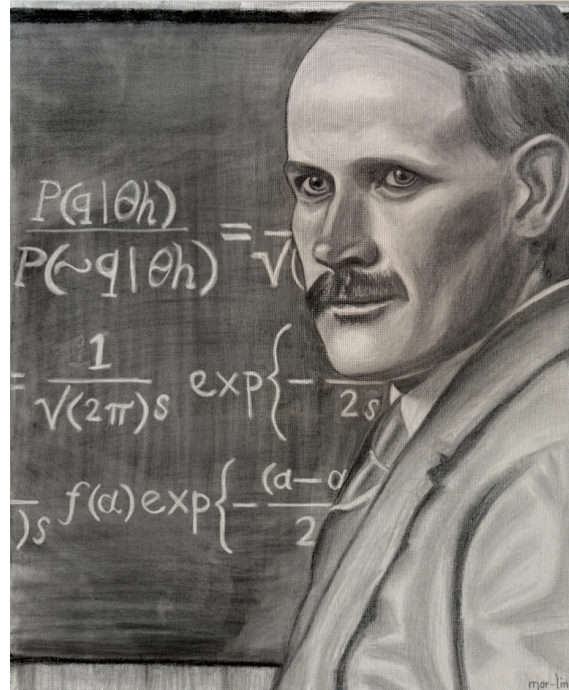


The Ockham Fraction



Harold Jeffreys
Painting, CC-BY:
Marlijn Bouwman

E.-J. Wagenmakers



Bio

- ◆ Psychological Methods Unit @ UvA
- ◆ Main interests:
 - Bayesian inference
 - Philosophy of science
 - Open science
 - Open-source statistical software (JASP)



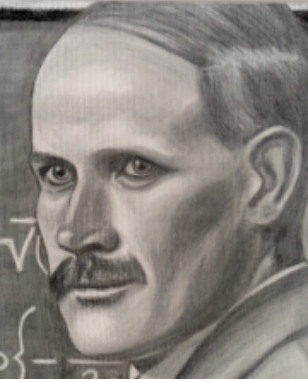
Outline

- ◆ Introductory remarks
- ◆ Senn's stubborn mule
- ◆ Ockham & prior-data conflict
- ◆ Ockham & the Jeffreys-Lindley paradox
- ◆ Ockham & the role of informed priors
- ◆ The more pressing problem



Outline

- ◆ **Introductory remarks**
- ◆ Senn's stubborn mule
- ◆ Ockham & prior-data conflict
- ◆ Ockham & the Jeffreys-Lindley paradox
- ◆ Ockham & the role of informed priors
- ◆ The more pressing problem



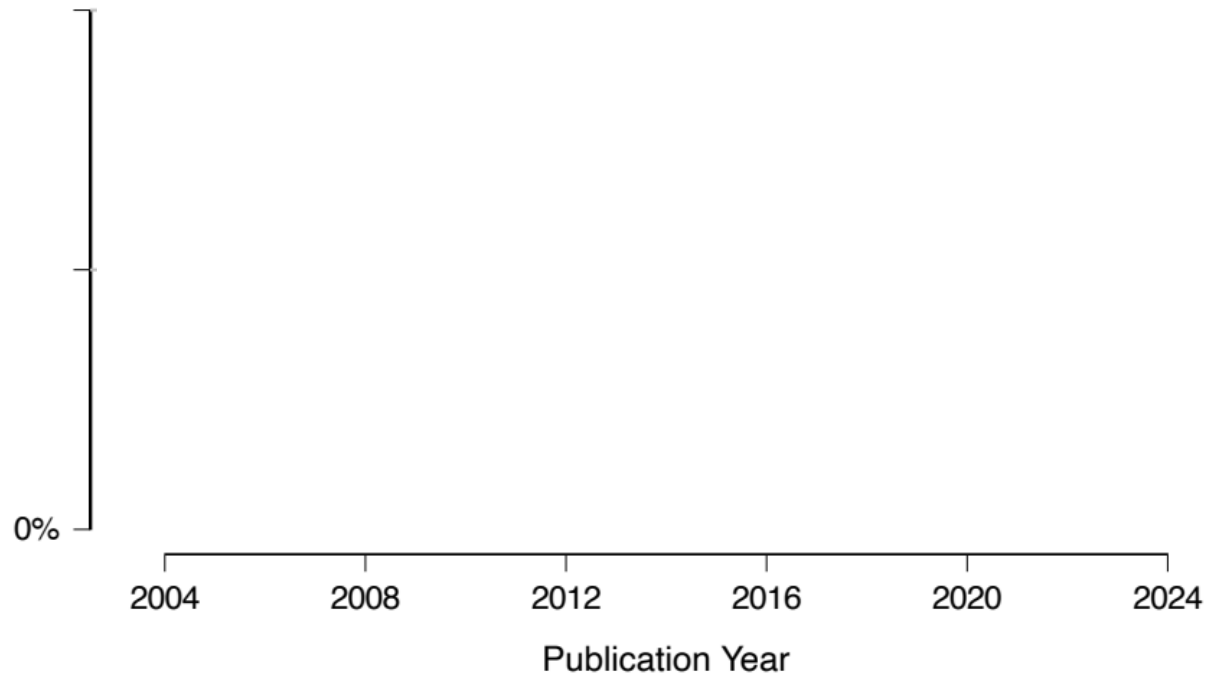
The Bayes Factor

- ◆ The Bayes factor quantifies the extent to which the observed data are better predicted under the null hypothesis than under an alternative hypothesis.
- ◆ “This approach is rarely used in the empirical sciences” (EJ in Munich, 2023, based on previous experience)



