

# Why does preregistration increase the persuasiveness of evidence? A Bayesian rationalization.

## Abstract

Preregistration is becoming increasingly popular in the psychological sciences. At the same time, many researchers struggle to make precise predictions and feel overwhelmed by the need to prespecify every detail of their analysis plan. Should they surrender and discard the idea of preregistration altogether? Not at all. We argue for the benefits of preregistration beyond strictly confirmatory studies. Grounded in Bayesian reasoning, we define a formal objective for preregistration that neither declares changes to a preregistration to be sinful nor punishes rigour. This objective rests on the relevance of “theoretical risk”, a generalization of type-I error rate, for judging the evidential support for theories. We propose that preregistration should aim to reduce the uncertainty in judging the theoretical risk instead of directly limiting the type-I error rate. We argue that this reformulation extends the utility of preregistration, provides principled justification for preregistration and is more closely aligned with researchers’ intuition. If we do not know how the researchers arrived at their evidence, we have great uncertainty about the theoretical risk they took, which severely limits the utility of the researchers’ evidence. Preregistration effectively reduces this uncertainty and therefore generates more persuasive evidence.

A successful prediction should lend more credibility to a theory than a postdiction, all else equal. Much has been theorized on the vital distinction between exploration and confirmation, discovery and justification, hypothesis-generating and hypothesis-testing and so forth (see Hoyningen-Huene 2006). To honour this distinction and produce results that are considered confirmatory, many researchers opt to preregister their studies (Nosek et al. 2018). Indeed, rigorous application of preregistration prevents researchers from reporting a simple confirmatory story for a set of results produced by an arduous process of trial and error (Wagenmakers et al. 2012). The scientific community widely agrees that actively concealing such “trying for results” is misconduct — but how exactly is it impeding scientific progress? We believe that a rigorous answer to this question enables us to clarify common misunderstandings that lead researchers to deem preregistration as unfit for their purposes. These misunderstandings are interesting from a conceptual perspective but also lead to decreased efficiency of the scientific process if researchers misguidedly decide that preregistration is not applicable for their intended study. Asking researchers to invest time and resources into preregistration demands a principled justification of its utility.

One reason why researchers are hesitant to preregister is that they feel restricted by the boundaries specified in their preregistration (Goldin-Meadow 2016). Even researchers who have been persuaded of its advantages sometimes apply strategies to skirt these self-imposed restrictions. One strategy is to write a loose preregistration, to begin with;

another is to deviate from the preregistration afterwards. To complicate matters further, both strategies may be used with compelling scientific reasoning or with the self-serving intent of generating desirable results no matter the nature of the phenomenon under study. To summarize, the conundrum is that the current conceptualization of preregistration deems a study “properly” preregistered even if it contains a recipe for p-hacking before data collection but “improper” if the researcher did technically deviate afterwards. This, of course, goes against most researchers’ intuition about what is considered exploratory and confirmatory. We take these issues as signals of a misunderstanding about how preregistration contributes to scientific progress. In particular, we want to challenge the common view that preregistration somehow cleanly separates confirmatory from exploratory.

Deciding if a study is confirmatory or exploratory is a judgment made by the researcher. The issues outlined above arise by delegating this judgment to a simple decision rule. That is, if something is preregistered, it is considered confirmatory, and if it is not preregistered, it is considered exploratory. This rule of thumb is often applied implicitly. Other times it is stated as the explicit goal of preregistration (Mellor and Nosek 2018). In both cases, it leads to confusion when applied outside specific bounds. Preregistering an inherently exploratory analysis (like testing dozens of relations) does not make it confirmatory, nor will a carefully conducted confirmatory study become exploratory if the researcher deviates from the preregistration in minor details. To uphold the simplicity of the decision rule, it is routinely suggested only to preregister when a study is expected to have a low type I error rate by design and will run/be analyzed without changes no matter what (Simmons, Nelson, and Simonsohn 2021; Bakker et al. 2020). Under these conditions, the rule holds, but such restricted use makes preregistration a niche solution unable to match the greater problem of replicability in psychology and elsewhere.

