

University of Plymouth
Cognition Institute conference
18.06.2018

Handout by
Christopher Germann
Marie Curie Fellow
PhD Candidate
CogNovo
Cognition Institute
University of Plymouth
www.cognovo.eu/christopher-germann
www.cognovo.eu/project-14



The seemingly never-ending tyranny of mindless null hypothesis testing rituals is a fundamental problem for psychology and the biomedical sciences which has far-reaching detrimental academic and societal ramifications.

What can we as a scientifically-minded community do to end it??? This university is unfortunately not responding to my numerous attempts to address this pivotal issue! (That is, my repeated attempts to communicate with the graduate school were simply ignored even though I invested a lot of time and effort in meticulously documenting the problem at hand while also providing potential solutions.) As a consequence, the same mindless statistical rituals are still handed-down to students (the next generation of researchers who will impact and shape society). The saga of ignorance continues...

Null hypothesis significance testing (NHST) is unfortunately still the predominant *modus operandi* in psychology and the biomedical sciences. However, research shows that the vast majority of researchers are unable to interpret even a simple *t*-test correctly (Gigerenzer, 1993, 2004; Haller & Krauss, 2002). That is, *p*-values (and associated confidence intervals (Hoekstra, Morey, Rouder, & Wagenmakers, 2014)) are constantly misinterpreted which leads to fallacious logical inferences which in turn seriously impede scientific progress and have far-reaching real-world ramifications. The unavoidable conclusion is: Psychologists need to change the way they analyse their data! The Aristotelian syllogistic logic which underlies null-hypothesis testing clearly shows that *p*-values have very limited informational value (Cohen, 1994, 1995) and the Popperian logic of scientific hypothesis testing (i.e., falsification (Popper, 1959)) is constantly violated by current research practices.

Research indicates that the “inverse probability problem” is the most persistent statistical illusion (wishful Bayesian thinking) which is perpetuated by the majority of researchers, editors, and even textbooks (Gigerenzer, 2004; Haller & Krauss, 2002; Loftus, 1996).

$p(D|H_0) \neq p(H_0|D)$

The psychology curriculum needs to foster “rational intelligence” (RQ≠IQ; Stanovich, 1999, 2012; Stanovich & West, 2014) and true statistical thinking skills — not uncritical conformity to “mindless mechanical statistical rituals” (Gigerenzer, 2004). This is also a matter of character (Fromm, 1976). Intrinsic motivation on part of the research community plays a crucial role. Research is not about publishing and making a career (publish or perish) but about real scientific progress. However, the current climate of “hypercompetition and perverse incentives” (Edwards & Roy, 2017) is antagonistic towards genuine intrinsically motivated science and rewards ego-driven and unethical behaviour, viz., intrinsic motivation and altruistic behaviour (e.g., authentic concerns regarding scientific progress and its impact on society) are not incentivised by the current system and its perverted reward structure.

Currently, the prevalence of statistical reporting errors in the field of psychology is staggering as metanalytic studies clearly show (Nuijten, Hartgerink, van Assen, Epskamp, & Wicherts, 2016). Psychology as a science needs to radically¹ change its methods otherwise it will lose all its credibility.

However, it does not have to be this way! There are numerous alternative (much more powerful and flexible) statistical methods like Bayes Factor analysis (both subjective and objective Bayes (Berger, 2006; Goldstein, 2006)) and Bayesian parameter estimation techniques using Markov chain Monte Carlo methods (Hastings, 1970). Preregistration of studies is another important step in the right direction (Nosek, Ebersole, DeHaven, & Mellor, 2018). In addition, full datasets (Boulton et al., 2012) and associated analysis scripts (Sakaluk, Williams, & Biernat, 2014) should be published whenever this is feasible. Instead of playing the game we need to change it! (Chambers, Feredoes, Muthukumaraswamy, & Etchells, 2014). Scientific integrity and non-conformist critical thinking à la Marie Curie are crucial scientific virtues in this context (see also Edwards & Roy, 2017). However, well-intended attempts to catalyse positive change are not without serious risks. Numerous social psychology indicate that whoever dares to challenge deeply-engrained group-norms and behaviours runs the risk of being marginalised, ostracised (the social death penalty), and punished in various ways (Williams, 2007). The study of group-dynamics is of great importance in this context. Neuroscientific fMRI studies demonstrate that social exclusion activates the same brain regions (viz., anterior cingulate cortex, right ventral prefrontal cortex) as does physical pain (Eisenberger, Lieberman, & Williams, 2003). However, social exclusion and other forms of psychological aggression do not leave any visible physical marks and are therefore difficult to objectively verify *post hoc*. However, social consequences can be much more devastating to the individual than physical violence. Furthermore, various forms of *ad hominem* arguments and other “microaggressions” are widely utilised ways to antagonise admonishers. Braking the spell of statistical illusions is thus no easy task and a real challenge for psychologists. However, the multifarious social risks outlined above do not stop all individuals from clearly pointing out that “the emperor wears no clothes!” (to employ Hans Christian Andersen's well-known parable), as the following commendable examples demonstrate:

