



UvA-DARE (Digital Academic Repository)

A Two-Stage Bayesian Sequential Assessment of Exploratory Hypotheses

Stefan, A.M.; Lengersdorff, L.L.; Wagenmakers, E.-J.

DOI

[10.1525/collabra.40350](https://doi.org/10.1525/collabra.40350)

Publication date

2022

Document Version

Final published version

Published in

Collabra: Psychology

License

CC BY

[Link to publication](#)

Citation for published version (APA):

Stefan, A. M., Lengersdorff, L. L., & Wagenmakers, E.-J. (2022). A Two-Stage Bayesian Sequential Assessment of Exploratory Hypotheses. *Collabra: Psychology*, 8, [40350]. <https://doi.org/10.1525/collabra.40350>

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, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Methodology and Research Practice

A Two-Stage Bayesian Sequential Assessment of Exploratory Hypotheses

Angelika M. Stefan^{1,2} ^a, Lukas L. Lengersdorff^{2,3}, Eric-Jan Wagenmakers² ¹ Department of Psychology, Universität der Bundeswehr München, Germany, ² Department of Psychology, University of Amsterdam, The Netherlands, ³ Department of Cognition, Emotion, and Methods in Psychology, University of Vienna, Austria

Keywords: Bayes factor, Hypothesis testing, Bayesian updating, Sequential designs, Confirmatory trial, Preregistration

<https://doi.org/10.1525/collabra.40350>

Collabra: Psychology

Vol. 8, Issue 1, 2022

Separating confirmatory and exploratory analyses is vital for ensuring the credibility of research results. Here, we present a two-stage Bayesian sequential procedure that combines a maximum of exploratory freedom in the first stage with a strictly confirmatory regimen in the second stage. It allows for flexible sampling schemes and a statistically coherent carry-over of information from the exploratory to the confirmatory stage. We believe that this procedure will facilitate preregistration as well as the formulation of precise hypotheses in the field of psychology and can be integrated elegantly into the registered report publishing framework. We demonstrate the methodology with a simulated application example from the field of social neuroscience.

A transparent distinction between exploratory and confirmatory analyses is vital for ensuring the credibility of research results (Wagenmakers et al., 2012). New publication formats, such as Registered Reports and Exploratory Reports, as well as the ascent of preregistration in the social sciences are founded on this premise (Chambers, 2013; McIntosh, 2017; Nosek & Lindsey, 2018). However, many researchers are still hesitant to commit to a single analysis pipeline in the form of a preregistration before seeing the data. This is particularly the case in disciplines such as neuroscience, where elaborate data preprocessing procedures and complex data structures make it challenging to decide on the most appropriate analysis method *a priori* (Poldrack et al., 2017).

Here we present a two-stage Bayesian sequential procedure that combines a maximum of exploratory freedom in the first stage with a strictly confirmatory regimen in the second stage, while allowing for flexible sampling schemes and a statistically coherent carry-over of information. This procedure may benefit researchers who are faced with the dilemma on whether (a) to shoot from the hip by running a confirmatory study with inadequate planning, risking severe deviations from the preregistration plan (Sarafoglou et al., 2022), or (b) to sacrifice resources to an initial exploratory study that does not allow for hypothetico-deductive inference (Jebb et al., 2017).

The Two-Stage Bayesian Sequential Procedure

[Figure 1](#) illustrates the two-stage sequential Bayesian process. At the first stage, researchers can explore a variety of different analysis plans, including alternative preprocessing techniques, statistical models and hypotheses, outcome variables, and participant inclusion criteria. Researchers are able to do this as data roll in, sequentially updating the knowledge about competing models or hypotheses (Schönbrodt et al., 2017). In the exploratory stage, there is only a single rule: “anything goes”. For example, in neuroscience, researchers may test the same conceptual hypothesis using different voxel-based definitions of a brain region (Poldrack, 2007), or different methods to select electrode channels in EEG analyses (Alotaiby et al., 2015). The exploratory stage can be stopped as soon as the researcher has identified a hypothesis and associated analysis method that is deemed sufficiently promising for a strictly confirmatory test. At this freely chosen point in time (i.e., T_1 in [Figure 1](#)), the researcher enters the confirmatory part of the study.

