UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Combine Statistical Thinking With Open Scientific Practice: A Protocol of a Bayesian Research Project

Sarafoglou, A.; van der Heijden, A.; Draws, T.; Cornelisse, J.; Wagenmakers, E.-J.; Marsman, M.

DOI
10.1177/14757257221077307

Publication date
2022

Document Version
Final published version

Published in
Psychology Learning and Teaching

License
CC BY

Link to publication

Citation for published version (APA):
Sarafoglou, A., van der Heijden, A., Draws, T., Cornelisse, J., Wagenmakers, E.-J., & Marsman, M. (2022). Combine Statistical Thinking With Open Scientific Practice: A Protocol of a Bayesian Research Project. *Psychology Learning and Teaching*, *21*(2), 138-150. https://doi.org/10.1177/14757257221077307

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.
UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

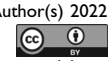
Download date:17 Oct 2026

Check for updates

Report

PSYCHOLOGY LEARNING & TEACHING

Combine Statistical Thinking With Open Scientific Practice: A Protocol of a Bayesian Research Project

Psychology Learning & Teaching
2022, Vol. 21(2) 138–150
© The Author(s) 2022

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/14757257221077307
journals.sagepub.com/home/plj

Alexandra Sarafoglou
Department of Psychology, University of Amsterdam, The Netherlands

Anna van der Heijden
Department of Psychology, University of Amsterdam, The Netherlands
Currently at De Groene Lobby, The Netherlands

Tim Draws
Faculty of Engineering, Mathematics and Computer Science, Delft University of Technology, The Netherlands

Joran Cornelisse
Department of Psychology, University of Amsterdam, The Netherlands
Currently at SocialDatabase, The Netherlands

Eric-Jan Wagenmakers and Maarten Marsman
Department of Psychology, University of Amsterdam, The Netherlands

Abstract

Current developments in the statistics community suggest that modern statistics education should be structured holistically, that is, by allowing students to work with real data and to answer concrete statistical questions, but also by educating them about alternative frameworks, such as Bayesian inference. In this article, we describe how we incorporated such a holistic structure in a Bayesian research project on ordered binomial probabilities. The project was conducted with a group of three undergraduate psychology students who had basic knowledge of Bayesian statistics and programming, but lacked formal mathematical training. The research project aimed to (1) convey the basic mathematical concepts of Bayesian inference; (2) have students experience the entire empirical cycle including collection, analysis, and interpretation of data and (3) teach students open science practices.

Corresponding author:
Alexandra Sarafoglou, Nieuwe Achtergracht 129B, 1001 NK Amsterdam, The Netherlands.
Email: alexandra.sarafoglou@gmail.com

Keywords
Bayesian inference, education, Bayes factor, encompassing – prior, replication, open science

Introduction

The curriculum guidelines of the American Statistical Association (ASA) argue that statistics education in undergraduate programmes should not be primarily focused on teaching statistical methods and mathematical foundations, but also emphasize scientific practice, that is, study design, data collection, programming skills, and data analysis (American Statistical Association, 2014; Horton & Hardin, 2015; Wasserstein & Lazar, 2016). In general, students should learn to "think with and about data" (Cobb, 2015, p. 267) and thus develop a holistic understanding of statistics (Horton and Hardin, 2015).

This holistic understanding of statistics also includes learning and understanding alternatives to classical inference based on *p*-values. Bayesian inference is becoming increasingly popular and its adoption has been advocated for both scientific practice (Wasserstein & Lazar, 2016) and statistics education (Cobb, 2015). Recent examples for undergraduate courses and in-class demonstrations on Bayesian methods that require only little or no mathematical or statistical training are described in Witmer (2017; on teaching Markov chain Monte Carlo methods), Rouder and Morey (2018; on teaching Bayes' rule), and van Doorn, Matzke and Wagenmakers (2020; on teaching the key concepts of Bayesian inference). However, little attention has been paid to the design and structure of Bayesian research projects that can be conducted with a small group of students, for instance, in the context of a thesis or internship project, or a seminar. These formats, as opposed to standard courses, allow for more extensive research projects, since supervisors can offer individual support for students, and can dedicate more time to the execution of the project.

We believe that a research project on Bayesian inference should take advantage of the rather long project duration and the small group size by introducing students in detail to the theoretical and practical aspects of Bayesian inference. Theoretical aspects of Bayesian inference entail that by the end of the project students should feel comfortable with the standard terminology, be able to understand how to assign a prior distribution, specify a likelihood function, derive a posterior distribution, and compute a marginal likelihood. The practical aspects entail that students should be able to apply their theoretical knowledge to address a concrete research question, and experience the entire empirical cycle, including study planning, preregistration, data collection and analysis, and interpretation of the results. These teaching goals resulted in three guiding principles for structuring the project, listed below.