“I suggest to you that Sir Ronald has befuddled us, mesmerized us, and led us down the primrose path. I believe that the almost universal reliance on merely refuting the null hypothesis is one of the worst things that ever happened in the history of psychology.” (Meehl, 1978, p. 817; Former President of the American Psychological Association, *inter alia*)

The eminent and highly influential statistician Jacob Cohen argues that null hypothesis significance testing „*not only fails to support the advance of psychology as a science but also has seriously impeded it.*“ (Cohen, 1997, p. 997; * 1923; † 1998; Fellow of the American Association for the Advancement of Science, *inter alia*)

¹ The term “radical” is etymologically derived from the Latin word “radix” meaning “root” (cf. the radical sign $\sqrt{\quad}$ in mathematics). That is, radical change means to “change from the roots”).

“Few researchers are aware that their own heroes rejected what they practice routinely. Awareness of the origins of the ritual and of its rejection could cause a virulent cognitive dissonance, in addition to dissonance with editors, reviewers, and dear colleagues. Suppression of conflicts and contradicting information is in the very nature of this social ritual.” (Gigerenzer, 2004, p. 592; Director Emeritus of the Center for Adaptive Behavior and Cognition at the Max Planck Institute for Human Development, *inter alia*)

References

- Berger, J. O. (2006). The case for objective Bayesian analysis. *Bayesian Analysis*, *1*(3), 385–402. <https://doi.org/10.1214/06-BA115>
- Boulton, G., Campbell, P., Collins, B., Elias, P., Hall, W., Graeme, L., ... Walport, M. (2012). *Science as an open enterprise*. *Science*. [https://doi.org/ISBN 978-0-85403-962-3](https://doi.org/ISBN%20978-0-85403-962-3)
- Chambers, C. D., Feredoes, E., Muthukumaraswamy, D., & Etchells, J. (2014). Instead of “playing the game” it is time to change the rules: Registered Reports at AIMS Neuroscience and beyond. *AIMS Neuroscience*, *1*(1), 4–17. <https://doi.org/10.3934/Neuroscience.2014.1.4>
- Cohen, J. (1994). The Earth Is Round ($p < .05$). *American Psychologist*, *49*(12), 997–1003. <https://doi.org/https://doi.org/10.1037/0003-066X.49.12.997>
- Cohen, J. (1995). The earth is round ($p < .05$): Rejoinder. *American Psychologist*, *50*(12), 1103–1103. <https://doi.org/10.1037/0003-066X.50.12.1103>
- Dunjko, V., & Briegel, H. J. (2017). Machine learning & artificial intelligence in the quantum domain. *ArXiv E-Prints*, 1709.02779. Retrieved from <https://arxiv.org/pdf/1709.02779.pdf%0Ahttp://arxiv.org/abs/1709.02779>
- Edwards, M. A., & Roy, S. (2017). Academic Research in the 21st Century: Maintaining Scientific Integrity in a Climate of Perverse Incentives and Hypercompetition. *Environmental Engineering Science*, *34*(1), 51–61. <https://doi.org/10.1089/ees.2016.0223>
- Eisenberger, N. I., Lieberman, M. D., & Williams, K. D. (2003). Does rejection hurt? An fMRI study of social exclusion. *Science*. <https://doi.org/10.1126/science.1089134>
- Fromm, E. (1976). *To Have Or To Be?* Bloomsbury. Retrieved from <https://books.google.co.uk/books?id=VUVMAQAAQBAJ>
- Gigerenzer, G. (1993). The superego, the ego, and the id in statistical reasoning. In *A handbook for data analysis in the behavioral sciences* (pp. 311–339). <https://doi.org/10.1017/CBO9780511542398>
- Gigerenzer, G. (2004). Mindless statistics. *Journal of Socio-Economics*, *33*(5), 587–606. <https://doi.org/10.1016/j.socec.2004.09.033>
- Goldstein, M. (2006). Subjective Bayesian Analysis : Principles and Practice Applied subjectivism. *Bayesian Analysis*, *1*(3), 403–420.
- Gottesman, D., & Chuang, I. L. (1999). Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations. *Nature*, *402*(6760), 390–393. <https://doi.org/10.1038/46503>

