Bayes factors for research workers

Ly, A.

Publication date
2018
Document Version
Other version
License
Other

References

Abramowitz, M. and Stegun, I. A. (1964). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55 of *National Bureau of Standards Applied Mathematics Series*. Dover Publications. 108, 157, 160

Ahn, W.-Y., Busemeyer, J. R., Wagenmakers, E.-J., and Stout, J. C. (2008). Comparison of decision learning models using the generalization criterion method. *Cognitive Science*, 32:1376–1402. 195

Ahn, W.-Y., Krawitz, A., Kim, W., Busemeyer, J. R., and Brown, J. W. (2011). A model-based fMRI analysis with hierarchical Bayesian parameter estimation. *Journal of Neuroscience Psychology and Economics*, 4:95–110. 173, 201

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723. 231

Aldrich, J. (2005). The statistical education of Harold Jeffreys. *International Statistical Review*, 73(3):289–307. 225

Amari, S.-I., Barndorff-Nielsen, O. E., Kass, R. E., Lauritzen, S. L., and Rao, C. R. (1987). *Differential geometry in statistical inference*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 10. Institute of Mathematical Statistics, Hayward, CA. 253

Andraszewicz, S., Scheibehenne, B., Rieskamp, J., Grasman, R. P. P. P., Verhagen, A. J., and Wagenmakers, E.-J. (2015). An introduction to Bayesian hypothesis testing for management research. *Journal of Management*, 41:521–543. 37

Andrews, M. and Baguley, T. (2013). Prior approval: The growth of Bayesian methods in psychology. *British Journal of Mathematical and Statistical Psychology*, 66:1–7. 171

Anscombe, F. J. (1973). Graphs in statistical analysis. *The American Statistician*, 27(1):17–21. 37, 128

Atkinson, C. and Mitchell, A. (1981). Rao's distance measure. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 345–365. 253

Bailey, W. N. (1964). *Generalized hypergeometric series*. Cambridge Tracts in Mathematics and Mathematical Physics, No. 32. Stechert-Hafner, Inc., New York. 162

Baker, M. (2016). Is there a reproducibility crisis? *Nature*, 533:452–454. 4, 301

Balasubramanian, V. (1996). A geometric formulation of Occam's razor for inference of parametric distributions. *arXiv preprint adap-org/9601001*. 232, 241

Banterle, M., Grazian, C., Lee, A., and Robert, C. P. (2015). Accelerating Metropolis-Hastings algorithms by delayed acceptance. *arXiv preprint arXiv:1503.00996*. 243

Barber, T. X. (1976). *Pitfalls in Human Research: Ten Pivotal Points*. Pergamon Press Inc., New York. 117

Bark, R., Dieckmann, S., Bogerts, B., and Northoff, G. (2005). Deficit in decision making in catatonic schizophrenia: An exploratory study. *Psychiatry Research*, 134:131–141. 194

Bartlett, M. S. (1957). A comment on D. V. Lindley's statistical paradox. *Biometrika*, 44:533–534. 18, 103

Batchelder, W. H. and Riefer, D. M. (1980). Separation of storage and retrieval factors in free recall of clusterable pairs. *Psychological Review*, 87:375–397. 213

Batchelder, W. H. and Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, 6:57–86. 232

Bateman, H., Erdélyi, A., Magnus, W., Oberhettinger, F., Tricomi, F. G., Bertin, D., Fulks, W. B., Harvey, A. R., Thomsen Jr, D. L., Weber, M. A., Whitney, E. L., and Stampfel, R. (1953). *Higher transcendental functions*, volume 2. McGraw-Hill New York. 157

Bateman, H., Erdélyi, A., Magnus, W., Oberhettinger, F., Tricomi, F. G., Bertin, D., Fulks, W. B., Harvey, A. R., Thomsen Jr, D. L., Weber, M. A., Whitney, E. L., and Stampfel, R. (1954). *Tables of integral transforms*, volume 1. McGraw-Hill. 86, 152, 157

Bayarri, M. J. (1981). Inferencia Bayesiana sobre el coeficiente de correlación de una población normal bivariante. *Trabajos de estadística y de investigación operativa*, 32(3):18–31. 150, 151

Bayarri, M. J., Benjamin, D. J., Berger, J. O., and Sellke, T. M. (2016). Rejection odds and rejection ratios: A proposal for statistical practice in testing hypotheses. *Journal of Mathematical Psychology*, 72:90–103. 116, 171

Bayarri, M. J. and Berger, J. O. (2000). P values for composite null models. *Journal of the American Statistical Association*, 95(452):1127–1142. 58

Bayarri, M. J. and Berger, J. O. (2013). Hypothesis testing and model uncertainty. In Damien, P., Dellaportas, P., Polson, N. G., and Stephens, D. A., editors, *Bayesian Theory and Applications*, pages 361–394. Oxford University Press, Oxford. 103

Bayarri, M. J., Berger, J. O., Forte, A., and García-Donato, G. (2012). Criteria for Bayesian model choice with application to variable selection. *The Annals of Statistics*, 40(3):1550–1577. 25, 42, 47, 83, 103, 222, 268

Bayarri, M. J. and Mayoral, A. M. (2002). Bayesian analysis and design for comparison of effect–sizes. *Journal of Statistical Planning and Inference*, 103:225–243. 130

Bechara, A., Damasio, A. R., Damasio, H., and Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, 50:7–15. 174, 193, 194, 205

Bechara, A., Damasio, H., Damasio, A. R., and Lee, G. P. (1999). Different contributions of the human amygdala and ventromedial prefrontal cortex to decision-making. *Journal of Neuroscience*, 19:5473–5481. 194

Bechara, A., Damasio, H., Tranel, D., and Anderson, S. W. (1998). Dissociation of working memory from decision making within the human prefrontal cortex. *Journal of Neuroscience*, 18:428–437. 194

Bechara, A., Damasio, H., Tranel, D., and Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, 275:1293–1295. 194

Bechara, A., Tranel, D., and Damasio, H. (2000). Characterization of the decision-making deficit of patients with ventromedial prefrontal cortex lesions. *Brain*, 123:2189–2202. 194

Bennett, C. H. (1976). Efficient estimation of free energy differences from Monte Carlo data. *Journal of Computational Physics*, 22:245–268. 171, 173

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis*. Springer Verlag. 37, 43, 242

Berger, J. O. (2006). Bayes factors. In Kotz, S., Balakrishnan, N., Read, C., Vidakovic, B., and Johnson, N. L., editors, *Encyclopedia of Statistical Sciences, vol. 1 (2nd ed.)*, pages 378–386. Wiley, Hoboken, NJ. 10, 127, 131, 173

Berger, J. O., Bernardo, J. M., and Sun, D. (2015). Overall objective priors. *Bayesian Analysis*, 10(1):189–221. 150

Berger, J. O. and Berry, D. A. (1988). The relevance of stopping rules in statistical inference. In Gupta, S. S. and Berger, J. O., editors, *Statistical Decision Theory and Related Topics: Vol. 4*, pages 29–72. Springer Verlag, New York. 82

Berger, J. O. and Delampady, M. (1987). Testing precise hypotheses. *Statistical Science*, pages 317–335. 120

Berger, J. O. and Molina, G. (2005). Posterior model probabilities via path-based pairwise priors. *Statistica Neerlandica*, 59:3 – 15. 172

Berger, J. O. and Pericchi, L. R. (2001). Objective Bayesian methods for model selection: Introduction and comparison. *Lecture Notes-Monograph Series*, pages 135–207. 47

Berger, J. O., Pericchi, L. R., and Varshavsky, J. A. (1998). Bayes factors and marginal distributions in invariant situations. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 307–321. 45, 103, 256

Berger, J. O. and Sun, D. (2008). Objective priors for the bivariate normal model. *The Annals of Statistics*, pages 963–982. 73, 150, 151

Berger, J. O. and Wolpert, R. L. (1988). *The Likelihood Principle (2nd ed.)*. Institute of Mathematical Statistics, Hayward (CA). 116

Berkson, J. (1938). Some difficulties of interpretation encountered in the application of the chi-square test. *Journal of the American Statistical Association*, 33(203):526–536. 14

Bernardo, J. M. (2015). An overall prior for the five-parameter normal distribution. Paper presented at the 11th International Workshop on Objective Bayes Methodology dedicated to Susie Bayarri. 149

Bickel, P. J. (2006). Regularization in statistics. *Test*, 15(2):271–344. With discussion by Li, Tsybakov, van de Geer, Yu, Valdés, Rivero, Fan, van der Vaart, and a rejoinder by the author. 64

Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press Baltimore. 219, 263

Bickel, P. J. and Kleijn, B. J. K. (2012). The semiparametric Bernstein–von Mises theorem. *The Annals of Statistics*, 40(1):206–237. 46, 119, 242

Blair, R. J. R., Colledge, E., and Mitchell, D. G. V. (2001). Somatic markers and response reversal: Is there orbitofrontal cortex dysfunction in boys with psychopathic tendencies? *Journal of Abnormal Child Psychology*, 29:499–511. 194

Boekel, W., Wagenmakers, E.-J., Belay, L., Verhagen, A. J., Brown, S. D., and Forstmann, B. (2015). A purely confirmatory replication study of structural brain-behavior correlations. *Cortex*, 66:115–133. 34

Bolt, B. (1982). The constitution of the core: seismological evidence. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 306(1492):11–20. 11

Borgwardt, K. M. and Ghahramani, Z. (2009). Bayesian two-sample tests. *arXiv preprint arXiv:0906.4032*. 64, 71

Brown, C. E. (1998). Coefficient of variation. In *Applied multivariate statistics in geohydrology and related sciences*, pages 155–157. Springer. 193

Brown, L. D., Cai, T. T., and DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statistical Science*, pages 101–117. 221, 249

Burbea, J. (1984). Informative geometry of probability spaces. Technical report, DTIC Document. 253

Burbea, J. and Rao, C. R. (1982). Entropy differential metric, distance and divergence measures in probability spaces: A unified approach. *Journal of Multivariate Analysis*, 12(4):575–596. 253

Burbea, J. and Rao, C. R. (1984). Differential metrics in probability spaces. *Probability and mathematical statistics*, 3(2):241–258. 253

Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information–Theoretic Approach (2nd ed.)*. Springer Verlag, New York. 231

Busemeyer, J. R. and Stout, J. C. (2002). A contribution of cognitive decision models to clinical assessment: Decomposing performance on the Bechara gambling task. *Psychological Assessment*, 14:253–262. 174, 193, 195, 196, 197, 198, 199, 201, 203, 204, 205

Butler, R. W. and Wood, A. T. A. (2002). Laplace approximations for hypergeometric functions with matrix argument. *The Annals of Statistics*, 30(4):1155–1177. 110

Calin-Jageman, R. J. and Caldwell, T. L. (2014). Replication of the superstition and performance study by Damisch, Stoberock, and Mussweiler (2010). *Social Psychology*, 45:239–245. 137

Campbell, L. L. (1965). A coding theorem and Rényi's entropy. *Information and Control*, 8(4):423–429. 260

Carlin, B. P. and Chib, S. (1995). Bayesian model choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 473–484. 174, 206

Cavedini, P., Riboldi, G., D'Annucci, A., Belotti, P., Cisima, M., and Bellodi, L. (2002a). Decision-making heterogeneity in obsessive-compulsive disorder: Ventromedial prefrontal cortex function predicts different treatment outcomes. *Neuropsychologia*, 40:205–211. 194

Cavedini, P., Riboldi, G., Keller, R., D'Annucci, A., and Bellodi, L. (2002b). Frontal lobe dysfunction in pathological gambling patients. *Biological Psychiatry*, 51:334–341. 194

Chambers, C. D. (2013). Registered Reports: A new publishing initiative at Cortex. *Cortex*, 49:609–610. 99, 117

Chandramouli, S. and Shiffrin, R. M. (2016). Extending Bayesian induction. *Journal of Mathematical Psychology*, 72:38–42. 41, 51, 52, 54, 57

Chechile, R. A. (1973). *The Relative Storage and Retrieval Losses in Short–Term Memory as a Function of the Similarity and Amount of Information Processing in the Interpolated Task*. PhD thesis, University of Pittsburgh. 213

Chen, M.-H., Shao, Q.-M., and Ibrahim, J. G. (2012). *Monte Carlo methods in Bayesian computation*. Springer Science & Business Media. 173

Chen, Y. and Hanson, T. E. (2014). Bayesian nonparametric k-sample tests for censored and uncensored data. *Computational Statistics & Data Analysis*, 71:335–346. 71

Chernoff, H. and Savage, I. R. (1958). Asymptotic normality and efficiency of certain nonparametric test statistics. *The Annals of Mathematical Statistics*, pages 972–994. 72

Chib, S. and Jeliazkov, I. (2001). Marginal likelihood from the Metropolis–Hastings output. *Journal of the American Statistical Association*, 96(453):270–281. 207

Colonius, H. (2016). An invitation to coupling and copulas: With applications to multisensory modeling. *Journal of Mathematical Psychology*, 74:2–10. 74

Consonni, G. and Veronese, P. (2008). Compatibility of prior specifications across linear models. *Statistical Science*, 23(3):332–353. 45

Cook, A. (1990). Sir Harold Jeffreys. 2 April 1891-18 March 1989. *Biographical Memoirs of Fellows of the Royal Society*, 36:302–333. 11

Cover, T. M. and Thomas, J. A. (2006). *Elements of information theory*. John Wiley & Sons. 257