We show that this simple decision rule is a special case of a more general conceptualization under Bayesian reasoning. To that end, we first introduce some tools of Bayesian philosophy of science and map the exploration/confirmation distinction onto a dimensional quantity we call “theoretical risk” (we borrow the term from Meehl 1978), which is inversely related to type I error rate. We then outline two interpretations of how theoretical risk is impacted by preregistration. The first interpretation corresponds to the traditional application of preregistration to research paradigms that focus on confirmation by maximizing the theoretical risk or equivalently by limiting type I error. The second interpretation is our main contribution and demonstrates the broad applicability of preregistration for both exploratory and confirmatory studies that are implemented as preregistered or have undergone changes after preregistration. Following this interpretation, the theoretical risk is not necessarily directly maximized by preregistration, but rather the uncertainty in judging the theoretical risk is minimized.

To arrive at this interpretation, we rely on three arguments. The first is that theoretical risk is vital for judging theories. This argument comes natural to a (Neo-)Popperian or falsificationist, but is not self-evident for a confirmationist. The second argument is that

the theoretical risk for a given study is generally uncertain. The third and last assumption is that this uncertainty is reduced by applying preregistration.

When we put these arguments into the Bayes' formula, we conclude that because preregistration decreases uncertainty about the theoretical risk, which in turn increases our expectation to gain evidence for or against a theory, preregistration is warranted for any study. We will later discuss how our argument relates to a (Neo-)Popperian philosophy of science. For now, we stick to a Bayesian rationale since quantifying uncertainty is crucial for our argument, which would require considerable thought in a falsificationist framework.

## **Epistemic value and the Bayesian rationale**

Let us start by defining what we call expected epistemic value. If researchers plan to conduct a study, they usually hope it will change their assessment of some theory's verisimilitude (truthlikeness). In other words, they hope to learn something from conducting the study. The amount of knowledge researchers gain from a particular study concerning the verisimilitude of a specific theory is what we call epistemic value. While researchers can not know what exactly they will learn from a study, they can form an expectation that helps them decide which study to conduct. This expectation is what we term expected epistemic value. We assume three things about this estimation process and how it relates to choosing a study to conduct.

1. Researchers judge the evidence for or against a hypothesis rationally.
2. They expect other researchers to apply the same rational process.
3. All else equal, researchers try to maximize the expected epistemic value for other researchers.

The first assumption leads to our adoption of a Bayesian framework. Our rationale is as follows. Researchers who decide to conduct a study are akin to choosing a study to bet on. They have to "place the bet" by conducting the study, therefore, invest resources and stand to gain epistemic value with some probability. This conceptualization of choosing a study as a betting problem allows us to apply a "Dutch Book" argument (Christensen 1991). This argument states that any better must follow the axioms of probability to avoid being "irrational", i.e., accepting bets that lead to sure losses. Fully developing a Dutch book argument for this problem requires careful consideration of what kind of studies to include as possible bets, defining a conversion rate from the stakes to the reward, and modelling what liberties researchers have in choosing a study. Without deliberating these concepts further, we find it persuasive that researchers should not violate the axioms of probability if they have some expectation about what they stand to gain with some likelihood from conducting a study. The axioms of probability are sufficient to derive the Bayes formula, on which we will heavily rely for our further arguments. The argument is not sufficient, however, to warrant conceptualizing the kind of epistemic value we reason about in terms of probability; that remains a leap of faith. Please note that our decision

to adopt this aspect of the Bayesian philosophy of science does not imply anything about the statistical methods researchers use. In fact, this conceptualization is purposefully reductionistic to be compatible with a wide range of philosophies of science and statistical methods researchers might subscribe to.

## Epistemic Value and Theoretical Risk

Our first argument is that theoretical risk is crucial for judging theories. Put simply, risky predictions create persuasive evidence if they turn out to be correct. This point is crucial because we attribute much of the appeal of preregistration to this fact.

Let us make some simplifying assumptions and define notation. We restrict ourselves to evidence of a binary nature (either exists or does not) since continuous evidence would lead to some quite involved derivations. We denote the probability of a hypothesis before observing evidence as  $P(H)$  and its complement as  $P(\neg H) = 1 - P(H)$ . The probability of observing evidence under some hypothesis is  $P(E|H)$ . We can calculate the probability of the hypothesis after observing the evidence with help from the Bayes formula:

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

$P(H|E)$  is of great relevance since  $P(H|E)$  is often used directly or indirectly as a measure of corroboration of a hypothesis. In the tradition of Carnap (XXX), in its direct use, it is called corroboration as firmness; in its relation to  $P(H)$ , it is called increase in firmness. We refrain from discussing specific measures of corroboration since no measure shows universally better properties than others. However, it is generally expected that any measure of corroboration increases monotonically with an increase in  $P(H|E)$ .

