Preregistration as a Tool to Evaluate the Severity of a Test

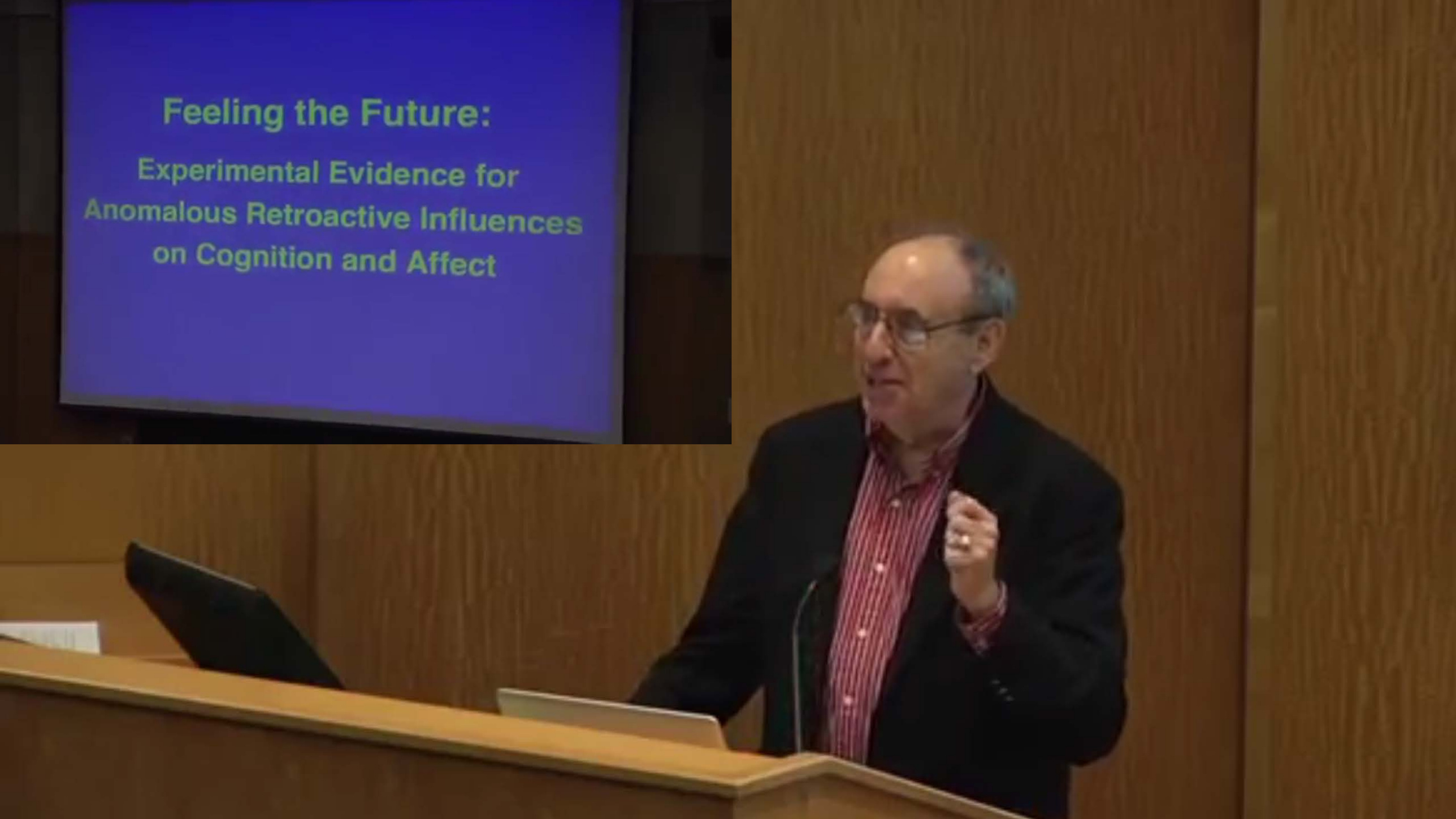
Daniël Lakens

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

NWO Nederlandse Organisatie voor Wetenschappelijk Onderzoek

Now here there was a difficulty. The test of significance is not nearly so automatic an inference process as had been thought. It is manifestly contingent on the decision of the investigator as to whether to run a one- or a two-tailed test. And somehow, making the decision *after* the data were collected and the means computed, seemed like "cheating." How should this be handled? Should there be some central registry in which one registers one's decision to run a one- or two-tailed test before collecting the data? Should one, as one eminent psychologist once suggested to me, send oneself a letter so that the postmark would prove that one had pre-decided to run a one-tailed test?

