Yes, but Don't Underestimate Estimation

Reply to Morey, Rouder, Verhagen, and Wagenmakers (2014)

Fiona Fidler¹, Geoff Cumming²

¹Australian Centre of Excellence for Environmental Decisions (CEED), The University of Melbourne

²School of Psychological Science, La Trobe University

Morey, Rouder, Verhagen, and Wagenmakers (2014) make a useful contribution by emphasizing that hypothesis testing has an important role in science, quite apart from null-hypothesis significance testing, which they agree should not be used. We are encouraged that they describe estimation as essential (p. 1289), and part of what is “becoming the conventional wisdom in psychology” (p. 1289), even if a touch of irony is intended in that wording. However, we do not agree that “the benefits of estimation have been overstated” (p. 1289).

Estimation can contribute in many ways, including to model evaluation and theory development. Morey et al. state that “hypothesis testing, not estimation, is necessary for testing the quantitative predictions of theories” (p. 1290). We agree that testing such predictions is necessary, but any approach to assessing a quantitative prediction presumably requires comparison with a relevant quantitative statistic derived from data, which is an estimate by another name. Morey et al. need to explain how what they refer to as hypothesis testing can assess quantitative predictions without some contribution from estimation. If estimation cannot contribute, their concept of hypothesis testing may be narrower than ours.

Morey et al. make several bold, prescriptive statements—in addition to this one—about the way science (including hypothesis testing and theory development) should operate. They have little space to expand on their philosophy of science or definition of hypothesis testing, but without more justification, their statements are unconvincing. Cumming (2014, p. 21) referred with approval to a study by Velicer et al. (2008), who used estimation to evaluate the transtheoretical model of behavior change. Velicer et al. judged that 11 of 15 predictions falling within 95% confidence intervals based on data provided overall support for the model.

Morey et al. are critical of the arbitrariness of basing the conclusion on 11 of 15 correct predictions and also of the neglect of results expected if the model under test were false. However, judgments about the strengths and weaknesses of a model are inevitably and properly somewhat subjective and therefore somewhat arbitrary. Velicer et al. considered the transtheoretical model without explicit reference to any alternative because no other model with anything like the same promise has been proposed. In our view, they made appropriate use of estimation to evaluate the model and guide its further development. Such knowledgeable judgment in context is a necessary part of any evaluation of a model against data. For example, Bayesian model comparison, which Morey et al. advocate (p. 1290), uses the Bayes factor to quantify the relative strength of evidence that a data set provides for two competing models. Researchers need to judge how Bayes factor values are used to guide conclusions—with or without reference to suggested benchmarks for which values of the

Bayes factor provide “barely worth mentioning,” “substantial,” or “strong” evidence for one model over another (Jeffreys, 1961, p. 432). In fact, an often-cited advantage of Bayesian methods is that they make salient the subjectivity inherent in scientific inference (Berger & Berry, 1988; Wagenmakers, Lee, Lodewyckx, & Iverson, 2008). Similarly, the information

theoretic approach of Burnham and Anderson (2002), mentioned by Morey et al. (p. 1290), provides Akaike information criterion (AIC) values that allow a quantitative comparison of two or more models, given a set of data. In this case, also, researchers need to take knowledgeable account of the full research context when they decide what conclusions are justified by particular AIC values.

Estimation also often contributes via meta-analysis. Testing the hypothesis that ESP exists is perhaps the strongest example of a case in which effect size is relatively unimportant: Convincing evidence of any nonzero ESP effect would be a dramatic finding. Yet evaluation of the evidence requires integration over studies, and metaanalysis based on point and interval estimates is typically at the center of such evaluation. Even if researchers are interested only in a dichotomous test of a hypothesis, they should report estimation information so that their results can be included in a future meta-analysis.

Estimation can also play an important role in theory development. Meta-analysis often includes moderator analysis, which examines the possible contribution of moderator variables to variation in estimates found by individual studies. Such analysis can contribute to theoretical advance by identifying variables to be included in better models. Cumming (2012, p. 249) cited the example of Wilson and McKenzie (1998), who used meta-analysis to identify important but previously little-recognized variables in the study of clumsy children. Our meta-analysis examples are frequentist, but we expect that any quantitative approach to integration of evidence over studies needs estimates of some kind from the individual studies. Morey et al. allow that “estimation is necessary in pre theoretical work before clear predictions can be made, and is also necessary in post theoretical work for theory revision” (p. 1290).

Given that wide brief for estimation, and the contributions that estimation can make (as we have just described), we are puzzled why Morey et al. see such a limited role for estimation. We also wonder why Morey et al. make such a sharp distinction between pretheoretical and post theoretical work, given ongoing philosophy-of-science debates about such things. They are declared Bayesians, but what Bayesian philosophy makes such a sharp distinction and rejects subjectivity?

To progress, psychological science needs improved theories that are more quantitative and are always being tested against empirical findings, using whatever methods give greatest insight. We agree with Morey et al. that model comparison and evaluation are centrally important.

We contend, however, that estimation can play a valuable role in such model development and, more generally, in theoretical as well as empirical progress, regardless of the underlying philosophical foundation of one’s methods.

Author Contributions F. Fidler and G. Cumming jointly wrote and revised the manuscript. Both authors approved the final version for submission.

Declaration of Conflicting Interests The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Reference

- References Berger, J. O., & Berry, D. A. (1988). Statistical analysis and the illusion of objectivity. *American Scientist*, 76, 159–165.
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information theoretic approach* (2nd ed.). New York, NY: Springer.
- Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York, NY: Routledge.
- Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, 25, 7–29. doi:10.1177/0956797613504966
- Jeffreys, H. (1961). *The theory of probability* (3rd ed.). Oxford, England: Oxford University Press.
- Morey, R. D., Rouder, J. N., Verhagen, J., & Wagenmakers, E.-J. (2014). Why hypothesis tests are essential for psychological science: A comment on Cumming (2014). *Psychological Science*, 25, 1289–1290.
- Velicer, W. F., Cumming, G., Fava, J. L., Rossi, J. S., Prochaska, J. O., & Johnson, J. (2008). Theory testing using quantitative predictions of effect size. *Applied Psychology: An International Review*, 57, 589–608. doi:10.1111/j.1464-0597.2008.00348.x
- Wagenmakers, E.-J., Lee, M. D., Lodewyckx, T., & Iverson, G. (2008). Bayesian versus frequentist inference. In H. Hoijtink, I. Klugkist, & P. A. Boelen (Eds.), *Bayesian evaluation of informative hypotheses in psychology* (pp. 181–207). New York, NY: Springer.
- Wilson, P. H., & McKenzie, B. E. (1998). Information processing deficits associated with developmental coordination disorder: A meta-analysis of research findings. *Journal of Child Psychology and Psychiatry*, 39, 829–840.