Cox, D. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–39. 256

Cramér, H. (1946). Methods of mathematical statistics. *Princeton University Press*, 23. 249

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, 25:7–29. 124

Cvitković, M., Smith, A.-S., and Pande, J. (2017). Asymptotic expansions of the hypergeometric function with two large parameters-application to the partition function of a lattice gas in a field of traps. *Journal of Physics A: Mathematical and Theoretical*, 50(26). 112

Dai, J., Kerestes, R., Upton, D. J., Busemeyer, J. R., and Stout, J. C. (2015). An improved cognitive model of the Iowa and Soochow gambling tasks with regard

to model fitting performance and tests of parameter consistency. *Frontiers in Psychology*, 6:229. 195

Dai, X., Wertenbroch, K., and Brendl, C. M. (2008). The value heuristic in judgments of relative frequency. *Psychological Science*, 19:18–19. 139, 140

Damisch, L., Stoberock, B., and Mussweiler, T. (2010). Keep your fingers crossed! How superstition improves performance. *Psychological Science*, 21:1014–1020. 137, 138

Dass, S. C. and Lee, J. (2004). A note on the consistency of Bayes factors for testing point null versus non-parametric alternatives. *Journal of statistical planning and inference*, 119(1):143–152. 65

Dawid, A. P. (1977). Further comments on some comments on a paper by Bradley Efron. *The Annals of Statistics*, 5(6):1249. 253

Dawid, A. P. (2011). Posterior model probabilities. In Gabbay, D. M., Bandyopadhyay, P. S., Forster, M. R., Thagard, P., and Woods, J., editors, *Handbook of the Philosophy of Science*, volume 7, pages 607–630. Elsevier, North-Holland. 222, 242

Dawid, A. P. and Lauritzen, S. L. (2001). Compatible prior distributions. *Bayesian Methods with Applications to Science, Policy and Official Statistics*, pages 109–118. 45

Dawid, A. P., Stone, M., and Zidek, J. (1973). Marginalization paradoxes in Bayesian and structural inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 189–233. 151

de Rooij, S. and Grünwald, P. D. (2011). Luckiness and regret in Minimum Description Length inference. In Gabbay, D. M., Bandyopadhyay, P. S., Forster, M. R., Thagard, P., and Woods, J., editors, *Handbook of the Philosophy of Science*, volume 7, pages 865–900. Elsevier, North-Holland. 231, 232

DiCiccio, T. J., Kass, R. E., Raftery, A., and Wasserman, L. (1997). Computing Bayes factors by combining simulation and asymptotic approximations. *Journal of the American Statistical Association*, 92(439):903–915. 174, 183, 184, 205

Dickey, J. M. (1971). The weighted likelihood ratio, linear hypotheses on normal location parameters. *The Annals of Mathematical Statistics*, pages 204 – 223. 193, 203, 206

Dickey, J. M. and Lientz, B. P. (1970). The weighted likelihood ratio, sharp hypotheses about chances, the order of a Markov chain. *The Annals of Mathematical Statistics*, 41:214–226. 22, 70, 98, 133, 193, 203, 206

Didelot, X., Everitt, R. G., Johansen, A. M., and Lawson, D. J. (2011). Likelihood-free estimation of model evidence. *Bayesian Analysis*, 6(1):49–76. 172

Diebolt, J. and Robert, C. P. (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 363–375. 48

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psycholology*, 5:781. 132

Dvoretzky, A., Kiefer, J., and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *The Annals of Mathematical Statistics*, pages 642–669. 64

Ebersole, C., Atherton, O., Belanger, A., Skulborstad, H., Allen, J., Banks, J., Baranski, E., Bernstein, M., Bonfiglio, D., Boucher, L., Brown, E., Budiman, N.,

Cairo, A., Capaldi, C., Chartier, C., Chung, J., Cicero, D., Coleman, J., Conway, J., Davis, W., Devos, T., Fletcher, M., German, K., Grahe, J., Hermann, A., Hicks, J., Honeycutt, N., Humphrey, B., Janus, M., Johnson, D., Joy-Gaba, J., Juzeler, H., Keres, A., Kinney, D., Kirshenbaum, J., Klein, R., Lucas, R., Lustgraaf, C., Martin, D., Menon, M., Metzger, M., Moloney, J., Morse, P., Prislin, R., Razza, T., Re, D., Rule, N., Sacco, D., Sauerberger, K., Shrider, E., Shultz, M., Siemsen, C., Sobocko, K., Sternglanz, R., Summerville, A., Tskhay, K., van Allen, Z., Vaughn, L., Walker, R., Weinberg, A., Wilson, J., Wirth, J., Wortman, J., and Nosek, B. (2016). Many Labs 3: Evaluating participant pool quality across the academic semester via replication. *Journal of Experimental Social Psychology*, 67:68–82. 4, 129, 301

Edwards, W. (1965). Tactical note on the relation between scientific and statistical hypotheses. *Psychological Bulletin*, 63:400–402. 120

Edwards, W., Lindman, H., and Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review*, 70:193–242. 15, 46, 120

Efron, B. (1975). Defining the curvature of a statistical problem (with applications to second order efficiency). *The Annals of Statistics*, 3(6):1189–1242. With a discussion by C. R. Rao, Don A. Pierce, D. R. Cox, D. V. Lindley, Lucien LeCam, J. K. Ghosh, J. Pfanzagl, Niels Keiding, A. Philip Dawid, Jim Reeds and with a reply by the author. 253

Etz, A. and Vandekerckhove, J. (2017). Introduction to Bayesian inference for psychology. *Psychonomic Bulletin & Review*. Manuscript accepted for publication. 132, 133

Etz, A. and Wagenmakers, E.-J. (2017). J. B. S. Haldane's contribution to the Bayes factor hypothesis test. *Statistical Science*, 32(2):313–329. 43, 47, 83, 102, 124, 131, 173, 225

Ferguson, T. S., Genest, C., and Hallin, M. (2000). Kendall's tau for serial dependence. *Canadian Journal of Statistics*, 28:587–604. 74, 75

Festinger, L. and Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *The Journal of Abnormal and Social Psychology*, 58(2):203. 123, 125, 126

Feynman, R. (1998). *The Meaning of it All: Thoughts of a Citizen–Scientist*. Perseus Books, Reading, MA. 117

Fisher, R. A. (1912). On an absolute criterion for fitting frequency curves. *Messenger of Mathematics*, 41:155–160. 218

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, 10(4):507–521. 149

Fisher, R. A. (1920). A mathematical examination of the methods of determining the accuracy of an observation by the mean error, and by the mean square error. *Monthly Notices of the Royal Astronomical Society*, 80:758–770. 149, 216

Fisher, R. A. (1921). On the "probable error" of a coefficient of correlation deduced from a small sample. *Metron*, 1:3–32. 149

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 222:309–368. 149, 218, 247, 248

Fisher, R. A. (1925). Theory of statistical estimation. *Mathematical Proceedings of the Cambridge Philosophical Society*, 22(5):700–725. 218

Forstmann, B. U., Dutilh, G., Brown, S., Neumann, J., Von Cramon, D. Y., Ridderinkhof, K. R., and Wagenmakers, E.-J. (2008). Striatum and pre-SMA facilitate decision-making under time pressure. *Proceedings of the National Academy of Sciences*, 105(45):17538–17542. 269

Fraser, D. A. (1961). On fiducial inference. *The Annals of Mathematical Statistics*, pages 661–676. 151

Fréchet, M. (1943). Sur l'extension de certaines evaluations statistiques au cas de petits echantillons. *Revue de l'Institut International de Statistique*, pages 182–205. 249

Frieden, B. R. (2004). *Science from Fisher information: A unification*. Cambridge University Press. 242

Friston, K. J. and Penny, W. (2011). *Post hoc* Bayesian model selection. *NeuroImage*, 56:2089–2099. 49

Frühwirth-Schnatter, S. (2004). Estimating marginal likelihoods for mixture and Markov switching models using bridge sampling techniques. *The Econometrics Journal*, 7:143–167. 174, 182, 186, 187, 192, 193, 205

Galton, F. (1907). Vox populi. *Nature*, 75:450–451. 87

Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman & Hall/CRC, Boca Raton, FL. 37, 172, 174, 178

Geisser, S. (1980). The contributions of Sir Harold Jeffreys to Bayesian inference. In Zellner, A. and Kadane, Joseph, B., editors, *Bayesian Analysis in Econometrics and Statistics: Essays in Honor of Harold Jeffreys*, pages 13–20. Amsterdam: North-Holland. 11, 37

Gelfand, A. E. and Dey, D. K. (1994). Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 501–514. 182, 183

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian data analysis*. Chapman & Hall/CRC, London, 3rd edition. 73

Gelman, A. and Meng, X.-L. (1998). Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical science*, pages 163–185. 172, 206

Genest, C. and Favre, A.-C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, 12:347–368. 74

Ghosal, S., Ghosh, J. K., and Ramamoorthi, R. (1997). Non-informative priors via sieves and packing numbers. In *Advances in statistical decision theory and applications*, pages 119–132. Springer. 64, 243, 259

Ghosal, S., Ghosh, J. K., and Van Der Vaart, A. W. (2000). Convergence rates of posterior distributions. *Annals of Statistics*, 28(2):500–531. 64

Ghosal, S., Lember, J., and Van Der Vaart, A. (2008). Nonparametric Bayesian model selection and averaging. *Electronic Journal of Statistics*, 2:63–89. 64

Ghosh, J. K. (1985). Efficiency of estimates–part I. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 310–325. 248

Ghosh, M. (2011). Objective priors: An introduction for frequentists. *Statistical Science*, pages 187–202. 103

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., editors (1996). *Markov chain Monte Carlo in Practice*. Chapman & Hall/CRC, Boca Raton (FL). 37

Gino, F. and Wiltermuth, S. S. (2014). Evil genius? How dishonesty can lead to greater creativity. *Psychological Science*, 4:973–981. 21, 22, 23, 24

Girolami, M. and Calderhead, B. (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(2):123–214. 243

Goldacre, B. (2009). *Bad Science*. Fourth Estate, London. 117

Golden, R. M. (1995). Making correct statistical inferences using the wrong probability model. *Journal of Mathematical Psychology*, 39:3–20. 242

Golden, R. M. (2000). Statistical tests for comparing possibly misspecified and nonnested models. *Journal of Mathematical Psychology*, 44(1):153–170. 242

Good, I. J. (1980). The contributions of Jeffreys to Bayesian statistics. In Zellner, A., editor, *Bayesian Analysis in Econometrics and Statistics: Essays in Honor of Harold Jeffreys*, pages 21–34. North-Holland Publishing Company, Amsterdam, The Netherlands. 11

Gradshteyn, I. S. and Ryzhik, I. M. (2007). *Table of Integrals, Series, and Products*. Academic Press, 7 edition. 32, 38, 109, 153, 155, 156, 162, 166, 167

Grazian, C. and Liseo, B. (2014). Approximate integrated likelihood via ABC methods. *arXiv preprint arXiv:1403.0387*. 243

Grazian, C. and Robert, C. P. (2015). Jeffreys' priors for mixture estimation. In *Bayesian Statistics from Methods to Models and Applications*, pages 37–48. Springer. 51, 243

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4):711–732. 174, 206

Greiner, R. (1909). Über das Fehlersystem der Kollektivmasslehre. *Zeitschift für Mathematik und Physik*, pages 121–158. 73

Griffin, H. (1958). Graphic computation of tau as a coefficient of disarray. *Journal of the American Statistical Association*, 53:441–447. 70

Gronau, Q. F., Ly, A., and Wagenmakers, E.-J. (2017a). Informed Bayesian t-tests. *arXiv preprint arXiv:1704.02479*. 126, 128, 137, 222

Gronau, Q. F., Sarafoglou, A., Matzke, D., Ly, A., Boehm, U., Marsman, M., Leslie, D. S., Forster, J. J., Wagenmakers, E.-J., and Steingroever, H. (2017b). A tutorial on bridge sampling. *Journal of Mathematical Psychology*, 81:80–97. 243

Gronau, Q. F., van Erp, S., Heck, D. W., Cesario, J., Jonas, K. J., and Wagenmakers, E.-J. (2017c). 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*, 2:123–138. 140

Gronau, Q. F. and Wagenmakers, E.-J. (2017). Bayesian evidence accumulation in experimental mathematics: A case study of four irrational numbers. *Experimental Mathematics*. Manuscript accepted for publication. 98, 132

Grünwald, P. (2016). Safe probability. *arXiv preprint arXiv:1604.01785*. 249

Grünwald, P. and van Ommen, T. (2014). Inconsistency of Bayesian inference for misspecified linear models, and a proposal for repairing it. *arXiv preprint arXiv:1412.3730*. 249, 268

Grünwald, P. D. (1998). *The Minimum Description Length Principle and Reasoning under Uncertainty*. PhD thesis, ILLC and University of Amsterdam. 241

Grünwald, P. D. (2007). *The Minimum Description Length Principle*. MIT Press, Cambridge, MA. 231, 232

Grünwald, P. D. (2017). Safe probability. *Journal of Statistical Planning and Inference*. Manuscript accepted for publication. 268

Grünwald, P. D. and Mehta, N. A. (2016). Fast rates with unbounded losses. *arXiv preprint arXiv:1605.00252*. 242

Grünwald, P. D., Myung, I. J., and Pitt, M. A., editors (2005). *Advances in Minimum Description Length: Theory and Applications*. MIT Press, Cambridge, MA. 231