The second, confirmatory, stage starts by preregistering the exact analysis pipeline that was selected based on the exploratory analyses. Preregistration is now straightforward since the researcher already has analysis scripts detailing the exact analysis procedure that are based on the acquired knowledge of data structures from the first stage. The goal of the confirmatory stage is to test the concrete hypothesis extracted from the exploratory stage. The hy-

^a Correspondence:
angelika.m.stefan@gmail.com

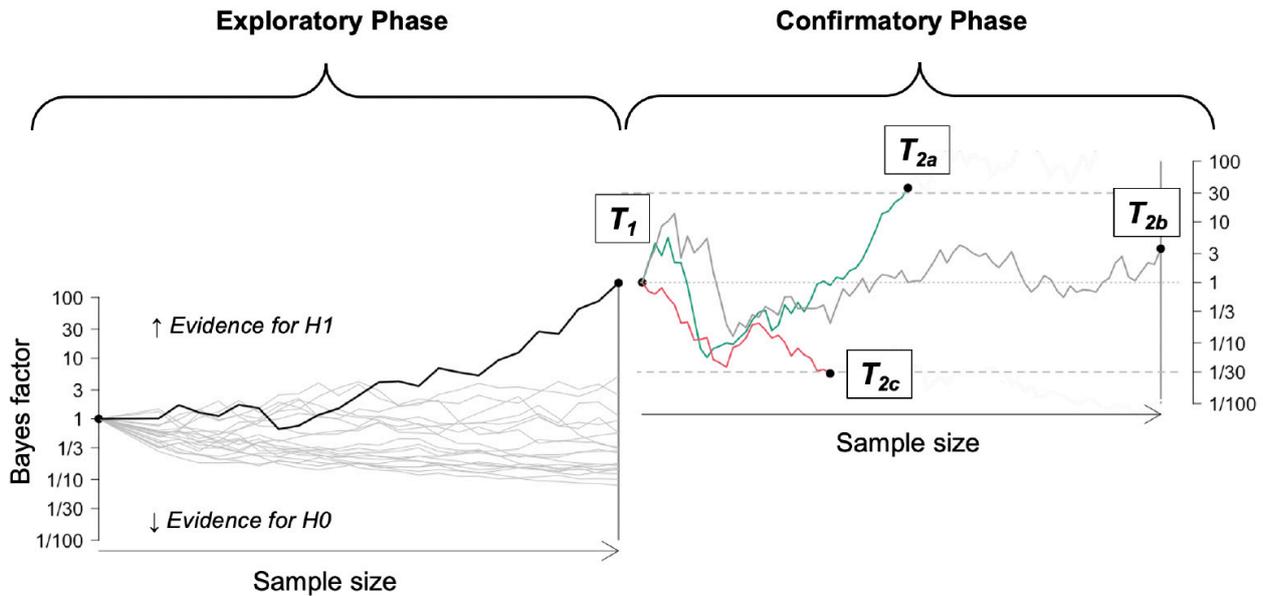


Figure 1. The two-stage sequential Bayesian procedure. In the exploratory phase, sampling continues until a promising analysis method has been selected (T_1). In the confirmatory phase the selected hypothesis is put to the test. Sampling stops once sufficient evidence has accrued for the alternative hypothesis (T_{2a}) or for the null hypothesis (T_{2c}), or until resources are depleted (T_{2b}).

hypothesis test can be conducted in a sequential manner again, where the researcher stops data collection as soon as sufficient evidence has been obtained for the null or alternative hypothesis, or a maximum sample size has been reached (e.g., Stefan et al., 2019). Evidence is quantified by means of the Bayes factor, with Bayes factors larger than one indicating evidence for the alternative hypothesis and evidence smaller than one indicating evidence for the null hypothesis (Jeffreys, 1961; Rouder et al., 2018).

Notably, the Bayesian approach allows the coherent carry-over of information from the exploratory to the confirmatory stage. Following the principle “today’s posterior is tomorrow’s prior” (Lindley, 1972, p. 2), information from the exploratory stage can be used to formulate prior distributions on all model parameters in the confirmatory stage. This can be viewed as enriching the hypothesis based on prior knowledge, or putting probabilistic constraints on the parameter space to make the models more informative (Lee & Vanpaemel, 2017). The easiest way to do this is to use the posterior distributions from the exploratory stage as priors for the confirmatory stage (Ly et al., 2018; Verhagen & Wagenmakers, 2014). However, researchers who worry that their exploratory results may be overoptimistic may adopt a more cautionary approach and discount the information from the first stage to some degree – for instance by using power priors (Chen et al., 2000) or by incorporating knowledge about the results from the alternative, less promising analyses; doing so will shrink the prior distribution for the confirmatory stage toward the null value.