The first principle is to introduce students to the mathematics underlying Bayesian statistics. In our own teaching of Bayesian methods in undergraduate psychology courses, we usually hide the mathematics and instead aim to provide students with an intuition about how Bayesians use distributions to quantify uncertainty about model parameters and hypotheses. This approach helps students interpret posterior distributions, credible intervals, and Bayes factors (for a gentle technical introduction to Bayesian inference without mathematical derivations see Etz and Vandekerckhove, 2018). However, for students who want to specialize in research methods and statistics it is important to go beyond an intuitive understanding and be introduced to the mathematics behind these key concepts. Without the mathematical foundations, students will find the statistical literature difficult to understand.

The second principle is to let students experience scientific practice. In line with the ASA guidelines on statistics education (American Statistical Association, 2014), we believe that students learn

most when they are given the opportunity to gain hands-on experience on how to apply the methods taught to a real data example. We therefore set up a Bayesian replication study that demonstrates a series of Bayesian benefits. For instance, in contrast to frequentist analyses, the Bayesian framework allows students (1) to discriminate between "absence of evidence" and "evidence of absence" of the effect in the replication study (Dienes, 2014; Keysers et al., 2020; Verhagen & Wagenmakers, 2014); (2) to experience Bayesian learning by incorporating prior knowledge – such as data from previous experiments – to construct a more informative test (Verhagen & Wagenmakers, 2014); (3) to monitor evidence as the data accumulates (Rouder, 2014). In addition, it allows students to learn how conclusions from significant *p*-values differ from conclusions drawn from Bayes factors by conducting a Bayesian reanalysis of the results of the original experiment.[1]

The third and final principle is to convey open science practices. Reproducibility and replicability are core scientific values, but yet psychological science is currently facing a crisis of confidence as a disappointing proportion of key findings appear to be reproducible (Baker, 2016; Camerer et al., 2018; Klein et al., 2014; Klein et al., 2018; Nature Publishing Group, 2016; Open Science Collaboration, 2015; Pashler & Wagenmakers, 2012). To a large extent, the low rate of reproducible findings can be attributed to the great flexibility in data analysis in combination with selective reporting of significant results (Simmons et al., 2011), the high prevalence of questionable research practices (John et al., 2012), the reluctance to conduct direct replication studies (Pashler & Harris, 2012; Schmidt, 2009), and the poor availability of research data (Houtkoop et al., 2018; Wicherts et al., 2006; for a special issue on data sharing see Simons, 2018). To address these problems, psychological science today relies on numerous open scientific practices, such as preregistration and Registered Reports, large-scale collaborations, and sharing of data, materials, and code (e.g., Chambers, 2013; Chambers & Tzavella, 2021; Kidwell et al., 2016; Morey et al., 2016; Moshontz et al., 2018; Nosek et al., 2015). However, to truly integrate these practices into the research culture, it is necessary to introduce the principles of open science to students at an early stage (Chopik et al., 2018; Funder et al., 2014; Morling & Calin-Jageman, 2020; Munafò et al., 2017; Sarafoglou et al., 2020). Since thesis projects often require detailed design and analysis plans we view them as a good opportunity for supervisors to teach them both the philosophy behind open science and the practical skills needed to apply open science practices. Therefore, we set up a preregistered replication study, and have students publish the analysis code, and share the data and materials on the Open Science Framework (OSF; Center for Open Science, 2021).

The purpose of this article is to share our experiences on designing and supervising a Bayesian thesis project for undergraduate psychology students. Lecturers who intend to offer a Bayesian research project for a small group of students which emphasizes mathematical training as well as practical experience with real data might find helpful advice on what focal points to set when planning their project. In addition, the described project can serve as illustrative example in a classroom setting, to teach students Bayesian learning, and a simple method to evaluate ordinal expectations. In the following, we will describe the course structure, the theoretical and the practical part of the project in more detail.

Supplemental Material

Interested readers can visit our OSF project folder (https://osf.io/zfhbc/) to access the following information: the study preregistration, the analysis code, all data and materials, and the student evaluations. Furthermore, it contains the results of the Bayesian reanalysis of the original studies and the formal description of the mathematical model for multiple independent binomial probabilities.

Project Overview

Here we describe the thesis project titled "A Bayesian View on 'Science versus the Stars': Bayes factor analysis for ordered binomial probabilities" at the University of Amsterdam. The topic of the thesis project was the Bayesian analysis of ordinal expectations of multiple binomial probabilities. We chose this topic due to both its relevance in the psychological literature and the simplicity of the statistical model. Ordinal expectations of binomial probabilities are common in the area of psychometrics and theories on rational decision making (see e.g., Cavagnaro & Davis-Stober, 2014; Davis-Stober, 2009; Guo & Regenwetter, 2014; Haaf et al., 2020; Heck & Davis-Stober, 2019; Myung et al., 2005; Regenwetter et al., 2011; Regenwetter et al., 2018; Tijmstra et al., 2015). For instance, a psychometrician who evaluates whether a test for cognitive performance can be measured on an interval scale needs to test the assumption that the probability to solve a given item is non-decreasing for the ability of a person. One argument to use Bayesian methods for these problems is that we can easily incorporate ordinal expectations of the binomial probabilities in the respective prior distributions (Klugkist et al., 2010). This makes the corresponding statistical model particularly simple and enables students to derive the method even without formal mathematical training.

