

Selected Bibliography: a selection of articles with links. (*Numbers in parentheses after an item refer to a summer seminar meeting day on the schedule.*)

Amrhein, Greenland, & McShane (2019). "Comment: Retire Statistical Significance", *Nature* 567: [305-308](#).

Achinstein ([2010](#)). Mill's Sins or Mayo's Errors? (E&I: 170-188). (10)

Bacchus, Kyburg, & Thalos ([1990](#)). Against Conditionalization, *Synthese* (85): 475-506. (11) (14)

Barnett ([1999](#)). Comparative Statistical Inference (Chapter 6: Bayesian Inference), John Wiley & Sons. (1) (11) (14)

Begley & Ellis ([2012](#)) Raise standards for preclinical cancer research. *Nature* 483: 531-533. (9)

Bem ([2011](#)). Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect, *Journal of Personality and Social Psychology* 100(3), 407-425. (9)

Bem, Utts & Johnson ([2011](#)). Must Psychologists Change the Way They Analyze Their Data? *Journal of Personality and Social Psychology*, 101(4), 716-719. (9)

Benjamin, Berger, Johannesson et al ([2017](#)) Redefine Statistical Significance, *Nature Human Behaviour* 2, 6-10. (10)

Benjamini & Hochberg ([1995](#)). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, *Journal of The Royal Statistical Society*. (9)

Berger, J. ([2003](#)). Could Fisher, Jeffreys and Neyman have Agreed on Testing? *Stat Sci* 18: 1-12. (1) (5) (6) (7) (9)

Berger, J. ([2006](#)). The Case for Objective Bayesian Analysis and [Rejoinder](#), *Bayesian Analysis* 1(3), 385-402; 457-64. (11)

Berger, J. & Sellke ([1987](#)). Testing a Point Null Hypothesis: The Irreconcilability of P Values and Evidence (with Discussion and Rejoinder), *Journal of the American Statistical Association* 82(397), 112-22; 135-9. (6) (7) (9) (10).

Bernardo, J. ([1997](#)). Non-informative Priors Do Not Exist: A Dialogue with Jose M. Bernardo, *Journal of Statistical Planning and Inference* 65(1), 159-77. (6) (7)

Bernardo, J. ([2010](#)). Integrated Objective Bayesian Estimation and Hypothesis Testing (with discussion), *Bayesian Statistics* 9, 1-68. (10)

Brown, E. N. and Kass, R. E. (2009). What is Statistics? (with discussion), The American Statistician 63, 105–23. (1), (11) (14)

Birnbaum, A. (1970), Statistical Methods in Scientific Inference (letter to the Editor), Nature 225(5237): 1033 (1)

For extensive Birnbaum references see this [post](#) on Error Statistics Philosophy Blog

Casella & R. Berger (1987a). Reconciling Bayesian and Frequentist Evidence in the One-sided Testing Problem, Journal of the American Statistical Association 82(397), 106–11. (10)

Casella, G. and Berger, R. (1987b). Comment on Testing Precise Hypotheses by J. O. Berger and M. Delampady, Statistical Science 2(3), 344–7. (10)

Colquhoun, D. (2014). ‘An Investigation of the False Discovery Rate and the Misinterpretation of P-values’, Royal Society Open Science 1(3), (16 pages). (12)

Cousins, R. (2017). ‘The Jeffreys-Lindley Paradox and Discovery Criteria in High Energy Physics’, Synthese 194, 395–432. (6) (7)

Cox, D. (1977). The Role of Significance Tests (with Discussion), Scandinavian Journal of Statistics 4, 49–70. (3) (4) (5)

Cox, D. (2006a). Principles of Statistical Inference, CUP.

Cox & Mayo (2010). Objectivity and Conditionality in Frequentist Inference (E&I: 276-304). (6) (7) (9)

Cox & Mayo (2011) A Statistical Scientist Meets a Philosopher of Science: A Conversation between Sir David Cox and Deborah Mayo (as recorded, June 2011). Rationality, Markets and Morals (RMM), 2, Special Topic: Statistical Science and Philosophy of Science, 103-114. (11)

Crupi & Tentori (2010). Irrelevant Conjunction: Statement and Solution of a New Paradox, Phil Sci, 77, 1–13. (2)

Earman, J. and Glymour, C. (1980). ‘Relativity and Eclipses: The British Eclipse Expeditions of 1919 and Their Predecessors’, Historical Studies in the Physical Sciences 11(1), 49–85. (5)

Edwards, Lindman & Savage E, L, & S (1963). Bayesian Statistical Inference for Psychological Research, Psychological Review 70(3), 193–242. (1)