In short, we want to increase  $P(H|E)$ . Increases in  $P(H|E)$  are associated with increased epistemic value, of which we want to maximize the expectation. So how can we increase  $P(H|E)$ ? The Bayes formula gives us three options to investigate, namely  $P(H)$ ,  $P(E|H)$  and  $P(E)$ . The first option leads us to the unsurprising conclusion that higher  $P(H)$  leads to higher  $P(H|E)$ . However, the a priori probability of a hypothesis is nothing our study design can change. The second option is similar commonsensical; that is, an increase in  $P(E|H)$  leads to higher  $P(H|E)$ . Consequently, researchers should ensure that their study design allows them to find evidence for their hypothesis, in case it is true. While this truism is strongly related to study design, it is unrelated to preregistration. Thus,  $P(E)$  remains to be considered. Since  $P(E)$  is the denominator, increasing it will decrease  $P(H|E)$ : The more unlikely it is to observe evidence, the more it increases the probability of the hypothesis if we do observe it. In other words, high risk, high reward.

If we equate riskiness with a low probability of obtaining evidence, the Bayesian rationale

perfectly aligns with the observation that risky predictions lead to persuasive evidence. This tension between high risk leading to high reward is central to our consideration of preregistration. A high risk, high reward strategy is bound to result in many losses that are eventually absorbed by the high gains. Sustaining many “failed” studies is not exactly aligned with the incentive structure under which many if not most researchers operate. Consequently, researchers have an incentive to appear to take more risks than they actually do, which misleads their readers to give their claims more credence than they deserve. It is at this juncture that the practice and mispractice of preregistration comes into play. We argue that the main function of preregistration is to enable proper judgment of the riskiness of a study.

To better understand how preregistrations can do that, let us take a closer look at what contributes to  $P(E)$ . Using the law of total probability, we can split  $P(E)$  into two terms:

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

We already have noted that there is not much to do about  $P(H)$  (and hence its counter probability  $P(\neg H)$ ), and that it is common sense to increase  $P(E|H)$ . The real lever to pull is, therefore,  $P(E|\neg H)$ . This probability tells us how likely it is that we find evidence when in fact, the theory is not true. Its counter probability  $P(\neg E|\neg H) = 1 - P(E|\neg H)$  is what we call “theoretical risk”, because it is the risk a theory takes on in predicting the occurrence of particular evidence in its favour.

Let us note some interesting properties of theoretical risk  $P(\neg E|\neg H)$ . First, increasing theoretical risk leads to higher  $P(H|E)$  (our objective). Second, if the theoretical risk is smaller than  $P(E|H)$  it follows that  $P(H|E)$  must decrease when observing the evidence. Third, if the theoretical risk equals zero, then  $P(H|E)$  can at best be equal to  $P(H)$  but only if  $P(H|E)$  is one. In other words, observing a sure fact does not lend credence to a hypothesis.

This sounds like a truism but is directly related to Popper’s seminal criterion of demarcation. He stated that if it is impossible to prove a hypothesis false ( $P(E|\neg H) = 1$  or  $P(\neg E|\neg H) = 0$ ), it can not be considered a scientific hypothesis (Popper 2002, 18). We note these relations to underline that the Bayesian rational we apply here is able to reconstruct many commonly held views on riskiness and epistemic value.

## Preregistration as a Means to Increase Theoretical Risk?

After we discussed that increasing the theoretical risk will increase the epistemic value, it is intuitive to task preregistration with maximizing theoretical risk. Indeed, limiting type I error rate is commonly stated as a goal of preregistration. We argue that while such a conclusion is plausible, we must first consider at least two constraints that place an upper bound on the theoretical risk.