- Haller, H., & Krauss, S. (2002). Misinterpretations of significance: a problem students share with their teachers? *Methods of Psychological Research Online*, 7(1), 1–20. <https://doi.org/http://www.mpr-online.de>
- Handsteiner, J., Friedman, A. S., Rauch, D., Gallicchio, J., Liu, B., Hosp, H., ... Zeilinger, A. (2017). Cosmic Bell Test: Measurement Settings from Milky Way Stars. *Physical Review Letters*, 118(6). <https://doi.org/10.1103/PhysRevLett.118.060401>
- Hastings, W. K. (1970). Monte carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1), 97–109. <https://doi.org/10.1093/biomet/57.1.97>
- Hoekstra, R., Morey, R. D., Rouder, J. N., & Wagenmakers, E.-J. (2014). Robust misinterpretation of confidence intervals. *Psychonomic Bulletin & Review*, 21(5), 1157–1164. <https://doi.org/10.3758/s13423-013-0572-3>
- Loftus, G. R. (1996). Psychology will be a much better science when we change the way we analyze data. *Current Directions in Psychological Science*, 5(6), 161–171. <https://doi.org/10.1111/1467-8721.ep11512376>
- Lum, D. J., Howell, J. C., Allman, M. S., Gerrits, T., Verma, V. B., Nam, S. W., ... Lloyd, S. (2016). Quantum enigma machine: Experimentally demonstrating quantum data locking. *Physical Review A*. <https://doi.org/10.1103/PhysRevA.94.022315>
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, 201708274. <https://doi.org/10.1073/pnas.1708274114>
- Nuijten, M. B., Hartgerink, C. H. J., van Assen, M. A. L. M., Epskamp, S., & Wicherts, J. M. (2016). The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior Research Methods*, 48(4), 1205–1226. <https://doi.org/10.3758/s13428-015-0664-2>
- Pirandola, S., & Braunstein, S. L. (2016). Physics: Unite to build a quantum Internet. *Nature*. <https://doi.org/10.1038/532169a>
- Popper, K. R. (1959). The logic of scientific discovery. *London: Hutchinson*, 268(3), 244. [https://doi.org/10.1016/S0016-0032\(59\)90407-7](https://doi.org/10.1016/S0016-0032(59)90407-7)
- Ren, J. G., Xu, P., Yong, H. L., Zhang, L., Liao, S. K., Yin, J., ... Pan, J. W. (2017). Ground-to-satellite quantum teleportation. *Nature*, 549(7670), 70–73. <https://doi.org/10.1038/nature23675>
- Sakaluk, J., Williams, A., & Biernat, M. (2014). Analytic Review as a Solution to the Misreporting of Statistical Results in Psychological Science. *Perspectives on Psychological Science*, 9(6), 652–660. <https://doi.org/10.1177/1745691614549257>
- Stanovich, K. E. (1999). Who Is Rational? Studies of Individual Differences in Reasoning. *Pp. Xvi*, 1–296. https://doi.org/10.1207/S15327566IJCE0404_5
- Stanovich, K. E. (2012). On the Distinction Between Rationality and Intelligence: Implications for Understanding Individual Differences in Reasoning. In *The Oxford Handbook of Thinking and Reasoning*. <https://doi.org/10.1093/oxfordhb/9780199734689.013.0022>
- Stanovich, K. E., & West, R. F. (2014). The Assessment of Rational Thinking: IQ ≠ RQ. *Teaching of Psychology*. <https://doi.org/10.1177/0098628314537988>
- Williams, K. D. (2007). Ostracism. *Annual Review of Psychology*, 58(1), 425–452.

