neg H)$ ); therefore, it should only slightly raise your posterior if they do. On a  
699 substantive matter, we believe both kinds of research are common in psychology. It is,  
700 therefore, mostly a disagreement on terminology. This disagreement only highlights why  
701 using a mathematical framework to investigate such things is so useful and ultimately  
702 indispensable because we can clearly see where and how we differ in our reasoning.

703 We believe that our reasoning is quite similar to Höfler et al. (2022), who call for  
704 transparent exploration using preregistration. We could be more sure of our agreement, if  
705 they had formulated their arguments within a mathematical framework, which would also  
706 have helped to dissolve an apparent conflict in their definitions of confirmation, exploration,  
707 and transparency. On the one hand, they define “The principle difference between  
708 confirmation and exploration is that confirmation adheres to an evidential norm for the  
709 test of a hypothesis to pass.”, but then suggest that transparent exploration can be  
710 conducted using inferences tests as a filtering mechanism. Their distinction between  
711 confirmation, intransparent and transparent exploration are otherwise just as well placed  
712 along the dimensions, theoretical risk and uncertainty about theoretical risk.

713 With the goal to facilitate rigorous exploration, we have proposed a workflow for  
714 preregistration called *preregistration as code* (PAC) elsewhere (Peikert et al., 2021). In a  
715 PAC, researchers use computer code for the planned analysis as well as a verbal description  
716 of theory and methods for the preregistration. This combination is facilitated by dynamic  
717 document generation, where the results of the code, such as numbers, figures, and tables,



**Declarations**

742

743 All code and materials required to reproduce this article are available under  
744 <https://github.com/aaronpeikert/bayes-prereg> (Peikert & Brandmaier, 2023a). The  
745 authors have no competing interests to declare that are relevant to the content of this  
746 article.



## References

- 747  
748 Akker, O. van den, Bakker, M., Assen, M. A. L. M. van, Pennington, C. R., Verweij, L.,  
749 Elsherif, M., Claesen, A., Gaillard, S. D. M., Yeung, S. K., Frankenberger, J.-L.,  
750 Krautter, K., Cockcroft, J. P., Kreuer, K. S., Evans, T. R., Heppel, F., Schoch, S. F.,  
751 Korbmacher, M., Yamada, Y., Albayrak-Aydemir, N., ... Wicherts, J. (2023, May 10).  
752 *The effectiveness of preregistration in psychology: Assessing preregistration strictness*  
753 *and preregistration-study consistency*. <https://doi.org/10.31222/osf.io/h8xjw>
- 754 Bakker, M., Veldkamp, C. L. S., Assen, M. A. L. M. van, Cromptoets, E. A. V., Ong, H.  
755 H., Nosek, B. A., Soderberg, C. K., Mellor, D., & Wicherts, J. M. (2020). Ensuring the  
756 quality and specificity of preregistrations. *PLOS Biology*, *18*(12), e3000937.  
757 <https://doi.org/10.1371/journal.pbio.3000937>
- 758 Baumeister, R. F. (2016). Charting the future of social psychology on stormy seas:  
759 Winners, losers, and recommendations. *Journal of Experimental Social Psychology*, *66*,  
760 153–158. <https://doi.org/10.1016/j.jesp.2016.02.003>
- 761 Brandmaier, A. M., Oertzen, T. von, Ghisletta, P., Hertzog, C., & Lindenberger, U. (2015).  
762 LIFESPAN: A tool for the computer-aided design of longitudinal studies. *Frontiers in*  
763 *Psychology*, *6*, 272.
- 764 Brandmaier, A. M., von Oertzen, T., McArdle, J. J., & Lindenberger, U. (2013). Structural  
765 equation model trees. *Psychological Methods*, *18*(1), 71–86.  
766 <https://doi.org/10.1037/a0030001>
- 767 Cagan, R. (2013). San Francisco Declaration on Research Assessment. *Disease Models &*  
768 *Mechanisms*, dmm.012955. <https://doi.org/10.1242/dmm.012955>
- 769 Carnap, R. (1950). *Logical Foundations of Probability*. Chicago, IL, USA: Chicago  
770 University of Chicago Press.
- 771 Chan, A.-W., Hróbjartsson, A., Haahr, M. T., Gøtzsche, P. C., & Altman, D. G. (2004).  
772 Empirical Evidence for Selective Reporting of Outcomes in Randomized  
773 TrialsComparison of Protocols to Published Articles. *JAMA*, *291*(20), 2457–2465.