First, the theory itself limits theoretical risk: Some theories simply do not make risky predictions, and preregistration will not change that. Consider the case of a researcher contemplating the relation between two sets of variables. Suppose each set by itself is well studied, and strong theories tell the researcher how the variables within the set relate. However, our imaginary researcher considers the relation between these two sets. For lack of a better theory, they assume that some relation between any variables of the two sets exists. This is not a risky prediction to make in psychology, even without statistical issues like alpha inflation (Orben and Lakens 2020). However, we would consider it a success if the researcher would use the evidence to develop a more precise (and therefore risky) theory, e.g. by specifying which variables from one set relate to which variables from the other set, to what extent, in which direction, with which functional shape etc. We will later show that preregistration increases the belief one can stake in the further specified theory, though it remains low till substantiated by testing it again. The point, however, is that we want to show that preregistration increases the expected epistemic value without regard to the theory being tested.

Second, available resources limit theoretical risk. Increasing  $P(\neg E|\neg H)$  will usually decrease  $P(E|H)$  unless more resources are invested. In other words, one can not increase power while maintaining the same type I error rate without increasing the invested resources. Tasking preregistration with an increase in theoretical risk makes it difficult to balance this trade-off. Mindlessly maximizing theoretical risk would either never produce evidence or require huge amounts of resources.

## Uncertainty about Theoretical Risk

We established that higher theoretical risk leads to more persuasive evidence. In other words, we have reconstructed the interpretation that preregistrations supposedly work by restricting the researchers, which in turn increases the theoretical risk (or equivalently limit type I error rate) and thereby creates more persuasive evidence. Nevertheless, there are trade-offs for increasing theoretical risk. However, employing a mathematical framework allows us to navigate the trade-offs more effectively and move towards a second, more favourable interpretation. To that end, we incorporate uncertainty into our framework.

## Statistical Methods

Preregistration is often discussed in relation to statistical methods, so let us note some relations between  $P(E|\neg H)$  and statistical methods. If you consider the overly simplistic case where the research hypothesis is equal to the statistical alternative hypothesis, then the null-hypothesis is  $\neg H$ , and therefore  $P(E|\neg H)$  is equivalent to the type I error rate. Researchers who choose a smaller type I error rate can hence be more sure of their results, if significant. However, the research hypothesis is seldomly equal to the statistical null hypothesis. We argue that  $P(E|\neg H)$  (and hence its inverse, theoretical risk)

also encompasses factors outside the statistical realm, most notably the study design and broader analytical strategies.

Statistical methods stand out among these factors because we have a large toolbox for assessing and controlling their contribution to  $P(E|\neg H)$ . Examples of our ability to exert this control are the setting of type I error rate, the use of corrected fit measures (i.e., adjusted  $R^2$ ), information criteria or cross-validation in machine learning. These are all tools that help us account for biases in statistical methods that increase the likelihood of signifying results even when there are none. They are thus closely related to theoretical risk.

The point is that the contribution of statistical methods to  $P(E|\neg H)$  can be formally assessed. For many statistical models it can be analytically computed under some assumptions. For those models or assumptions where this is impossible, one can employ Monte Carlo simulation to estimate the contribution to  $P(E|\neg H)$ . The precision with which researchers can discuss statistical contributions to  $P(E|\neg H)$  has lured the community concerned with research methods into ignoring other factors that are much more uncertain. We can not hope to resolve this uncertainty; but we have to be aware of its implications. These are relayed in the following.

### **Causes of Uncertainty**

As we noted, it is possible to quantify how statistical models affect the theoretical risk based on mathematical considerations and simulation. However, other factors in the broader context of the study are much harder to quantify. If one chooses to focus only on the contribution of statistical methods to  $P(E|\neg H)$ , one is bound to underestimate it. Take, for example, a t-test. Under ideal circumstances (assumption of independence, normality of residuals, equal variance), it stays true to its type I error rate. However, researchers might do many very reasonable things in the broader context of the study that affect theoretical risk: They might exclude outliers, choose to drop an item, enlarge their definition of the population to be sampled, translate their questionnaires, impute missing values, or any number of other things. All of these decisions carry a small risk that they increase the likelihood of obtaining evidence despite the underlying hypothesis being false. Even if the t-test itself perfectly maintains its type I error rate, these factors must be added to  $P(E|\neg H)$ . While, in theory, these factors may leave  $P(E|\neg H)$  unaffected or even decrease it, we argue that this is not the case in practice. Whether researchers want to or not, except under strict blinding, they continuously process information about how the study is going. While one can hope that processing this information does not affect their decision making either way, this cannot be secured. We, therefore, conclude that statistical properties only guarantee a lower bound to the quantity we seek to minimize. The only thing we can conclude with some certainty is that  $P(E|\neg H)$  is not lower than what the statistical model guarantees without knowledge about the other factors at play.