The confirmatory stage can result in three qualitatively distinct outcomes: The exploratory results from the first stage are supported, as indicated by compelling evidence in favor of the alternative hypothesis (e.g., T_{2a} in Figure

1); the exploratory results are disconfirmed, as indicated by compelling evidence in favor of the null hypothesis (e.g., T_{2c} in Figure 1), or the data remain ambiguous with regard to the tested hypotheses (e.g., T_{2b} in Figure 1). The latter outcome can occur when resource constraints prohibit the continuation of data collection and true parameter values fall somewhere in between the values postulated in the null and alternative model.

Neural Correlates of Perspective Taking: A Simulated Application Example

In the following, we briefly illustrate the practical use of the two-stage Bayesian sequential procedure with a simulated application example from the field of social neuroscience. The simulation code and a detailed documentation can be found on <https://osf.io/z3ckm/>.

A group of fictitious researchers conduct a functional MRI study in which participants complete a task measuring the neural correlates of perspective taking (Lamm et al., 2019; Schurz et al., 2014). The participants complete this task under two different conditions. In one condition, they are subjected to some treatment that should increase their perspective taking abilities. In the other condition, this treatment is absent.¹ The goal of the study is to determine whether the treatment causes a change in task-related activation in the right parietal cortex.

Based on previous findings, researchers expect that the effect of interest should be localized in one of the ten sub-areas of the right parietal cortex described in the brain atlas by Mars et al. (2011), see Figure 2A. However, due to the novelty of the treatment, they are uncertain which exact subarea might be affected. Moreover, they are not sure

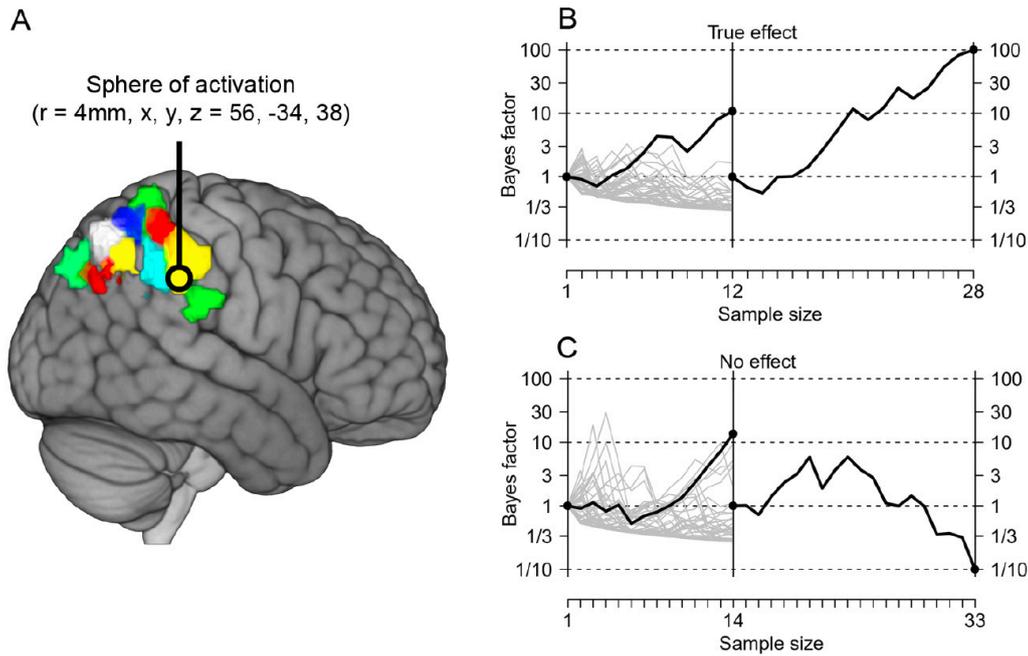


Figure 2. The simulated application example.