Gunel, E. and Dickey, J. (1974). Bayes factors for independence in contingency tables. *Biometrika*, 61:545–557. 138, 139

Haight, F. A. (1967). *Handbook of the Poisson distribution*. Wiley. 102

Hájek, J. (1970). A characterization of limiting distributions of regular estimates. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 14(4):323–330. 219, 248

Hájek, J. (1972). Local asymptotic minimax and admissibility in estimation. In *Proceedings of the sixth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 175–194. 242

Hald, A. (2008). *A history of parametric statistical inference from Bernoulli to Fisher, 1713-1935*. Springer Science & Business Media. 225

Haldane, J. B. S. (1932). A note on inverse probability. *Mathematical Proceedings of the Cambridge Philosophical Society*, 28:55–61. 82

Hammersley, J. M. and Handscomb, D. C. (1964). *Monte Carlo methods*. London: Methuen. 176

Hannig, J., Iyer, H., and Patterson, P. (2006). Fiducial generalized confidence intervals. *Journal of the American Statistical Association*, 101(473):254–269. 151

Harms, C. (2016). A Bayes factor for replications of ANOVA results. *arXiv preprint*. 130

Heck, D. W., Moshagen, M., and Erdfelder, E. (2014). Model selection by minimum description length: Lower-bound sample sizes for the Fisher information approximation. *Journal of Mathematical Psychology*, 60:29–34. 234

Heck, D. W., Wagenmakers, E.-J., and Morey, R. D. (2015). Testing order constraints: Qualitative differences between Bayes factors and normalized maximum likelihood. *Statistics & Probability Letters*, 105:157–162. 50

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics*, 19:293–325. 74

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, pages 382–401. 82, 173

Holmes, C. C., Caron, F., Griffin, J. E., and Stephens, D. A. (2015). Two-sample Bayesian nonparametric hypothesis testing. *Bayesian Analysis*, 10(2):297–320. 64, 71

Hotelling, H. and Pabs, M. (1936). Rank correlation and tests of significance involving no assumption of normality. *Annals of Mathematical Statistics*, 7:29–43. 72

Huzurbazar, V. S. (1950). Probability distributions and orthogonal parameters. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 46, pages 281–284. Cambridge University Press. 256

Huzurbazar, V. S. (1956). Sufficient statistics and orthogonal parameters. *Sankhyā: The Indian Journal of Statistics (1933-1960)*, 17(3):217–220. 256

Huzurbazar, V. S. (1991). Sir Harold Jeffreys: Recollections of a student. *Chance*, 4(2):18–21. 11

Inagaki, N. (1970). On the limiting distribution of a sequence of estimators with uniformity property. *Annals of the Institute of Statistical Mathematics*, 22(1):1–13. 219, 248

Ionides, E. L. (2008). Truncated importance sampling. *Journal of Computational and Graphical Statistics*, 17:295–311. 179

Jamil, T., Ly, A., Morey, R. D., Love, J., Marsman, M., and Wagenmakers, E.-J. (2017). Default "Gunel and Dickey" Bayes factors for contingency tables. *Behavior Research Methods*, 49:638–652. 138

JASP Team (2017). JASP (Version 0.8.2)[Computer software]. 82, 118, 124, 300

Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press, Cambridge. 11

Jeffreys, H. (1924). *The Earth, Its Origin, History and Physical Constitution*. Cambridge University Press. 11

Jeffreys, H. (1931). *Scientific inference*. Cambridge University Press. 11

Jeffreys, H. (1935). Some tests of significance, treated by the theory of probability. *Proceedings of the Cambridge Philosophical Society*, 31(2):203–222. 16, 102, 150

Jeffreys, H. (1938). Significance tests when several degrees of freedom arise simultaneously. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 165:161–198. 135

Jeffreys, H. (1939). *Theory of Probability*. Oxford University Press, Oxford, UK, 1st edition. 102, 112

Jeffreys, H. (1942). On the significance tests for the introduction of new functions to represent measures. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 180(982):256–268. 83

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 186(1007):453–461. 18, 43, 59, 62, 103, 226

Jeffreys, H. (1948). *Theory of Probability*. Oxford University Press, Oxford, UK, 2nd edition. 24, 82, 87, 112, 126

Jeffreys, H. (1955). The present position in probability theory. *The British Journal for the Philosophy of Science*, 5:275–289. 12

Jeffreys, H. (1961). *Theory of Probability*. Oxford University Press, Oxford, UK, 3rd edition. 3, 12, 13, 14, 15, 18, 20, 24, 25, 27, 33, 34, 35, 36, 37, 39, 43, 45,

47, 49, 69, 70, 73, 76, 82, 83, 84, 99, 102, 103, 105, 124, 127, 131, 135, 150, 151, 173, 186, 222, 225, 256

Jeffreys, H. (1973). *Scientific Inference*. Cambridge University Press, Cambridge, UK, 3 edition. 12, 13, 14

Jeffreys, H. (1980). Some general points in probability theory. In Zellner, A. and Kadane, Joseph, B., editors, *Bayesian Analysis in Econometrics and Statistics: Essays in Honor of Harold Jeffreys*, pages 451–453. Amsterdam: North-Holland. 13, 14, 37, 49

Jeffreys, H. and Jeffreys, B. S. (1946). *Methods of Mathematical Physics*. Cambridge University Press, Cambridge, UK. 11

Johnson, V. E. (2005). Bayes factors based on test statistics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(5):689–701. 69, 71, 79

Johnson, V. E. (2013). Revised standards for statistical evidence. *Proceedings of the National Academy of Sciences of the United States of America*, 110:19313–19317. 15, 120

Johnson, V. E. and Rossell, D. (2010). On the use of non-local prior densities in Bayesian hypothesis tests. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(2):143–170. 51

Kamary, K., Eun, J., and Robert, Christian, P. (2016). Non-informative reparameterisations for location-scale mixtures. *arXiv preprint arXiv:1601.01178*. 51

Kamary, K., Mengersen, K., Robert, C. P., and Rousseau, J. (2014). Testing hypotheses via a mixture estimation model. *arXiv preprint arXiv:1412.2044*. 48, 49, 50, 51

Kary, A., Taylor, R., and Donkin, C. (2016). Using Bayes factors to test the predictions of models: A case study in visual working memory. *Journal of Mathematical Psychology*, 72:210–219. 49

Kass, R. E. (1989). The geometry of asymptotic inference. *Statistical Science*, 4(3):188–234. 228, 232

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90:773–795. 70, 83, 102, 124, 131, 172, 186

Kass, R. E. and Vaidyanathan, S. (1992). Approximate Bayes factors and orthogonal parameters, with application to testing equality of two binomial proportions. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 129–144. 256

Kass, R. E. and Vos, P. W. (2011). *Geometrical foundations of asymptotic inference*, volume 908. John Wiley & Sons. 253

Kass, R. E. and Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, 91:1343–1370. 103

Kelman, H. C. (1953). Attitude change as a function of response restriction. *Human Relations*, 6(3):185–214. 125

Kendall, M. (1938). A new measure of rank correlation. *Biometrika*, 30:81–93. 69, 72

Kendall, M. and Gibbons, J. D. (1990). *Rank Correlation Methods*. Oxford University Press, New York. 69, 73, 74, 75

Keuken, M. C., Ly, A., Boekel, W., Wagenmakers, E.-J., Belay, L., Verhagen, A., Brown, S. D., and Forstmann, B. U. (2017). Corrigendum to "a purely confirmatory replication study of structural brain-behavior correlations"[cortex 66 (2015) 115–133]. *Cortex*, 93:229–233. 34

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., and Nosek, B. A. (2016). Badges to acknowledge open practices: A simple, low cost, effective method for increasing transparency. *PLoS Biology*, 14:e1002456. 118

Klaassen, C. A. J. and Lenstra, A. J. (2003). Vanishing Fisher information. *Acta Applicandae Mathematicae*, 78(1):193–200. 232

Klauer, K. C. and Kellen, D. (2011). The flexibility of models of recognition memory: An analysis by the minimum-description length principle. *Journal of Mathematical Psychology*, 55(6):430–450. 232

Kleijn, B. J. K. and van der Vaart, A. W. (2012). The Bernstein-von-Mises theorem under misspecification. *Electronic Journal of Statistics*, 6:354–381. 119

Kleijn, B. J. K. and Zhao, Y. Y. (2017). Criteria for posterior consistency. *arXiv preprint arXiv:1308.1263*. 64, 243

Klein, R., Ratliff, K., Vianello, M., Adams, Jr., R. B., Bahník, v., 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., Hasselman, F., Hicks, J., Hovermale, J., Hunt, S., Huntsinger, J., IJzerman, H., John, M.-S., Joy-Gaba, J., Barry Kappes, H., Krueger, L., Kurtz, J., Levitan, C., Mallett, R., Morris, W., Nelson, A., Nier, J., Packard, G., Pilati, R., Rutchick, A., Schmidt, K., Skorinko, J., Smith, R., Steiner, T., Storbeck, J., Van Swol, L., Thompson, D., van 't Veer, A., Vaughn, L., Vranka, M., Wichman, A., Woodzicka, J., and Nosek, B. (2014). Investigating variation in replicability: A "many labs" replication project. *Social Psychology*, 45:142–152. 4, 129, 301

Klugkist, I., Laudy, O., and Hoijtink, H. (2005). Inequality constrained analysis of variance: a Bayesian approach. *Psychological Methods*, 10(4):477. 23

Korostelev, A. P. and Korosteleva, O. (2011). *Mathematical statistics: Asymptotic minimax theory*, volume 119. American Mathematical Society. 242

Kotz, S., Kozubowski, T. J., and Podgorski, K. (2001). *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance*. Springer, New York. 221

Kraft, L. G. (1949). A device for quantizing, grouping, and coding amplitude-modulated pulses. Master's thesis, Massachusetts Institute of Technology. 257

Krishnamoorthy, K. and Thomson, J. (2004). A more powerful test for comparing two Poisson means. *Journal of Statistical Planning and Inference*, 119(1):23–35. 102

Krupenye, C., Kano, F., Hirata, S., Call, J., and Tomasello, M. (2016). Great apes anticipate that other individuals will act according to false beliefs. *Science*, 354:110–114. 132, 133, 134, 135, 136, 143, 144, 145

Kruskal, W. (1958). Ordinal measures of association. *Journal of the American Statistical Association*, 53:814–861. 69, 70, 73

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86. 259

Labadi, L. A., Masuadi, E., and Zarepour, M. (2014). Two-sample Bayesian nonparametric goodness-of-fit test. *arXiv preprint arXiv:1411.3427*. 64, 71

LeCam, L. (1970). On the assumptions used to prove asymptotic normality of maximum likelihood estimates. *The Annals of Mathematical Statistics*, 41(3):802–828. 219, 248

LeCam, L. (1990). Maximum likelihood: An introduction. *International Statistical Review/Revue Internationale de Statistique*, 58(2):153–171. 218

Lee, M. D. (2008). Three case studies in the Bayesian analysis of cognitive models. *Psychonomic Bulletin & Review*, 15:1–15. 172

Lee, M. D., Lodewyckx, T., and Wagenmakers, E.-J. (2015). Three Bayesian analyses of memory deficits in patients with dissociative identity disorder. In Raaijmakers, J. R., Criss, A., Goldstone, R., Nosofsky, R., and Steyvers, M., editors, *Cognitive Modeling in Perception and Memory: A Festschrift for Richard M. Shiffrin*, pages 189–200. Psychology Press. 49

Lee, M. D. and Wagenmakers, E.-J. (2013). *Bayesian Cognitive Modeling: A Practical Course*. Cambridge University Press, Cambridge. 10, 222

Lehmann, E. L. (2011). *Fisher, Neyman, and the creation of classical statistics*. Springer Science & Business Media. 225

Levelt, W. J., Drenth, P., and Noort, E. (2012). Flawed science: The fraudulent research practices of social psychologist Diederik Stapel. 4, 301

Lewis, S. M. and Raftery, A. E. (1997). Estimating Bayes factors via posterior simulation with the Laplace—Metropolis estimator. *Journal of the American Statistical Association*, 92(438):648–655. 10, 131, 173

Li, Y. and Clyde, M. A. (2015). Mixtures of g-priors in generalized linear models. *arXiv preprint arXiv:1503.06913*. 222

Liang, F., Paulo, R., Molina, G., Clyde, M. A., and Berger, J. O. (2008). Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association*, 103(481). 24, 25, 85, 222

Lindley, D. V. (1957). A statistical paradox. *Biometrika*, 44:187–192. 15, 18, 103

Lindley, D. V. (1965). *Introduction to Probability & Statistics from a Bayesian Viewpoint. Part 2. Inference*. Cambridge University Press, Cambridge. 150, 151

Lindley, D. V. (1972). *Bayesian Statistics, a Review*. SIAM, Philadelphia (PA). 87, 133

Lindley, D. V. (1977). The distinction between inference and decision. *Synthese*, 36:51–58. 43

Lindley, D. V. (1980). Jeffreys's contribution to modern statistical thought. In Zellner, A., editor, *Bayesian Analysis in Econometrics and Statistics: Essays in Honor of Harold Jeffreys*, pages 35–39. North-Holland Publishing Company, Amsterdam, The Netherlands. 11