## The effects of uncertainty

Before we ask how preregistration is influencing this uncertainty, we must consider the implications of being uncertain about the theoretical risk. Within the Bayesian framework, this is both straightforward and insightful. To get an expectation, we express uncertainty as a probability distribution and then integrate over it:

$$\mathbb{E}(p(H|E)) = \int \frac{p(H)p(E|H)}{p(H)p(E|H) + p(\neg H)p(E|\neg H)} d\mathbb{P}(p(E|\neg H))$$

To illustrate the effect of uncertainty, let  $p(E|H) = .8$  (e.g., power of 80%) and  $p(H) = .1$  and assume a uniform distribution for  $p(E|\neg H)$  of the form:

$$f(x) = \begin{cases} \frac{1}{1-\alpha} & \text{for } \alpha \leq x \leq 1, \\ 0 & \text{for } x < \alpha \text{ or } x > 1 \end{cases}$$

Where  $\alpha$  is, e.g., the statistical model's nominal type I error rate.

We chose this uniform distribution to capture our statement that a statistical model only guarantees a lower bound to  $p(E|\neg H)$  (and since its a probability the upper bound is 1) as it is the maximum entropy distribution under this assumption and conforms therefore with our Bayesian framework (Giffin and Caticha 2007).

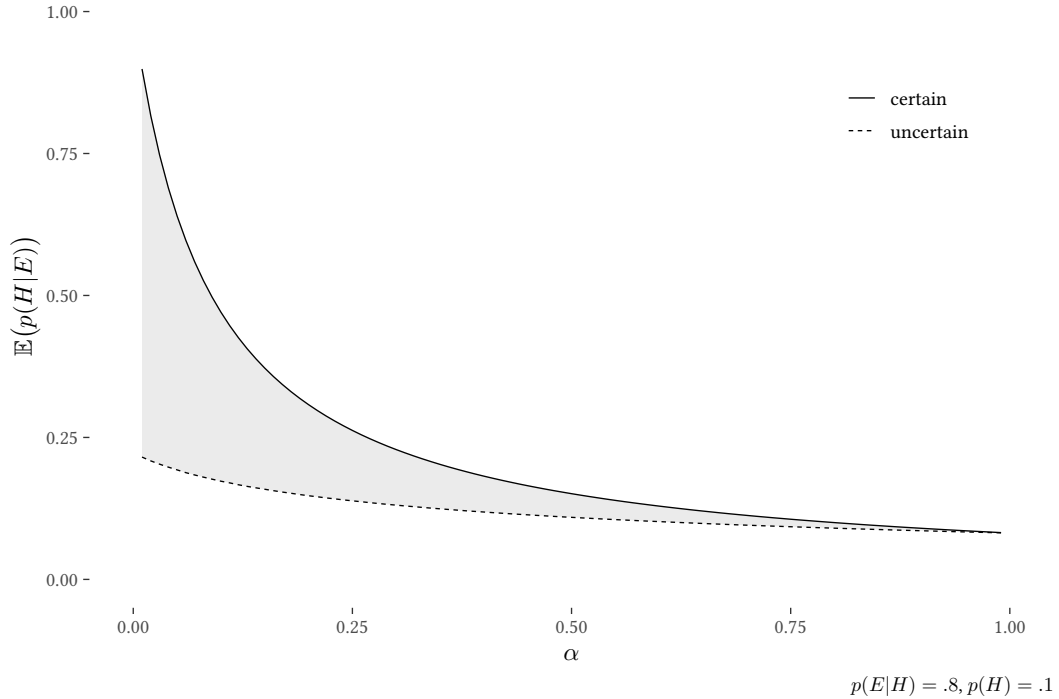


Figure 1: Corroboration as firmness ( $p(E|H)$ ) as a function of  $\alpha$ , where  $\alpha$  is either certain (solid line) or maximally uncertain (dotted line).