Julius Pfadt

University of Amsterdam



Articles with a Bayesian Analysis in Psychonomic Bulletin & Review

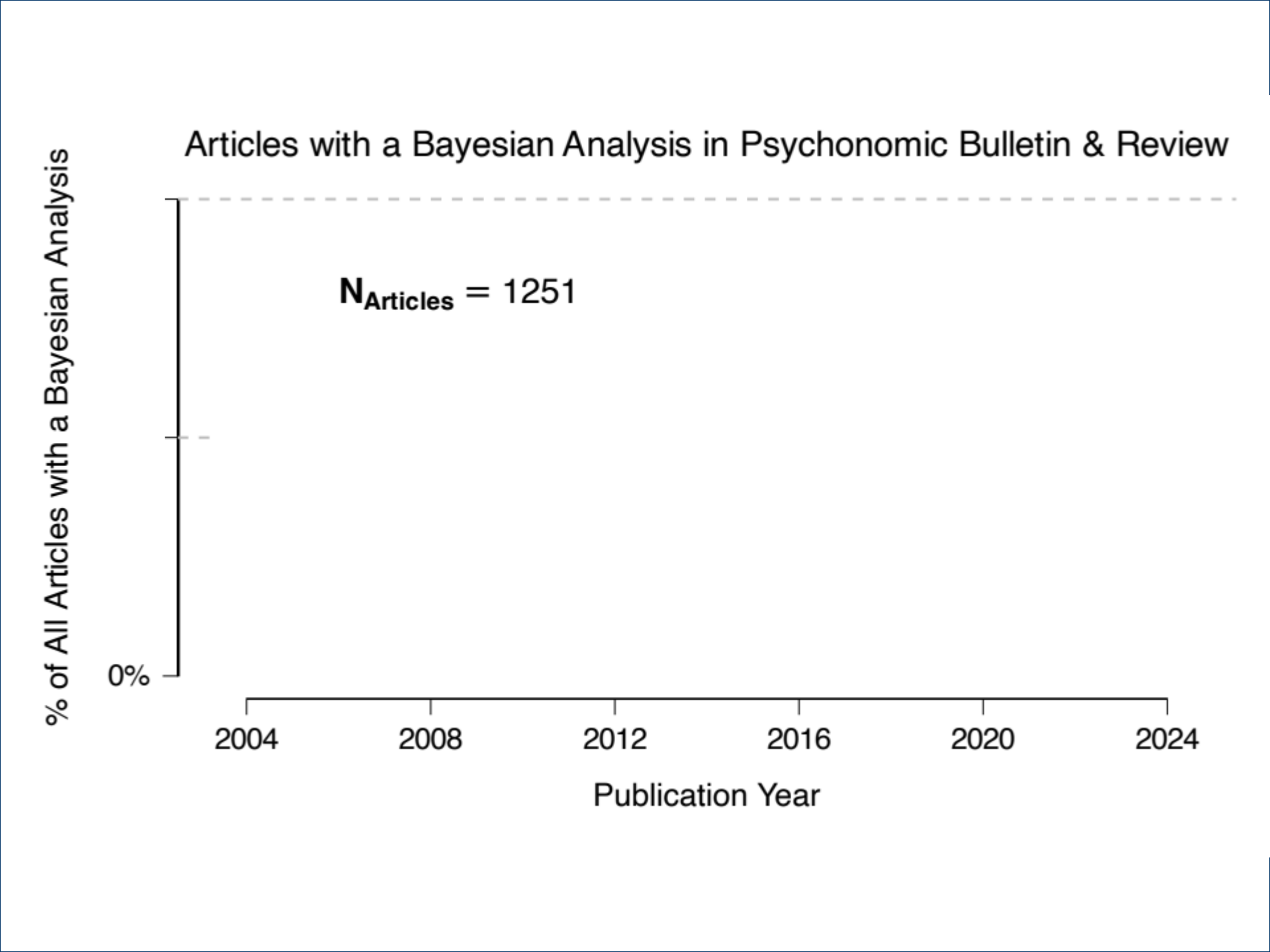
$N_{\text{Articles}} = 1251$

% of All Articles with a Bayesian Analysis

0%

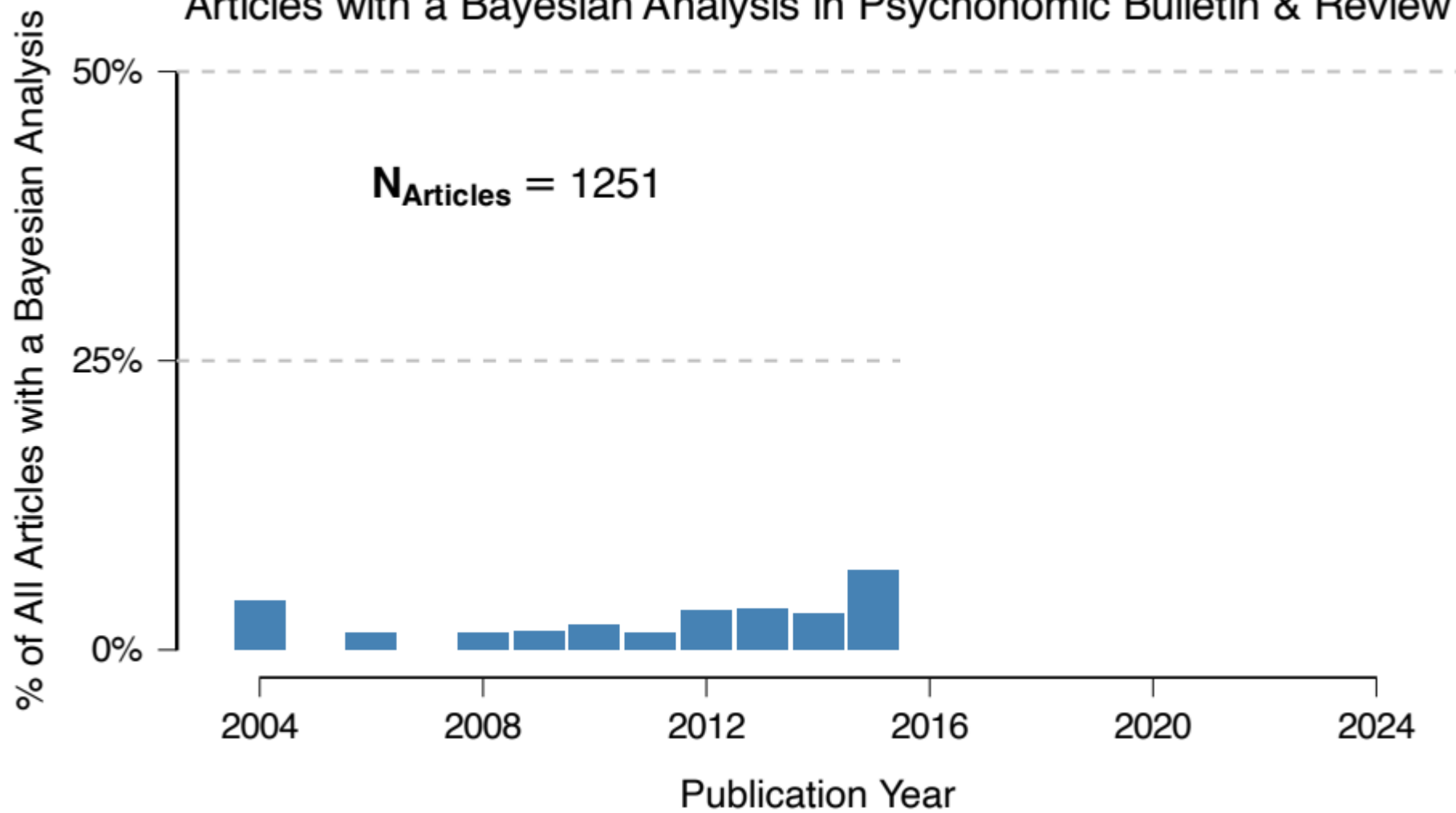
2004 2008 2012 2016 2020 2024

Publication Year



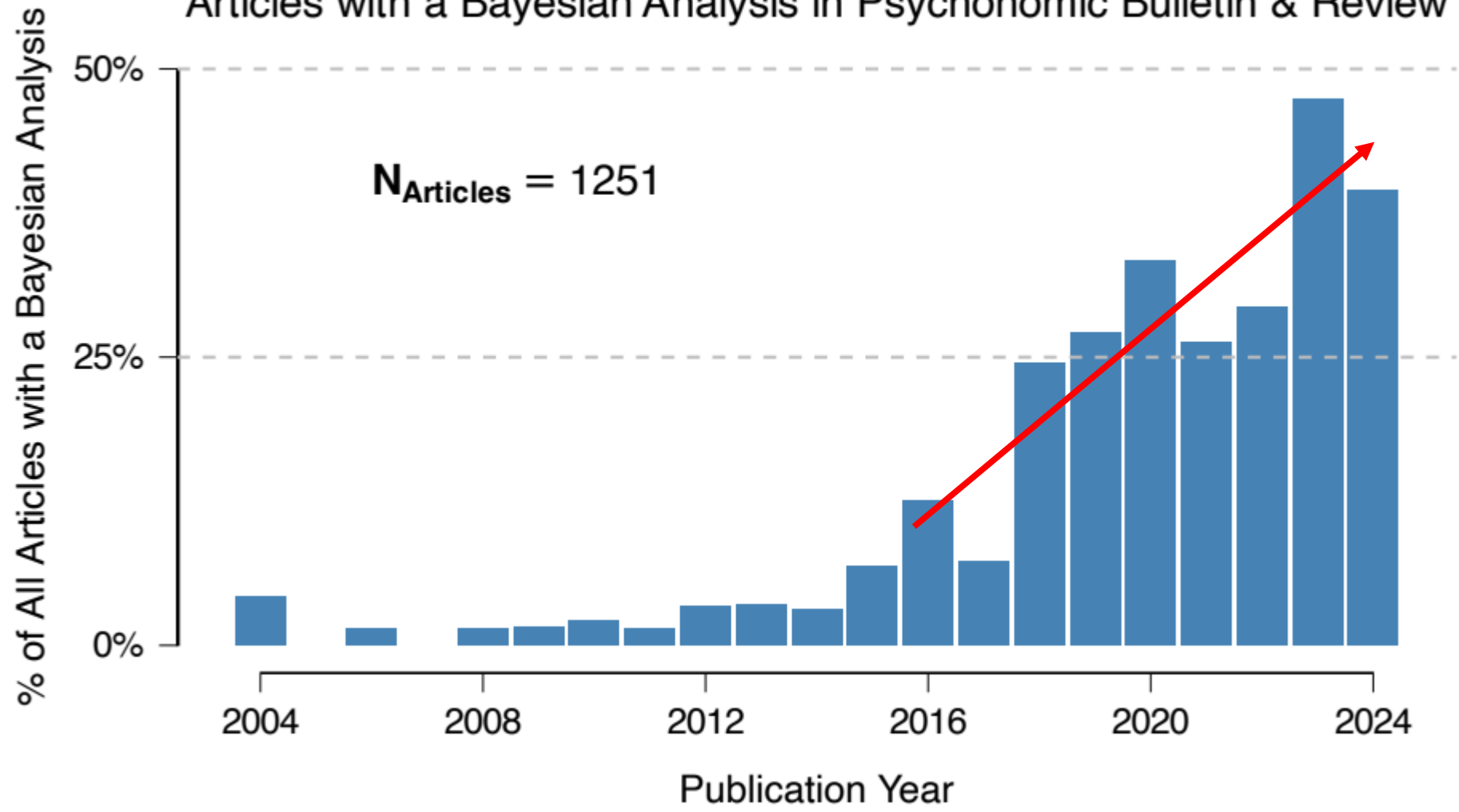
Articles with a Bayesian Analysis in Psychonomic Bulletin & Review

$N_{\text{Articles}} = 1251$



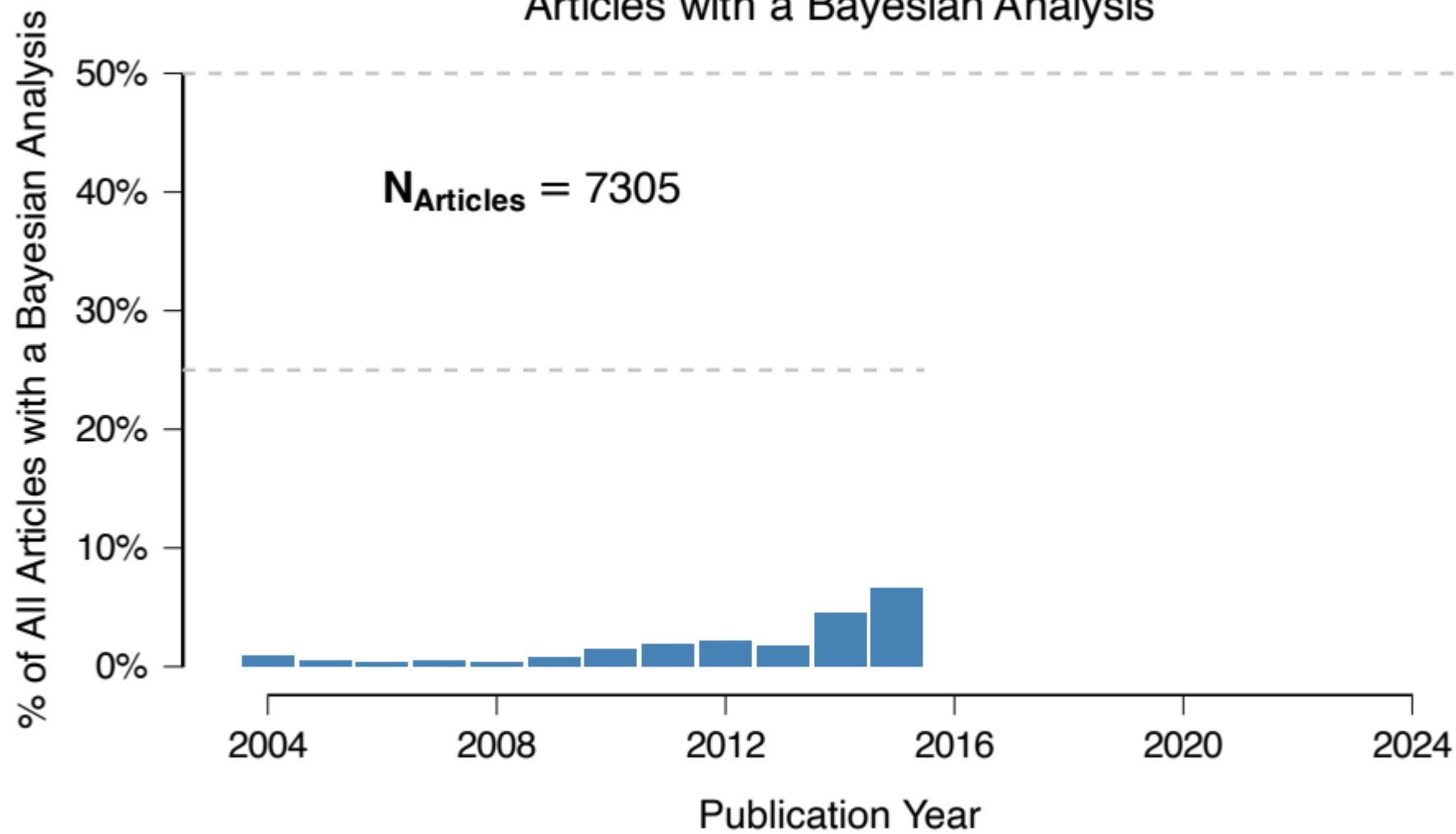
Articles with a Bayesian Analysis in Psychonomic Bulletin & Review