Lindley, D. V. (1985). *Making Decisions*. Wiley, London, 2 edition. 118, 119

Lindley, D. V. (1991). Sir Harold Jeffreys. *Chance*, 4(2):10–14, 21. 11

Lindley, D. V. (1997). Some comments on Bayes factors. *Journal of Statistical Planning and Inference*, 61(1):181–189. 63

Liu, C. C. and Aitkin, M. (2008). Bayes factors: Prior sensitivity and model generalizability. *Journal of Mathematical Psychology*, 52:362–375. 36, 120

López, J. L. and Pagola, P. J. (2011). A systematic "saddle point near a pole" asymptotic method with application to the Gauss hypergeometric function. *Studies in Applied Mathematics*, 127(1):24–37. 112

Luce, R. (1959). *Individual choice behavior*. Wiley, New York. 196

Ly, A., Boehm, U., Heathcote, A., Turner, B. M., Forstmann, B., Marsman, M., and Matzke, D. (2017a). A flexible and efficient hierarchical Bayesian approach to the exploration of individual differences in cognitive-model-based neuroscience. In Moustafa, A. A., editor, *Computational Models of Brain and Behavior*, pages 467–480. John Wiley & Sons. 269

Ly, A., Etz, A., Marsman, M., and Wagenmakers, E.-J. (2017b). Replication Bayes factors from evidence updating. *Manuscript in preparation*. 47, 119, 127, 242

Ly, A., Marsman, M., Verhagen, A. J., Grasman, R. P. P. P., and Wagenmakers, E.-J. (2017c). A tutorial on Fisher information. *Journal of Mathematical Psychology*, 80:40–55. 16, 18, 44, 52, 62, 63, 103, 138, 151, 175

Ly, A., Marsman, M., and Wagenmakers, E.-J. (2017d). Analytic posteriors for Pearson's correlation coefficient. *Statistica Neerlandica*. Manuscript accepted for publication. 86, 222

Ly, A., Raj, A., Etz, A., Marsman, M., Gronau, Q. F., and Wagenmakers, E.-J. (2017e). Bayesian reanalyses from summary statistics and the strength of statistical evidence. *Manuscript submitted for publication*. 133, 222

Ly, A., Verhagen, A. J., and Wagenmakers, E.-J. (2016a). Harold Jeffreys's default Bayes factor hypothesis tests: Explanation, extension, and application in psychology. *Journal of Mathematical Psychology*, 72:19–32. 42, 43, 45, 73, 76, 83, 84, 90, 99, 102, 112, 117, 126, 131, 173, 222, 225, 256

Ly, A., Verhagen, A. J., and Wagenmakers, E.-J. (2016b). An evaluation of alternative methods for testing hypotheses, from the perspective of Harold Jeffreys. *Journal of Mathematical Psychology*, 72:43–55. 112, 117, 126, 131, 173, 222, 225, 256

Marin, J.-M. and Robert, C. P. (2010). On resolving the Savage–Dickey paradox. *Electronic Journal of Statistics*, 4:643–654. 18, 22, 203

Marin, J.-M. and Robert, C. P. (2014). *Bayesian essentials with R*. Springer. 48

Marsman, M. (2014). *Plausible values in statistical inference*. PhD thesis, Universiteit Twente. 269

Marsman, M., Ly, A., and Wagenmakers, E.-J. (2016a). Four requirements for an acceptable research program. *Basic and Applied Social Psychology*, 38(6):308–312. 140, 242

Marsman, M., Maris, G., Bechger, T., and Glas, C. (2016b). What can we learn from plausible values? *Psychometrika*, 81(2):274–289. 269

Marsman, M., Schönbrodt, F. D., Morey, R. D., Yao, Y., Gelman, A., and Wagenmakers, E.-J. (2017). A Bayesian bird's eye view of 'Replications of important results in social psychology'. *Royal Society Open Science*, 4:160426. 82, 130

Marsman, M. and Wagenmakers, E.-J. (2017). Three insights from a Bayesian interpretation of the one–sided p value. *Educational and Psychological Measurement*, 77(3):529–539. 120

Martino, D. J., Bucay, D., Butman, J. T., and Allegri, R. F. (2007). Neuropsychological frontal impairments and negative symptoms in schizophrenia. *Psychiatry Research*, 152:121–128. 194

Matejka, J. and Fitzmaurice, G. (2017). Same stats, different graphs: Generating datasets with varied appearance and identical statistics through simulated annealing. *CHI 2017 Conference Proceedings: ACM SIGCHI Conference on Human Factors in Computing Systems*. 128

Matzke, D., Dolan, C. V., Batchelder, W. H., and Wagenmakers, E.-J. (2015a). Bayesian estimation of multinomial processing tree models with heterogeneity in participants and items. *Psychometrika*, 80:205–235. 172

Matzke, D., Nieuwenhuis, S., van Rijn, H., Slagter, H. A., van der Molen, M. W., and Wagenmakers, E.-J. (2015b). The effect of horizontal eye movements on free recall: A preregistered adversarial collaboration. *Journal of Experimental Psychology: General*, 144:e1–e15. 144

Matzke, D. and Wagenmakers, E.-J. (2009). Psychological interpretation of the ex-Gaussian and shifted Wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review*, 16(5):798–817. 172

McMillan, B. (1956). Two inequalities implied by unique decipherability. *IRE Transactions on Information Theory*, 2(4):115–116. 257

Mednick, S. (1962). The associative basis of the creative process. *Psychological Review*, 69(3):220–232. 22

Meng, X.-L. and Schilling, S. (2002). Warp bridge sampling. *Journal of Computational and Graphical Statistics*, 11:552–586. 187, 207

Meng, X.-L. and Wong, W. H. (1996). Simulating ratios of normalizing constants via a simple identity: A theoretical exploration. *Statistica Sinica*, 6(4):831–860. 5, 171, 173, 174, 187, 188, 205, 302

Mira, A. and Nicholls, G. (2004). Bridge estimation of the probability density at a point. *Statistica Sinica*, 14:603–612. 207

Mitchell, A. F. (1962). Sufficient statistics and orthogonal parameters. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 58, pages 326–337. Cambridge University Press. 256

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., and Zwaan, R. A. (2016). The peer reviewers' openness initiative: Incentivizing open research practices through peer review. *Royal Society Open Science*, 3:150547. 118

Morey, R. D. and Rouder, J. N. (2015). BayesFactor 0.9.11-1. Comprehensive R Archive Network. 82, 87, 126

Morey, R. D. and Wagenmakers, E.-J. (2014). Simple relation between Bayesian order-restricted and point-null hypothesis tests. *Statistics and Probability Letters*, 92:121–124. 22, 91

Morris, D. E., Oakley, J. E., and Crowe, J. A. (2014). A web-based tool for eliciting probability distributions from experts. *Environmental Modelling & Software*, 52:1–4. 91

Mulder, J. (2014). Bayes factors for testing inequality constrained hypotheses: Issues with prior specification. *British Journal of Mathematical and Statistical Psychology*, 67(1):153–171. 33

Mulder, J. (2016). Bayes factors for testing order-constrained hypotheses on correlations. *Journal of Mathematical Psychology*, 72:104–115. 33

Mulder, J., Hoijtink, H., and Klugkist, I. (2010). Equality and inequality constrained multivariate linear models: objective model selection using constrained posterior priors. *Journal of Statistical Planning and Inference*, 140(4):887–906. 22

Mulder, J. and Wagenmakers, E.-J. (2016). Editors' introduction to the special issue "Bayes factors for testing hypotheses in psychological research: Practical relevance and new developments". *Journal of Mathematical Psychology*, 72:1–5. 172

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, 47:90–100. 16, 214, 218

Myung, I. J., Balasubramanian, V., and Pitt, M. (2000a). Counting probability distributions: Differential geometry and model selection. *Proceedings of the National Academy of Sciences*, 97(21):11170–11175. 232

Myung, I. J., Forster, M. R., and Browne, M. W. (2000b). Guest editors' introduction: Special issue on model selection. *Journal of Mathematical Psychology*, 44:1–2. 172

Myung, I. J., Forster, M. R., and Browne, M. W. (2000c). Model selection [Special issue]. *Journal of Mathematical Psychology*, 44(1–2). 37, 231, 232

Myung, I. J. and Navarro, D. J. (2005). Information matrix. *Encyclopedia of Statistics in Behavioral Science*. 214

Myung, I. J., Navarro, D. J., and Pitt, M. A. (2006). Model selection by normalized maximum likelihood. *Journal of Mathematical Psychology*, 50:167–179. 231, 232

Myung, I. J. and Pitt, M. A. (1997). Applying Occam's razor in modeling cognition: A Bayesian approach. *Psychonomic Bulletin & Review*, 4(1):79–95. 10, 37, 124

Myung, I. J. and Pitt, M. A. (2016). Model comparison in psychology. In Wixted, J. and Wagenmakers, E.-J., editors, *The Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience (Fourth Edition)*, volume 5: Methodology. John Wiley & Sons, New York, NY. Manuscript accepted for publication. 232

Navarro, D. J., Griffiths, T. L., Steyvers, M., and Lee, M. D. (2006). Modeling individual differences using Dirichlet processes. *Journal of Mathematical Psychology*, 50:101–122. 173

Neal, R. M. (2001). Annealed importance sampling. *Statistics and Computing*, 11:125–139. 179

Nelsen, R. (2006). *An Introduction to Copulas*. Springer-Verlag New York, second edition. 74

Newton, M. A. and Raftery, A. E. (1994). Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 3–48. 178, 182, 183

Ng, H., Gu, K., and Tang, M. (2007). A comparative study of tests for the difference of two Poisson means. *Computational Statistics & Data Analysis*, 51(6):3085–3099. 102

Nickerson, R. S. (2000). Null hypothesis statistical testing: A review of an old and continuing controversy. *Psychological Methods*, 5:241–301. 124

Noether, G. E. (1955). On a theorem of Pitman. *Annals of Mathematical Statistics*, 26:64–68. 72

Nosek, B. A., 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., Ishiyama, J., Karlan, D., Kraut, A., Lupia, A., Mabry, P., Madon, T. A., Malhotra, N., Mayo-Wilson, E., McNutt, M., Miguel, E., Levy Paluck, E., Simonsohn, U., Soderberg, C., Spellman, B. A., Turitto, J., VandenBos, G., Vazire, S., Wagenmakers, E.-J., Wilson, R., and Yarkoni, T. (2015). Promoting an open research culture. *Science*, 348:1422–1425. 118

Nosek, B. A. and Lakens, D. (2014). Registered reports: A method to increase the credibility of published results. *Social Psychology*, 45:137–141. 4, 117, 129, 301

Ntzoufras, I. (2009). Bayesian model and variable evaluation. In *Bayesian modeling using WinBUGS*, pages 389–433. John Wiley & Sons. 172, 174, 178, 180

Nuijten, M. B., Hartgerink, C. H., Assen, M. A., Epskamp, S., and Wicherts, J. M. (2016). The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior research methods*, 48(4):1205–1226. 81

Oberhettinger, F. (1972). Hypergeometric functions. In Abramowitz, M. and Stegun, I. A., editors, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, pages 555–566. New York: Dover Publications. 38, 160

O'Hagan, A. and Forster, J. (2004). *Kendall's Advanced Theory of Statistics Vol. 2B: Bayesian Inference (2nd ed.)*. Arnold, London. 10, 131

Olver, F. W. J., Lozier, D. W., Boisvert, R. F., and Clark, C. W., editors (2010). *NIST handbook of mathematical functions*. U.S. Department of Commerce, National Institute of Standards and Technology, Washington, DC; Cambridge University Press, Cambridge. 161, 163

Open Science Collaboration (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science*, 7(6):657–660. 117

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251):aac4716. 4, 129, 139, 301

Overstall, A. M. and Forster, J. J. (2010). Default Bayesian model determination methods for generalised linear mixed models. *Computational Statistics & Data Analysis*, 54:3269–3288. 184, 186, 187, 189, 209

Owen, A. and Zhou, Y. (2000). Safe and effective importance sampling. *Journal of the American Statistical Association*, 95:135–143. 179

Pajor, A. (2016). Estimating the Marginal Likelihood Using the Arithmetic Mean Identity. *Bayesian Analysis*, pages 1–27. 178

Pashler, H. and Wagenmakers, E.-J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, 7:528–530. 4, 117, 129, 301

Peirce, C. S. (1878). Deduction, induction, and hypothesis. *Popular Science Monthly*, 13:470–482. 117

Peirce, C. S. (1883). A theory of probable inference. In Peirce, C. S., editor, *Studies in Logic*, pages 126–181. Little & Brown, Boston. 117

Pericchi, L. R., Liu, G., and Torres, D. (2008). Objective Bayes factors for informative hypotheses: "Completing" the informative hypothesis and "splitting" the

Bayes factor. In Hoijtink, H., Klugkist, I., and Boelen, P. A., editors, *Bayesian Evaluation of Informative Hypotheses*, pages 131–154. Springer Verlag, New York. 23

Pitt, M., Myung, I. J., and Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review*, 109(3):472–491. 172, 231, 241

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing*, volume 124, pages 1–8. Vienna. 201

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, 6:7–11. 193

Poirier, D. J. (2006). The growth of Bayesian methods in statistics and economics since 1970. *Bayesian Analysis*, 1:969–979. 171

Przyborowski, J. and Wilenski, H. (1940). Homogeneity of results in testing samples from Poisson series with an application to testing clover seed for dodder. *Biometrika*, 31(3-4):313–323. 102, 104

Ractliffe, J. F. (1964). The significance of the difference between two Poisson variables: An experimental investigation. *Applied Statistics*, pages 84–86. 102