Neither absolute certainty nor uncertainty are realistic scenarios but represent the boundary conditions that all realistic conditions fall in. Depending on the distribution that





Figure 2: Several measures for corroboration as increase in firmness as a function of  $\alpha$ , where  $\alpha$  is either certain (solid line) or maximally uncertain (dotted line).

expresses our state of knowledge about the theoretical risk a study took on, our expectation for what to gain varies considerably. Uncertainty about theoretical risk is expressed through the variance (or rather entropy) of the distribution. Generally, we expect that increases in uncertainty (expressed as more entropic distributions) lead to a decreased expected epistemic value.

## Preregistration as a Means to Decreasing Uncertainty about the Theoretical Risk

We hope to have persuaded you to accept two arguments: First, the theoretical risk is important for judging evidential support for theories. Second, the theoretical risk is inherently uncertain, which diminishes the persuasiveness of the gathered evidence. The last argument is that preregistrations reduces this uncertainty.

Recollect, our three assumptions:

1. Researchers judge the evidence for or against a hypothesis rationally.
2. They expect other researchers to apply the same rational process.
3. All else equal, researchers try to increase the expected epistemic value for other researchers.

The point we make with these assumption is that the authors aim to persuade other researchers, their readers. Unfortunately, the case for lack of insight into the myriad of factors that influence theoretical risk is particularly strong for the reader of the resulting

article (or, more generally, the consumer of the research product). We have to remember that the process of weighing evidence for or against a theory extends beyond the original authors to all the people they hope to persuade. While the authors may have deep insight into what they did and how it might influence the theoretical risk they took, their readers have much greater uncertainty about these factors. In particular, they never know what relevant factors the authors, intentionally or not, failed to disclose. From the perspective of an ultimate sceptic, they may claim maximum uncertainty.

Communicating clearly what the authors did to arrive at the evidence is crucial for judging the theoretical risk they took. Preregistrations are ideal to communicate just that, because any description after the fact is suspect to be incomplete. The authors could have decided to exclude a number of analytic strategies they tried out etc. That is not to say that any study that was not preregistered was subjected to practices of p hacking. The point is, we can not exclude this and a myriad of other possibilities and, hence, are left uncertain. If the researcher do describe what they intend to do beforehand, and then report they did exactly that, the readers can be certain, that they have received a complete account of the situation. They still might be uncertain about the actual theoretical risk the authors took but much less so. Remaining sources of uncertainty might be unfamiliarity with statistical methods or experimental paradigms, the probability of plain error in analyses etc. In any case a well written preregistration should aim to reduce the uncertainty about the theoretical risk and hence increase the persuasiveness of evidence.

## Discussion

We started out with the observation that preregistrations do not always cleanly divide preregistration from conformation. Enforcing this distinction would mean to use preregistrations as a means to increase the theoretical risk researcher take on. No doubt this results in more trustworthy evidence and may even increase the effectiveness of scientific undertakings in general. However, as the term implies, directly maximizing theoretical risk increases the likelihood that a study fails to deliver results favoring the theory. As we showed, minimizing the uncertainty around the theoretical risk shares the beneficial properties without increasing the likelihood of a “failed” study. Let us note that a study that does not result in the evidence a researcher is hoping for is still precious for the scientific process. Under the current incentive structure, however, many researchers have reason to avoid such outcomes and few, if any, have the privilege to operate outside of these incentives.

The advantage of this perspective that puts uncertainty about theoretical risk front and center is twofold. First, aiming to reduce uncertainty is beneficial across a wide range of potential theoretical risks. That is, researchers benefit from preregistrating their study, regardless whether or not the study is “exploratory”. If researchers conduct a more exploratory study they can clearly communicate how exploratory they aim to be and their results can be judged accordingly. Second, if researchers realize after the fact, that it

would be unwise to follow their preregistration in certain details, they may change it. Allowing to deviate from the preregistration is controversial (XXX) because the authors might sift through all the possibilities and then select the most beneficial for their theory. This argument, in its fullest consequence, might result in the same uncertainty we argued is appropriate for not preregistered studies. However, if the authors offer a convincing argument for their deviation, their readers might be compelled to increase their uncertainty only a limited amount.