$N_{\text{Articles}} = 1251$



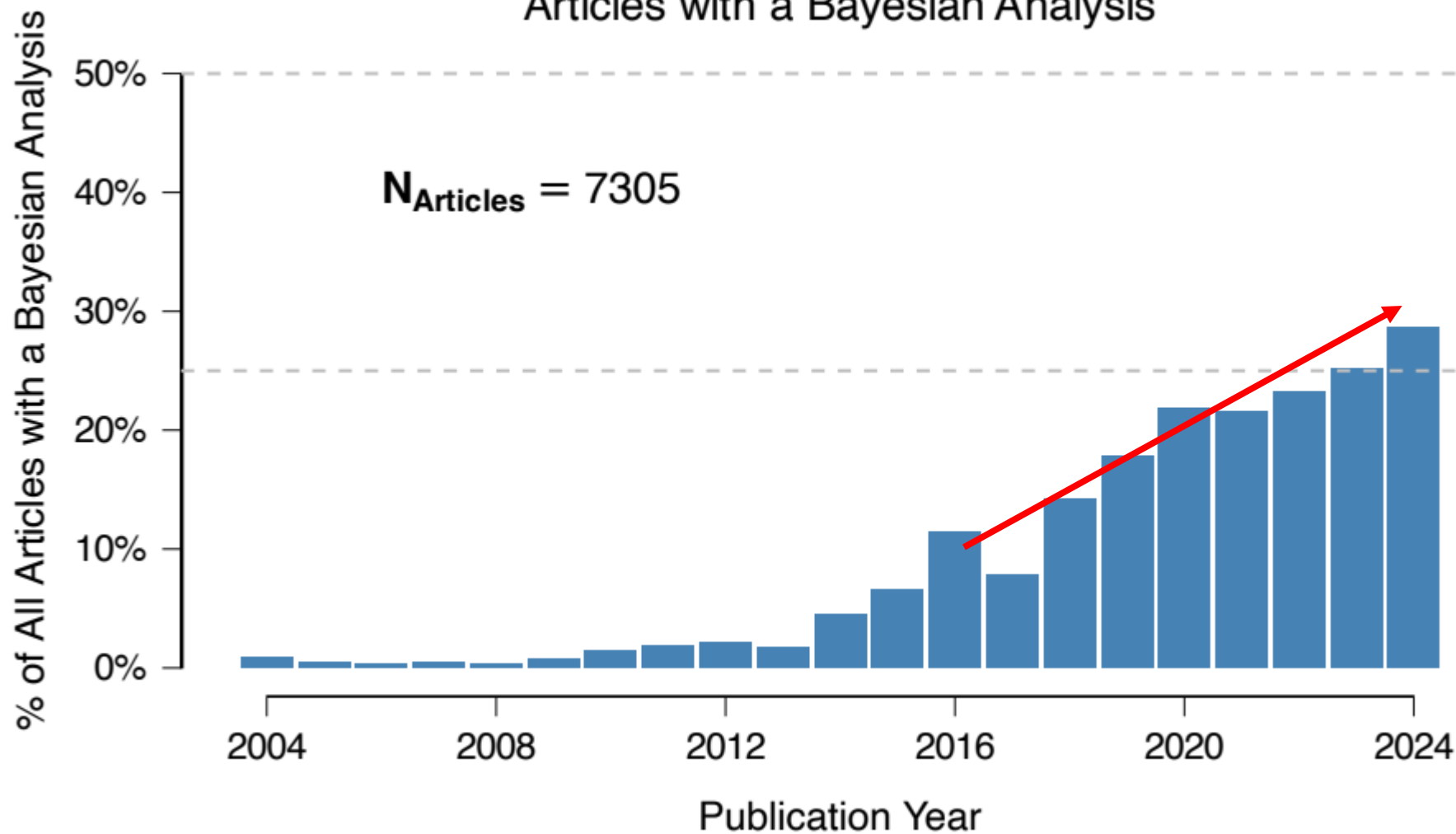
Articles with a Bayesian Analysis

$N_{\text{Articles}} = 7305$

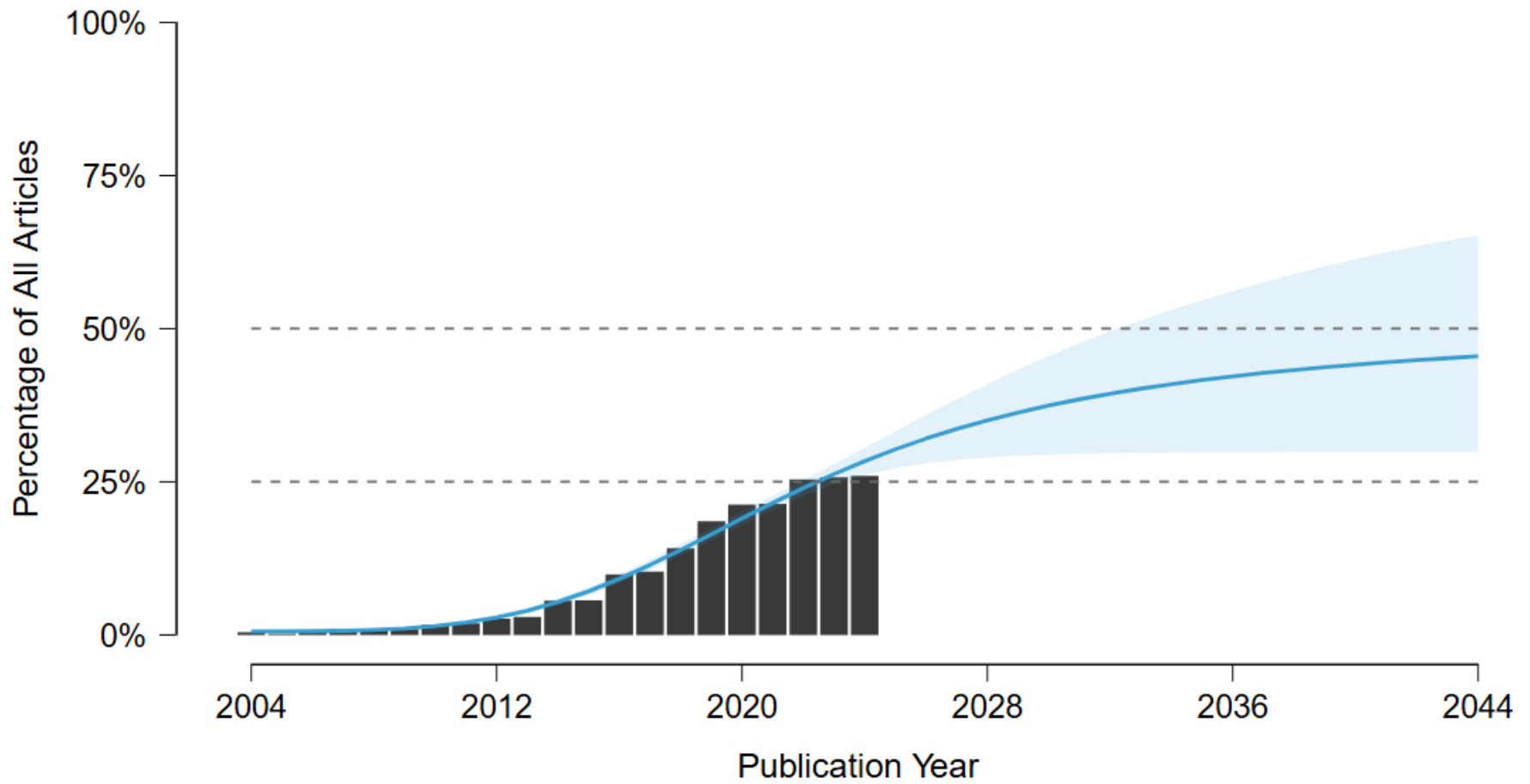


Articles with a Bayesian Analysis

$N_{\text{Articles}} = 7305$





*Observed and Predicted Percentage of **Bayesian** Articles Across Six Psychology Journals from 2004–2044*





Harold Jeffreys (1891-1989)

Painting, CC-BY: Marlijn Bouwman



*Simple models
tend to make precise
predictions*

MacKay





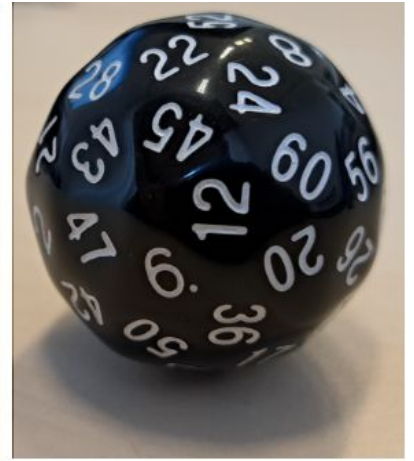
D3



D6



D12



D60



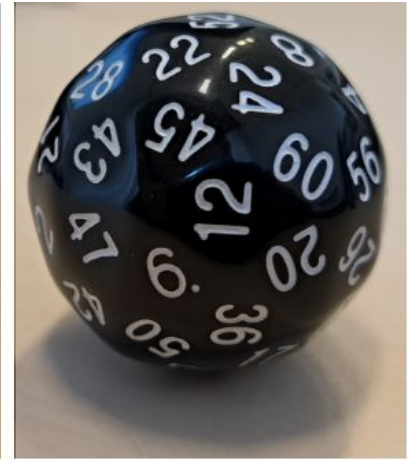
D3



D6



D12

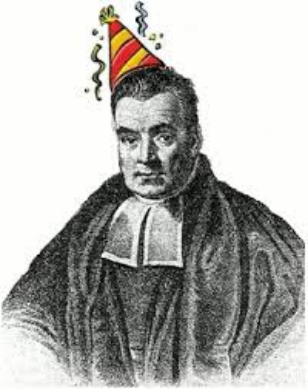


D60

You see the following outcomes:

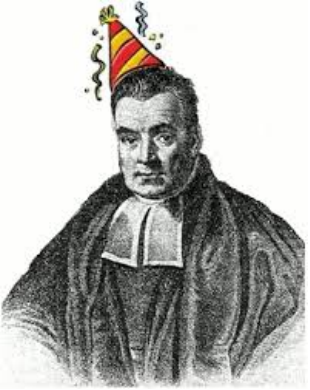
2, 3, 3, 3, 1, 1, 2, 3

What die do you think generated these outcomes?



Outline

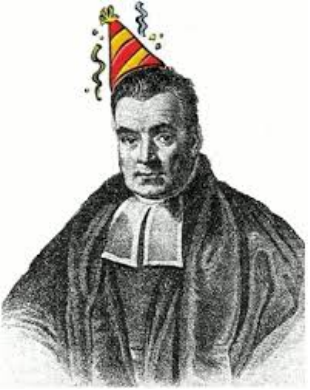
- ◆ Introductory remarks
- ◆ **Senn's stubborn mule**
- ◆ Ockham & prior-data conflict
- ◆ Ockham & the Jeffreys-Lindley paradox
- ◆ Ockham & the role of informed priors
- ◆ The more pressing problem



Being Stubborn and Wrong

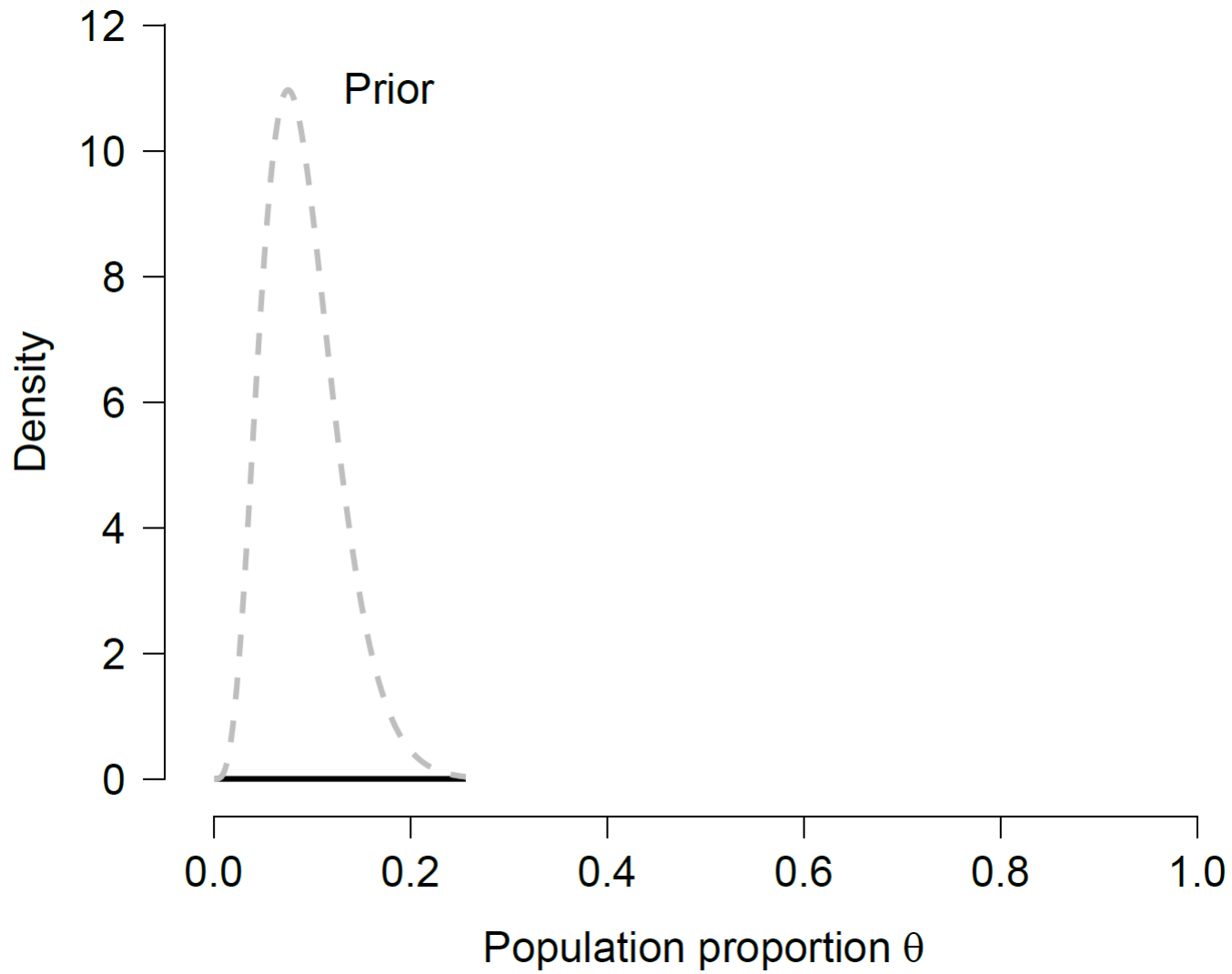
In a nutshell, a Bayesian will perform poorly if he/she is both misguided (with prior mean far from the true value of the parameter) and stubborn (placing a good deal of weight near the prior mean).