A: View of the right hemisphere of the brain. Marked in different colors are the ten subareas of the right parietal cortex as described by Mars et al. (2011). For each of these ten subareas, there exist four different size definitions (not depicted here), resulting in a total of 40 different possible regions of interest for analysis. The black circle marks the position and size of the sphere of activation as used in the simulation study. B and C: Example trajectories of Bayes factors obtained in the simulated application example, with B) a true effect of the treatment present, and C) no effect present. In the exploratory phase (left part of each plot), researchers monitor the Bayes factors for each of the different ROIs. After a minimum of ten participants, the researchers stop the exploratory phase as soon as one ROI has given a Bayes factor larger than 10 in favor of their hypothesis. In the confirmatory phase (right part of each plot), the researchers sequentially test their research hypothesis using the ROI definition identified in the exploratory phase. The researchers sample participants until they achieve a Bayes factor larger than 100 (as depicted in B), which they consider compelling evidence for their hypothesis; or until they achieve a Bayes factor smaller than 1/10 (as depicted in C), which they consider enough evidence against their hypothesis; or until they run out of resources (not depicted here).

of the size of the region of interest (ROI) they should use in their analysis. For each subarea, the brain atlas offers four different size definitions, but the researchers have no prior knowledge about which definition might be the best one. Thus, in total there are 40 different possible definitions of the ROI to choose from, and the researchers are struggling to preregister the specific ROI that they wish to use in their analysis. To overcome this difficulty, the researchers follow the Bayesian two-stage design, using the following decision strategies:

- In the exploratory phase, the researchers monitor the Bayesian hypothesis test for each ROI. They decide to measure at least ten participants, and then stop the exploratory phase as soon as one of the ROIs results in a Bayes factor larger than 10. Then, they choose this ROI for their confirmatory study, and preregister it accordingly. If no ROI definition results in a BF larger than 10 after 30 participants, the researchers cancel the study without a confirmatory phase.
- In the confirmatory phase, the researchers first measure ten participants, and then monitor the Bayes factors until one of three events occurs that cause the

study to be discontinued: Either they observe a BF larger than 100, which they consider compelling evidence for an effect of the treatment on brain activity; or they observe a BF smaller than 1/10, which they consider enough evidence that the treatment has no substantial effect on brain activity after all; or they run out of time and financial resources for the study.

We simulated 100 runs of the described two-stage procedure for two different scenarios. The simulation code and figures depicting all of these runs can be found on <https://osf.io/z3ckm/>. In the first scenario, there was a sphere of increased activity due to the treatment in one of the specified ROIs (radius = 4mm, MNI coordinates: $x = 56$, $y = -34$, $z = 38$, cf. Figure 2A). In 79 of the simulated runs, the researchers ended up with substantial evidence in favor of the treatment effect in the confirmatory phase (cf. Figure 2B). In 14 runs, the researchers identified a promising ROI in the exploratory phase, but then obtained strong evidence for a null effect in the confirmatory phase. Finally, seven runs yielded inconclusive evidence. In the second simulation scenario, the treatment had no effect on brain activity in any of the ROIs. In 45 of the runs, the study was stopped

¹ We remain deliberately vague about the exact nature of the task and the treatment to emphasize that this scenario is completely fictional.

after the exploratory phase, because no ROI reached a $BF > 10$. Of the remaining runs, 51 runs resulted in strong evidence for the null hypothesis (cf. [Figure 2C](#)), only one run resulted in strong evidence for the alternative hypothesis, and 3 runs resulted in inconclusive evidence in the confirmatory phase.

In summary, the two-stage procedure appears to be a practical research method in this fictitious application example. In most cases where there really is a treatment effect, the exploratory stage allows the researchers to quickly specify the most appropriate ROI, and then find convincing evidence for the effect in the confirmatory stage, using no more resources than necessary. Despite its high degree of flexibility, the procedure also nearly never leads the researchers to erroneously claim the existence of an effect if a true effect is absent. Although it can happen by random chance that the researchers identify an apparently promising ROI in the exploratory phase, this spurious result would rarely achieve a convincing level of evidence in the confirmatory phase. It is illustrative to contrast this with the situation that would arise if the researchers would forego the confirmatory phase, and base their conclusions solely on the results of their exploratory analyses. Then, the large number of researcher degrees of freedom would surely lead to a high false-positive rate (54% of cases in our example, if $BF > 10$ is used as a threshold). We would like to emphasize, though, that the error rates we report here are specific to the simulation scenario we used. Additional research will be needed to discern how often the two-stage procedure leads to the correct decision in different situations.