During the theoretical part of the project students familiarized themselves with the computation of Bayes factors for ordered binomial probabilities using the encompassing prior method (Klugkist et al., 2005). During the practical part of the project the students applied the methods in practice by conducting a preregistered reanalysis and replication study.

Course Structure

The full thesis project – starting from the first introductory lesson to submission of the research report – took 16 weeks. A weekly overview of the research project is provided in Table 1. Our students had to hand-in two writing assignments, create the preregistration of the empirical study, and write the final report. On average the students worked 22−23 h per week on the project for which they were rewarded with 12 ECTS credits. The following section describes these components in more detail.

Supervision

During the theoretical part of the project we supervised the students intensively; we had weekly group meetings that were structured as lectures, we gave students two writing assignments, and we reviewed and discussed these assignments with each student individually. During the practical part, the students then primarily worked independently with little need for supervision. The weekly group meetings were replaced by individual contact hours that gave students the opportunity to discuss details of their report.

Writing Assignments

We dedicated the first four sessions at the beginning of the project to the theoretical concepts of Bayesian inference. During these four weeks, students wrote two short 1–2 page reports explaining Bayesian parameter estimation and hypothesis testing. The first report concerned Bayesian inference for one binomial probability. In the second report, they had to generalize these concepts to multiple binomial probabilities. Students could incorporate these reports as part of the methods section in their final report.

Table 1. A week-by-week overview of our project "A Bayesian view on science versus the stars: Bayes factor analysis for ordered binomial probabilites".

Week	Goal	Activities
1	Reiterating knowledge	Bayesian parameter estimation and hypothesis testing for the beta-binomial model
		Write methods section of research report
2	Establishing knowledge	Generalize concepts to multiple binomials
		Write methods section of research report
3	Establishing knowledge	Derive and apply Savage-Dickey density ratio
		Write methods section of research report
4	Establishing knowledge	Derive and apply encompassing prior approach
		Bayesian reanalysis of Carlson (1985) and Wyman and Vyse (2008)
		Write introduction of research report
5	Writing	Finalize the methods section of the research report
		Write introduction of research report
6–7	Preregister study	Plan replication study
		Create preregistration document
8	Preregister study	Print all necessary documents, prepare data collection (e.g, book lab)
		Finalize preregistration
9–10	Data collection 1	Participants fill out NEO-FFI and report date and place of birth
11	Create study materials	Generate personality descriptions
		Prepare follow-up data collection
12–13	Data collection 2	Participants perform choice task
14	Analyzing data	Analyze data and upload the dataset to the OSF
		Write results section of research report
15–16	Finalizing project	Finalize research report Prepare 20-min presentation

Preregistration

Our students had three weeks to create the preregistration document. Since our students answered the same research question, we let them create the preregistration together. The preregistration featured the following components: the study design; the sampling plan, sampling plan rationale, and stopping rule for data collection; exclusion criteria; the description of the materials and procedure; the research question and hypotheses (including the expected direction of the effect); details on the statistical model and analysis plan, including specifications for prior distributions, number of samples drawn, inference criteria, and handling of missing data.

Grading Criteria

For the most part, grading was based on the individual research reports. We assessed whether students were able to (1) justify the proposed research question and methods; (2) describe the Bayesian concepts accurately by using the specific terminology; (3) discuss and interpret the results correctly; and (4) adopt a scientific writing style. In addition, students could receive a pass or fail both on their final presentation and on their learning progress. The writing assignments and the preregistration were not graded.

The Theoretical Part: Bayesian Parameter Estimation and Hypothesis Testing For Multiple Binomial Probabilities

The goal for the theoretical part of the project was to teach students when and how the encompassing prior approach is used, and how it is derived. To ease the students into this topic, we asked them to reiterate the basic mathematical concepts in Bayesian inference by means of one binomial success probability, that is, Bayesian parameter estimation (including Bayes' rule, the prior distribution, the likelihood function, marginal likelihood, and the posterior distribution) and Bayesian hypothesis testing (including prior model odds, the Bayes factor, and posterior model odds). Subsequently, students had to generalize these concepts to multiple binomial success probabilities.