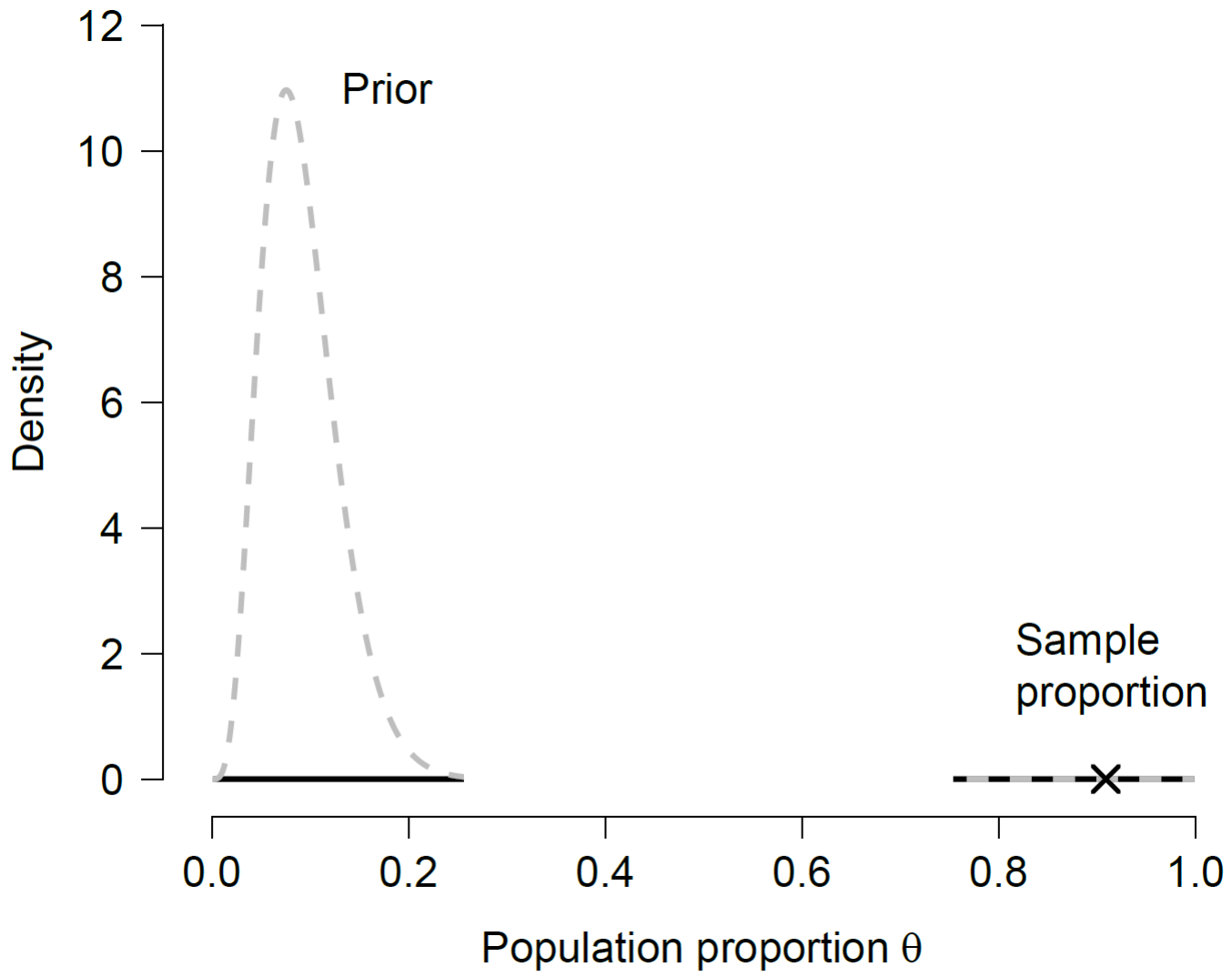
Samaniego, 2013

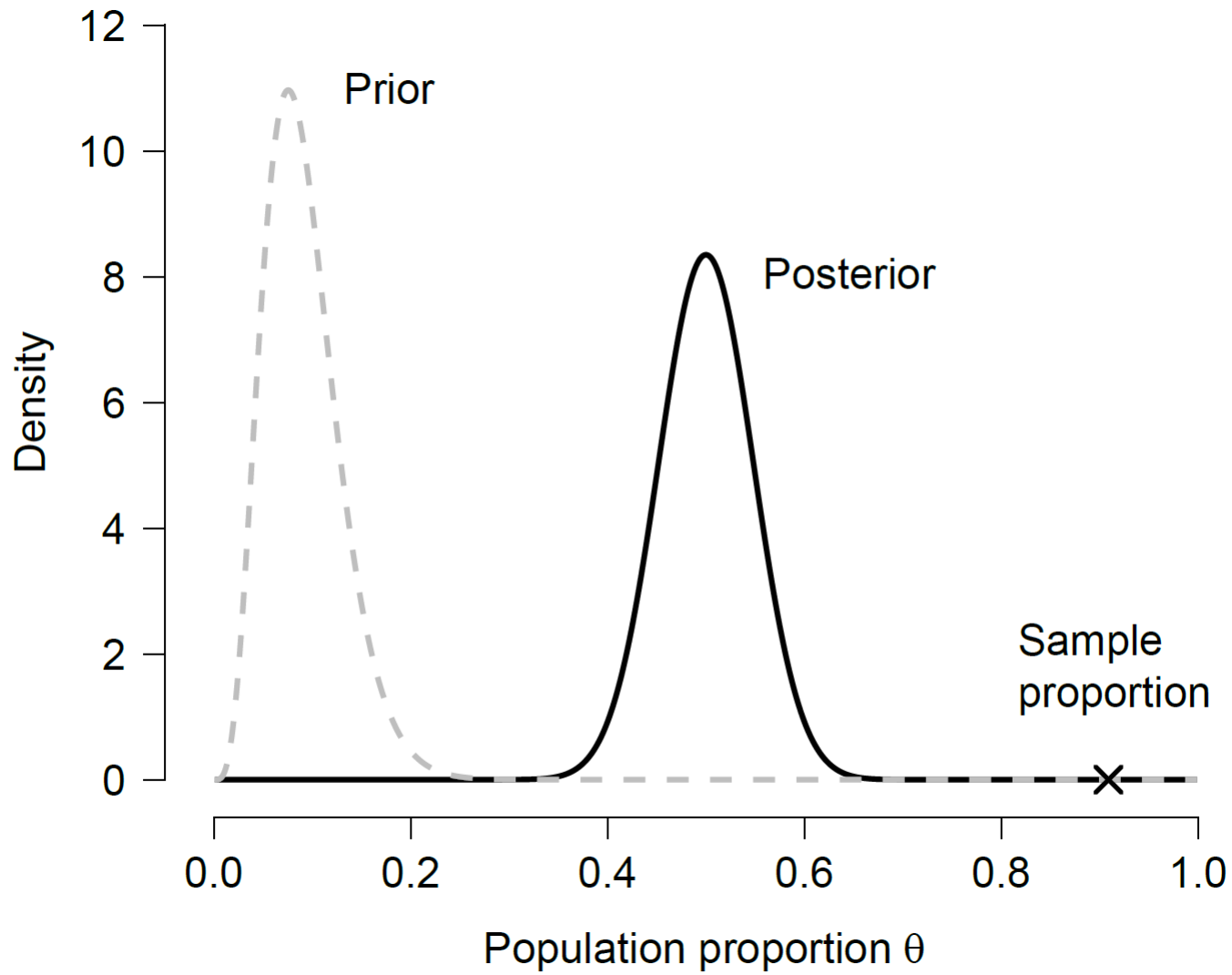


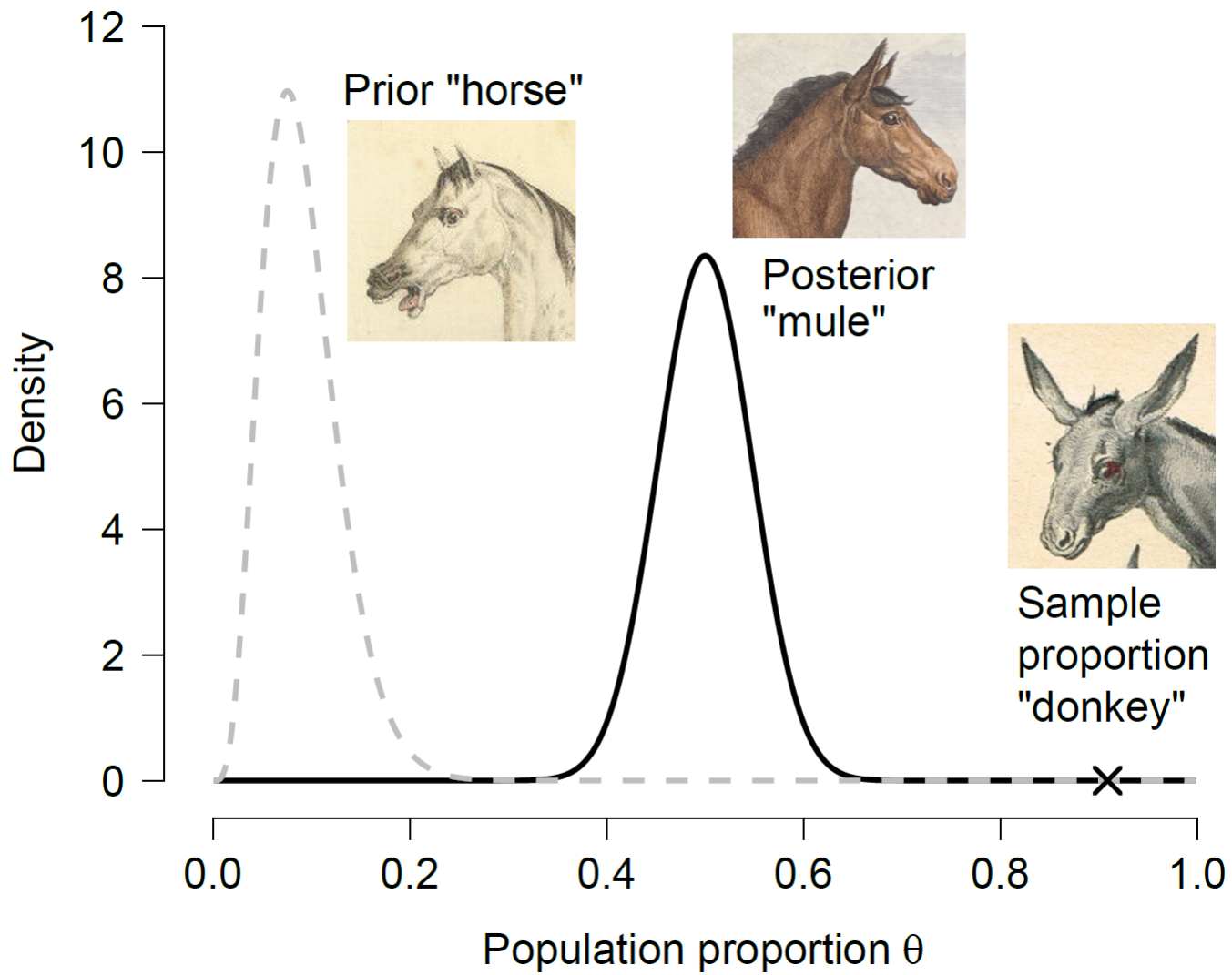
Definition of a Bayesian (Adjusted from Senn, 2007)

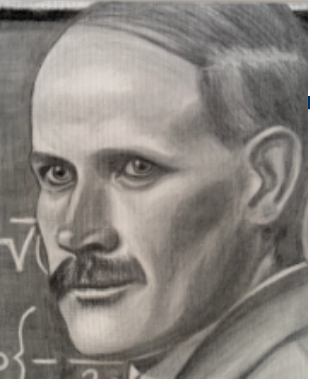
“One who, strongly expecting a horse and clearly viewing a donkey, confidently asserts having seen a mule.”