Raftery, A. E. (1995). Bayesian model selection in social research. In Marsden, P. V., editor, *Sociological Methodology*, pages 111–196. Blackwells, Cambridge. 231

Raftery, A. E. and Banfield, J. D. (1991). Stopping the Gibbs Sampler, the Use of Morphology, and Other Issues in Spatial Statistics (Bayesian image restoration, with two applications in spatial statistics)–(Discussion). *Annals of the Institute of Statistical Mathematics*, 43:32–43. 176

Ramsey, F. P. (1926). Truth and probability. In Braithwaite, R. B., editor, *The Foundations of Mathematics and Other Logical Essays*, pages 156–198. Kegan Paul, London. 12

Randles, R. H. and Wolfe, D. A. (1979). *Introduction to the Theory of Nonparametric Statistics*. Springer Texts in Statistics. Wiley, New York. 71

Rao, C. R. (1945). Information and accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, 37(3):81–91. 249, 253

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85:59–108. 213

Ray, K. and Schmidt-Hieber, J. (2016). Minimax theory for a class of nonlinear statistical inverse problems. *Inverse Problems*, 32(6):065003. 243

Rényi, A. (1961). On measures of entropy and information. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 547–561. 260

Rescorla, R. A. and Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In Black, A. H. and Prokasy, W. F., editors, *Classical conditioning II: Current research and theory*, pages 64–99. New York. 195

Rissanen, J. (1996). Fisher information and stochastic complexity. *IEEE Transactions on Information Theory*, 42:40–47. 232, 233, 240

Rivoirard, V. and Rousseau, J. (2012). Bernstein–von Mises theorem for linear functionals of the density. *The Annals of Statistics*, 40(3):1489–1523. 242

Robert, C. P. (1993). A note on Jeffreys-Lindley paradox. *Statistica Sinica*, 3(2):601–608. 43

Robert, C. P. (2007). *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Science & Business Media. 43

Robert, C. P. (2014). On the Jeffreys-Lindley paradox. *Philosophy of Science*, 81(2):216–232. 43

Robert, C. P. (2015). The Metropolis–Hastings algorithm. *arXiv preprint arXiv:1504.01896*. 50, 180

Robert, C. P. (2016). The expected demise of the Bayes factor. *Journal of Mathematical Psychology*, 72:33–37. 41, 42, 48, 50, 51, 173, 222

Robert, C. P., Chopin, N., and Rousseau, J. (2009). Harold Jeffreys's Theory of Probability revisited. *Statistical Science*, pages 141–172. 11, 29, 34, 37, 42, 83, 150, 225

Roberts, S. and Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing in psychology. *Psychological Review*, 107:358–367. 232

Rouder, J. N. (2014). Optional stopping: No problem for Bayesians. *Psychonomic Bulletin & Review*, 21:301–308. 82, 132

Rouder, J. N. and Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review*, 12:573–604. 172, 173

Rouder, J. N., Lu, J., Morey, R. D., Sun, D., and Speckman, P. L. (2008). A hierarchical process-dissociation model. *Journal of Experimental Psychology: General*, 137:370–389. 173

Rouder, J. N., Lu, J., Speckman, P., Sun, D., and Jiang, Y. (2005). A hierarchical model for estimating response time distributions. *Psychonomic Bulletin & Review*, 12:195–223. 172, 173

Rouder, J. N., Lu, J., Sun, D., Speckman, P., Morey, R., and Naveh-Benjamin, M. (2007). Signal detection models with random participant and item effects. *Psychometrika*, 72:621–642. 172

Rouder, J. N., Morey, R. D., Verhagen, A. J., Province, J. M., and Wagenmakers, E.-J. (2016a). Is there a free lunch in inference? *Topics in Cognitive Science*, 8:520–547. 99, 117, 124

Rouder, J. N., Morey, R. D., and Wagenmakers, E.-J. (2016b). The interplay between subjectivity, statistical practice, and psychological science. *Collabra*, 2:1–12. 99

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., and Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16(2):225–237. 10, 21, 24, 82, 84, 90, 99, 118

Rozeboom, W. W. (1960). The fallacy of the null–hypothesis significance test. *Psychological Bulletin*, 57:416–428. 42, 119

Rudin, W. (1991). *Functional analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill, Inc., New York, second edition. 228

Rutherford, E., Geiger, H., and Bateman, H. (1910). LXXVI. the probability variations in the distribution of α particles. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 20(118):698–707. 102

Salomond, J.-B. (2013). Bayesian testing for embedded hypotheses with application to shape constrains. *arXiv preprint arXiv:1303.6466*. 64

Salomond, J.-B. (2014). Adaptive Bayes test for monotonicity. In *The Contribution of Young Researchers to Bayesian Statistics*, pages 29–33. Springer. 64

Savage, L. J. (1961). The foundations of statistics reconsidered. In Neyman, J., editor, *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1*, pages 575–586. University of California Press, Berkely, CA. 82

Scheibehenne, B., Gronau, Q. F., Jamil, T., and Wagenmakers, E.-J. (2017). Fixed or random? A resolution through model-averaging. Reply to Carlsson, Schimmack, Williams, and Burkner. *Psychological Science*, 28(11):1698–1701. 140

Scheibehenne, B., Jamil, T., and Wagenmakers, E.-J. (2016). Bayesian evidence synthesis can reconcile seemingly inconsistent results: The case of hotel towel reuse. *Psychological Science*, 27(7):1043–1046. 115, 118, 130, 140

Scheibehenne, B. and Pachur, T. (2015). Using Bayesian hierarchical parameter estimation to assess the generalizability of cognitive models of choice. *Psychonomic Bulletin & Review*, 22:391–407. 173

Schönbrodt, F. D. and Wagenmakers, E.-J. (2017). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review*. Manuscript accepted for publication. 132

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6:461–464. 206, 231

Scott, J. G. and Berger, J. O. (2006). An exploration of aspects of Bayesian multiple testing. *Journal of statistical planning and inference*, 136(7):2144–2162. 268

Scott, J. G. and Berger, J. O. (2010). Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem. *The Annals of Statistics*, 38(5):2587–2619. 43, 268

Sellke, T., Bayarri, M. J., and Berger, J. O. (2001). Calibration of p values for testing precise null hypotheses. *The American Statistician*, 55(1):62–71. 120

Senn, S. (2009). Comment. *Statistical Science*, 24(2):185–186. 11

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York. 79

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423. 258

Shiffrin, R. M., Chandramouli, S., and Grünwald, P. D. (2016). Bayes factors, relations to minimum description length, and overlapping model classes. *Journal of Mathematical Psychology*, 72:56–77. 50, 63

Shiffrin, R. M., Lee, M. D., Kim, W., and Wagenmakers, E.-J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science*, 32:1248–1284. 172, 173

Shiue, W.-K. and Bain, L. J. (1982). Experiment size and power comparisons for two-sample Poisson tests. *Applied Statistics*, 31(2):130–134. 102

Sinharay, S. and Stern, H. S. (2005). An empirical comparison of methods for computing Bayes factors in generalized linear mixed models. *Journal of Computational and Graphical Statistics*, 14:415–435. 192

Steegen, S., Dewitte, L., Tuerlinckx, F., and Vanpaemel, W. (2014). Measuring the crowd within again: A pre-registered replication study. *Frontiers in Psychology*, 5:786. 87, 88, 89

Steingroever, H., Pachur, T., Smíra, M., and Lee, M. D. (2016a). Bayesian techniques for analyzing group differences in the Iowa gambling task: A case study of intuitive and deliberate decision makers. *Psychonomic Bulletin & Review*. Manuscript accepted for publication. 194, 195

Steingroever, H., Wetzels, R., Horstmann, A., Neumann, J., and Wagenmakers, E.-J. (2013a). Performance of healthy participants on the Iowa gambling task. *Psychological Assessment*, 25:180–193. 194

Steingroever, H., Wetzels, R., and Wagenmakers, E.-J. (2013b). A comparison of reinforcement-learning models for the Iowa gambling task using parameter space partitioning. *The Journal of Problem Solving*, 5:Article 2. 194, 195

Steingroever, H., Wetzels, R., and Wagenmakers, E.-J. (2013c). Validating the PVL-Delta model for the Iowa gambling task. *Frontiers in Psychology*, 4: 898. 194

Steingroever, H., Wetzels, R., and Wagenmakers, E.-J. (2014). Absolute performance of reinforcement-learning models for the Iowa gambling task. *Decision*, 1:161–183. 194, 195, 196

Steingroever, H., Wetzels, R., and Wagenmakers, E.-J. (2016b). Bayes factors for reinforcement-learning models of the Iowa Gambling Task. *Decision*, 3(2):115. 49, 193, 194, 195, 197, 198, 199, 200

Stephens, M. and Balding, D. J. (2009). Bayesian statistical methods for genetic association studies. *Nature Reviews Genetics*, 10:681–690. 43

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, 64(3):153–181. 213

Stigler, S. (1973). Studies in the history of probability and statistics. XXXII Laplace, Fisher, and the discovery of the concept of sufficiency. *Biometrika*, 60(3):439–445. 216

Stigler, S. (1986). *The history of statistics: The measurement of uncertainty before 1900*. Belknap Press. 225

Stigler, S. (2007). The epic story of maximum likelihood. *Statistical Science*, 22(4):598–620. 149

Stirzaker, D. (2000). Advice to hedgehogs, or, constants can vary. *The Mathematical Gazette*, 84(500):197–210. 102

Stone, C. J., Hansen, M. H., Kooperberg, C., and Truong, Y. K. (1997). Polynomial splines and their tensor products in extended linear modeling: 1994 Wald memorial lecture. *The Annals of Statistics*, 25:1371–1470. 203

Strack, F., Martin, L. L., and Stepper, S. (1988). Inhibiting and facilitating conditions of the human smile: A nonobtrusive test of the facial feedback hypothesis. *Journal of Personality and Social Psychology*, 54:768–777. 90, 91, 95

Stulp, G., Buunk, A. P., Verhulst, S., and Pollet, T. V. (2013). Tall claims? Sense and nonsense about the importance of height of US presidents. *The Leadership Quarterly*, 24(1):159–171. 30, 31, 32, 33

Sun, D. and Berger, J. O. (2007). Objective Bayesian analysis for the multivariate normal model. *Bayesian Statistics*, 8:525–562. 151

Surowiecki, J. (2004). *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. Doubleday, New York. 87

Swirles, B. (1991). Harold Jeffreys: Some reminiscences. *Chance*, 4(2):22–23, 26. 11

Temme, N. M. (2003). Large parameter cases of the Gauss hypergeometric function. *Journal of Computational and Applied Mathematics*, 153(1):441–462. 110, 112

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Ann. Statist.*, pages 1701–1728. 156

Trafimow, D. and Marks, M. (2015). Editorial. *Basic And Applied Social Psychology*, 37:1–2. 120

Tribus, M. and McIrvine, E. C. (1971). Energy and information. *Scientific American*, 225(3):179–188. 258

Turner, B. M., Sederberg, P. B., and McClelland, J. L. (2016). Bayesian analysis of simulation-based models. *Journal of Mathematical Psychology*, 72:191–199. 49

van der Pas, S. and Grünwald, P. D. (2014). Almost the best of three worlds: Risk, consistency and optional stopping for the switch criterion in single parameter model selection. *arXiv preprint arXiv:1408.5724*. 242

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press. 119, 228, 242, 248, 263

van der Vaart, A. W. (2002). The statistical work of Lucien Le Cam. *Annals of Statistics*, pages 631–682. 248

van Doorn, J., Ly, A., Marsman, M., and Wagenmakers, E.-J. (2017). Bayesian estimation of Kendall's tau using a latent normal approach. *arXiv preprint arXiv:1703.01805*. 268

van Erven, T., Grünwald, P., and De Rooij, S. (2012). Catching up faster by switching sooner: A predictive approach to adaptive estimation with an application to the AIC–BIC dilemma. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(3):361–417. 37, 242

van Erven, T. and Harremos, P. (2014). Rényi divergence and Kullback-Leibler divergence. *IEEE Transactions on Information Theory*, 60(7):3797–3820. 260

van Ommen, T., Koolen, W. M., Feenstra, T. E., and Grünwald, P. D. (2016). Robust probability updating. *International Journal of Approximate Reasoning*, 74:30–57. 249

Vandekerckhove, J., Matzke, D., and Wagenmakers, E.-J. (2015). Model comparison and the principle of parsimony. In Busemeyer, J., Townsend, J., Wang, Z. J., and Eidels, A., editors, *Oxford Handbook of Computational and Mathematical Psychology*. Oxford: Oxford University Press. 179

Vandekerckhove, J., Rouder, J. N., and Kruschke, J. K. (2017). Beyond the new statistics: Bayesian inference for psychology [special issue]. *In preparation for Psychonomic Bulletin & Review*. 124

Vanpaemel, W. (2010). Prior sensitivity in theory testing: An apologia for the Bayes factor. *Journal of Mathematical Psychology*, 54:491–498. 36, 99

Vanpaemel, W. (2016). Prototypes, exemplars and the response scaling parameter: A Bayes factor perspective. *Journal of Mathematical Psychology*, 72:183–190. 171

Verdinelli, I. and Wasserman, L. (1995). Computing Bayes factors using a generalization of the Savage-Dickey density ratio. *Journal of the American Statistical Association*, 90(430):614–618. 203