The Practical Part: Reanalysis and Replication of Wyman and Vyse

We searched for empirical studies which involved hypotheses about the ordering of multiple binomial probabilities. The study by Wyman and Vyse (2008) is a suitable candidate for a replication study, for several reasons. First, the study had an engaging research question, that is, whether the accuracy of psychological personality descriptions is similar to the accuracy of astrological natal charts. Second, the dependent variables in Wyman and Vyses' study allowed for the formulation of an ordinal expectation. Third, replicating the study did not require knowledge about sophisticated concepts such as item response theory (Birnbaum, 1968; Rasch, 1960). Fourth, the experimental setup for the study was straightforward which made the planning and execution of a preregistered replication study feasible for our time frame. Note that the study by Wyman and Vyse is itself a conceptual replication of a study conducted by Carlson (1985). We chose to replicate the study by Wyman and Vyse (2008), however, since the authors had a clearer setup and material that was easier to reproduce. We aimed to replicate the study by Wyman and Vyse (2008) as closely as possible which meant that we adapted the original research design with only a few practical changes.

Methods

Sampling Plan. We preregistered to collect data from 50 participants or stop data collection by June 1st, 2018. The target sample size was based on the number of participants in the original studies which was 56 participants in Carlson (1985) and 52 participants in Wyman and Vyse (2008). Unfortunately, we were not able to reach the preregistered target sample before our testing period ended. We were only able to recruit 31 participants. Of those, 2 participants did not attend the second meeting, leaving us with a final sample of 29 participants.

Materials. In their study, Wyman and Vyse used the NEO Five Factor inventory (NEO-FFI, Costa & McCrae, 1985, 1992) to create psychological personality descriptions and the software Astrolabe (Astrolabe Inc, 2018) to create astrological natal charts for each participant. Then, an experimenter gave each participant their own psychological personality description and a psychological personality description belonging to another participant. The participant was then asked to decide which of the two personality descriptions was their own. This procedure was then repeated for the astrological personality description.

Procedure. The research design required two testing periods that were one week long and approximately two weeks apart. During the first testing period, the students assessed participants with the

NEO-FFI personality inventory and collected information, that is, date and place of birth, that allowed them to create astrological natal charts for each participant with the free version of the software used by the original authors. In the second testing period the participants had to perform a simple choice task; they were asked to identify both their own psychological personality description and their astrological natal chart out of two descriptions each (i.e., a chance level of 50%).

Hypotheses. For the replication, our students took into account the direction of the original results and thus tested the ordinal hypothesis H_r that the success probability for psychological personality descriptions is *higher* than that for astrological personality descriptions. This hypothesis was then tested against a point-null hypothesis H_0 that both success probabilities are equal to chance. Furthermore, as will be explained in the next section, the calculation of the Bayes factor required another hypothesis – referred to as the encompassing hypothesis H_e – that both success probabilities can vary freely.

Analysis Plan. The students assigned a beta prior distribution to the model parameters and used the data from Wyman and Vyse (2008) to inform their prior beliefs. Specifically, based on Wyman and Vyse's data, the students assigned a Beta(42, 12) prior distribution to the probability of correctly identifying one's own psychological personality description and a Beta(25, 29) prior distribution to the success probability of correctly identifying one's own astrological personality description. That is, the prior for psychological personality descriptions favours success probabilities well above chance level while for astrological personality descriptions success probabilities at chance level are favoured, with medians and 95% credible intervals of 0.77 [0.65, 0.87] and 0.46 [0.34, 0.59], respectively. To compare H_0 versus H_r, the students first had to take a two-step approach. First, they needed to compute the Bayes factor between H_0 and H_e, denoted as BF_{0e}, using the Savage-Dickey density ratio (Dickey & Lientz, 1970; Wagenmakers et al., 2010) and the Bayes factor between H_r and H_e, denoted as BF_{re}, using the encompassing prior approach (Klugkist et al., 2005). The students then obtained BF_{r0} through transitivity, that is: $BF_{r0} = BF_{re} \times BF_{e0}$. A detailed description of the statistical model is available in the online appendix.

Results of the Replication Study

In our replication study, out of 29 participants, 25 correctly identified their own psychological personality description and 18 participants correctly identified their own astrological personality description (see Table 2). Given our data and the prior knowledge provided by the Wyman and Vyse study, the result suggests extreme evidence (i.e., $BF_{r0} = 1884$) in favour of the hypothesis that people recognize their psychological personality description more reliably than their astrological personality description.

Considerations for Lecturers

Our experience with this project suggests that three considerations warrant special attention. First, lectures should be aware of the prior knowledge of their students. The described project was designed for students who have some knowledge of Bayesian statistics, but also a basic background in the programming language R (R Core Team, 2021). However, despite their familiarity with the key concepts of Bayesian inference, our students found the mathematical parts of the project particularly challenging. Therefore, we recommend lectures to allow enough time to reiterate necessary mathematical components.

Table 2. Data from the current research project, as well as from Wyman and Vyse (2008) and Carlson (1985), where x_{psy} and x_{astro} denote the number of participants who correctly identified their psychological personality description and their astrological personality description, respectively, n_{psy} and n_{astro} denote the respective total number of observations, and $\hat{\theta}_{psy}$ and $\hat{\theta}_{astro}$ denote the sample proportions of correctly identifying one's own personality description.