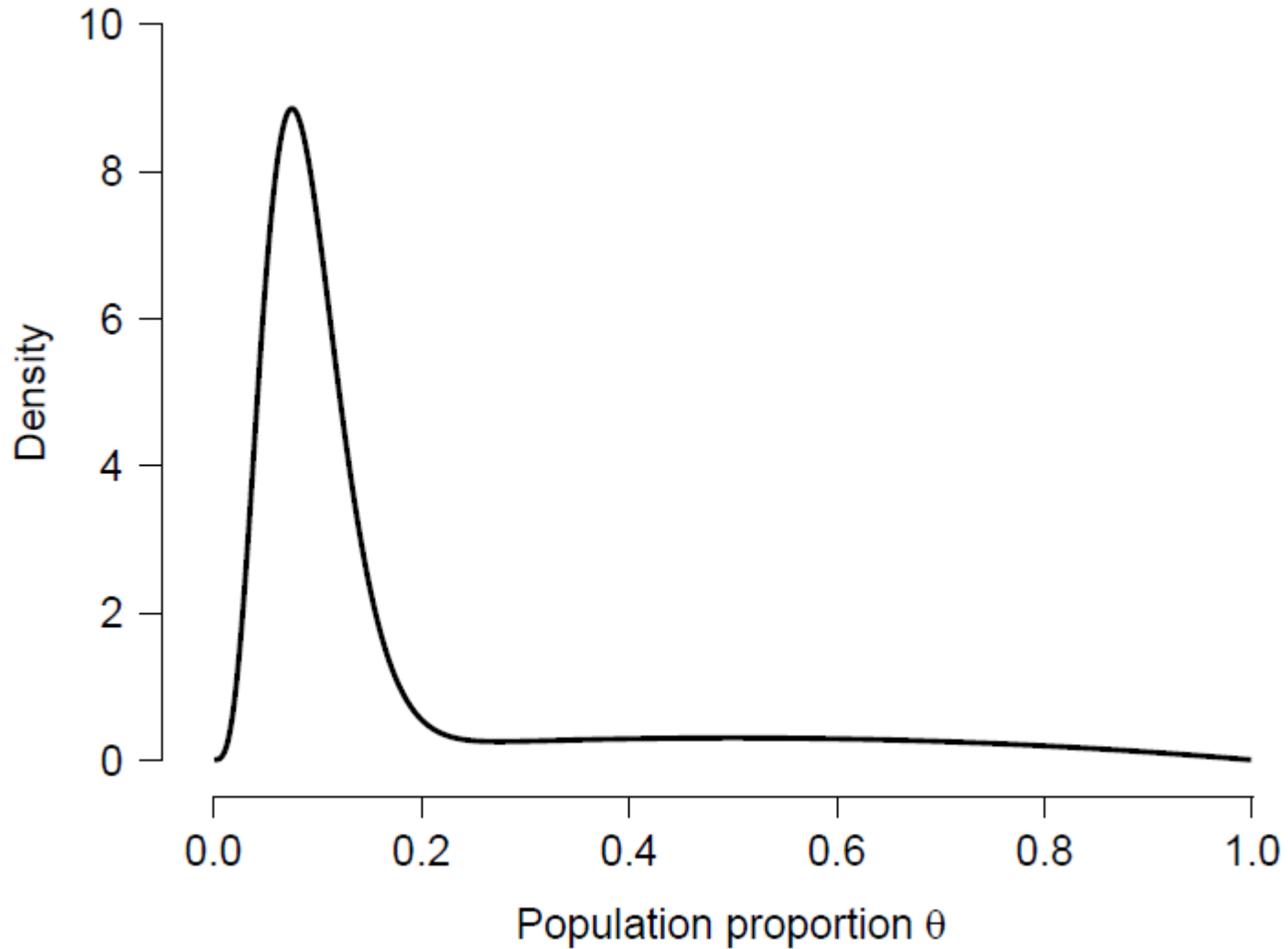




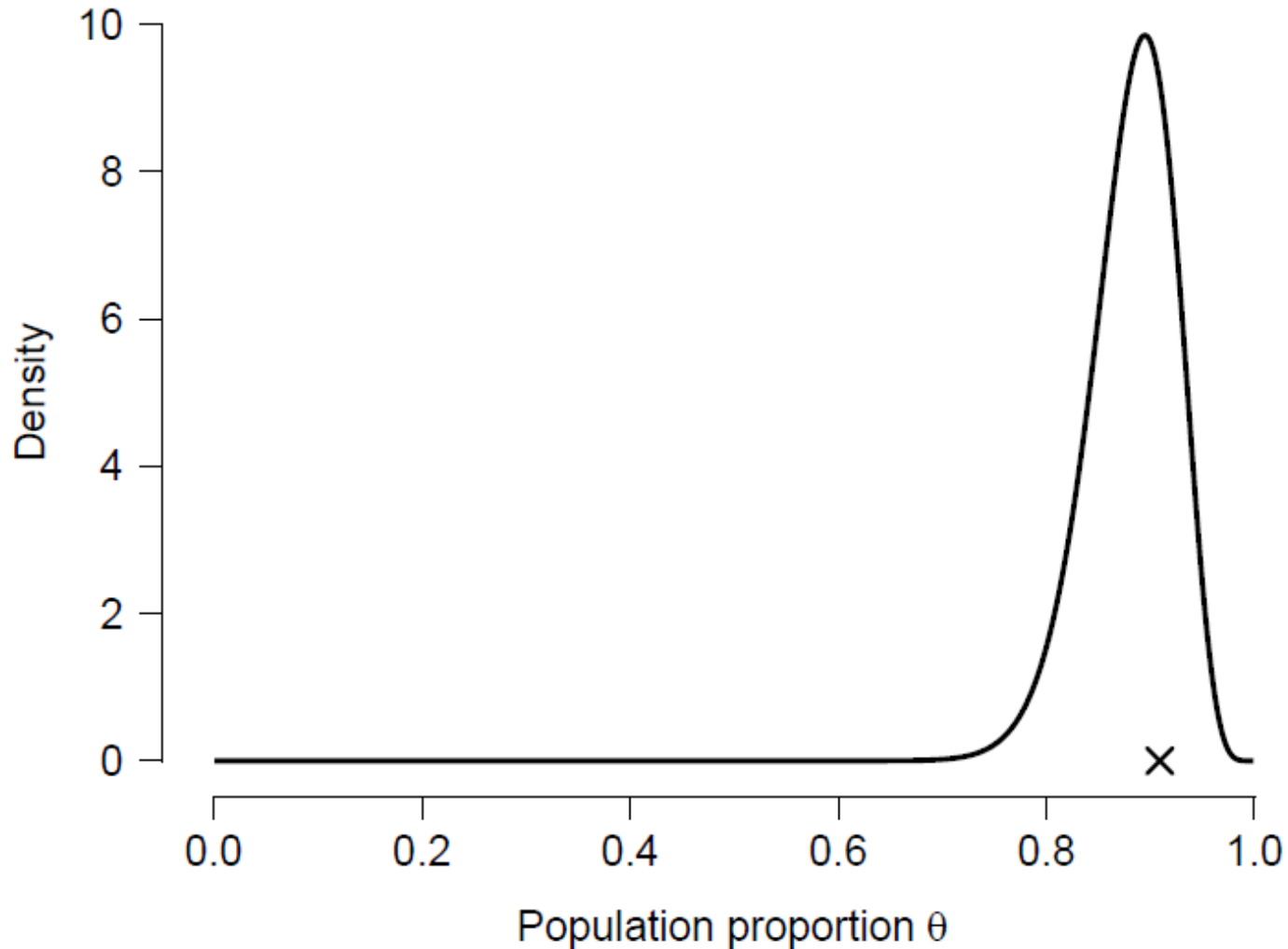
A Remedy: Robust Mixture Priors

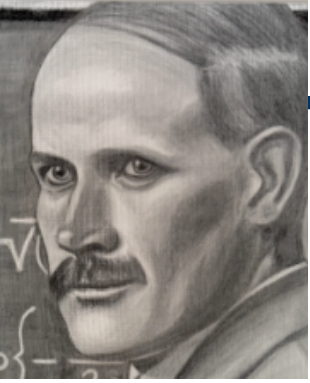
- ◆ Invest some prior mass on a relatively vague “insurance prior”.
- ◆ If prior-data conflict should occur, the insurance prior will kick in and take over the inference.
- ◆ Robustifies the inference by adding an epistemic safety net.

A robust mixture prior



A robust mixture posterior;
the mule has bolted. Mixture weights
are the *posterior* probabilities.





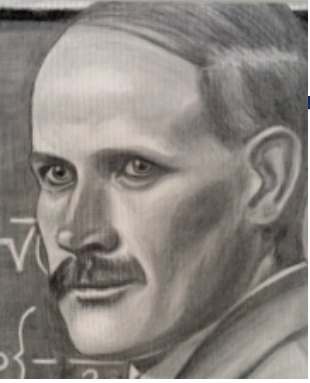
Robust Mixture Priors for Hypothesis Testing

- ◆ Same prior structure applies.
- ◆ The BF for the mixture is a weighted average of the BFs for each component, with the mixture weight the *prior probability* for that component.



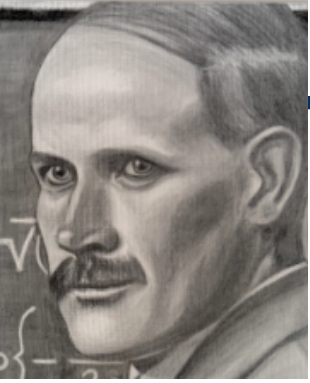
Outline

- ◆ Introductory remarks
- ◆ Senn's stubborn mule
- ◆ **Ockham & prior-data conflict**
- ◆ Ockham & the Jeffreys-Lindley paradox
- ◆ Ockham & the role of informed priors
- ◆ The more pressing problem



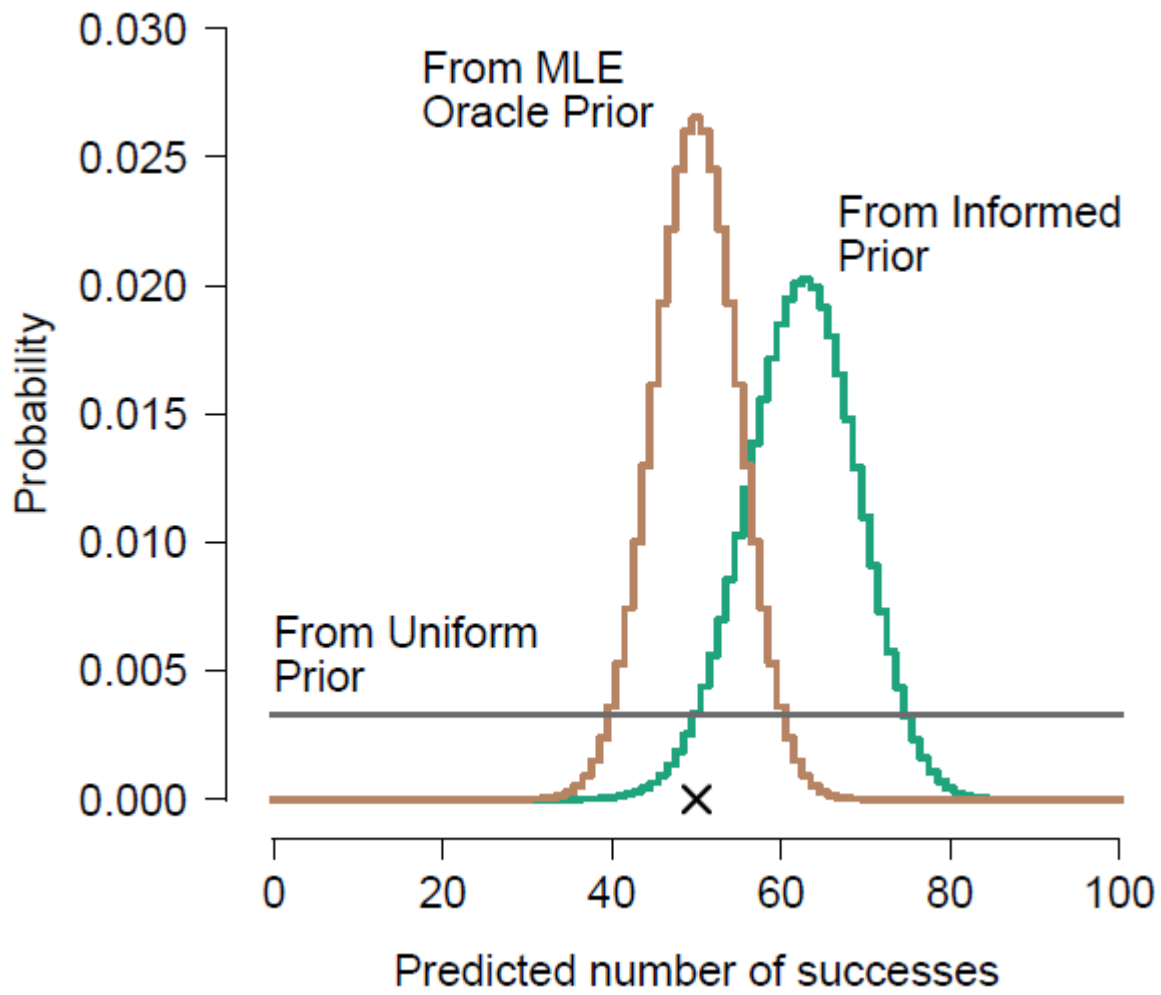
Prior-Data Conflict

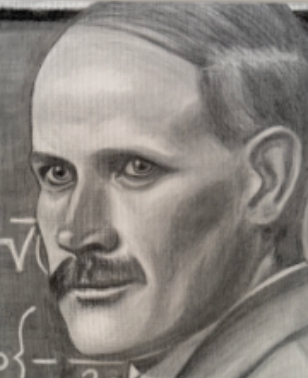
- ◆ No prior is free from prior-conflict.
- ◆ You can be stubborn and wrong, but also needlessly vague. In both cases, *prior mass is wasted*.