Efron, B. (1986). Why Isn’t Everyone a Bayesian?, The American Statistician 40(1), 1–5. (3) (4)

Efron, B. (1998). R. A. Fisher in the 21st Century and Rejoinder, *Statistical Science* 13(3), 95–114; 121–2. (9)

Efron (2013) A 250-Year Argument: Belief, Behavior, and the Bootstrap, *Bulletin of the American Mathematical Society* 50(1), 126–46. (11) (14)

Feynman (1974). Cargo Cult Science (Graduation Speech) (1) (3) (4)

Fisher (1930). Inverse Probability, *Mathematical Proceedings of the Cambridge Philosophical Society* 26(4), 528–35. (6) (7)

Fisher (1934). Two New Properties of Mathematical Likelihood, *Proceedings of the Royal Society of London Series A* 144 (852), 285–307. (6) (7)

Fisher (1935a)/(1947). *The Design of Experiments*, 1st ed., Edinburgh: Oliver and Boyd. Reprinted in Fisher 1990. (Lady Tasting Tea) (1)

Fisher, R. A. (1936), Uncertain Inference, *Proceedings of the American Academy of Arts and Sciences* 71, 248–58. (6) (7)

Fisher (1955), *Statistical Methods and Scientific Induction*, *J R Stat Soc (B)* 17: 69–78. (1), (5) (6) (7)

Fitelson & Hawthorne (2004). Re-Solving Irrelevant Conjunction with Probabilistic Independence, *Phil Sci* 71: 505–514. (2)

Gelman (2011). Induction and Deduction in Bayesian Data Analysis, *RMM* 2, 67–78. (10)

Gelman & Carlin (2014). Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors, *Perspectives on Psychological Science* 9, 641–51. (10)

Gelman & Hennig (2017). Beyond Subjective and Objective in Statistics, *Journal of the Royal Statistical Society: Series A* 180(4), 967–1033. (11) (14)

Gelman & Loken (2014). The Statistical Crisis in Science, *American Scientist* 2, 460–5. (3) (4)

Gelman & Shalizi (2013). Philosophy and the Practice of Bayesian Statistics (with discussion), *Brit. J. Math. Stat. Psy.* 66(1): 5–64. (11) (14)

Gigerenzer and Marewski (2017). Surrogate Science: The Idol of a Universal Method for Scientific Inference, *Journal of management* 41(2), 421–40. (11)

Gonick & Smith (1992). *The Cartoon Guide to Statistics*, HarperPerennial.

Goodman ([1993](#)). P-values, Hypothesis Tests, and Likelihood-Implications for Epidemiology of a Neglected Historical Debate, *American Journal of Epidemiology* 137(5), 485–96. (11)

Goodman ([1999](#)). Toward Evidence-Based Medical Statistics. 2: The Bayes Factor, *Annals of Internal Medicine*, 130(12), 1005–13. (9)

Greenland ([2012](#)). Nonsignificance Plus High Power Does Not Imply Support for the Null Over the Alternative, *Annals of Epidemiology* 22, 364–8. (12)

Greenland & Poole ([2013](#)). Living with P Values: Resurrecting a Bayesian Perspective on Frequentist Statistics and [Rejoinder](#): Living with Statistics in Observational Research, *Epidemiology* 24(1), 62–8; 73–8. Gelman [comment](#). (10)

Greenland, Senn, Rothman et al. ([2016](#)). Statistical Tests, P values, Confidence Intervals, and Power: A Guide to Misinterpretations, *European Journal of Epidemiology* 31(4), 337–50. (10)

Hacking ([1972](#)). Review: Likelihood, *The British Journal for the Philosophy of Science* 23(2), 132–7. (1)

Hacking ([1980](#)). The Theory of Probable Inference: Neyman, Peirce and Braithwaite, in Mellor, D. (ed.), *Science, Belief and Behavior: Essays in Honour of R. B. Braithwaite*, Cambridge: Cambridge University Press, pp. 141–60. (1) (2)

Haig, B. ([2016](#)). ‘Tests of Statistical Significance Made Sound’, *Educational and Psychological Measurement* 77(3) 489–506. (10)

Hardwicke and Ioannidis (forthcoming). Petitions in scientific argumentation: dissecting the request to retire statistical significance ([preprint](#)). *European Journal of Clinical Investigation*.