Discussion

The two-stage Bayesian sequential procedure offers multiple advantages: (1) The proposed procedure facilitates the specification of precise hypotheses and their translation to statistical models; (2) Analyses in the confirmatory stage can be fixed to the precise setup that was piloted in the exploratory stage, making preregistration straightforward and potentially reducing the risk of unplanned deviations; (3) The Bayesian framework allows for a seamless integration of knowledge obtained from exploratory analyses into the confirmatory trial; (4) The sampling plan is flexible and efficient. Sequential testing has repeatedly been shown to be about 50% more efficient than conducting fixed- N trials with the same power (Schnuerch & Erdfelder, 2020; Schönbrodt et al., 2017), and the Bayesian approach allows for ad-hoc adjustments of the sampling plan, for example to react to changes in available resources (Rouder, 2014); (5) Overconfidence based on exploratory results leads to decreased predictive accuracy of the alternative hypothesis in the confirmatory stage and is naturally penalized in the Bayes factor. This means that researchers are motivated to formulate realistic expectations and be conscious about their modeling choices.

Conceptually, the Bayesian two-stage design is similar to a classic pilot-study design where an exploratory pilot phase can be used to gain experience with the materials, procedure, and data environment, for the purpose of subsequently conducting a confirmatory trial (Leon et al., 2011).

However, the proposed two-stage design possesses two important features that distinguish it from classic pilot study designs. First, data collection in both stages is conducted in a sequential manner, leading to substantial efficiency benefits compared to traditional pilot study designs where sample sizes (at least in the confirmatory stage) are typically fixed in advance (Schönbrodt et al., 2017). Second, data are not discarded after the pilot study, but instead used to enrich the tested hypotheses in the confirmatory phase in a mathematically coherent manner that takes the uncertainty about population parameters into account. This does not only make the confirmatory tests more specific, but can also increase their efficiency with regard to expected sample sizes (Stefan et al., 2019).

In our view, the two-stage sequential Bayesian procedure is particularly well-suited for publication in the Registered Report format (Chambers, 2013). The first stage of the Registered Report can serve as a platform to transparently report the analysis results conducted in the exploratory phase and to determine the exact analysis path for the confirmatory phase in consultation with the reviewers. In the second stage of the Registered Report, the confirmatory phase can be executed and reported. Thus, the procedure of Registered Reports closely mirrors the two-stage Bayesian sequential design. However, may also be beneficial to publish the two stages of the design separately, particularly if exploratory analyses or the derived experimental design are sufficiently complex to make an independent contribution to the literature. In this case, the first stage can, for example, follow the format of an exploratory report (McIntosh, 2017) or of a study protocol submission, as it is common for clinical trials (Li et al., 2016).

It is important to note that the Bayesian sequential two-stage procedure does not relieve researchers of their due diligence. Although “anything goes” in the exploratory stage, it is still important to distinguish exploratory from confirmatory results. If analyses are optimized for unambiguous hypothesis testing results in the exploratory stage, parameter estimates resulting from these analyses might be inflated (Simonsohn, 2014). Moreover, researchers need to be careful to avoid the double-use of data: The prior distribution formulated based on exploratory data should not be re-used to analyze the same exploratory data (or a combined data set). Lastly, the prior distribution under the alternative hypothesis should not be adjusted if the evidence in the first stage pointed towards the null hypothesis (Jeffreys, 1961). This is due to the fact that under a true null effect the posterior under the alternative model will mimic the null model, which makes the models virtually indistinguishable in the confirmatory stage. Additionally, researchers need to be aware in this case that if their inferential goals changed to confirming a null result based on initial evidence for the null model, the criteria for selecting an analysis pathway based on the exploratory results may change as well. For example, while they might find it worthwhile to focus on a subgroup of participants in which an effect seems to be strongest to demonstrate the existence of a phenomenon, they might find that confirming the absence

of said effect in the whole population might be more interesting.

