Water Tasting: A Bayesian Multiverse Analysis

Suzanne Hoogeveen◉

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

suzanne.j.hoogeveen@gmail.com

Abstract

Varying data inclusion criteria and prior settings, a Bayesian multiverse analysis failed to provide robust evidence for the claim that people prefer Chaudfontaine over tap water.

Target article

E.-J. Wagenmakers and F. Bartoš, *People prefer the taste of Belgian mineral water Chaudfontaine over Amsterdam tap water,* PsyArXiv (2025), doi:10.31234/osf.io/2u84v

1 Goal

In order to assess the robustness of the evidence for a preference of Chaudfontaine over tap water [1], we conducted a Bayesian multiverse analysis [2], see also [3–5]. In a multiverse analysis, the researcher can evaluate different potential constellations of the data (e.g., exclusions, theoretically relevant subgroups), priors settings, and included predictors without committing to one—perhaps arbitrarily chosen—analysis path.

2 Methods

Here we conducted a Bayesian multiverse analysis using all 48 unique paths based on 16 data constellations and 3 prior settings. In particular, we systematically combined inclusion criteria related to discriminatory ability (excluding participants who reported no difference in taste between the two cups), reasonable (i.e., modest) taste differences (excluding participants who reported a difference of more than 15 points on the 100-pt scale), reasonable expectation (excluding participants who reported an *expected* difference of more than 15 points on the 100-pt scale), source identification (excluding participants who failed to identify the source of the water correctly), resulting in 16 unique data sets. In addition, we assessed robustness to prior settings by calculating the evidence in favor of the key hypothesis of interest under three different directional priors, namely a "default" zero-centered Cauchy distribution with scale of 0.707 [6], the Oosterwijk prior (a t-distribution with location = 0.35, scale = 0.102, and 3 degrees of freedom [7, 8], and the Vohs prior (a normal distribution with M = 0.3 and
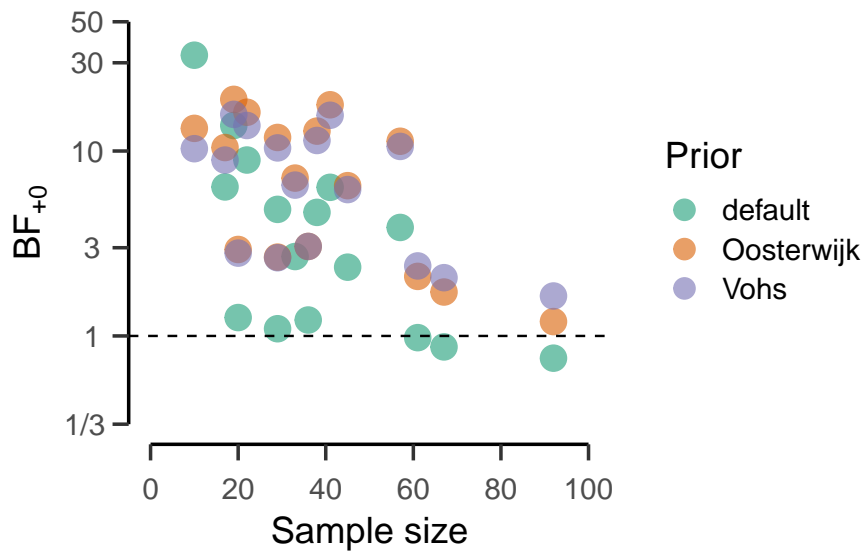
Figure 1: Results from the Bayesian multiverse analysis. Bayes factors in favor of a Chaudfontaine preference effect are above the dashed horizontal line; Bayes factors against a Chaudfontaine preference effect are below the horizontal line. The *x*-axis reflects to the number of subjects the analysis is based on. The color of the points refers to the different prior settings. Overall, the multiverse indicates a lack of robustness in the evidence for the hypothesis. Most evidence comes from the path that included data only from subjects who (1) tasted a difference between the two cups, (2) reported and, (3) expected a realistic difference in taste between the two cups, and (4) correctly identified the two cups, under the default prior for the analysis ($n = 10$). Least evidence comes from the path that included data from the full sample, also under the default prior for the analysis ($n = 92$).

SD $= 0.15$ [9].

Across all multiverse paths, a one-sided Bayesian paired-samples *t*-test was conducted to test the hypothesis that people prefer Chaudfontaine mineral water over Amsterdam tap water. The `jaspTTests` module in R was used to run the analyses [10].