Verhagen, A. J., Levy, R., Millsap, R. E., and Fox, J.-P. (2015). Evaluating evidence for invariant items: A Bayes factor applied to testing measurement invariance in IRT models. *Journal of Mathematical Psychology*, 72:171–182. 171

Verhagen, A. J. and Wagenmakers, E.-J. (2014). Bayesian tests to quantify the result of a replication attempt. *Journal of Experimental Psychology: General*, 143(4):1457–1475. 47, 87, 115, 119, 127, 129, 130, 137, 138, 140

Vul, E. and Pashler, H. (2008). Measuring the crowd within probabilistic representations within individuals. *Psychological Science*, 19:645–647. 87, 88

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, 14:779–804. 58

Wagenmakers, E.-J., Beek, T., Dijkhoff, L., Gronau, Q. F., Acosta, A., Adams, R., Albohn, D. N., Allard, E. S., Benning, S. D., Blouin–Hudon, E.-M., Bulnes, L. C., Caldwell, T. L., Calin–Jageman, R. J., Capaldi, C. A., Carfagno, N. S., Chasten, K. T., Cleeremans, A., Connell, L., DeCicco, J. M., Dijkstra, K., Fischer, A. H., Foroni, F., Hess, U., Holmes, K. J., Jones, J. L. H., Klein, O., Koch, C., Korb, S., Lewinski, P., Liao, J. D., Lund, S., Lupiáñez, J., Lynott, D., Nance, C. N., Oosterwijk, S., Özdoğru, A. A., Pacheco–Unguetti, A. P., Pearson, B., Powis, C., Riding, S., Roberts, T.-A., Rumiati, R. I., Senden, M., Shea–Shumsky, N. B., Sobocko, K., Soto, J. A., Steiner, T. G., Talarico, J. M., van Allen, Z. M., Vandekerckhove, M., Wainwright, B., Wayand, J. F., Zeelenberg, R., Zetzer, E. E., and Zwaan, R. A. (2016a). Registered Replication Report: Strack, Martin, & Stepper (1988). *Perspectives on Psychological Science*, 11:917–928. 90, 91, 95

Wagenmakers, E.-J., Grünwald, P. D., and Steyvers, M. (2006). Accumulative prediction error and the selection of time series models. *Journal of Mathematical Psychology*, 50:149–166. 37, 121, 242

Wagenmakers, E.-J., Lee, M. D., Rouder, J. N., and Morey, R. D. (2014). Another statistical paradox, or why intervals cannot be used for model comparison. *Manuscript submitted for publication*. 36

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., and Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology*, 60:158–189. 22, 133, 193, 203

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, A. J., Selker, R., Gronau, Q. F., Dropmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., van Kesteren, E.-J., van Doorn, J., Šmíra, M., Epskamp, S., Etz, A., Matzke, D., de Jong, T., van den Bergh, D., Sarafoglou, A., Steingroever, H., Derks, K., Rouder, J. N., and Morey, R. D. (2017a). Bayesian statistical inference for psychological science. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*. Manuscript accepted for publication. 126, 132

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, A. J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., and Morey, R. D. (2017b). Bayesian statistical inference for psychological science. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*. Manuscript accepted for publication. 124, 132

Wagenmakers, E.-J., Morey, R. D., and Lee, M. D. (2016b). Bayesian benefits for the pragmatic researcher. *Current Directions in Psychological Science*, 25:169–176. 116, 124, 126, 131

Wagenmakers, E.-J., Verhagen, A. J., and Ly, A. (2016c). How to quantify the evidence for the absence of a correlation. *Behavior Research Methods*, pages 413–426. 34, 115, 119, 130

Wagenmakers, E.-J., Verhagen, A. J., Ly, A., Bakker, M., Lee, M. D., Matzke, D., Rouder, J. N., and Morey, R. D. (2015a). A power fallacy. *Behavior Research Methods*, 47:913–917. 116

Wagenmakers, E.-J., Verhagen, A. J., Ly, A., Matzke, D., Steingroever, H., Rouder, J. N., and Morey, R. D. (2017c). The need for Bayesian hypothesis testing in psychological science. In Lilienfeld, S. O. and Waldman, I., editors, *Psychological Science Under Scrutiny: Recent Challenges and Proposed Solutions*, pages 123–138. John Wiley and Sons. 14, 116

Wagenmakers, E.-J. and Waldorp, L. (2006a). Editors' introduction. *Journal of Mathematical Psychology*, 50:99–100. 172

Wagenmakers, E.-J. and Waldorp, L. (2006b). Model selection: Theoretical developments and applications [Special issue]. *Journal of Mathematical Psychology*, 50(2). 37, 232

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., Kievit, R., and van der Maas, H. L. J. (2015b). A skeptical eye on psi. In May, E. and Marwaha, S., editors, *Extrasensory Perception: Support, Skepticism, and Science*, pages 153–176. ABC-CLIO. 117, 135

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., and van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi. *Journal of Personality and Social Psychology*, 100:426–432. 43

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., and Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, 7:627–633. 117

Wald, A. (1949). Statistical decision functions. *The Annals of Mathematical Statistics*, pages 165–205. 243

Waldorp, L. (2009). Robust and unbiased variance of GLM coefficients for misspecified autocorrelation and hemodynamic response models in fMRI. *International Journal of Biomedical Imaging*, 2009:723912. 242

Waldorp, L., Christoffels, I., and van de Ven, V. (2011). Effective connectivity of fMRI data using ancestral graph theory: Dealing with missing regions. *NeuroImage*, 54(4):2695–2705. 242

Waldorp, L., Huizenga, H., and Grasman, R. (2005). The Wald test and Cramér–Rao bound for misspecified models in electromagnetic source analysis. *IEEE Transactions on Signal Processing*, 53(9):3427–3435. 242

Wang, L. and Meng, X.-L. (2016). Warp bridge sampling: The next generation. *arXiv preprint arXiv:1609.07690*. 187

Wasserman, L. (2006). *All of Nonparametric Statistics*. Springer Texts in Statistics. Springer Science and Business Media, New York. 69

Wasserstein, R. L. and Lazar, N. A. (2016). The ASA's statement on p-values: Context, process, and purpose. *The American Statistician*. 124

Wetzels, R., Grasman, R. P. P. P., and Wagenmakers, E.-J. (2010a). An encompassing prior generalization of the Savage–Dickey density ratio test. *Computational Statistics & Data Analysis*, 54:2094–2102. 63, 193, 203

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., and Wagenmakers, E.-J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 t tests. *Perspectives on Psychological Science*, 6:291–298. 15, 22, 82, 95, 96, 97, 120

Wetzels, R., Raaijmakers, J. G., Jakab, E., and Wagenmakers, E.-J. (2009). How to quantify support for and against the null hypothesis: A flexible WinBUGS implementation of a default Bayesian t test. *Psychonomic Bulletin & Review*, 16(4):752–760. 24, 84, 99

Wetzels, R., Tutschkow, D., Dolan, C., van der Sluis, S., Dutilh, G., and Wagenmakers, E.-J. (2016). A Bayesian test for the hot hand phenomenon. *Journal of Mathematical Psychology*, 72:200–209. 171

Wetzels, R., Vandekerckhove, J., Tuerlinckx, F., and Wagenmakers, E.-J. (2010b). Bayesian parameter estimation in the Expectancy Valence model of the Iowa gambling task. *Journal of Mathematical Psychology*, 54:14–27. 173, 201

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1):1–25. 238, 242

Whittaker, E. (1902). On the functions associated with the parabolic cylinder in harmonic analysis. *Proceedings of the London Mathematical Society*, 1(1):417–427. 157

Wijsman, R. (1973). On the attainment of the Cramér-Rao lower bound. *The Annals of Statistics*, 1(3):538–542. 249

Willerman, L., Schultz, R., Rutledge, J. N., and Bigler, E. D. (1991). In vivo brain size and intelligence. *Intelligence*, 15:223–228. 76, 78

Witte, E. H. and Zenker, F. (2016). Reconstructing recent work on macro-social stress as a research program. *Basic and Applied Social Psychology*, 38(6):301–307. 4, 115, 116, 117, 118, 119, 120, 121, 301

Worthy, D. A. and Maddox, W. T. (2014). A comparison model of reinforcement-learning and win-stay-lose-shift decision-making processes: A tribute to W. K. Estes. *Journal of Mathematical Psychology*, 59:41–49. 195

Worthy, D. A., Pang, B., and Byrne, K. A. (2013). Decomposing the roles of perseveration and expected value representation in models of the Iowa gambling task. *Frontiers in Psychology*, 4. 195

Wrinch, D. and Jeffreys, H. (1919). On some aspects of the theory of probability. *Philosophical Magazine*, 38:715–731. 49, 225

Wrinch, D. and Jeffreys, H. (1921). On certain fundamental principles of scientific inquiry. *Philosophical Magazine*, 42:369–390. 13, 49, 83, 226

Wrinch, D. and Jeffreys, H. (1923). On certain fundamental principles of scientific inquiry. *Philosophical Magazine*, 45:368–375. 49, 226

Wu, H., Myung, I. J., and Batchelder, W. H. (2010). Minimum description length model selection of multinomial processing tree models. *Psychonomic Bulletin & Review*, 17:275–286. 232

Yang, G. L. (1999). A conversation with Lucien Le Cam. *Statistical Science*, pages 223–241. 248

Yang, G. L. and Le Cam, L. (2000). *Asymptotics in Statistics: Some Basic Concepts*. Springer-Verlag, Berlin. 46, 242

Yuan, Y. and Johnson, V. E. (2008). Bayesian hypothesis tests using nonparametric statistics. *Statistica Sinica*, 18:1185–1200. 71, 72

Zellner, A. (1980). Introduction. In Zellner, A., editor, *Bayesian Analysis in Econometrics and Statistics: Essays in Honor of Harold Jeffreys*, pages 1–10. North-Holland Publishing Company, Amsterdam, The Netherlands. 11

Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior distributions. *Bayesian inference and decision techniques: Essays in Honor of Bruno De Finetti*, 6:233–243. 24

Zellner, A. and Siow, A. (1980). Posterior odds ratios for selected regression hypotheses. In Bernardo, Jose, M., DeGroot, Morris, H., Lindley, Dennis, V., and Smith, Adrian, F., editors, *Bayesian Statistics: Proceedings of the First International Meeting Held in Valencia*, volume 1, pages 585–603. Springer. 24, 25

Zwaan, R. A., Etz, A., Lucas, R. E., and Donnellan, B. M. (2017). Making replications mainstream. *Behavioral and Brain Sciences*, pages 1–50. 129

Nederlandse Samenvatting

De Bayesiaanse hypothese toetsen die wij ontwikkeld hebben zijn bedoeld om empirische onderzoekers te helpen met (i) het kwantificeren van evidentie voor of tegen een hypothese, en (ii) het leren en construeren van (statistische) modellen en theorieën op basis van geobserveerde data.

Een statistisch model geeft een versimpelde beschrijving van de werkelijkheid met een mathematische relatie $f(d \mid \theta)$ tussen observaties, de data, d en *parameters* θ. Zo kan d bijvoorbeeld refereren naar bloeddrukmetingen voor en na een behandeling van een steekproef van patiënten, θ refereert dan naar de effectgrootte, en f is in de meeste gevallen een normale verdeling om er rekening mee te houden dat de metingen slechts een steekproef zijn uit een grotere populatie patiënten.

Om te toetsen of de behandeling effectief is vergelijken we het *nul model* $\mathcal{M}_0$, het statistisch model waarbij de effectgrootte op nul wordt gezet $\theta = 0$, met het alternatief model $\mathcal{M}_1$, het model waarin de effectgrootte elke reëele waarde kan aannemen.

De *a priori plausibiliteit* van de effectiviteit hangt af van welke behandeling de patiënt krijgt voorgeschreven. De a priori kans dat de behandeling effectief is, is relatief hoog, zeg, $P(\mathcal{M}_1) = 0.9$ en $P(\mathcal{M}_0) = 0.1$, wanneer de patiënten worden voorgeschreven om pillen in te nemen met een actieve stof ontwikkeld om bloeddruk te verlagen. Een ander, maar equivalente, manier om deze a priori model kansen te beschrijven is met behulp van de *a priori model kansverhouding*, in dit geval, negen staat tot één, dus $\frac{P(\mathcal{M}_1)}{P(\mathcal{M}_0)} = 9$. Op basis van de geobserveerde data kunnen we de a priori model kansverhouding bijwerken tot *a posteriori* model kansverhouding $\frac{P(\mathcal{M}_1 \mid d)}{P(\mathcal{M}_0 \mid d)}$ met behulp van de *regel van Bayes* en leidt tot de volgende cruciale vergelijking:

$$\frac{P(\mathcal{M}_1 \mid d)}{P(\mathcal{M}_0 \mid d)} = \underbrace{\frac{p(d \mid \mathcal{M}_1)}{p(d \mid \mathcal{M}_0)}}_{\mathrm{BF}_{10}(d)} \frac{P(\mathcal{M}_1)}{P(\mathcal{M}_0)}, \tag{.0.1}$$

waar $P(\mathcal{M}_i \mid d)$ refereert naar de *a posteriori model kans* van $\mathcal{M}_i$ gegeven de observaties, en $p(d \mid \mathcal{M}_i)$ refereert naar de marginale waarschijnlijkheid van model $\mathcal{M}_i$. De term $\mathrm{BF}_{10}(d)$ is de zogeheten factor van Bayes, oftewel, *Bayes factor*,

en beschrijft hoe de a priori model kansverhouding wordt bijgewerkt tot de a posteriori model kansverhouding gegeven de observaties d.