Data							
Study	x_{psy}	n_{psy}	$\hat{\theta}_{psy}$	x_{astro}	n_{astro}	$\hat{\theta}_{astro}$	Chance level
Current project	25	29	0.86	18	29	0.62	0.50
Wyman and Vyse (2008)	41	52	0.79	24	52	0.46	0.50
Carlson (1985)	25	56	0.45	28	83	0.38	0.33

Second, when students are required to independently draft the preregistration we recommend the use of preregistration templates. For instance, the OSF offers preregistration templates for standard empirical research, but also replication studies (see https://osf.io/ zab38/wiki/home/ for an overview of all preregistration forms). The transparency checklist by Aczel et al. (2020) is another highly accessible tool which covers the most important aspects for achieving transparency and openness in preregistrations and manuscripts.

Finally, in the described project, we based our target sample size on the number of participants in the original studies. Alternatively, lecturers could based their target sample size on a Bayesian design analysis (Stefan et al., 2019). A Bayesian design analysis is considered the Bayesian version of a frequentist power analysis and allows researchers to determine the minimum number of participants needed to achieve compelling evidence either in favour or against the hypothesis. Lecturers could also choose to do sequential testing, that is, monitor the evidence as the data accumulates and stop data collection as soon as the evidence is sufficiently compelling (e.g., Rouder, 2014).

Summary

The discussed research project allowed students to learn a relevant Bayesian method to compute Bayes factors for ordinal expectations (i.e., the encompassing prior approach), and increase their understanding of the underlying mathematical concepts of Bayesian inference. We believe that this learning success was primarily due to the simplicity of the discussed statistical model which enabled the students to formulate the likelihood function, assign a prior distribution, derive the posterior distribution, and understand the encompassing prior approach even without strong mathematical background.

In addition, students gained practical experience through designing and conducting a reanalysis and replication study. Through this experience the students learned the advantages of Bayesian statistics in the context of replication research, for instance, by being able to quantify evidence for the absence of the predicted effect, but also by incorporating prior knowledge into their analyses and hence draw more informed decisions. In addition, the project gave students the opportunity to practice open research practices by letting them preregister their study, that is, create an analysis plan prior to data collection, and share their data, materials, and code. The confrontation with real data challenged the students to think in broader terms, that is, by discovering how different methods (i.e., the Savage-Dickey density ratio and the encompassing prior approach) can be utilized to answer specific research questions.

We believe that a research project is an ideal opportunity to integrate the theory and mathematics of Bayesian inference with hands-on experience, and confront students with all aspects of the

empirical cycle. This experience gives students valuable insights into scientific practice, and equips them with problem solving skills that are necessary when they pursue their careers as psychological researchers and methodologists.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (grant number 406-17-568, 451-17-017, 743086 UNIFY).

ORCID iDs

Alexandra Sarafoglou https://orcid.org/0000-0003-0031-685X
Tim Draws https://orcid.org/0000-0001-5053-4674

Supplemental material

Supplemental material for this article is available online.

Note

1. We refer the interested reader to Wagenmakers (2007) and Wagenmakers et al. (2018) for a more detailed discussion on the benefits of Bayesian inference.

References

Aczel, B., Szaszi, B., Sarafoglou, A., Kekecs, Z., Kucharský, Š, Benjamin, D., Chambers, C. D., Fisher, A., Gelman, A., Gernsbacher, M. A., Ioannidis, J. P., Johnson, E., Jonas, K., Kousta, S., Lilienfeld, S. O., Lindsay, D. S., Morey, C. C., Munafò, M., Newell, B. R., …, Wagenmakers, E.-J. (2020). A consensus-based transparency checklist. *Nature Human Behaviour*, *4*(1), 4–6. https://doi.org/10.1038/s41562-019-0772-6

American Statistical Association. (2014). Curriculum guidelines for undergraduate programs in statistical science.

Astrolabe Inc. (2018). Astrolabe [Computer software].

Baker, M. (2016). 1,500 Scientists lift the lid on reproducibility. *Nature News*, *533*(7604), 452. https://doi.org/10.1038/533452a

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. Lord & M. Novick (Eds.), *Statistical theories of mental test scores* (pp. 395–479). Addison-Wesley.

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirch-ler, M., Nave, G., Nosek, B., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Heikensten, E., Hummer, L., Imai, T., Isaksson, S.,…, Wu, H. (2018). Evaluating replicability of social science experiments in *Nature* and *Science*. *Nature Human Behaviour*, *2*(6280), 637–644. https://doi.org/10.1038/s41562-018-0399-z

Carlson, S. (1985). A double-blind test of astrology. *Nature*, *318*, 419–425. https://doi.org/10.1038/318419a0

Cavagnaro, D. R. & Davis-Stober, C. P. (2014). Transitive in our preferences, but transitive in different ways: An analysis of choice variability. *Decision*, *1*(2), 102–122. https://doi.org/10.1037/dec0000011

Center for Open Science (2021). Open Science Framework.