Howson ([1997](#)). A Logic of Induction, *Phil Sci*64(2): 268-290. (11) (14)

Howson ([2017](#)). Putting on the Garber Style? Better Not, *Philosophy of Science* 84(4), 659-76. (1)

Howson & Urbach ([1993](#)) Chapter 15, ([2006](#)) Chapter 5. *Scientific Reasoning: The Bayesian Approach*, 2nd & 3rd(Chapter 5) eds. Open court. (9)

Hubbard & Bayarri ([2003](#)). Confusion Over Measures of Evidence versus Errors and Rejoinder, *The American Statistician*57(3), 171-8; 181-2. (6) (7) (9)

Ioannidis ([2005](#)). Why most published research findings are false. *PLoS Med* 2(8): e124. (12)

Kadane ([2016](#)). Beyond Hypothesis Testing, *Entropy* 18(5), article 199, 1–5. (6) (7) (9)

Kass ([2011](#)). Statistical Inference: The Big Picture (with discussion and [rejoinder](#)), *Statistical Science* 26(1), 1–20. (11) (14)

Kass & Wasserman ([1996](#)). The Selection of Prior Distributions by Formal Rules, *Journal of the American Statistical Association* 91, 1343–70. (11) (14)

Lakens et al ([2018](#)) Justify Your Alpha *Nature Human Behaviour* 2, 168-71. (10)

Lambert & Black ([2012](#)). Learning From Our GWAS Mistakes: From Experimental Design to Scientific Method, *Biostatistics* 13(2), 195–203. (9)

Lehmann ([1993a](#)). ‘The Bertrand-Borel Debate and the Origins of the Neyman-Pearson Theory’, in Ghosh, J., Mitra, S., Parthasarathy, K. and Prak Ma Rao, L. (eds.), *Statistics and Probability: A Raghu Raj Bahadur Festschrift*, New Delhi: Wiley Eastern, 371–80. Reprinted in Lehmann 2012, pp. 965–74. (9)

Levelt Committee, Noort Committee, Drenth Committee ([2012](#)). Flawed Science: The Fraudulent Research Practices of Social Psychologist Diederik Stapel, *Stapel Investigation: Joint Tilburg/Groningen/Amsterdam investigation of the publications by Mr. Stapel* (commissielevelt.nl/). (3) (4)

Lindley ([2000](#)). The Philosophy of Statistics (with Discussion), *Journal of the Royal Statistical Society: Series D* 49(3), 293–337. (11) (14)

Mayo general [bibliography](#)

Mayo ([1996](#)). *Error and the Growth of Experimental Knowledge*, U of Chicago P.

Mayo ([1997](#)). Response to Howson and Laudan, *Phil Sci* 64(2): 323-333. (11) (14)

Mayo ([2003](#)). Commentary on J. Berger’s Fisher Address, *Stat Sci* 18: 19-24. (1), (5), (6) (7) (9)

Mayo ([2004](#)). An Error-Statistical Philosophy of Evidence in *The Nature of Scientific Evidence: Statistical, Philosophical & Empirical Considerations*. (Taper & Lele eds.), UCP: 79-118. (1)

Mayo ([2005](#)). Philosophy of Statistics in Sarkar & Pfeifer (eds.) *Philosophy of Science: An Encyclopedia*, Routledge: 802-815. (1)

Mayo ([2010b](#)). An Error in the Argument from Conditionality and Sufficiency to the Likelihood Principle (E&I: 305-14). (6) (7) (9)

Mayo (2010c). Sins of the Epistemic Probabilist: Exchanges with Achinstein (E&I: 189-201). (10)

Mayo (2010e). Learning from Error: The Theoretical Significance of Experimental Knowledge, The Modern Schoolman. Guest editor, Kent Staley. 87(3/4), (March/ May 2010). Experimental and Theoretical Knowledge, The Ninth Henle Conference in the History of Philosophy, 191–217.

Mayo (2013) Presented Version: On the Birnbaum Argument for the Strong Likelihood Principle. In JSM Proceedings, Section on Bayesian Statistical Science. Alexandria, VA: American Statistical Association, 440-453. (6) (7) (9)

Mayo (2013). Comments on A. Gelman and C. Shalizi, Brit. J. Math. Stat. Psy. 66(1): 57-64. (11) (14)

Mayo (2014). On the Birnbaum Argument for the Strong Likelihood Principle, (with discussion) Statistical Science 29(2) pp. 227-239, 261-266. (6) (7) (9)

Mayo (2016). Don't Throw Out the Error Control Baby with the Bad Statistics Bathwater: A Commentary on Wasserstein, R. L. and Lazar, N. A. 2016, The ASA's Statement on p-Values: Context, Process, and Purpose, The American Statistician 70(2) (supplemental materials). (1), (6) (7), (11) (14)