Bakan, 1966


Feeling the Future:
Experimental Evidence for
Anomalous Retroactive Influences
on Cognition and Affect

Results and Discussion

Across all 100 sessions, participants correctly identified the future position of the erotic pictures significantly more frequently than the 50% hit rate expected by chance: 53.1%, $t(99) = 2.51$, $p = .01$, $d = 0.25$.[3] In contrast, their hit rate on the nonerotic pictures did not differ significantly from chance: 49.8%, $t(99) = -0.15$, $p = .56$. This was true across all types of nonerotic pictures: neutral pictures, 49.6%; negative pictures, 51.3%; positive pictures, 49.4%; and romantic but nonerotic pictures, 50.2%.

Bem, 2011

Do you trust Daryl Bem predicted an effect of erotic pictures? And if not, why should anyone trust you?

Who cares if the effect was predicted? **Severe testers care.**

“The main way a theory gets money in the bank is by predicting facts that, absent the theory, would be antecedently improbable.”

Meehl, 1990

Why do we care about prediction? Do we want **temporal** novelty?

No.

Mayo, 2018

3) A scientific prediction may be made *with respect to events in the past as well as in the present or in the future.* In this connection psychologists sometimes differentiate between 'prediction' and 'postdiction,' but this is not a fundamental distinction. The justification of the term prediction (= foretelling) is that, in principle, the verifying procedure is situated in the future and has a still unknown outcome, which can therefore be predicted. The outcome itself, however, may well be the result of events in the remote past. A historian, for example, will predict that, *upon investigation* of hitherto imperfectly studied texts, *it will be found* that in the year... Charlemagne attempted...; or a geologist will predict on the strength of a theory that, in certain strata of certain localities, certain fossils so many millennia old will be found.

De Groot, 1969

Why do we care about prediction? Do we want **theoretical** novelty? **No**.

Mayo, 2018

Why do we care about prediction? Do we want **use** novelty? **Maybe,** but it’s a bit vague.

Mayo, 2018

The real underlying question is: Is the severity of the test violated?