Chambers, C. D. (2013). Registered reports: A new publishing initiative at *Cortex*. *Cortex*, *49*(3), 609–610. https://doi.org/10.1016/j.cortex.2012.12.016

Chambers, C. D. & Tzavella, L. (2021). The past, present and future of registered reports. *Nature Human Behaviour*, *6*, 29–42. https://doi.org/10.1038/s41562-021-01193-7

Chopik, W., Bremner, R., Defever, A. & Keller, V. (2018). How (and whether) to teach undergraduates about the replication crisis in psychological science. *Teaching of Psychology*, *45*(2), 158–163. https://doi.org/10.1177/0098628318762900

Cobb, G. (2015). Mere renovation is too little too late: We need to rethink our undergraduate curriculum from the ground up. *The American Statistician*, *69*(4), 266–282. https://doi.org/10.1080/00031305.2015.1093029

Costa, P. T. & McCrae, R. R. (1985). *The NEO personality inventory*. Psychological Assessment Resources.

Costa, P. T. & McCrae, R. R. (1992). *NEO-PI-R and NEO-FFI professional manual* (Vol. 38).

Davis-Stober, C. P. (2009). Analysis of multinomial models under inequality constraints: Applications to measurement theory. *Journal of Mathematical Psychology*, *53*(2), 1–13. https://doi.org/10.1016/j.jmp.2008.08.003

Dickey, J. M. & Lientz, B. (1970). The weighted likelihood ratio, sharp hypotheses about chances, the order of a Markov chain. *The Annals of Mathematical Statistics*, *41*, 214–226. https://doi.org/10.1214/aoms/1177697203

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*, 781. https://doi.org/10.3389/fpsyg.2014.00781

Etz, A. & Vandekerckhove, J. (2018). Introduction to Bayesian inference for psychology. *Psychonomic Bulletin & Review*, *25*(1), 5–34. https://doi.org/10.3758/s13423-017-1262-3

Funder, D. C., Levine, J. M., Mackie, D. M., Morf, C. C., Sansone, C., Vazire, S. & West, S. G. (2014). Improving the dependability of research in personality and social psychology: Recommendations for research and educational practice. *Personality and Social Psychology Review*, *18*(1), 3–12. https://doi.org/10.1177/1088868313507536

Guo, Y. & Regenwetter, M. (2014). Quantitative tests of the perceived relative argument model: Comment on loomes (2010). *Psychological Review*, *121*, 696–705. https://doi.org/10.1037/a0036095

Haaf, J. M., Merkle, E. C. & Rouder, J. N. (2020). Do items order? The psychology in IRT models. *Journal of Mathematical Psychology*, *98*, 102398. https://doi.org/10.1016/j.jmp.2020.102398

Heck, D. W. & Davis-Stober, C. P. (2019). Multinomial models with linear inequality constraints: overview and improvements of computational methods for Bayesian inference. *Journal of Mathematical Psychology*, *91*, 70–87. https://doi.org/10.1016/j.jmp.2019.03.004

Horton, N. J. & Hardin, J. S. (2015). Teaching the next generation of statistics students to "think with data": Special issue on statistics and the undergraduate curriculum. *The American Statistician*, *69*(4), 259–265. https://doi.org/10.1080/00031305.2015.1094283

Houtkoop, B. L., Chambers, C., Macleod, M., Bishop, D. V., Nichols, T. E. & Wagenmakers, E.-J. (2018). Data sharing in psychology: A survey on barriers and preconditions. *Advances in Methods and Practices in Psychological Science*, *1*(1), 70–85. https://doi.org/10.1177/2515245917751886

John, L., Loewenstein, G. & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*(5), 524–532. https://doi.org/10.1177/0956797611430953

Keysers, C., Gazzola, V. & Wagenmakers, E.-J. (2020). Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence. *Nature Neuroscience*, *23*(7), 788–799. https://doi.org/10.1038/s41593-020-0660-4

Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L.-S., Kennett, C., Slowik, A., Sonnleitner, C., Hess-Holden, C., Errington, T. M., Fiedler, S. & Nosek, B. A. (2016). Badges to acknowledge open practices: A simple, low cost, effective method for increasing transparency. *PLOS Biology*, *14*(5), e1002456. https://doi.org/10.1371/journal.pbio.1002456

Klein, R., Ratliff, K., Vianello, M., Adams Jr., R. B., Bahník, Š, Bernstein, M., Bocian, K., Brandt, M., Brooks, B., Brumbaugh, C., Cemalcilar, Z., Chandler, J., Cheong, W., Davis, W., Devos, T., Eisner, M.,

Frankowska, N., Furrow, D., Galliani, E. M., …, Nosek, B. (2014). Investigating variation in replicability: A "many labs" replication project. *Social Psychology*, *45*(3), 142–152. https://doi.org/10.1027/1864-9335/a000178