Prior-Data Conflict

- ◆ No prior is free from prior-conflict.
- ◆ You can be stubborn and wrong, but also needlessly vague. In both cases, *prior mass is wasted*.
- ◆ Exception: the *oracle prior* – all prior mass centered at the MLE.



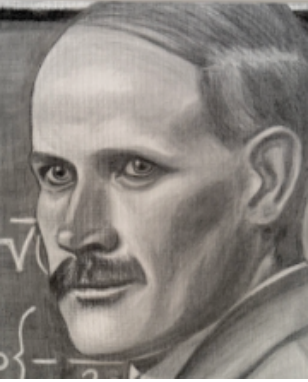


Prior-Data Conflict

- ◆ Hence a measure of prior-data conflict may be the BF for \mathcal{H}_1 (the model with the prior of interest) vs. the MLE:

$$\mathcal{F} \equiv \frac{p(\text{data} \mid \mathcal{H}_1)}{p(\text{data} \mid \hat{\theta}, \mathcal{H}_1)} = \frac{\int p(\text{data} \mid \theta, \mathcal{H}_1) p(\theta \mid \mathcal{H}_1) d\theta}{\max_{\theta} p(\text{data} \mid \theta, \mathcal{H}_1)} \in (0, 1].$$

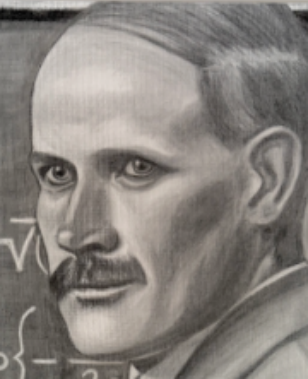
- ◆ \mathcal{F} near 0 implies that a lot of prior mass has been wasted.



Prior-Data Conflict

- ◆ For the normal-normal model with $H_1: \theta \sim N(\mu_0, \sigma_0^2)$ vs. the MLE we have:

$$\mathcal{F} = \underbrace{\frac{\text{se}(\hat{\theta})}{\sqrt{\sigma_0^2 + \text{se}(\hat{\theta})^2}}}_{\mathcal{F}_{\text{spread}}} \times \underbrace{\exp\left(-\frac{1}{2} \frac{(\hat{\theta} - \mu_0)^2}{\sigma_0^2 + \text{se}(\hat{\theta})^2}\right)}_{\mathcal{F}_{\text{alignment}}}$$

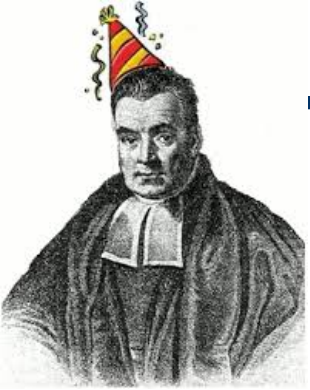


Prior-Data Conflict

- ◆ When the SE is much smaller than the prior scale this yields:

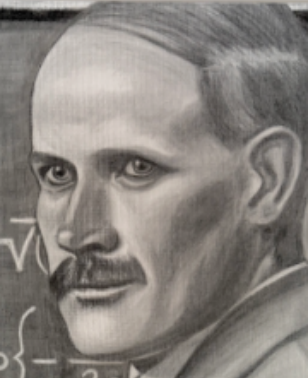
$$\mathcal{F} \approx \underbrace{\frac{\text{se}(\hat{\theta})}{\sigma_0}}_{\mathcal{F}_{\text{spread}}} \times \underbrace{\exp\left(-\frac{1}{2} \frac{(\hat{\theta} - \mu_0)^2}{\sigma_0^2}\right)}_{\mathcal{F}_{\text{alignment}}}$$

- ◆ This attributes “prior waste” to *vagueness*, *wrongness*, and *stubbornness*.



Outline

- ◆ Introductory remarks
- ◆ Senn's stubborn mule
- ◆ Ockham & prior-data conflict
- ◆ **Ockham & the Jeffreys-Lindley paradox**
- ◆ Ockham & the role of informed priors
- ◆ The more pressing problem

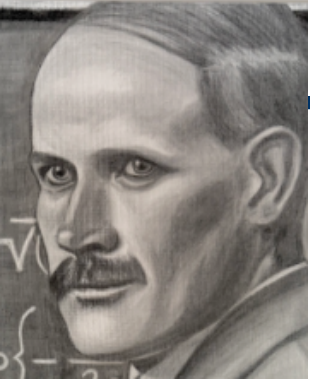


The Jeffreys-Lindley Paradox

- ◆ Usually “F” is discussed in the context of Bayesian model selection.
- ◆ See MacKay (1992) and Jaynes (2002), but perhaps Gull (1988) was there first:

BAYESIAN INDUCTIVE INFERENCE AND MAXIMUM ENTROPY

Stephen F. Gull
Mullard Radio Astronomy Observatory
Cavendish Laboratory
Madingley Road
Cambridge CB3 0HE, United Kingdom



The Jeffreys-Lindley Paradox

- ◆ We use the Chib-Besag insight and switch the role of marginal likelihood and posterior probability:

$$p(\theta \mid \text{data}, \mathcal{H}_1) = \frac{p(\text{data} \mid \theta, \mathcal{H}_1) p(\theta \mid \mathcal{H}_1)}{p(\text{data} \mid \mathcal{H}_1)}$$



$$p(\text{data} \mid \mathcal{H}_1) = \frac{p(\text{data} \mid \theta, \mathcal{H}_1) p(\theta \mid \mathcal{H}_1)}{p(\theta \mid \text{data}, \mathcal{H}_1)},$$

$$p(\theta \mid \text{data}, \mathcal{H}_1) = \frac{p(\text{data} \mid \theta, \mathcal{H}_1) p(\theta \mid \mathcal{H}_1)}{p(\text{data} \mid \mathcal{H}_1)}$$



$$p(\text{data} \mid \mathcal{H}_1) = \frac{p(\text{data} \mid \theta, \mathcal{H}_1) p(\theta \mid \mathcal{H}_1)}{p(\theta \mid \text{data}, \mathcal{H}_1)},$$

This holds for *any* theta.
Here we just pick the MLE...

$$\begin{aligned}
p(\text{data} \mid \mathcal{H}_1) &= \frac{p(\text{data} \mid \hat{\theta}, \mathcal{H}_1) p(\hat{\theta} \mid \mathcal{H}_1)}{p(\hat{\theta} \mid \text{data}, \mathcal{H}_1)} \\
&= p(\text{data} \mid \hat{\theta}, \mathcal{H}_1) \times \underbrace{\frac{p(\hat{\theta} \mid \mathcal{H}_1)}{p(\hat{\theta} \mid \text{data}, \mathcal{H}_1)}}_{\text{BF}_{1\hat{\theta}}} \\
&= p(\text{data} \mid \hat{\theta}, \mathcal{H}_1) \times \mathcal{F} \\
&= \frac{p(\text{data} \mid \hat{\theta}, \mathcal{H}_1)}{\text{BF}_{\hat{\theta}_1}}.
\end{aligned}$$

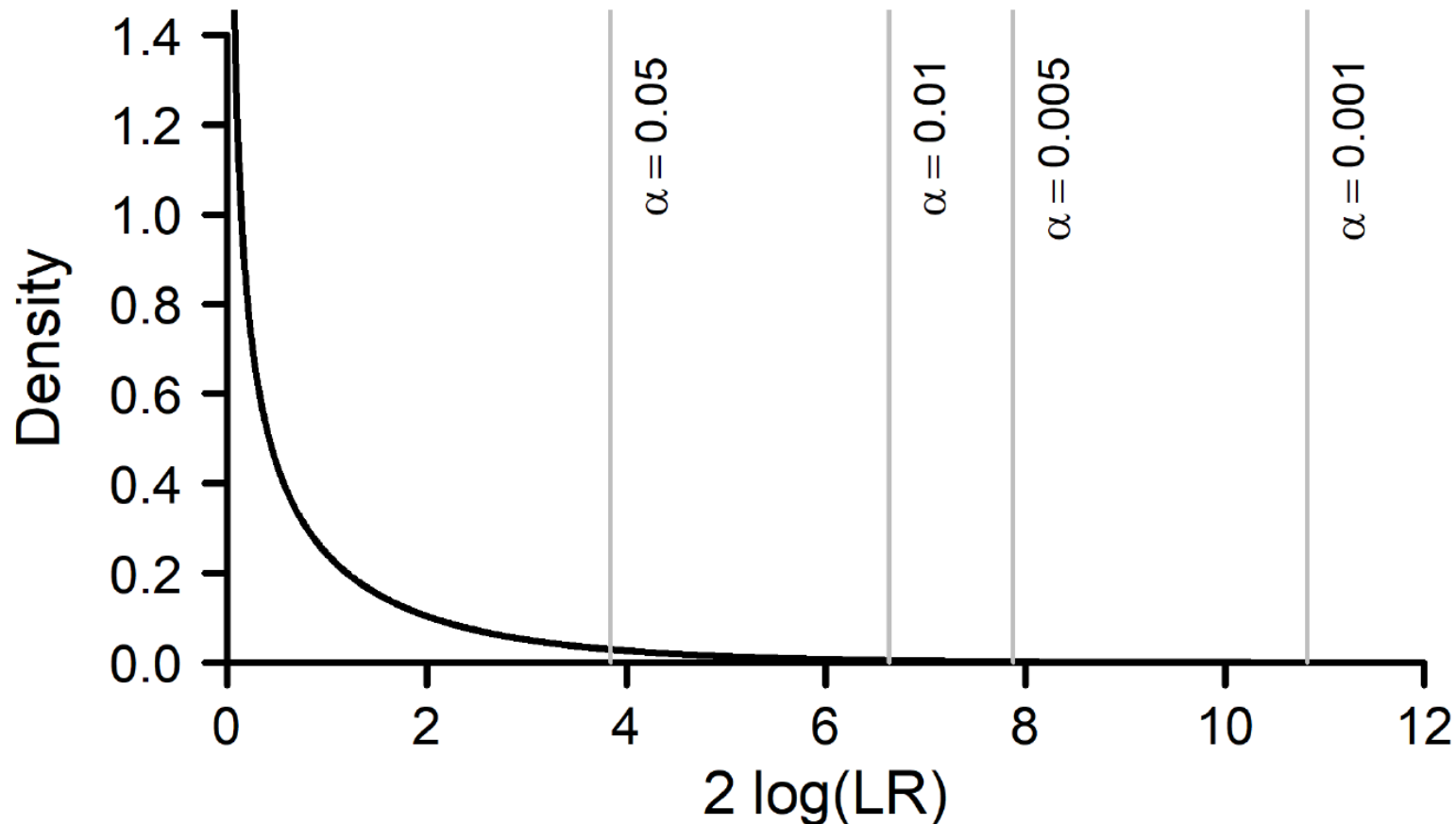
$$\begin{aligned}
p(\text{data} \mid \mathcal{H}_1) &= \frac{p(\text{data} \mid \hat{\theta}, \mathcal{H}_1) p(\hat{\theta} \mid \mathcal{H}_1)}{p(\hat{\theta} \mid \text{data}, \mathcal{H}_1)} \\
&= p(\text{data} \mid \hat{\theta}, \mathcal{H}_1) \times \underbrace{\frac{p(\hat{\theta} \mid \mathcal{H}_1)}{p(\hat{\theta} \mid \text{data}, \mathcal{H}_1)}}_{\text{BF}_{1\hat{\theta}}} \\
&= p(\text{data} \mid \hat{\theta}, \mathcal{H}_1) \times \mathcal{F} \\
&= \frac{p(\text{data} \mid \hat{\theta}, \mathcal{H}_1)}{\text{BF}_{\hat{\theta}_1}}.
\end{aligned}$$