Mayo & Cox (2006). Frequentist Statistics as a Theory of Inductive Inference, Optimality: The Second Erich L. Lehmann Symposium (ed. J. Rojo), Lecture Notes-Monograph series, Institute of Mathematical Statistics (IMS), Vol. 49: 77-97. (5)

Mayo & Spanos (2004). Methodology in Practice: Statistical Misspecification Testing, Phil Sci 71: 1007-1025. (12)

Mayo & Spanos (2006). [Severe Testing as a Basic Concept in a Neyman-Pearson Philosophy of Induction](#), Brit. J. Phil. Sci., 57: 323-357. (5), (11)

Mayo & Spanos (eds) (2010). Error and Inference: Recent Exchanges on Experimental Reasoning, Reliability and the Objectivity and Rationality of Science, CUP. (E&I)

Mayo & Spanos (2011). Error Statistics in Philosophy of Statistics, Handbook of Philosophy of Science 7, Philosophy of Statistics, (Gabbay, Thagard & Woods (eds); Bandyopadhyay & Forster (Vol eds.)) Elsevier: 1-46. (1)

Mayo, Spanos & Staley (Guest eds.) (2011-2012): Rationality, Markets and Morals: Studies at the Intersection of Philosophy and Economics, (Albert, Kliemt, Lahno eds.). Special Topic: Statistical Science and Philosophy of Science: Where Do (Should) They Meet in 2011 and Beyond? (Complete collection of papers).

Meehl ([1978](#)). Theoretical Risks and Tabular Asterisks: Sir Karl, Sir Ronald, and the Slow Progress of Soft Psychology, *Journal of Consulting and Clinical Psychology* 46: 806-834. (3) (4)

Neyman, J. ([1934](#)). 'On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection', *The Journal of the Royal Statistical Society* 97(4), 558-625. Reprinted 1967 *Early Statistical Papers of J. Neyman*, 98-141.

Neyman ([1956](#)). Note on an Article by Sir Ronald Fisher, *J R Stat Soc(B)* 18: 288-294. (6) (7)

Neyman ([1957b](#)). The Use of the Concept of Power in Agricultural Experimentation, *Journal of the Indian Society of Agricultural Statistics* IX(1), 9-17.

Neyman ([1962](#)). Two Breakthroughs in the Theory of Statistical Decision Making, *Revue De l'Institut International De Statistique / Review of the International Statistical Institute*, 30(1), 11-27. (6) (7)

Neyman ([1976](#)). Tests of Statistical Hypotheses and Their Use in Studies of Natural Phenomena', *Communications in Statistics: Theory and Methods* 5(8), 737-51. (5)

Neyman ([1977](#)). Frequentist Probability and Frequentist Statistics, *Synthese* 36(1), 97-131. (9)

Neyman & Pearson ([1928](#)). On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference: Part I, *Biometrika* 20A(1/2), 175-240. Reprinted in *Joint Statistical Papers*, 1-66. (6) (7) (9)

Neyman & Pearson ([1933](#)) On the Problem of the Most Efficient Tests of Statistical Hypotheses, *Philosophical Transactions of the Royal Society of London Series A* 231, 289-337. Reprinted in *Joint Statistical Papers*, 140-85. (9)

Pearson ([1947](#)). The Choice of Statistical Tests Illustrated on the Interpretation of Data Classed in a 2 x 2 Table, *Biometrika* 34 (1/2), 139-167. Reprinted 1966 in *The Selected Papers of E. S. Pearson*, pp. 169-200. (5)

Pearson ([1955](#)). Statistical Concepts in Their Relation to Reality, *J R Stat Soc(B)* 17: 204-207. (6) (7)

Pearson & Chandra Sekar ([1936](#)). 'The Efficiency of Statistical Tools and a Criterion for the Rejection of Outlying Observations', *Biometrika* 28 (3/4), 308-20. Reprinted 1966 in *The Selected Papers of E. S. Pearson*, pp. 118-30. (9)

Pearson & Neyman ([1930](#)). 'On the Problem of Two Samples,' *Bulletin of the Academy of Polish Sciences*, 73-96. Reprinted 1966 in *Joint Statistical Papers*, 99-115. (1)

Peng, Dominici & Zeger (2006). Reproducible Epidemiologic Research American Journal of Epidemiology 163 (9), 783-789. (3) (4) (9)

Popper (1962). Conjectures and Refutations: The Growth of Scientific Knowledge. Basic Books. (3) (4)

Ratcliff & Oishi (2013). Gender Differences in Implicit Self-Esteem. Following a Romantic Partner's Success or Failure, Journal of Personality and Social Psychology 105(4), 688–702. (3) (4)

Reid & Cox (2015). 'On Some Principles of Statistical Inference', International Statistical Review 83(2), 293–308. (1) (11) (14)

SavageForum(1962) The Foundations of Statistical Inference: A Discussion, London: Methuen. (11) (14)

Senn (2001b). 'Two Cheers for P-values?' Journal of Epidemiology and Biostatistics 6(2), 193–204.