Klein, R., Vianello, M., Hasselman, F., Adams, B., Adams, R., Alper, S., Aveyard, M., Axt, J., Bahník, S., Batra, R., Berkics, M., Bernstein, M., Berry, D., Bialobrzeska, O., Binan, E., Bocian, K., Brandt, M., Busching, R., Redei, A., …, Nosek, B. (2018). Many labs 2: investigating variation in replicability across sample and setting. *Advances in Methods and Practices in Psychological Science*, *1*(4), 443–490. https://doi.org/10.1177/2515245918810225

Klugkist, I., Kato, B. & Hoijtink, H. (2005). Bayesian Model selection using encompassing priors. *Statistica Neerlandica*, *59*(1), 57–69. https://doi.org/10.1111/j.1467-9574.2005.00279.x

Klugkist, I., Laudy, O. & Hoijtink, H. (2010). Bayesian Evaluation of inequality and equality constrained hypotheses for contingency tables. *Psychological Methods*, *15*(3), 281–299. https://doi.org/10.1037/a0020137

Morey, R. D., Chambers, C. D., Etchells, P. J., Harris, C. R., Hoekstra, R., Lakens, D., Lewandowsky, S., Coker Morey, C., Newman, D. P., Schönbrodt, F. D., Vanpaemel, W., Wagenmakers, E.-J. & Zwaan, R. A. (2016). The peer reviewers' openness initiative: Incentivizing open research practices through peer review. *Royal Society Open Science*, *3*(1), 150547. https://doi.org/10.1098/rsos.150547

Morling, B. & Calin-Jageman, R. J. (2020). What psychology teachers should know about open science and the new statistics. *Teaching of Psychology*, *47*(2), 169–179. https://doi.org/10.1177/0098628320901372

Moshontz, H., Campbell, L., Ebersole, C. R., IJzerman, H., Urry, H. L., Forscher, P. S., Grahe, J. E., McCarthy, R. J., Musser, E. D., Antfolk, J., Castille, C. M., Evans, T. R., Fiedler, S., Flake, J. K., Forero, D. A., Janssen, S. M. J., Keene, J. R., Protzko, J., Aczel, B., …, Chartier, C. R. (2018). The Psychological Science Accelerator: Advancing psychology through a distributed collaborative network [Publisher: SAGE Publications Sage CA: Los Angeles, CA]. *Advances in Methods and Practices in Psychological Science*, *1*(4), 501–515. https://doi.org/10.1177/2515245918797607

Munafò, M., Nosek, B., Bishop, D., Button, K., Chambers, C., Du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J. & Ioannidis, J. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), 0021. https://doi.org/10.1038/s41562-016-0021

Myung, J., Karabatsos, G. & Iverson, G. J. (2005). A Bayesian approach to testing decision making axioms. *Journal of Mathematical Psychology*, *49*(3), 205–225. https://doi.org/10.1016/j.jmp.2005.02.004

Nature Publishing Group (2016). Reality check on reproducibility [editorial]. *Nature*, *533*, 437.

Nosek, B., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M.,…, Yarkoni, T. (2015). Promoting an open research culture. *Science (New York, N.Y.)*, *348*(6242), 1422–1425. https://doi.org/10.1126/science.aab2374

Open Science Collaboration, (2015). Estimating the reproducibility of psychological science. *Science (New York, N.Y.)*, *349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Pashler, H. & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science*, *7*(6), 531–536. https://doi.org/10.1177/1745691612463401

Pashler, H. & Wagenmakers, E.-J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, *7*(6), 528–530. https://doi.org/10.1177/1745691612465253

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. The Danish Institute for Educational Research.

Regenwetter, M., Cavagnaro, D. R., Popova, A., Guo, Y., Zwilling, C., Lim, S. H. & Stevens, J. R. (2018). Heterogeneity and parsimony in intertemporal choice. *Decision*, *5*(2), 63–94. https://doi.org/10.1037/dec0000069

Regenwetter, M., Dana, J. & Davis-Stober, C. P. (2011). Transitivity of preferences. *Psychological Review*, *118*(1), 42–56. https://doi.org/10.1037/a0021150

Rouder, J. N. (2014). Optional stopping: No problem for Bayesians. *Psychonomic Bulletin & Review*, *21*(2), 301–308. https://doi.org/10.3758/s13423-014-0595-4

Rouder, J. N. & Morey, R. D. (2018). Teaching Bayes' theorem: Strength of evidence as predictive accuracy. *The American Statistician*, 1–5. Advance online publication. https://doi.org/10.1080/00031305.2017.1341334

Sarafoglou, A., Hoogeveen, S., Matzke, D. & Wagenmakers, E.-J. (2020). Teaching good research practices: protocol of a research master course. *Psychology Learning & Teaching*, *19*(1), 46–59. https://doi.org/10.1177/1475725719858807