Correction for selection!

$$\begin{aligned}\text{BF}_{10} &= \frac{p(\text{data} \mid \mathcal{H}_1)}{p(\text{data} \mid \mathcal{H}_0)} = \frac{p(\text{data} \mid \hat{\theta})}{p(\text{data} \mid \theta_0)} \times \frac{p(\hat{\theta} \mid \mathcal{H}_1)}{p(\hat{\theta} \mid \text{data}, \mathcal{H}_1)} \\ &= \text{LR}_{10} \times \mathcal{F}.\end{aligned}$$

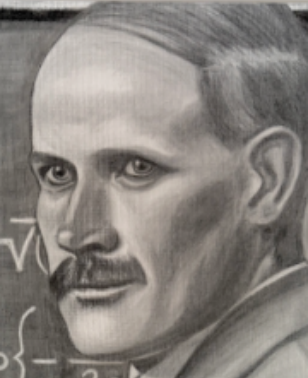
$$\begin{aligned} \text{BF}_{10} &= \frac{p(\text{data} \mid \mathcal{H}_1)}{p(\text{data} \mid \mathcal{H}_0)} = \frac{p(\text{data} \mid \hat{\theta})}{p(\text{data} \mid \theta_0)} \times \frac{p(\hat{\theta} \mid \mathcal{H}_1)}{p(\hat{\theta} \mid \text{data}, \mathcal{H}_1)} \\ &= \text{LR}_{10} \times \mathcal{F}. \end{aligned}$$

$\chi^2(1)$ distribution with LRT cutoffs



$$\begin{aligned}\text{BF}_{10} &= \frac{p(\text{data} \mid \mathcal{H}_1)}{p(\text{data} \mid \mathcal{H}_0)} = \frac{p(\text{data} \mid \hat{\theta})}{p(\text{data} \mid \theta_0)} \times \frac{p(\hat{\theta} \mid \mathcal{H}_1)}{p(\hat{\theta} \mid \text{data}, \mathcal{H}_1)} \\ &= \text{LR}_{10} \times \mathcal{F}.\end{aligned}$$

Suppose LR10 remains constant as sample size increases. The posterior distribution contracts around the MLE, driving the Ockham fraction down to zero.

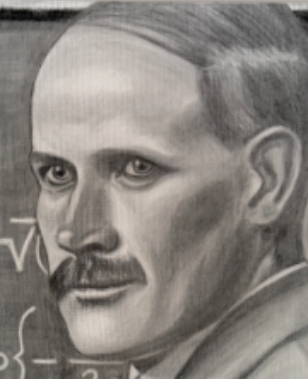


The Jeffreys-Lindley Paradox

- ◆ Also, recall:

$$\mathcal{F} \approx \underbrace{\frac{\text{se}(\hat{\theta})}{\sigma_0}}_{\mathcal{F}_{\text{spread}}} \times \underbrace{\exp\left(-\frac{1}{2} \frac{(\hat{\theta} - \mu_0)^2}{\sigma_0^2}\right)}_{\mathcal{F}_{\text{alignment}}}$$

- ◆ Suppose the prior is centered on the MLE, and its width is the sampling standard deviation (i.e., the “unit-information prior”).

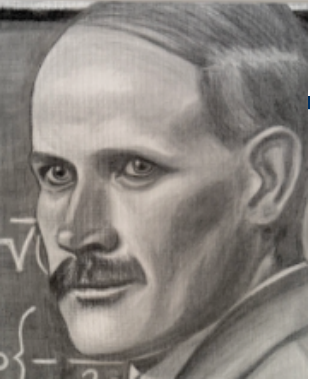


The Jeffreys-Lindley Paradox

- ◆ Then we have:

$$\mathcal{F} \approx \frac{\text{se}(\hat{\theta})}{\sigma_0} = \frac{\sigma / \sqrt{n}}{\sigma_0} = \frac{1}{\sqrt{n}}$$

- ◆ This also underlies the BIC. All of this was anticipated by Jeffreys (of course).



Prior-Data Conflict

- ◆ As $2\log(\text{LR}) \sim \chi^2_1$, we can connect p -values to UI Bayes factors. Specifically, when $p < .10$ we have (cf. Wagenmakers, 2022):

$$\text{BF}_{10} \approx \frac{1}{3p \cdot \sqrt{n}} = \frac{1/p}{3\sqrt{n}}$$

- ◆ This is another way of highlighting the Jeffreys-Lindley paradox.



A Fresh Way to Learn Bayesian Statistics



Amsterdam,
August 27 & 28, 2026
Hybrid

Workshop: Theory and Practice of Bayesian Hypothesis Testing with JASP

August 27 – 28, 2026

Event Description

This workshop can be attended either on-site (in Amsterdam) or online.

The main purpose of this workshop is to familiarize participants with key Bayesian concepts in hypothesis testing. Concrete examples illustrate how to compute, report, and interpret Bayesian hypothesis tests for popular statistical models such as correlation, regression, t-test, ANOVA, and contingency tables. To facilitate the learning process we use JASP, a program whose attractive graphical user interface allows us to focus on core Bayesian concepts and principles, unburdened by the need to explain the detailed workings of an unfamiliar software program such as WinBUGS.

Workshop: A Crash Course in Machine Learning with JASP

August 26, 2026

Event Description

This workshop can be attended either on-site (in Amsterdam) or online.

The main purpose of this workshop is to familiarize participants with key concepts in machine learning, assisted by the machine learning module in JASP (for a quick impression see the blogposts [here](#), [here](#), [here](#), and [here](#)). The workshop covers popular machine learning techniques such as k-nearest neighbors (KNN), random forests, and boosted regression trees. The techniques are then applied to concrete data sets. Additionally, the workshop explains how machine learning models, once trained, can generate predictions for new data sets. The workshop is designed to provide an accessible introduction to machine learning in JASP, emphasizing its user-friendly interface and powerful analytical capabilities. Through a combination of hands-on exercises and guided discussions, participants will learn how to implement, interpret, and evaluate machine learning models with ease.

Workshop: State-of-the-Art Meta-Analysis using JASP

August 25, 2026

Event Description

This workshop can be attended either on-site (in Amsterdam) or online.

The main purpose of this workshop is to equip participants with state-of-the-art meta-analytic techniques using JASP. Through concrete examples and hands-on exercises, participants will learn how to perform advanced meta-analyses, including meta-regression, multivariate/multilevel meta-analyses, and Bayesian publication bias-adjusted meta-analyses. JASP's intuitive graphical interface allows us to focus on the underlying methodological principles and practical interpretation without getting bogged down in complex coding. Real-world datasets will be used to illustrate how to carry out these analyses, interpret the output, and report the findings accurately.

Workshop: Discovering Statistics Using JASP

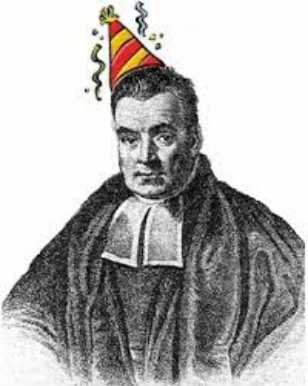
August 24, 2026

Event Description

This workshop can be attended either on-site (in Amsterdam) or online.

The main purpose of this workshop is to familiarize participants with the key contents of "[Discovering Statistics Using JASP](#)", the [new textbook](#) by [Andy Field](#), [Johnny van Doorn](#), and [EJ Wagenmakers](#). The workshop shows how JASP can be used to explain key statistical concepts such as confidence intervals, standard errors, p -values, and power. In addition, the workshop will feature a speedrun through the textbook, with regular pauses for hands-on exercises.





Outline

- ◆ Introductory remarks
- ◆ Senn's stubborn mule
- ◆ Ockham & prior-data conflict
- ◆ Ockham & the Jeffreys-Lindley paradox
- ◆ **Ockham & the role of informed priors**
- ◆ The more pressing problem

$$\begin{aligned}\text{BF}_{10} &= \frac{p(\text{data} \mid \mathcal{H}_1)}{p(\text{data} \mid \mathcal{H}_0)} = \frac{p(\text{data} \mid \hat{\theta})}{p(\text{data} \mid \theta_0)} \times \frac{p(\hat{\theta} \mid \mathcal{H}_1)}{p(\hat{\theta} \mid \text{data}, \mathcal{H}_1)} \\ &= \text{LR}_{10} \times \mathcal{F}.\end{aligned}$$

The more mass the prior puts on the MLE,
the higher the Bayes factor will be.



Example I: Testing a Universal Generalization

- ◆ $H_0: \theta = 1$ (“all ravens are black”)
- ◆ $H_1: \theta \sim \text{beta}(1, 1)$
- ◆ We observe an unbroken string of black ravens.
- ◆ The MLE under H_1 equals the null value.
- ◆ Hence, $\text{BF}_{01} = F$.

Example I: Testing a Universal Generalization

- ◆ Consider a sequential analysis:

$$\begin{aligned}
 \text{BF}_{10} &= \frac{1}{\prod_{\alpha=1}^n \left[\frac{1}{\alpha} + 1 \right]} \\
 &= \frac{1}{2} \times \frac{2}{3} \times \frac{3}{4} \times \cdots \times \frac{n}{n+1} \\
 &= \frac{1}{n+1}.
 \end{aligned}$$



Example I: Testing a Universal Generalization

- ◆ Consider a sequential analysis:

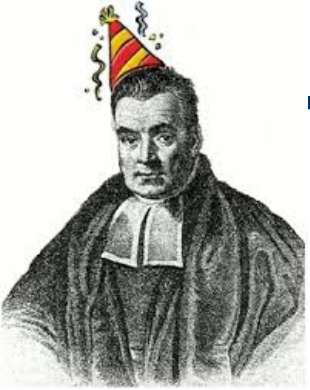
$$\begin{aligned} \text{BF}_{10} &= \frac{1}{\prod_{\alpha=1}^n \left[\frac{1}{\alpha} + 1 \right]} \\ &= \frac{1}{2} \times \frac{2}{3} \times \frac{3}{4} \times \cdots \times \frac{n}{n+1} \\ &= \frac{1}{n+1}. \end{aligned}$$

- ◆ The Fs approach 1 because H1 assigns more and more mass to the MLE.



Example II: Coin Tossing

- ◆ $H_0: \theta = 0.50$
- ◆ $H_1: \theta \sim \text{beta}(1,1)$
- ◆ $H_2: \text{“}\theta \text{ is about } 0.51\text{”}$ (Diaconis et al.)
- ◆ Suppose we have $s = 515, f = 485$.



Outline

- ◆ Introductory remarks
- ◆ Senn's stubborn mule
- ◆ Ockham & prior-data conflict
- ◆ Ockham & the Jeffreys-Lindley paradox
- ◆ Ockham & the role of informed priors
- ◆ **The more pressing problem**



Likelihood-Data Conflict

- ◆ How to deal the possibility that the likelihood is (seriously) misspecified?
- ◆ Gelman: people choke on the gnat of the prior while swallowing the camel of the likelihood.

Likelihood-Data Conflict

- ◆ Common strategy: adjust the model until misspecification can no longer be detected.
- ◆ Sometimes this may be the best one can do.
- ◆ But now the models are being cherry-picked, from an unknown prior distribution.
- ◆ Is there a Bayesian motivation for this?

Likelihood-Data Conflict

- ◆ I prefer to increase the model space and do Bayesian model-averaging.
- ◆ Another informative approach are many-analyst projects (although the cherry-picking problem remains).