Mayo, 2018

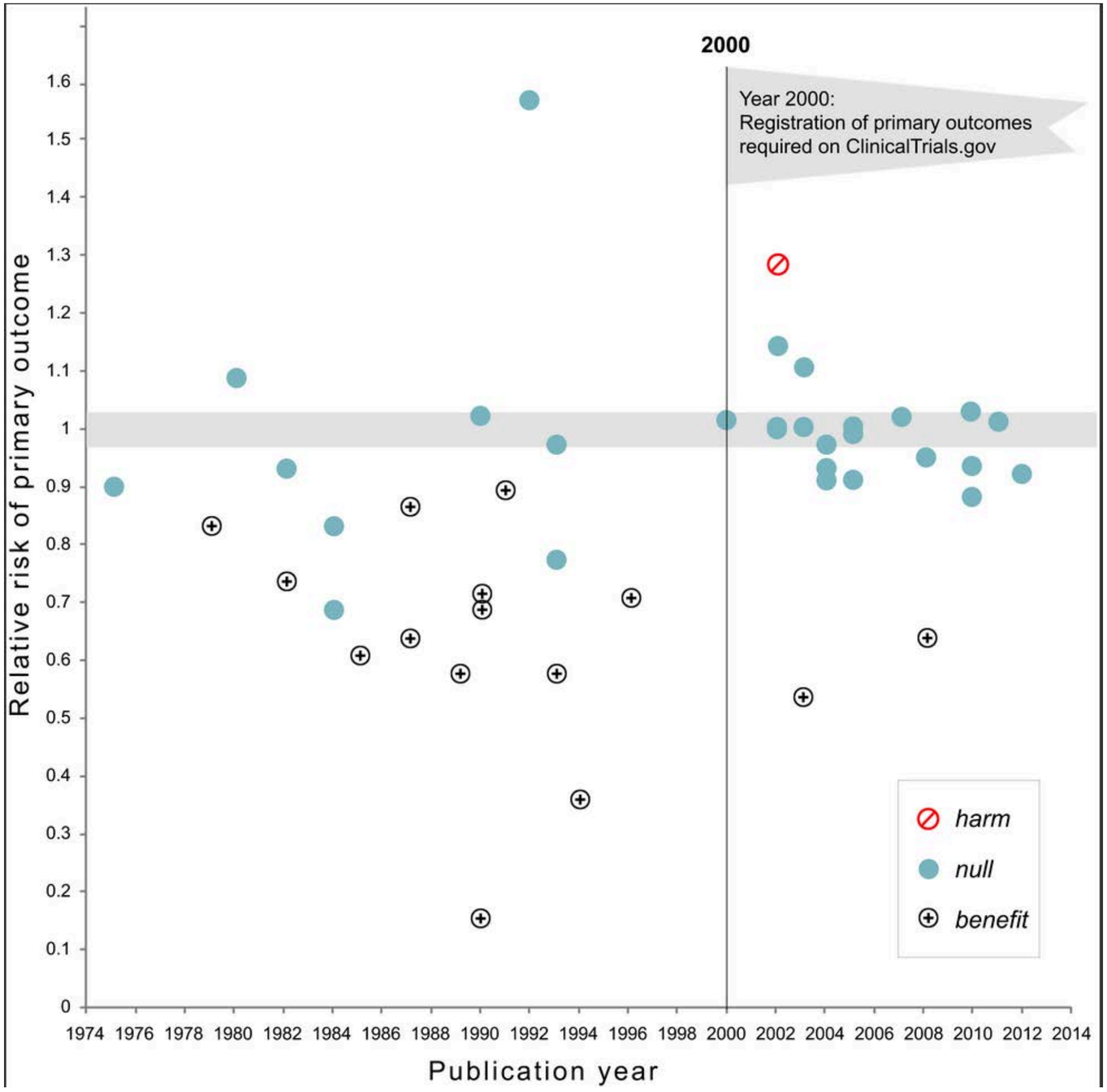
Preregistration has a long history in science – but the replication crisis and rise of the internet created momentum in psych.

Wiseman, Watt, & Kornbrot, 2019

Preregistration

Pre-registration formalizes error control.

Often raised in the context of Type 1 errors, but just as relevant for Type 2 errors (e.g., not finding effects if they exist).



Prior to 2000, 17/30 studies the number of null results in large National Heart Lung, and Blood Institute funded trials showed an effect. After pre-registration became required, only 2/25.

Kaplan & Irvin, 2015

Whenever an investigation is partly designed for hypothesis testing and partly of an exploratory nature — which is a not infrequent occurrence (cp. 4;2;3) — a strict differentiation should be maintained between these two elements. In particular, this applies to the publication of results. It is a serious offense against the social ethics of science to pass off an exploration as a genuine testing procedure. Unfortunately, this can be done quite easily by making it appear as if the hypotheses had already been formulated before the investigation started. Such misleading practices strike at the roots of 'open' communication among scientists.

De Groot, 1969

“In order to provide hard evidence for or against an empirical proposition, one has to resort to strictly confirmatory studies. The degree to which the scientific community will accept semiconfirmatory studies as evidence depends partly on the plausibility of the claim under scrutiny.”