De Bayes factor is makkelijk interpreteerbaar: $\text{BF}_{10}(d) = 7$ indiceert dat de observaties 7 keer zo waarschijnlijk zijn onder $\mathcal{M}_1$ als onder $\mathcal{M}_0$, en $\text{BF}_{10}(d) = .2$ indiceert dat de observaties 5 keer zo waarschijnlijk zijn onder $\mathcal{M}_0$ als onder $\mathcal{M}_1$. Gegeven de observaties is de Bayes factor $\text{BF}_{10}(d)$ altijd een niet-negatief getal en hoe hoger (lager) dit getal, hoe meer (minder) evidentie er is voor $\mathcal{M}_1$ ten opzichte van $\mathcal{M}_0$. Op een gelijksoortige manier, wanneer ook de activiteitsniveau van de patiënten is gemeten, kunnen we toetsen of de behandeling mensen moe maakt. Op deze manier krijgen we geleidelijk meer inzicht in de effecten van de behandeling op de populatie van patiënten.

De Bayes factor is een ratio van de marginale waarschijnlijkheid $p(d \mid \mathcal{M}_i)$ die aangeeft hoe goed het model op de geobserveerde data past. Deze marginale waarschijnlijkheid wordt berekend door de relatie $f_i(d \mid \theta)$ van model $\mathcal{M}_i$ gegeven de observaties d te evalueren op elke mogelijke parameter waarde θ en te middelen ten opzichte van een *a priori verdeling* $\pi_i(\theta)$:

$$p(d \mid \mathcal{M}_i) = \int f_i(d \mid \theta)\pi_i(\theta)\mathrm{d}\theta. \tag{.0.2}$$

Gegeven twee modellen, dus, de relaties $f_1(d \mid \theta)$ en $f_0(d \mid \theta)$, is het de taak van de statisticus om twee a priori verdelingen, namelijk, $\pi_0(\theta)$ en $\pi_1(\theta)$ te kiezen om daarmee een Bayes factor te construeren. Voor een gebruiksvriendelijke Bayes factor moet de statisticus er ook voor zorgen dat deze uit te rekenen is voor elke data set d. In dit proefschrift beschreven we hoe men a priori verdelingen voor Bayes factoren moet selecteren en berekenen. De resulterende Bayes factoren zijn of worden nog geïmplementeerd in het gratis software-pakket vernoemd naar Harold Jeffreys, *Jeffreys's Amazing Statistics Program*, `JASP`, (url: `https://jasp-stats.org/`, JASP Team, 2017).

Deel I. De onderliggende principes van de Bayes factor

Het eerste gedeelte van dit proefschrift richtte zich op de filosofie, de motivering en de constructie van zogeheten *Jeffreys's Bayes factoren*.

In Hoofdstuk 2 bespraken we de onderliggende principes van de Bayes factor, hoe deze te interpreteren, en gaven we een beschrijving van de algemene constructie waarmee Jeffreys a priori verdelingen selecteerde voor Bayes factoren. In deze constructie is het van belang om een Bayes factor te ontwerpen dat *predictief geijkt* en *informatie consistent* is. Een predictief geijkte Bayes factor is één wanneer de steekproefgrootte te klein en daarom ambigu is, terwijl een informatie consistente Bayes factor oneindig is wanneer de observaties overweldigend wijzen naar het bestaan van een effect. De constructie waarmee Jeffreys Bayes factoren ontwerpt is ontleend uit hoe hij zijn Bayesiaanse t-toets opzet. Deze constructie hebben we gebruikt om een Jeffreys's Bayes factor af te leiden voor de product-moment correlatiecoëfficiënt van Pearson. De resulterende Bayes factor is analytisch en makkelijk te gebruiken.

In Hoofdstuk 3 reageren wij op twee discussie artikelen op ons werk over Harold Jeffreys. In dit hoofdstuk belichtten wij de zogeheten *Jeffreys-Lindley-Bartlett*

paradox, het verschil tussen inferentie en besluitvorming, en het verschil tussen schatten en toetsen toe.

Deel II. Bayes factoren voor veelgebruikte statistische analyses

Het tweede deel van dit proefschrift richtte zich op Bayes factoren die wij geconstrueerd hebben voor bepaalde veelgebruikte statistische analyses.

In Hoofdstuk 4 zetten we een Bayesiaanse methode uiteen voor het schatten en toetsen van de rangcorrelatiecoëfficiënt τ van Kendall. Voor deze methode modelleerden we de toets statistiek die we daarna gecombineerd hebben met het analytische resultaat voor de correlatiecoëfficiënt van Pearson.

In Hoofdstuk 5 hebben we de afleiding van het analytische resultaat voor de correlatiecoëfficiënt van Pearson gebruikt om een geïnformeerde Bayesiaanse t-toets te construeren. De klasse van a priori verdelingen die wij hiervoor gebruikten is een veralgemenisering van de verdelingen die Harold Jeffreys aandroeg voor dit probleem, maar laat de locatie en schaal op de a priori verdeling van de effectgrootte vrij. Hierdoor kunnen onderzoekers wanneer ze substantiële voorkennis hebben deze gebruiken in hun t-toetsen.

In Hoofdstuk 6 introduceerden we *limiet-consistentie* als een desideratum voor het selecteren van a priori verdelingen voor twee-steekproef toetsen. Voor dit desideratum bekijken we de hypothetische scenario waarin de dataverzameling voor een process vroegtijdig wordt beëindigd, terwijl de dataverzameling van het tweede proces voor een onbepaalde tijd doorgaat. In dergelijke gevallen zou de Bayes factor moeten convergeren naar een eindige limiet. We constateren dat de Bayes factor die Jeffreys voorstelde voor het twee-steekproef Poisson probleem limiet-inconsistent is. Als oplossing generaliseren wij de Bayes factor van Jeffreys zodat deze wel limiet-consistent is.

Deel III. Wetenschappelijk kennis vergaren met Bayes factoren

Het derde deel van dit proefschrift richtte zich op het gebruik van Bayes factoren in de empirische wetenschappen als een instrument voor wetenschappelijk leren. In het bijzonder bespreken wij de rol van de Bayes factor in de "replicatie- en reproduceerbaarheidscrisis" (Baker, 2016, Levelt et al., 2012, Pashler and Wagenmakers, 2012).

In Hoofdstuk 7 bespraken wij kort hoe psychologen zich hebben ingezet om de reproduceerbaarheid van het veld te vergroten met grootschalige replicatie initiatieven, zoals het "Reproducibility Project: Psychology" (Open Science Collaboration, 2015), de speciale replicatie editie van *Social Psychology* (Nosek and Lakens, 2014) en de vele ManyLabs experimenten (Ebersole et al., 2016; Klein et al., 2014). Dit hoofdstuk is een commentaar op het werk van Witte and Zenker (2016). Zij beweren dat een "ander" gebruik van standaard statistische methoden op basis van p-waarden een oplossing is voor de replicatie- en reproduceerbaarheidscrisis. Ons standpunt is dat deze crisis veel omvattender is dan een discussie over de statistische methoden. Wij pleiten er namelijk voor om confirmatieve studies te preregistreren. Door te preregistreren worden termen beter gedefinieerd en vermijdt men het probleem van achteraf kanskapitalisatie. Daarnaast vinden

wij dat wetenschap open en transparant moet zijn waarbij onzekerheid gerapporteerd wordt, omdat dit een betere en eerlijke beeld geeft van het wetenschappelijke proces.

In Hoofdstuk 8 beschreven wij het gemak waarmee men een Bayesiaanse heranalyse kan doen, zelfs wanneer de volledige dataset niet beschikbaar is. Dit is relevant voor onderzoekers die naast p-waarden ook een Bayes factor willen rapporteren. Een Bayesiaanse heranalyse is ook handig voor redacteuren, recensenten, lezers en verslaggevers, omdat zij in één oogopslag de evidentie kunnen bepalen in gerapporteerde statistieken. Daarnaast demonstreren we hoe gevoelig de evidentie is voor veranderingen in de a priori verdelingen door middel van een robuustheidsanalyse. De Bayesiaanse heranalyse leidt ook tot een a posteriori verdeling waaruit men kan concluderen welke gebieden in de parameterruimte plausibeler worden nadat we de observaties in ogenschouw nemen. Als laatste bespraken wij hoe de a posteriori verdeling gebruikt kan worden als voorkennis in een vervolgstudie.

In Hoofdstuk 9 bespraken wij een algemene methode om de evidentie te extraheren uit de observaties van een directe replicatiepoging gegeven de observaties van een oorspronkelijke studie. Deze algemene methode is ontworpen om onderzoekers te helpen modellen te bouwen en kennis te vergaren uit een groeiend aantal replicatiestudies.

Deel IV. Analytische resultaten

Het vierde deel van dit proefschrift richtte zich op verschillende analytische resultaten die zijn gebruikt voor de constructie van de Bayesiaanse toetsen in dit proefschrift.

In Hoofdstuk 10 leidden wij de analytische a posteriori verdeling af voor een grote klasse van a priori verdelingen op de product-moment correlatiecoëfficiënt van Pearson. Dit resultaat is gebruikt voor de analytische Bayes factor in Hoofdstuk 2 en de afleiding vormt de basis van Hoofdstuk 4 en 5.

In Hoofdstuk 11 leidden wij analytische a posteriori verdelingingen af voor modellen met discrete data. Het eerste resultaat is gebruikt in Hoofdstuk 6 om een limiet-consistente Bayes factor te construeren voor het twee-steekproef Poisson probleem. Dit resultaat kan ook gebruikt worden om een robuustheidsanalyse te definiëren voor een binomiaal toets. Daarnaast bevat dit hoofdstuk analytische uitdrukkingen voor de eenzijdige binomiaal Bayes factoren. Het laatste resultaat is een analytische uitdrukking voor de ratio van kansverhoudingen in een 2-keer-2 contingentie tabel.

Deel V. Twee handleidingen

Het vijfde en laatste deel van dit proefschrift richtte zich op hulpmiddelen bij het construeren van Bayes factoren en biedt verdieping in mathematische statistisch modelleren.

In Hoofdstuk 12 legden wij uit hoe *bridge sampling* (Meng and Wong, 1996) gebruikt kan worden om uitkomsten van een MCMC-procedure te transformeren in een schatting van de marginale waarschijnlijkheid van een model. De bridge

sampler is relevant voor complexe modellen met hiërarchische structuren die lastig te beschrijven zijn met standaard wiskundige functies.

In Hoofdstuk 13 gaven we een algemene beschrijving van mathematische statistiek en de rol van Fisher informatie voor statistische modellen in het bijzonder. In het frequentistische paradigma werd uiteengezet hoe men hypothese toetsen en betrouwbaarheidsintervallen kan construeren door Fisher informatie te combineren met maximale waarschijnlijkheidsschatters. In het Bayesiaanse paradigma werd uiteengezet hoe men een standaard a priori verdeling kan construeren uit Fisher informatie. In het minimale beschrijvingslengte paradigma werd uiteengezet hoe Fisher informatie gebruikt wordt om de mate van model complexiteit te beschrijven. De resultaten hangen af van bepaalde regulariteitscondities die gegeven zijn in de appendix. Wanneer men modellen construeert die voldoen aan deze condities zullen de standaard statistische methoden (asymptotisch) geldig zijn.

Acknowledgements — Dankwoord

This dissertation would not have been completed without the help and support of many people. First and foremost, I would like to thank my supervisor Eric-Jan Wagenmakers, co-supervisor Maarten Marsman, the members of my defence committee: Jim Berger, Michael Lee, Peter Grünwald, Han van der Maas, Lourens Waldorp, and Raoul Grasman, as well the European Research Council who provided financial support for this project.

EJ, thank you for the interesting discussions, your guidance, generosity, and your optimistic demeanour with which you lead an ever expanding lab. With such a stimulating and inspiring surrounding, it was easy for me to grow as an independent researcher. I am proud to call myself a member of your lab, and I am grateful for the opportunity to have led the JASP team whenever you were too busy.

Marsman, thank you for being the unanticipated voice of reason, for calming me down at the right moment, and for being a great example of perseverance and intelligence. Dora, your cheerful presence as well as your role as the glue that binds the lab together made it easy for me to adapt to the group. It is a great honour to have you and Sacha as my paranymphs. Sacha, thank you for being a great friend, for dragging me to the gym, and the best King's days. Helen, for introducing new phrases into the Dutch language, and for honouring me with a star for good behaviour. Udo, for your display of strength by doing both a PhD and mathematics bachelor at the same time. Johnny, for being so chilled out with all the pressure I put on you, and your ability to surpass yourself every time. Quentin, thank you for being my external memory, for making sense of my equations, and for building on them with great success. Etz, for being so expressive, and infectiously happy and free whenever you are around.

Bruno, thank you for being such a great down-to-earth lead programmer and for coping with my stupidity once every while. Frans, for being so patient with me, and for sharing your Gall and Gall card. Tim de Jong, thank you for your unbreakable (I tried) spirit, your snide remarks, and for totally exceeding all expectations. Akash, thank you for being such a miracle worker and for making my work as CTO so much easier. Erik-Jan, for your enthusiasm, realism, and for being so knowledgeable. Koen, for doing much of the leg work in our software

project, and for your drive for learning. Alexandra, for your creativity and for keeping Marsman at bay. Tim Draws, for the passion with which you make our work known to the general public. Don, for your courage by going out of your comfort zone to solve any problem. Louise, for always being so helpful and the best gossips of course. Michael, thank you for being a laksa regular, a friendly latent supervisor around which we plan our best activities.