## Old Material

As far as I am aware, Mayo's severity argument currently provides one of the few philosophies of science that allows for a coherent conceptual analysis of the value of preregistration. Borsboom

We, therefore, propose an interpretation of preregistration that increases the expected epistemic value of studies without necessarily increasing the theoretical risk.

If you are a follower of a hypo-deductive rational this is self evident, if you are a positivist we hope our

The other two assumptions are later necessary to connect individual decisions that affect epistemic value to the research community as a whole.

Observe, that since  $P(E|\neg H)$ .

Please note, that we adopt a Bayesian rational for the meta scientific process of preregistration but that this does not imply any ties to the methods a researcher uses.

However, we aim to show that preregistration is indispensable for adequately judging a much broader range of studies. To that end, we first rationalize the appreciation for confirmation using Bayesian reasoning. Equipped with some of the basic tools of Bayesian Philosophy of Science, we can move beyond a simple dichotomy of exploration and confirmation.

To that end we show that the appreciation for the distinction between exploration and confirmatory neatly falls in line with Bayesian reasoning connected through a quantity we call *theoretical risk*. Then we proceed to show that preregistration impacts the theoretical risk via two pathways. Only considering the one path rationalizes recommendation to employ preregistration only for confirmatory studies, taking both paths together into consideration leads to a much wider applicability of preregistration.

But what exactly have you been cheated of? This appreciation of confirmatory results falls neatly in line with Bayesian reasoning (XXX). In order to apply Bayesian reasoning we have to cast confirmation and exploration in terms of probability. Consider the hypothesis that people who regularly engage in fitness activities are healthier. A confirmatory story might operationalize regularly as weekly, fitness activity as light jogging

and healthy as blood cholesterol level. A more exploratory study might consider several definitions of “regularly”, “fitness activities”, and “healthy”, e.g. “monthly”, “weekly”, “daily”, or “marathon”, “light jogging”, “intense cleaning”, or “cholesterol”, “self rated health”, “ability to play tag”. Both studies have some probability to turn up with positive results even in a world where fitness activities do not lead to healthier people, because there are other plausible explanations for the results. However, for the second study the number of explanations that could lead to evidence in favor, aside from the theory is much greater. Suppose a researcher actually conducted the second study but is reporting it as if they conducted the first study. You would feel cheated. But what exactly have you been cheated of? In fact, you have been cheated out of the information about all the possibilities that could have brought about this evidence. We summarize these possibilities as a probability. Therefore, we argue that confirmatory studies have a low probability of observing evidence in favor of their theory if we assume the theory to be false. Exploratory studies on the other hand are characterized by a higher probability that they find evidence, even when their hypothesis turns out to be wrong.

We connect the question of exploration vs confirmation to Bayesian reasoning via a quantity we call theoretical risk (we borrow the term from Meehl).

Theoretical risk is the inverse of the probability that one observes evidence in favor of a theory when we assume that the theory does not hold. We assume that confirmatory studies take a high theoretical risk, compared to exploratory studies which take a low theoretical risk. The tighter your definition of “evidence in favor” is, the more the probability of observing the evidence is tied to your theory holding.

Let us formalize the judgment about a hypothesis (H) given some evidence (E) as the conditional probability  $p(H|E)$ .

Using Bayes’ rule we arrive at:

$$p(H|E) = \frac{p(H)p(E|H)}{p(H)p(E|H) + p(\neg H)p(E|\neg H)}$$

Note the connection to null hypothesis testing. If H represents the alternative Hypothesis and  $\neg H$  represents the Null hypothesis, then  $p(E|H)$  is the power, while  $p(E|\neg H)$  represents the type I/alpha error. However, generally a theory is richer in content, than equating it with the alternative hypothesis in Null hypothesis testing.

The first paradigm is embraced by the open science community and conceptualizes pre-registration as a tool to reduce the type I error rate (XXXX). As we will later show, reducing the type I error rate is indeed a powerful lever to gather persuasive evidence. In practice, however, researchers can not reduce the type I error rate at will. They face three limiting factors. First, psychological theories tend to be vague. Some decisions are just not derivable from theory, which makes it extraordinarily difficult to maintain a type I error rate near zero. Second, researchers are constrained by resource limitations.