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, *13*(2), 90–100. https://doi.org/10.1037/a0015108

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

Simons, D. (Ed.) (2018). Challenges in making data available. *Advances in Methods and Practices in Psychological Science*, *1*. https://doi.org/10.1177/2515245918757424

Stefan, A. M., Gronau, Q. F., Schönbrodt, F. D. & Wagenmakers, E.-J. (2019). A tutorial on Bayes factor design analysis using informed priors. *Behavior Research Methods*, *51*(3), 1042–1058. https://doi.org/10.3758/s13428-018-01189-8

Tijmstra, J., Hoijtink, H. & Sijtsma, K. (2015). Evaluating manifest monotonicity using Bayes factors. *Psychometrika*, *80*(4), 880–896. https://doi.org/10.1007/s11336-015-9475-8

van Doorn, J., Matzke, D. & Wagenmakers, E.-J. (2020). An in-class demonstration of Bayesian inference. *Psychology Learning & Teaching*, *19*(1), 36–45. https://doi.org/10.1177/1475725719848574

Verhagen, J. & Wagenmakers, E.-J. (2014). Bayesian test to quantify the result of a replication attempt. *Journal of Experimental Psychology: General*, *143*(4), 1457–1475. https://doi.org/10.1037/a0036731

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of *p*-values. *Psychonomic Bulletin & Review*, *14*(5), 779–804. https://doi.org/10.3758/BF03194105

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H. & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology*, *60*(3), 158–189. https://doi.org/10.1016/j.cogpsych.2009.12.001

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q., Smira, M., Epskamp, S., Matzke, D., Rouder, J. & Morey, R. (2018). Bayesian Inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, *25*(1), 35–57. https://doi.org/10.3758/s13423-017-1343-3

Wasserstein, R. L. & Lazar, N. A. (2016). The ASA statement on p-values: context, process, and purpose. *The American Statistician*, 70(2), 129–133.

Wicherts, J., Borsboom, D., Kats, J. & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist*, *61*(7), 726–728. https://doi.org/10.1037/0003-066X.61.7.726

Witmer, J. (2017). Bayes and MCMC for undergraduates. *The American Statistician*, *71*(3), 259–264. https://doi.org/10.1080/00031305.2017.1305289

Wyman, A. J. & Vyse, S. (2008). Science versus the stars: A double-blind test of the validity of the neo five-factor inventory and computer-generated astrological natal charts. *The Journal of General Psychology*, *135*(3), 287–300. https://doi.org/10.3200/GENP.135.3.287-300

Author Biographies

Ms. Alexandra Sarafoglou is PhD candidate at the department of Psychological Methods at the University of Amsterdam. She has founded of the Open Science Community Amsterdam and is a member of the JASP development team (https://jasp-stats.org). Ms. Sarafoglou has received an NWO research talent grant to develop open science related research methods such as analysis analysis blinding, validate them empirically and make them accessible in the open-source softwares R

and JASP and through teaching materials. Her teaching interests include Bayesian inference, and open scientific practices.

Ms. Anna van der Heijden is an MSc. graduate from the University of Amsterdam. She specialized in psychological research methods during her bachelor's, did her master's in Behavioural Data Science and graduated with distinction. She is currently working as a project manager and research analyst at De Groene Lobby, a company that is working towards a more sustainable world focusing on litter prevention.

Mr. Tim Draws is PhD candidate in the Web Information Systems group at the Delft University of Technology. His research interest lies in the intersection between human behaviour, statistics, and computer science. During his PhD, Mr. Draws researchers how algorithmic and cognitive biases in web search affect user attitudes on debated topics. His teaching interests include research methods, and human computer interaction.

Mr. Joran Cornelisse is an MSc. graduate from the University of Amsterdam. He specialized in psychological research methods and informatics. He helped in developing the machine learning module of JASP during his master thesis. Currently, he works as a data scientist, where he also supervises Data Science & AI students who do their master thesis studies at SocialDatabase. Mr. Cornelisse's research and teaching interests primarily lie in Graph Neural Networks, Multimodal deep learning, and Natural Language Processing.

Prof. Eric-Jan Wagenmakers is full professor at the department of Psychological Methods at the University of Amsterdam. He has participated in drafting the Transparency and Openness Promotion (TOP) guidelines and has been a strong advocate of preregistration. He was also a member of the 2018 KNAW committee on replication research. His teaching interests include Bayesian inference, and open scientific practices.

Dr. Maarten Marsman is assistant professor at the department of Psychological Methods at the University of Amsterdam. Dr. Marsman is specialized in Bayesian inference, computational statistical methodology, psychometrics and educational measurement. He also helped develop new Bayesian statistical methods for psychological research and JASP. Dr. Marsman has received an Innovational Research Incentives Scheme Veni grant awarded by the NWO for the project "The Psychometrics of Learning". His teaching interests include Bayesian inference, psychometrics and mathematical psychology.