Balazs Aczel



Alexandra
Sarafoglou

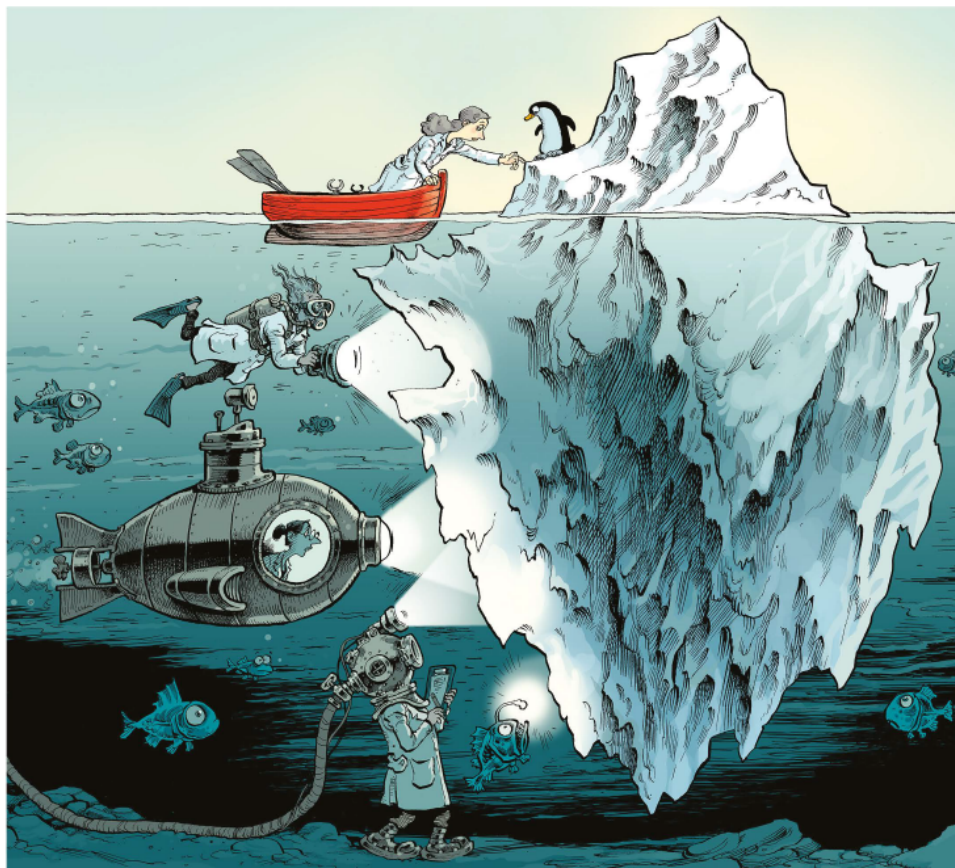


ILLUSTRATION BY DAVID PARKINS

One statistical analysis must not rule them all

Eric-Jan Wagenmakers, Alexandra Sarafoglou & Balazs Aczel

Any single analysis hides an iceberg of uncertainty. Multi-team analysis can reveal it.

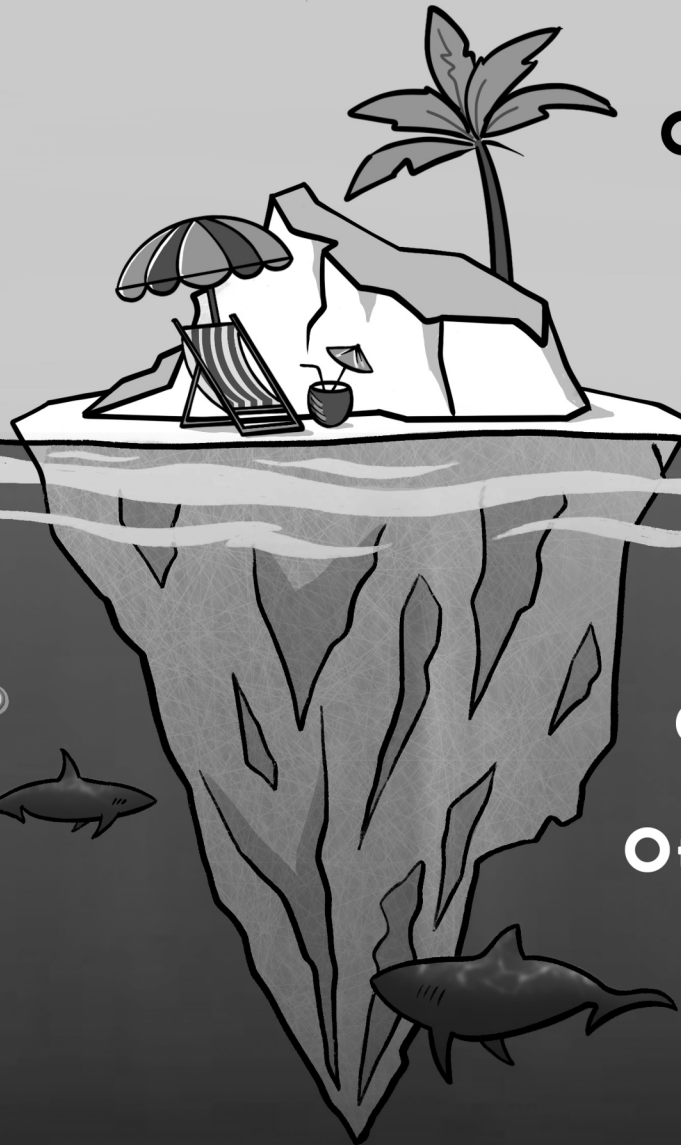
A typical journal article contains the results of only one analysis pipeline, by one set of analysts. Even in the best of circumstances, there is reason to think that judicious alternative analyses would yield different outcomes.

For example, in 2020, the UK Scientific Pandemic Influenza Group on Modelling asked nine teams to calculate the reproduction number R for COVID-19 infections¹. The

teams chose from an abundance of data (deaths, hospital admissions, testing rates) and modelling approaches. Despite the clarity of the question, the variability of the estimates across teams was considerable (see 'Nine teams, nine estimates').

On 8 October 2020, the most optimistic estimate suggested that every 100 people with COVID-19 would infect 115 others, but perhaps as few as 96, the latter figure implying that

*MAUCHLY'S TEST YIELDS $P < .05$.
DO WE USE HUYNH-FELDT
OR GREENHOUSE-GEISSER?*



One analyst

One pipeline

Other analysts

Other pipelines

Multi-Analysts Efforts

- ◆ Can provide crucial insight.
- ◆ Drawback: the amount of logistical effort makes routine application unfeasible.




Synchronous Robustness Reports

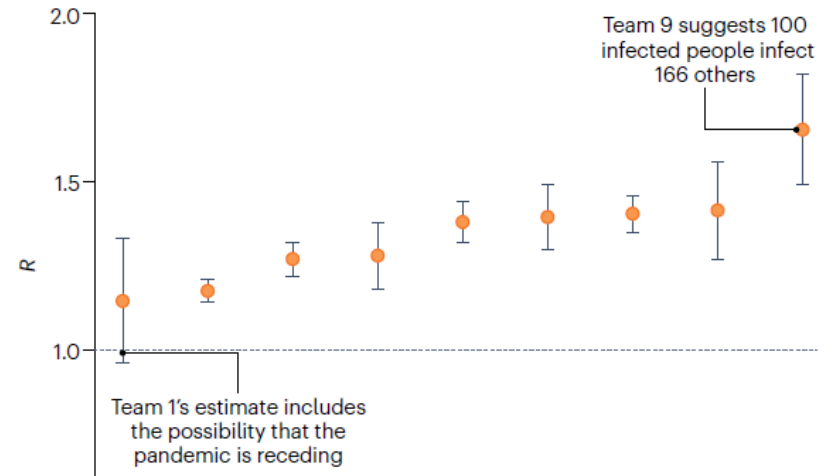
- ◆ Main idea: make it *easy* to publish “several-analyst” reports.
- ◆ Format: concise re-analyses published together with the main article as comments.
- ◆ Idea was recently published in *Nature Human Behaviour*...

Introducing synchronous robustness reports

František Bartoš, Alexandra Sarafoglou, Balazs Aczel, Suzanne Hoogeveen, Christopher D. Chambers & Eric-Jan Wagenmakers

 Check for updates

Most empirical research articles feature a single primary analysis that is conducted by the authors. However, different analysis teams usually adopt different analytical approaches and frequently reach varied conclusions. We propose synchronous robustness reports – brief reports that summarize the results of alternative analyses by independent experts – to strengthen the credibility of science.







Balazs Aczel



František
Bartoš



Alexandra
Sarafoglou



Chris Chambers



Suzanne Hoogeveen

Journal of Robustness Reports

About Journal of Robustness Reports

- [Description](#)
- [Scope](#)
- [Content](#)
- [Acceptance Criteria](#)
- [Submission and Editorial Process](#)
- [Editorial College \(Journal of Robustness Reports\)](#)
- [Genuine Open Access](#)

Useful further links

Page

[Author Guidelines](#)[Referee Guidelines](#)[Editorial Procedure](#)[Editorial Colleges](#)[Editorial College by-laws](#)

[SUBMIT JRR INQUIRY](#)[SUGGEST ARTICLE FOR REANALYSIS](#)

The Journal of Robustness Reports

The *Journal of Robustness Reports* (JRR) is a diamond open-access journal that focuses on the reanalysis of high-impact empirical findings. JRR contains collections of Robustness Reports; each report presents a 500-word reanalysis of the original findings using an alternative, plausible approach. The aim of JRR is to reveal the uncertainty surrounding scientific conclusions that usually remains hidden.

This companion website serves two purposes. First, the website features blog posts that highlight articles as suitable candidates for a robustness report. Second, the website features a [leaderboard of candidate articles](#); readers are invited to nominate new articles or vote on an existing one from the list.

[Learn more >](#)

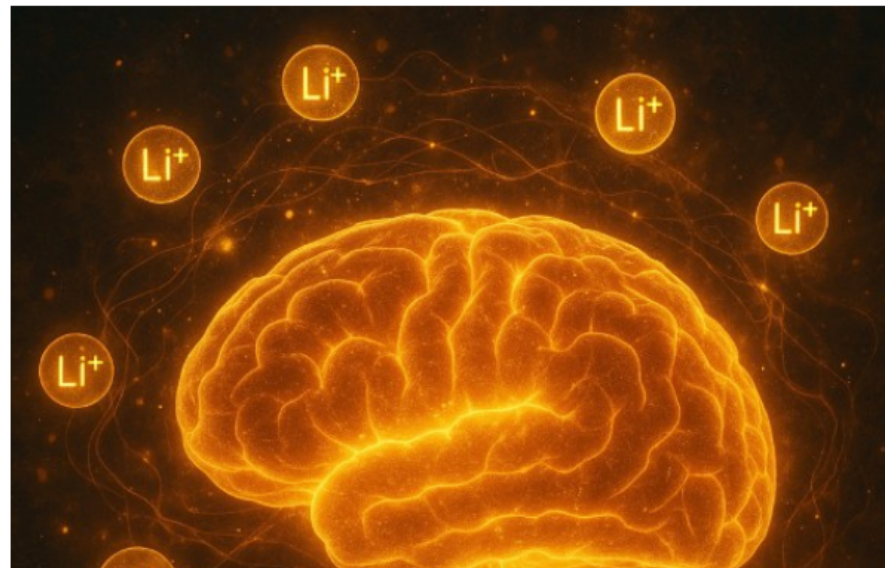
Recently Published

Title	Authors	Date
-------	---------	------

Blog Posts

Inaugural Post: "Lithium Deficiency and the Onset of Alzheimer's Disease"

🕒 July 5, 2025



Introducing the *Journal of Robustness Reports*

František Bartoš^{1*}, *Alexandra Sarafoglou*¹, *Balazs Aczel*²,
*Suzanne Hoogeveen*³, *Christopher D. Chambers*⁴ and *Eric-Jan Wagenmakers*¹

1 Psychological Methods Unit, University of Amsterdam, Amsterdam, the Netherlands

2 Institute of Psychology, ELTE Eotvos Lorand University, Budapest, Hungary

3 Department of Methodology and Statistics, Utrecht University, Utrecht, the Netherlands

4 School of Psychology, Cardiff University, Cardiff, United Kingdom

★ f.bartos96@gmail.com





Thanks for Your Attention!

*If the simple model
predicts better,
the more complex
model does not matter*

After Jeffreys