In general, researchers need to carefully consider their analysis choices and reflect on their theoretical implications. For example, simply choosing the analysis pathway yielding the strongest evidence in the exploratory phase may not yield the most severe test of a substantive theory. Additionally, every specific analysis pathway may limit the external validity of the study by restricting the generalizability to a specific context and procedure (Lin et al., 2021). It also needs to be considered that if the posterior from exploratory analyses is used as a prior, the tested alternative hypothesis in the confirmatory stage is an informed version of the alternative hypothesis tested in the exploratory stage, so the hypothesis tests in the two stages answer slightly different research questions (Etz et al., 2018).

Overall, we believe that the two-stage Bayesian sequential assessment of exploratory hypotheses addresses several methodological concerns. It facilitates preregistration, takes resource constraints seriously, elegantly connects exploratory and confirmatory aspects of research, motivates researchers to carefully consider their analytic choices, and allows researchers to quantify evidence in favor and against their focal hypothesis. We hope that the procedure will be a valuable addition to the methodological toolbox of many social science researchers.

.....

Contributions

AMS, LLL, and EJW conceptualized the manuscript. AMS wrote the first draft, AMS, LLL, and EJW edited the manuscript. AMS and LLL conceptualized the application example, LLL wrote the code underlying the simulations. AMS and LLL wrote the code underlying the figures. AMS, LLL, and EJW approved the submitted version for publication.

Competing interests

EJW leads the development of the open-source software package JASP (<https://jasp-stats.org>), a non-commercial, publicly-funded effort to make Bayesian and non-Bayesian statistics accessible to a broader group of researchers and students. The other authors declare no competing interests.

Data accessibility

No empirical data is associated with this manuscript. Code to reproduce the figures and the simulated application example can be found on <https://osf.io/z3ckm/>.

Funding

AMS was supported by the NWO Research Talent Programme (406.18.556). EJW was supported by an NWO Vici grant (016.Vici.170.083) and by a European Research Council Advanced Grant (743086 UNIFY). LLL was supported by an OeAD Marietta-Blau grant (MPC-2021-00158).