- 774 <https://doi.org/10.1001/jama.291.20.2457>
- 775 Chihara, C. S. (1987). Some Problems for Bayesian Confirmation Theory. *The British*  
776 *Journal for the Philosophy of Science*, 38(4), 551–560.
- 777 <https://doi.org/10.1093/bjps/38.4.551>
- 778 Christensen, D. (1991). Clever Bookies and Coherent Beliefs. *The Philosophical Review*,  
779 100(2), 229–247. <https://doi.org/10.2307/2185301>
- 780 Claesen, A., Gomes, S., Tuerlinckx, F., & Vanpaemel, W. (2021). Comparing dream to  
781 reality: An assessment of adherence of the first generation of preregistered studies.  
782 *Royal Society Open Science*, 8(10), 211037. <https://doi.org/10.1098/rsos.211037>
- 783 Dwan, K., Altman, D. G., Arnaiz, J. A., Bloom, J., Chan, A.-W., Cronin, E., Decullier, E.,  
784 Easterbrook, P. J., Elm, E. V., Gamble, C., Ghersi, D., Ioannidis, J. P. A., Simes, J., &  
785 Williamson, P. R. (2008). Systematic Review of the Empirical Evidence of Study  
786 Publication Bias and Outcome Reporting Bias. *PLOS ONE*, 3(8), e3081.  
787 <https://doi.org/10.1371/journal.pone.0003081>
- 788 Fetzer, J. H. (1974). Statistical Explanations. In K. F. Schaffner & R. S. Cohen (Eds.),  
789 *PSA 1972: Proceedings of the 1972 Biennial Meeting of the Philosophy of Science*  
790 *Association* (pp. 337–347). Springer Netherlands.  
791 [https://doi.org/10.1007/978-94-010-2140-1\\_23](https://doi.org/10.1007/978-94-010-2140-1_23)
- 792 Fiedler, K. (2017). What Constitutes Strong Psychological Science? The (Neglected) Role  
793 of Diagnosticity and A Priori Theorizing. *Perspectives on Psychological Science*, 12(1),  
794 46–61. <https://doi.org/10.1177/1745691616654458>
- 795 Finkel, E. J., Eastwick, P. W., & Reis, H. T. (2017). Replicability and other features of a  
796 high-quality science: Toward a balanced and empirical approach. *Journal of Personality*  
797 *and Social Psychology*, 113(2), 244–253. <https://doi.org/10.1037/pspi0000075>
- 798 Fried, E. I. (2020a). Lack of Theory Building and Testing Impedes Progress in The Factor  
799 and Network Literature. *Psychological Inquiry*, 31(4), 271–288.  
800 <https://doi.org/10.1080/1047840X.2020.1853461>

- 801 Fried, E. I. (2020b). Theories and Models: What They Are, What They Are for, and What  
802 They Are About. *Psychological Inquiry*, 31(4), 336–344.  
803 <https://doi.org/10.1080/1047840X.2020.1854011>
- 804 Giffin, A., & Caticha, A. (2007). Updating Probabilities with Data and Moments. *AIP*  
805 *Conference Proceedings*, 954, 74–84. <https://doi.org/10.1063/1.2821302>
- 806 Höfler, M., Scherbaum, S., Kanske, P., McDonald, B., & Miller, R. (2022). Means to  
807 valuable exploration: I. The blending of confirmation and exploration and how to  
808 resolve it. *Meta-Psychology*, 6. <https://doi.org/10.15626/MP.2021.2837>
- 809 Hoyningen-Huene, P. (2006). Context of Discovery Versus Context of Justification and  
810 Thomas Kuhn. In J. Schickore & F. Steinle (Eds.), *Revisiting Discovery and*  
811 *Justification: Historical and philosophical perspectives on the context distinction* (pp.  
812 119–131). Springer Netherlands. [https://doi.org/10.1007/1-4020-4251-5\\_8](https://doi.org/10.1007/1-4020-4251-5_8)
- 813 Hussey, I. (2021). A method to streamline p-hacking. *Meta-Psychology*, 5.  
814 <https://doi.org/10.15626/MP.2020.2529>
- 815 Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLOS*  
816 *Medicine*, 2(8), e124. <https://doi.org/10.1371/journal.pmed.0020124>
- 817 Kerr, N. L. (1998). HARKing: Hypothesizing After the Results are Known. *Personality*  
818 *and Social Psychology Review*, 2(3), 196–217.  
819 [https://doi.org/10.1207/s15327957pspr0203\\_4](https://doi.org/10.1207/s15327957pspr0203_4)
- 820 Koole, S. L., & Lakens, D. (2012). Rewarding Replications: A Sure and Simple Way to  
821 Improve Psychological Science. *Perspectives on Psychological Science*, 7(6), 608–614.  
822 <https://doi.org/10.1177/1745691612462586>
- 823 Kotzen, M. (2013). Multiple Studies and Evidential Defeat. *Noûs*, 47(1), 154–180.  
824 <http://www.jstor.org/stable/43828821>
- 825 Kukla, A. (1990). Clinical Versus Statistical Theory Appraisal. *Psychological Inquiry*, 1(2),  
826 160–161. [https://doi.org/10.1207/s15327965pli0102\\_9](https://doi.org/10.1207/s15327965pli0102_9)
- 827 Lakens, D. (2019). The value of preregistration for psychological science: A conceptual