One way to counter the first problem is conducting exploratory pilot studies, but running extensive pilot studies is expensive. Doubling the resources that go into investigating a research question is often just not tenable and mostly beyond the individual researcher to decide. Third, by reducing the type I error rate the researcher directly increases the likelihood that their study will result in disappointing null results. A study that fails to deliver results can be crippling to the researchers career. Depending on individual circumstances, e.g. the theory in question, resources allocated to a study, the career stage of the researcher, a researcher may simply feel unable to reduce the type I error rate to the threshold necessary for a confirmatory study. Constrained by these factors researchers tasked with writing a preregistration are incentivized to either give up or to muddle through by being intentionally vague. The latter practice leads to preregistrations stating unspecific assumptions such as “we expect some of the Xs to correlate with some of the Ys”.

Our conclusions can be applied to bayesian and frequentist methods alike. Of course, frequentist methods are traditionally viewed from a Popperian (or hypo-deductive) philosophy of science, but our conclusions are compatible with this view and notable later advancements like the error statistical view. However, error statistical view marries a statistical philosophy with specific statistical models and procedures. A marriage which leaves us with richer set of assumptions than necessary for the points we want to make about the value of preregistration.

## References

- Bakker, Marjan, Coosje L. S. Veldkamp, Marcel A. L. M. van Assen, Elise A. V. Cromptvoets, How Hwee Ong, Brian A. Nosek, Courtney K. Soderberg, David Mellor, and Jelte M. Wicherts. 2020. “Ensuring the Quality and Specificity of Preregistrations.” *PLOS Biology* 18 (12): e3000937. <https://doi.org/10.1371/journal.pbio.3000937>.
- Christensen, David. 1991. “Clever Bookies and Coherent Beliefs.” *The Philosophical Review* 100 (2): 229–47. <https://doi.org/10.2307/2185301>.
- Giffin, Adom, and Ariel Caticha. 2007. “Updating Probabilities with Data and Moments.” In *AIP Conference Proceedings*, 954:74–84. <https://doi.org/10.1063/1.2821302>.
- Goldin-Meadow, Susan. 2016. “Why Preregistration Makes Me Nervous.” *APS Observer* 29 (August).
- Hoyningen-Huene, Paul. 2006. “Context of Discovery Versus Context of Justification and Thomas Kuhn.” In *Revisiting Discovery and Justification: Historical and Philosophical Perspectives on the Context Distinction*, edited by Jutta Schickore and Friedrich Steinle, 119–31. Archimedes. Dordrecht: Springer Netherlands. [https://doi.org/10.1007/1-4020-4251-5\\_8](https://doi.org/10.1007/1-4020-4251-5_8).
- Meehl, Paul E. 1978. “Theoretical Risks and Tabular Asterisks: Sir Karl, Sir Ronald, and the Slow Progress of Soft Psychology.” *Journal of Consulting and Clinical Psychology* 46 (4): 806–34. <https://doi.org/10.1037/0022-006X.46.4.806>.
- Mellor, David T., and Brian A. Nosek. 2018. “Easy Preregistration Will Benefit Any Re-

- search.” *Nature Human Behaviour* 2 (2): 98–98. <https://doi.org/10.1038/s41562-018-0294-7>.
- Nosek, Brian A., Charles R. Ebersole, Alexander C. DeHaven, and David T. Mellor. 2018. “The Preregistration Revolution.” *Proceedings of the National Academy of Sciences* 115 (11): 2600–2606. <https://doi.org/10.1073/pnas.1708274114>.
- Orben, Amy, and Daniël Lakens. 2020. “Crud (Re)Defined.” *Advances in Methods and Practices in Psychological Science* 3 (2): 238–47. <https://doi.org/10.1177/2515245920917961>.
- Popper, Karl R. 2002. *The Logic of Scientific Discovery*. London; New York: Routledge.
- Simmons, Joseph P., Leif D. Nelson, and Uri Simonsohn. 2021. “Pre-Registration: Why and How.” *Journal of Consumer Psychology* 31 (1): 151–62. <https://doi.org/10.1002/jcpy.1208>.
- Wagenmakers, Eric-Jan, Ruud Wetzels, Denny Borsboom, Han L. J. van der Maas, and Rogier A. Kievit. 2012. “An Agenda for Purely Confirmatory Research.” *Perspectives on Psychological Science* 7 (6): 632–38. <https://doi.org/10.1177/1745691612463078>.