Senn (2002). 'A Comment on Replication, P-values and Evidence', S. N. Goodman, Statistics in Medicine 1992; 11:875-879', Statistics in Medicine 21(16), 2437–44. (10)

Senn (2011). You May Believe You Are a Bayesian But You Are Probably Wrong. RMM 2. (11) (14).

Simmons, Nelson & Simonsohn (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allow Presenting Anything as Significant, Psych. Sci., 22(11): 1359-1366.(1)

Simmons, Nelson & Simonsohn (2012). 'A 21 word solution', Dialogue: The Official Newsletter of the Society for Personality and Social Psychology 26(2), 4–7. (11)

Singh, Xie & Strawderman (2007). Confidence Distribution (CD) Distribution Estimator of a Parameter, IMS Lecture Notes–Monograph Series, Volume 54, Complex Datasets and Inverse Problems: Tomography, Networks and Beyond, pp. 132–50. (6) (7)

Spanos (2000). Revisiting Data Mining: "Hunting" with or without a License, Journal of Economic Methodology 7(2), 231–64.

Spanos (2007). Curve Fitting, the Reliability of Inductive Inference, and the Error-Statistical Approach, Philosophy of Science 74 (5), 1046-1066.

Spanos (2008a). Review of S. T. Ziliak and D. N. McCloskey's The Cult of Statistical Significance, Erasmus Journal for Philosophy and Economics 1(1), 154–64. (12)

Spanos ([2010a](#)). Akaike-type Criteria and the Reliability of Inference: Model Selection Versus Statistical Model Specification, *Journal of Econometrics* 158(2), 204–20. (12)

Spanos, A. ([2011b](#)). 'Foundational Issues in Statistical Modeling: Statistical Model Specification and Validation', *Rationality, Markets and Morals (RMM)* 2, 146–78.

Spanos ([2012](#)). Revisiting the Berger Location Model: Fallacious Confidence Interval or a Rigged Example? *Statistical Methodology*, 9, 555–61. (6) (7)

Spanos ([2013](#)). [Who Should Be Afraid of the Jeffreys-Lindley Paradox?](#) *Phi Sci* 80 (1):73-93. (8), (10)

Spiegelhalter ([2012](#)). Explaining 5 Sigma for the Higgs: How Well Did They Do?, Blogpost on Understandinguncertainty.org (8/7/2012).

Staley ([2017](#)). Pragmatic Warrant for Frequentist Statistical Practice: The Case of High Energy Physics, *Synthese* 194(2), 355–76 (6) (7)

Stapel ([2014](#)). *Faking Science: A True Story of Academic Fraud*. Translated by Brown, N. from the original 2012 Dutch *Ontsporing (Derailment)*. (3) (4)

Wagenmakers, ([2007](#)). A Practical Solution to the Pervasive Problems of P values, *Psychonomic Bulletin & Review* 14(5), 779–804. (9)

Wagenmakers & Grünwald ([2006](#)). A Bayesian Perspective on Hypothesis Testing: A Comment on Killeen (2005), *Psychological Science* 17(7), 641–2. (10)

Wagenmakers, Wetzels, Borsboom & van der Maas ([2011](#)). Why Psychologists Must Change the Way They Analyze Their Data: The Case of Psi: Comment on Bem (2011), *Journal of Personality and Social Psychology* 100, 426–32. (9)

Wasserstein & Lazar ([2016](#)). The ASA's Statement on P-values: Context, Process and Purpose, (and supplemental materials), *The American Statistician* 70(2), 129–33. (1) (6) (7) (11) (14)

Wasserstein, Schirm & Lazar ([2019](#)). Moving to a World Beyond " $p < 0.05$ ", *The American Statistician*, 73 sup 1, 1-19.

Zabell ([1992](#)). R. A. Fisher and Fiducial Argument, *Statistical Science* 7(3), 369–87. (6) (7)

These items are from the bibliography of [Statistical Inference as Severe Testing: How to Get Beyond the Statistics Wars](#) (Mayo 2018, CUP). Newer items will be added as we proceed.