- 828 analysis. , 62(3), 221–230. [https://doi.org/10.24602/sjpr.62.3\\_221](https://doi.org/10.24602/sjpr.62.3_221)
- 829 Lishner, D. A. (2015). A Concise Set of Core Recommendations to Improve the  
830 Dependability of Psychological Research. *Review of General Psychology*, 19(1), 52–68.  
831 <https://doi.org/10.1037/gpr0000028>
- 832 Mayo, D. G. (2018). *Statistical Inference as Severe Testing: How to Get Beyond the*  
833 *Statistics Wars* (First). Cambridge University Press.  
834 <https://doi.org/10.1017/9781107286184>
- 835 Mayo, D. G., & Spanos, A. (2011). Error Statistics. In *Philosophy of Statistics* (pp.  
836 153–198). Elsevier. <https://doi.org/10.1016/B978-0-444-51862-0.50005-8>
- 837 Meehl, P. E. (1990). Appraising and Amending Theories: The Strategy of Lakatosian  
838 Defense and Two Principles that Warrant It. *Psychological Inquiry*, 1(2), 108–141.  
839 [https://doi.org/10.1207/s15327965pli0102\\_1](https://doi.org/10.1207/s15327965pli0102_1)
- 840 Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the  
841 slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, 46(4),  
842 806–834. <https://doi.org/10.1037/0022-006X.46.4.806>
- 843 Mellor, D. T., & Nosek, B. A. (2018). Easy preregistration will benefit any research.  
844 *Nature Human Behaviour*, 2(2), 98–98. <https://doi.org/10.1038/s41562-018-0294-7>
- 845 Niiniluoto, I. (1998). Verisimilitude: The Third Period. *The British Journal for the*  
846 *Philosophy of Science*, 49(1), 1–29. <https://doi.org/10.1093/bjps/49.1.1>
- 847 Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration  
848 revolution. *Proceedings of the National Academy of Sciences*, 115(11), 2600–2606.  
849 <https://doi.org/10.1073/pnas.1708274114>
- 850 Oberauer, K. (2019). Preregistration of a forking path – What does it add to the garden of  
851 evidence? In *Psychonomic Society Featured Content*.
- 852 Oberauer, K., & Lewandowsky, S. (2019). Addressing the theory crisis in psychology.  
853 *Psychonomic Bulletin & Review*, 26(5), 1596–1618.  
854 <https://doi.org/10.3758/s13423-019-01645-2>

- 855 Open Science Collaboration. (2015). Estimating the reproducibility of psychological  
856 science. *Science*, *349*(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
- 857 Orben, A., & Lakens, D. (2020). Crud (Re)Defined. *Advances in Methods and Practices in*  
858 *Psychological Science*, *3*(2), 238–247. <https://doi.org/10.1177/2515245920917961>
- 859 Peikert, A., & Brandmaier, A. M. (2023a). *Supplemental materials for preprint: Why does*  
860 *preregistration increase the persuasiveness of evidence? A Bayesian rationalization.*  
861 Zenodo. <https://doi.org/10.5281/zenodo.7648471>
- 862 Peikert, A., & Brandmaier, A. M. (2023b). *Why does preregistration increase the*  
863 *persuasiveness of evidence? A Bayesian rationalization.* PsyArXiv; PsyArXiv.  
864 <https://doi.org/10.31234/osf.io/cs8wb>
- 865 Peikert, A., & Brandmaier, A. M. (2021). A Reproducible Data Analysis Workflow With R  
866 Markdown, Git, Make, and Docker. *Quantitative and Computational Methods in*  
867 *Behavioral Sciences*, 1–27. <https://doi.org/10.5964/qcmb.3763>
- 868 Peikert, A., van Lissa, C. J., & Brandmaier, A. M. (2021). Reproducible Research in R: A  
869 Tutorial on How to Do the Same Thing More Than Once. *Psych*, *3*(4), 836–867.  
870 <https://doi.org/10.3390/psych3040053>
- 871 Pham, M. T., & Oh, T. T. (2021). Preregistration Is Neither Sufficient nor Necessary for  
872 Good Science. *Journal of Consumer Psychology*, *31*(1), 163–176.  
873 <https://doi.org/10.1002/jcpy.1209>
- 874 Popper, K. R. (2002). *The logic of scientific discovery.* Routledge.
- 875 Rubin, M. (2020). Does preregistration improve the credibility of research findings? *The*  
876 *Quantitative Methods for Psychology*, *16*(4), 376–390.  
877 <https://doi.org/10.20982/tqmp.16.4.p376>
- 878 Salmon, W. C. (1970). Statistical Explanation. In *The Nature & function of scientific*  
879 *theories: Essays in contemporary science and philosophy* (pp. 173–232). University of  
880 Pittsburgh Press.
- 881 Schönbrodt, F., Gärtner, A., Frank, M., Gollwitzer, M., Ihle, M., Mischkowski, D., Phan, L.