Submitted: April 11, 2022 PST, Accepted: October 21, 2022 PST



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

References

- Alotaiby, T., El-Samie, F. E. A., Alshebeili, S. A., & Ahmad, I. (2015). A review of channel selection algorithms for EEG signal processing. *EURASIP Journal on Advances in Signal Processing*, 2015(1), 1–21. <https://doi.org/10.1186/s13634-015-0251-9>
- 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>
- Chen, M.-H., Ibrahim, J. G., & Shao, Q.-M. (2000). Power prior distributions for generalized linear models. *Journal of Statistical Planning and Inference*, 84(1), 121–137. [https://doi.org/10.1016/s0378-3758\(9\)00140-8](https://doi.org/10.1016/s0378-3758(9)00140-8)
- Etz, A., Haaf, J. M., Rouder, J. N., & Vandekerckhove, J. (2018). Bayesian inference and testing any hypothesis you can specify. *Advances in Methods and Practices in Psychological Science*, 1(2), 281–295. <https://doi.org/10.1177/2515245918773087>
- Jebb, A. T., Parrigon, S., & Woo, S. E. (2017). Exploratory data analysis as a foundation of inductive research. *Human Resource Management Review*, 27(2), 265–276. <https://doi.org/10.1016/j.hrmr.2016.08.003>
- Jeffreys, H. (1961). *Theory of probability* (Third). Oxford University Press.
- Lamm, C., Rütgen, M., & Wagner, I. C. (2019). Imaging empathy and prosocial emotions. *Neuroscience Letters*, 693, 49–53. <https://doi.org/10.1016/j.neulet.2017.06.054>
- Lee, M. D., & Vanpaemel, W. (2017). Determining informative priors for cognitive models. *Psychonomic Bulletin & Review*, 25(1), 114–127. <https://doi.org/10.3758/s13423-017-1238-3>
- Leon, A. C., Davis, L. L., & Kraemer, H. C. (2011). The role and interpretation of pilot studies in clinical research. *Journal of Psychiatric Research*, 45(5), 626–629. <https://doi.org/10.1016/j.jpsychires.2010.10.008>
- Li, T., Boutron, I., Salman, R. A.-S., Cobo, E., Fleming, E., Grimshaw, J. M., & Altman, D. G. (2016). Review and publication of protocol submissions to Trials – what have we learned in 10 years? *Trials*, 18(1), 34, s13063-016-1743-0. <https://doi.org/10.1186/s13063-016-1743-0>
- Lin, H., Werner, K. M., & Inzlicht, M. (2021). Promises and perils of experimentation: The mutual-internal-validity problem. *Perspectives on Psychological Science*, 16(4), 854–863. <https://doi.org/10.1177/1745691620974773>
- Lindley, D. V. (1972). Bayesian statistics: A review. *Bayesian Statistics*, 1–74. <https://doi.org/10.1137/1.9781611970654.ch1>
- Ly, A., Etz, A., Marsman, M., & Wagenmakers, E.-J. (2018). Replication Bayes factors from evidence updating. *Behavior Research Methods*, 51(6), 2498–2508. <https://doi.org/10.3758/s13428-018-1092-x>
- Mars, R. B., Jbabdi, S., Sallet, J., O'Reilly, J. X., Croxson, P. L., Olivier, E., Noonan, M. P., Bergmann, C., Mitchell, A. S., Baxter, M. G., Behrens, T. E. J., Johansen-Berg, H., Tomassini, V., Miller, K. L., & Rushworth, M. F. S. (2011). Diffusion-weighted imaging tractography-based parcellation of the human parietal cortex and comparison with human and macaque resting-state functional connectivity. *Journal of Neuroscience*, 31(11), 4087–4100. <https://doi.org/10.1523/jneurosci.5102-10.2011>
- McIntosh, R. D. (2017). Exploratory reports: A new article type for Cortex. *Cortex*, 96, A1–A4. <https://doi.org/10.1016/j.cortex.2017.07.014>
- Nosek, B. A. N., & Lindsey, S. E. (2018). Preregistration becoming the norm in psychological science. *APS Observer*, 31(3). <https://www.psychologicalscience.org/observer/preregistration-becoming-the-norm-in-psychological-science>
- Poldrack, R. A. (2007). Region of interest analysis for fMRI. *Social Cognitive and Affective Neuroscience*, 2(1), 67–70. <https://doi.org/10.1093/scan/nsm006>
- Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò, M. R., Nichols, T. E., Poline, J.-B., Vul, E., & Yarkoni, T. (2017). Scanning the horizon: Towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience*, 18(2), 115–126. <https://doi.org/10.1038/nrn.2016.167>
- 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., Haaf, J. M., & Vandekerckhove, J. (2018). Bayesian inference for psychology, part IV: Parameter estimation and Bayes factors. *Psychonomic Bulletin & Review*, 25(1), 102–113. <https://doi.org/10.3758/s13423-017-1420-7>
- Sarafoglou, A., Hoogeveen, S., & Wagenmakers, E.-J. (2022). Comparing analysis blinding with preregistration in the Many-Analysts Religion project. *PsyArXiv Preprint*. <https://doi.org/10.31234/osf.io/6dn8f>
- Schnuerch, M., & Erdfelder, E. (2020). Controlling decision errors with minimal costs: The sequential probability ratio test. *Psychological Methods*, 25(2), 206–226. <https://doi.org/10.1037/met0000234>
- Schönbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods*, 22(2), 322–339. <https://doi.org/10.1037/met0000061>
- Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews*, 42, 9–34. <https://doi.org/10.1016/j.neubiorev.2014.01.009>
- Simonsohn, U. (2014). Posterior hacking: Selective reporting invalidates Bayesian results also. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2374040>

Stefan, A. M., Gronau, Q. F., Schönbrodt, F. D., & Wagenmakers, E.-J. (2019). A tutorial on Bayes Factor Design Analysis using an informed prior. *Behavior Research Methods*, *51*(3), 1042–1058. <https://doi.org/10.3758/s13428-018-01189-8>

Verhagen, J., & Wagenmakers, E.-J. (2014). Bayesian tests 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., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, *7*(6), 632–638. <https://doi.org/10.1177/1745691612463078>

Supplementary Materials

Version 2 Response Letter

Download: https://collabra.scholasticahq.com/article/40350-a-two-stage-bayesian-sequential-assessment-of-exploratory-hypotheses/attachment/104198.pdf?auth_token=sAP2cSq8ovTFawGTZ8aj

Peer Review History

Download: https://collabra.scholasticahq.com/article/40350-a-two-stage-bayesian-sequential-assessment-of-exploratory-hypotheses/attachment/104199.docx?auth_token=sAP2cSq8ovTFawGTZ8aj