3   Results

The 48 Bayes factors (BF_{+0}) resulting from the multiverse analysis are visualized in Figure 1. The Bayes factors range from 0.758 to 32.96, with 93.8% (45/48) reflecting evidence in favor of the hypothesis (plotted above the horizontal line in Figure 1). However, for only 17 out of 48 paths the evidence can be qualified as strong (i.e., BF >10) yet these analyses are all based on fewer than 60 subjects. It is noteworthy that the strongest evidence is obtained from the smallest samples while samples with more than 60 subjects all indicate the absence of evidence for or against the preference for Chaudfontaine over tap water. Even though Bayes factors take the number of observations into account, this pattern emphasizes the risk of small samples and subsetting of data, which enhances the influence of extreme cases. Finally, the effect of prior settings mostly reflects the level of informativeness of the priors; effect sizes under the Oosterwijk and Vohs prior show stronger pulling towards the expected value in the prior, resulting in more conservative evidence in the smallest sample, yet more optimistic results for larger samples (reflecting the fact that the informed priors are more optimistic, that is, centered at a positive value rather than at zero).

4   Conclusion

The Bayesian multiverse analysis indicates that the evidence for a taste preference of Chaud-fontaine bottled water over Amsterdam tap water reported by [1] is fragile. Although most of the 48 explored analysis paths resulted in positive evidence for the effect, only 1/3 reflected strong evidence. Moreover, the results were rather sensitive to prior settings, which makes sense in light of the small sample. Since substantial evidence is solely obtained from paths that relied on (a combination of) strict exclusion criteria that resulted in discarding about half of the original sample, we do not consider the claim that people prefer Chaudfontaine water over tap water sufficiently supported in this particular experiment.

Acknowledgments and Disclosures

Reproducibility   We were able to computationally reproduce the original analysis and re-sults.

Code and Data Availability   Readers can access the the R code to conduct the multiverse analysis in our OSF folder at: https://osf.io/b5ndt/.

Funding   The author received no financial support for the research, authorship, and/or pub-lication of this article.

Conflicts of Interest   This Robustness Report accompanies the editorial [11] and is meant to illustrate the format and scope of a typical contribution to the *Journal of Robustness Reports*.

References

[1] E.-J. Wagenmakers and F. Bartoš, *People prefer the taste of Belgian mineral water Chaud-fontaine over Amsterdam tap water*, PsyArXiv (2025), doi:10.31234/osf.io/2u84v.

[2] S. Steegen, F. Tuerlinckx, A. Gelman and W. Vanpaemel, *Increasing transparency through a multiverse analysis*, Perspectives on Psychological Science **11**(5), 702 (2016), doi:https://doi.org/10.1177/1745691616658637.

[3] S. Hoogeveen, S. W. Berkhout, Q. F. Gronau, E.-J. Wagenmakers and J. M. Haaf, *Improv-ing statistical analysis in team science: The case of a Bayesian multiverse of Many Labs 4*, Advances in Methods and Practices in Psychological Science **6**(3), 25152459231182318 (2023), doi:https://doi.org/10.1177/25152459231182318.

[4] W. Tierney, J. Hardy, III, C. Ebersole, D. Viganola, E. Clemente, E. Gordon, S. Hoogeveen, J. M. Haaf, A. Dreber, M. Johannesson, T. Pfeiffer, H. Chapman *et al.*, *A creative destruction approach to replication: Implicit work and sex moral-ity across cultures.*, Journal of Experimental Social Psychology **93**, 104060 (2021), doi:https://doi.org/10.1016/j.jesp.2020.104060.

[5] W. Tierney, W. Cyrus-Lai, S. Hoogeveen, J. Haaf, B. Landy, H. J. III and E. Uhlmann, *Who respects an angry woman? A pre-registered re-examination of the relationships between gender, emotion expression, and status conferral*, Unpublished manuscript (2024).

[6] R. D. Morey and J. N. Rouder, *BayesFactor: Computation of Bayes factors for common designs*, R package version 0.9.12-4.3 (2015).

[7] Q. F. Gronau, A. Ly and E.-J. Wagenmakers, *Informed Bayesian t-tests*, The American Statistician **74**, 137 (2020).

[8] Q. F. Gronau, S. van Erp, D. W. Heck, J. Cesario, K. J. Jonas and E.-J. Wagenmakers, *A Bayesian model-averaged meta-analysis of the power pose effect with informed and default priors: The case of felt power*, Comprehensive Results in Social Psychology **2**(1), 123 (2017), doi:https://doi.org/10.1080/23743603.2017.1326760.

[9] K. D. Vohs, B. J. Schmeichel, S. Lohmann, Q. F. Gronau, A. J. Finley, S. E. Ainsworth, J. L. Alquist, M. D. Baker, A. Brizi, A. Bunyi, G. J. Butschek, C. Campbell *et al.*, *A multisite preregistered paradigmatic test of the ego-depletion effect*, Psychological Science **32**(10), 1566 (2021), doi:10.1177/0956797621989733.

[10] JASP Team, *jaspTTests: T-Test Module for JASP* (2022).

[11] F. Bartoš, A. Sarafoglou, B. Aczel, S. Hoogeveen, C. D. Chambers and E.-J. Wagenmakers, *Introducing the Journal of Robustness Reports*, Journal of Robustness Reports (2025).