- 882 V., Schmitt, M., Scheel, A. M., Schubert, A.-L., Steinberg, U., & Leising, D. (2022).  
883 *Responsible Research Assessment I: Implementing DORA for hiring and promotion in*  
884 *psychology*. PsyArXiv. <https://doi.org/10.31234/osf.io/rgh5b>
- 885 Shmueli, G. (2010). To Explain or to Predict? *Statistical Science*, 25(3), 289–310.  
886 <https://doi.org/10.1214/10-STS330>
- 887 Silagy, C. A., Middleton, P., & Hopewell, S. (2002). Publishing Protocols of Systematic  
888 Reviews Comparing What Was Done to What Was Planned. *JAMA*, 287(21),  
889 2831–2834. <https://doi.org/10.1001/jama.287.21.2831>
- 890 Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2021). Pre-registration: Why and How.  
891 *Journal of Consumer Psychology*, 31(1), 151–162. <https://doi.org/10.1002/jcpy.1208>
- 892 Sprenger, J., & Hartmann, S. (2019). *Bayesian Philosophy of Science*. Oxford University  
893 Press. <https://doi.org/10.1093/oso/9780199672110.001.0001>
- 894 Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing Transparency  
895 Through a Multiverse Analysis. *Perspectives on Psychological Science*, 11(5), 702–712.  
896 <https://doi.org/10.1177/1745691616658637>
- 897 Stefan, A. M., & Schönbrodt, F. D. (2023). Big little lies: A compendium and simulation  
898 of p-hacking strategies. *Royal Society Open Science*, 10(2).  
899 <https://doi.org/10.1098/rsos.220346>
- 900 Szollosi, A., Kellen, D., Navarro, D. J., Shiffrin, R., Rooij, I. van, Zandt, T. V., & Donkin,  
901 C. (2020). Is Preregistration Worthwhile? *Trends in Cognitive Sciences*, 24(2), 94–95.  
902 <https://doi.org/10.1016/j.tics.2019.11.009>
- 903 Tukey, J. W. (1972). Exploratory data analysis: As part of a larger whole. *Proceedings of*  
904 *the 18th Conference on Design of Experiments in Army Research and Development and*  
905 *Training*, 1–18. <https://apps.dtic.mil/sti/tr/pdf/AD0776910.pdf>
- 906 Tukey, J. W. (1980). We Need Both Exploratory and Confirmatory. *The American*  
907 *Statistician*, 34(1), 23–25. <https://doi.org/10.2307/2682991>
- 908 van Dongen, N., Sprenger, J., & Wagenmakers, E.-J. (2023). A Bayesian perspective on

- 909 severity: Risky predictions and specific hypotheses. *Psychonomic Bulletin & Review*,  
910 30(2), 516–533. <https://doi.org/10.3758/s13423-022-02069-1>
- 911 van Rooij, I., & Baggio, G. (2021). Theory Before the Test: How to Build  
912 High-Verisimilitude Explanatory Theories in Psychological Science. *Perspectives on*  
913 *Psychological Science*, 16(4), 682–697. <https://doi.org/10.1177/1745691620970604>
- 914 van Rooij, I., & Baggio, G. (2020). Theory Development Requires an Epistemological Sea  
915 Change. *Psychological Inquiry*, 31(4), 321–325.  
916 <https://doi.org/10.1080/1047840X.2020.1853477>
- 917 Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A.  
918 (2012). An Agenda for Purely Confirmatory Research. *Perspectives on Psychological*  
919 *Science*, 7(6), 632–638. <https://doi.org/10.1177/1745691612463078>
- 920 Willroth, E. C., & Atherton, O. E. (2023). *Best Laid Plans: A Guide to Reporting*  
921 *Preregistration Deviations* [Preprint]. PsyArXiv. <https://doi.org/10.31234/osf.io/dwx69>