Our lab could not thrive without the leadership of Han and Denny. I benefitted from interactions with members of their groups as well. Jonas, Lourens, Riet, Sharon, Claire, Gilles, Laura, Sacha, Udo, Johnny, Quentin, Adela, Raoul, Tahira, Pia, Don, Koen, Claudia, Joost, thanks for your active participation in my statistical learning reading seminar. It is a great pleasure to broaden my horizons with such a motivated group of inquisitive young researchers.

In addition to having a great group to work with at the Psychological Methods department, I was fortunate enough to have the opportunity to also explore other research interests in mathematical statistics, causality, and machine learning with other groups. Bas, Aad, Harry, Botond, Bartek, Frank, Paulo, Eduard, Richard, Paulo, Moritz, Shota, Joris, Johannes, Tim, Weimin, Jakob, Kolyan, JB, Stéphanie, Fengnan, Kevin, Suzanne, Dong, Gino, Nurzhan, Jarno, Jan, and Alice, thanks for giving me my mathematics fix at the Bayes Club, Statistics for Structures, or at the Van Dantzig seminar. It is an honour and a pleasure to (Bayes) club with you. Joris, Stephan, Philip, Sara, Patrick, Thijs, Mijung, Tom, and Tineke, thank you for letting me invade your weekly meetings, and for being highly correlated to all that I know about causality. Peter, Wouter, Nishant, Rianne, Tom, Judith, Allard, and Alice, thank you for welcoming in your group and teaching me about statistical learning.

My interest in research was mostly sparked by André Schram, Chris Klaassen, Henk Pijls, Ale Jan Homburg and Bas Kleijn, who mentored and inspired me during my time as a student. During my PhD I always felt the warm support from my friends, family and my dearest (the bestest) Gracia. The time I spent with friends such as Huub, Lai, Yaron, Fleur, Mark, Evan, Anannd, Vincent, Farhaz, Freek and Johan during this time were brief, but always well-appreciated. Words cannot express the amount of help I received from my grandfather, parents and sister, and I will be forever grateful for all the sacrifices they made on my behalf. They have always been a great example of diligence, hard work, and loving kindness. Gracia offered me love, care, support, great dinners, which kept on me on track to finish my PhD, and her presence makes everything better and so much more fun.

I am grateful to all co-authors who have contributed to the papers that form the basis of this dissertation. This thesis would have been impossible without your ideas, knowledge, and advice. It goes without saying that any remaining errors in this dissertation are all mine.

Publications

1. Wagenmakers, E.–J., Verhagen, A. J., **Ly, A.**, Bakker, M., Lee, M. D., Matzke, D., & Rouder, J. N. (2015). A power fallacy. *Behavior Research Methods*, *47*, 913–917.

2. Love, J.P., Selker, R., Verhagen, A.J., Marsman, M., Gronau, Q.F., Jamil, T., Šmíra, M., Epskamp, S., Wild, A., **Ly, A.**, Matzke, M., Morey, R.D., Rouder, J.N. & Wagenmakers, E.–J. (2015). Software to sharpen your stats. *APS Observer*, *28*, 27–29.

3. Wagenmakers, E.–J., Beek, T., Rotteveel, M., Gierholz, A., Matzke, D., Steingroever, H., **Ly, A.**, Verhagen, A. J., Selker, R., Sasiadek, A., Gronau, Q. F., Love, J., & Pinto, Y. (2015). Turning the hands of time again: A purely confirmatory replication study and a Bayesian analysis. *Frontiers in Psychology: Cognition*, *6*, 494.

4. Love, J., Selker, R., Marsman, M., Jamil, T. Dropmann, D., Verhagen, A.J., **Ly, A.**, Gronau, Q.F., Smira, M., Epskamp, S., Matzke, D., Wild, A., Knight, P., Rouder, J.N., Morey, R.D., & Wagenmakers, E.–J. (2015). JASP – Graphical statistical software for common statistical designs. Manuscript accepted for publication.

5. Tierney, W., Schweinsberg, M., Jordan, J., Kennedy, D., Qureshi, I., Sommer, S. A., Thornley, N., Madan, N., Vianello, M., Awtrey, E., Zhu, L., Diermeier, D., Heinze, J., Srinivasan, M., Tannenbaum, D., Bivolaru, E., Dana, J., Davis-Stober, C., du Plessis, C., Gronau, Q. F., Hafenbrack, A., Liao, E., **Ly, A.**, Marsman, M., Murase, T., Schaerer, M., Tworek, C., Wagenmakers, E.–J., Wong, L., Anderson, T., Bauman, C., Bedwell, W., Brescoll, V., Canavan, A., Chandler, J., Cheries, E., Cheryan, S., Cheung, F., Cimpian, A., Clark, M., Cordon, D., Cushman, F., Ditto, P., Amell, A., Frick, S., Gamez-Djokic, M., Grady, R., Graham, J., Gu, J., Hahn, A., Hanson, B., Hartwich, N., Hein, K., Inbar, Y., Jiang, L., Kellogg, T., Legate, N., Luoma, T., Maibeucher, H., Meindl, P., Miles, J., Mislin, A., Molden, D., Motyl, M., Newman, G., Ngo, H. H., Packham, H., Ramsay, P. S., Ray, J., Sackett, A., Sellier, A.-L., Sokolova, T., Sowden, W., Storage, D., Sun, X.,

van Bavel, J., Washburn, A., Wei, C., Wetter, E., Wilson, C., Darroux, S.-C., & Uhlmann, E. (2016). Data from a pre-publication independent replication initiative examining ten moral judgement effects. *Scientific Data*, *3*, 160082.

6. Wagenmakers, E.–J., Verhagen, A. J., & **Ly, A.** (2016). How to quantify the evidence for the absence of a correlation. *Behavior Research Methods*, *48*(2), 413–426.

7. **Ly, A.**, Verhagen, A. J., & Wagenmakers, E.–J. (2016). Harold Jeffreys's default Bayes factor hypothesis tests: Explanation, extension, and application in psychology. *Journal of Mathematical Psychology*, *72*, 19–32.

8. **Ly, A.**, Verhagen, A. J., & Wagenmakers, E.–J. (2016). An evaluation of alternative methods for testing hypotheses, from the perspective of Harold Jeffreys. *Journal of Mathematical Psychology*, *72*, 43–55.

9. Marsman, M., **Ly, A.**, & Wagenmakers, E.–J. (2016). Four requirements for an acceptable research program. *Basic and Applied Social Psychology*, *38*(6), 308–312.

10. van Doorn, J., **Ly, A.**, Marsman, M., & Wagenmakers, E.–J. (2016). Bayesian inference for Kendall's rank correlation coefficient. *The American Statistician*.

11. Dutilh, G., Vandekerckhove, J., **Ly, A.**, Matzke, D., Pedroni, A., Frey, R., Rieskamp, J., & Wagenmakers, E.–J. (2017). A test of the diffusion model explanation for the worst performance rule using preregistration and blinding. *Attention, Perception, & Psychophysics*, *79*(3), 713–725.

12. Wagenmakers, E.–J., Verhagen, A. J., **Ly, A.**, Matzke, D., Steingroever, H., Rouder, J. N., & Morey, R. D. (2017). The need for Bayesian hypothesis testing in psychological science. In Lilienfeld, S. O., & Waldman, I. (Eds.), *Psychological Science Under Scrutiny: Recent Challenges and Proposed Solutions*, 123–138, John Wiley and Sons.

13. **Ly, A.**, Marsman, M., & Wagenmakers, E.–J. (2017). Analytic posteriors for Pearson's correlation coefficient. *Statistica Neerlandica*.

14. Schweinsberg, M., Madan, N., Vianello, M., Sommer, S. A., Jordan, J., Tierney, W., Awtrey, E., Zhu, L., Diermeier, D., Heinze, J., Srinivasan, M., Tannenbaum, D., Bivolaru, E., Dana, J., Davis-Stober, C. P., Du Plessis, C. Gronau, Q. F., Hafenbrack, A. C., Liao, E. Y., **Ly, A.**, Marsman, M., Murase, T., Qureshi, I., Schaerer, M., Thornley, N., Tworek, C. M., Wagenmakers, E-J., Wong, L., Anderson, T., Bauman, C. W., Bedwell, W. L., Brescoll, V., Canavan, A., Chandler, J. J., Cheries, E., Cheryan, S., Cheung, F., Cimpian, A., Clark, M., Cordon, D., Cushman, F., Ditto, P. H., Donahue, T., Frick, S. E., Gamez-Djokic, M., Hofstein Grady, R., Graham, J., Gu, J., Hahn, A., Hanson, B. E., Hartwich, N. J., Hein, K., Inbar, Y., Jiang, L., Kellogg, T., Kennedy, D. M., Legate, N., Luoma, T. P., Maibeucher, H., Meindl, P., Miles, J., Mislin, A., Molden, D. C., Motyl, M., Newman,

G., Ngo, H. H., Packham, H., Ramsay, P. S., Ray, J. L., Sackett, A. M., Sellier, A-L., Sokolova, T., Sowden, W., Storage, D., Sun, X., Van Bavel, J. J., Washburn, A. N., Wei, C., Wetter, E., Wilson, C., Darroux, S-C., & Uhlmann, E. L. (2017). The pipeline project: Pre-publication independent replications of a single laboratory's research pipeline. *Journal of Experimental Social Psychology*. Manuscript accepted for publication.

15. Jamil, T., Marsman, M., **Ly, A.**, Morey, R. D., & Wagenmakers, E.–J. (2017). What are the odds' Modern relevance and Bayes factor solutions for MacAlister's problem from the 1881 Educational Times. *Educational and Psychological Measurement*, *77*(5), 819–830.

16. Jamil, T., **Ly, A.**, Morey, R. D., Love, J., Marsman, M., & Wagenmakers, E.–J. (2017). Default "Gunel and Dickey" Bayes factors for contingency tables. *Behavior Research Methods*, *49*(2), 638–652.

17. Wagenmakers, E.–J., Marsman, M., Jamil, T., **Ly, A.**, Verhagen, A. J., Love, J., Selker, R., Gronau, Q. F., Smira, M., Epskamp, S., Matzke, D., Rouder, J. N., Morey, R. D. (2017). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*. Manuscript accepted for publication.

18. Wagenmakers, E.–J., Love, J., Marsman, M., Jamil, T., **Ly, A.**, Verhagen, A. J., Selker, R., Gronau, Q. F., Dropmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., van Kesteren, E.-J., van Doorn, J., Smira, M., Epskamp, S., Etz, A., Matzke, D., Rouder, J. N., & Morey, R. D. (2017). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*. Manuscript accepted for publication.

19. Keuken, M. C., **Ly, A.**, Boekel, W. E., Wagenmakers, E.–J. , Belay, L., Verhagen, A. J., Brown, S. D., & Forstmann, B. U. (2017). Corrigendum for: A purely confirmatory replication study of structural brain-behavior correlations. *Cortex*, *93*, 229–233.

20. **Ly, A.**, Marsman, M., Verhagen, A. J., Grasman, R. P. P. P., Wagenmakers, E.–J. (2017). A tutorial on Fisher information. *Journal of Mathematical Psychology*, *80*, 40–55.

21. Matzke, D., **Ly, A.**, Selker, R., Weeda, W. D., Scheibehenne, B., Lee, M. D., & Wagenmakers, E.–J. (2017). Bayesian inference for correlations in the presence of measurement error and estimation uncertainty. *Collabra: Psychology*, 3(1), 25.

22. van Doorn, J., **Ly, A.**, Marsman, M., & Wagenmakers, E.–J. (2017). Bayesian estimation of Kendall's tau using a latent normal approach. Manuscript submitted for publication.

23. Gronau, Q. F., Sarafoglou, A., Matzke, D., **Ly, A.**, Boehm, U., Marsman, M., Leslie, D. S., Forster, J. J., Wagenmakers, E.–J., & Steingroever, H. (2017). A tutorial on bridge sampling. *Journal of Mathematical Psychology*, *81*, 80–97.

24. Gronau, Q. F., **Ly, A.**, & Wagenmakers, E.–J. (2017). Informed Bayesian *t*-tests. Manuscript submitted for publication.

25. **Ly, A.**, Raj, A., Etz, A., Marsman, M., Gronau, Q. F., & Wagenmakers, E.–J. (2017). Bayesian reanalyses from summary statistics: A guide for academic consumers. Manuscript submitted for publication.

26. **Ly, A.**, Etz, A., Marsman, M., & Wagenmakers, E.–J. (2017). Replication Bayes factors from evidence updating. Manuscript submitted for publication.

27. **Ly, A.**, Raj, A., Marsman, M., & Wagenmakers, E.–J. (2017). A limit-consistent Bayes factor for testing the equality of two Poisson rates. Manuscript in preparation.

28. **JASP team** (2017). JASP (Version 0.8.4.0)[Computer software].

29. **Ly, A.**, Boehm, U., Heathcote, A., Turner, B. M. , Forstmann, B., Marsman, M., & Matzke, D. (2017). A flexible and efficient hierarchical Bayesian approach to the exploration of individual differences in cognitive-model-based neuroscience. In Mousstafa, A. A., (Ed.), *Computational Models of Brain and Behavior*, 467–480.