Wagenmakers, Wetzels, Borsboom, van der Maas, 2012

Registered Reports do not prevent or discourage exploratory analysis. Rather, they make clear the distinction between confirmatory and exploratory analysis. This applies to registered reports whether they are conducted for replications or original research. Confirmatory results follow a preregistered analysis plan and thereby ensure interpretability of the reported *p*-values (Wagenmakers et al., 2012). In exploratory analysis, *p*-values lose their meaning due to an unknown inflation of the alpha-level. That does not mean that exploratory analysis is not valuable; it is just more tentative.

Nosek & Lakens, 2014

What Registered Reports Do Not Do

Preregistration and peer review in advance of data collection or analysis do not lead to definitive results. Even highly powered designs – like those in this issue – leave room for Type 1 and Type 2 errors. Furthermore, when registered reports are used for replication studies, different results between original and replication research could mean that there are unknown moderators or boundary conditions that differentiate the two studies. As such, the replication can raise more questions than it answers.

Nosek & Lakens, 2014

So initially, preregistration was presented purely as an approach to make it possible to ‘audit’ a *p*-value, and check if it they did not lose their “error probing capacity” (Mayo, 2018).

However, researchers also too easily assumed preregistered studies are always more compelling:

"This is particularly important if one wants to convince a skeptical audience of a controversial claim: After all, **confirmatory studies are much more compelling than exploratory studies**."

Wagenmakers, Wetzels, Borsboom, van der Maas, 2012

Taken together, these practices *[reducing p-hacking and publication bias, and power analysis]* will ensure that articles published as *Registered Reports* have a substantially **higher truth value** than regular studies. Such articles can therefore be expected to be more replicable and have a greater impact on the field.

Chambers, NeuroChambers blog, 2012

Preregistration clarifies the distinction between planned and unplanned research by reducing unnoticed flexibility. **This improves credibility of findings** and calibration of uncertainty.

Nosek, Beck, Campbell, Flake, Hardwicke, Mellor, van ‘t Veer, Vazire, 2019

In practice, confirmatory might be **much more compelling**, have **improved credibility of findings**, and **higher truth value.** They might also not.

By not presenting this as an *empirical possibility on average*, people promoting registered reports upset some of their peers. Multiple people have made a point along the lines of:

“Thus, the danger in focusing on the achievement of methodological precision is that we forget to consider critically whether we have actually examined the issues at hand in the best possible way, or whether we even are asking the appropriate questions. To the extent that this is the case, we are at risk of becoming **methodological fetishists**”

Ellemers, 2013

The discussion about costs and benefits of preregistration has been hindered by a lack of a conceptual analysis of what preregistration aims to accomplish.

So what is preregistration for?

Preregistration adds value for people who, based on their philosophy of science, increase their trust in claims that are supported by severe tests and predictive successes.

Lakens, 2020

Preregistration has the goal to allow others to transparently evaluate the capacity of a test to support or falsify a prediction.

Lakens, 2020

Preregistration itself does not make a study better or worse compared to a non-preregistered study.

Lakens, 2020

For example, preregistering a tautological prediction (the theory of reasoned action predicts reasoned action) is not a severe test.

Fiedler, 2004

Let alone that many preregistrations are of very low quality (which social norms currently do not allow you to say out loud, because “they tried”).

Let alone that many preregistrations are of very low quality (which social norms currently do not allow you to say out loud, because “they tried”).

However, preregistration makes it *relatively* more likely that tests have desired long run error rates, and thus makes it *relatively* more likely that tests are more severe.

But Daniël, surely that is not the only thing preregistration for?

Yes, that is all it does. Preregistration *only* has the goal to allow others to transparently evaluate the capacity of a test to support or falsify a prediction.

Sure, completing a preregistration has other consequences. Thinking through your study might improve it.

But there is no need to publicly post the preregistration to reap that benefit. They are simply **positive externalities**.

But there is no need to publicly post the preregistration to reap that benefit. They are simply **positive externalities**.

Why does this distinction matter? If the use of a tool is detached from a philosophy of science it risks becoming a heuristic.

Researchers should not choose to preregister because it has become a new norm, they should preregister because it supports their goal: **severe tests**.

Thank you

@Lakens

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

NWO Nederlandse Organisatie voor Wetenschappelijk Onderzoek