



# Robust Diversity in Cognitive Science

Joachim Vandekerckhove<sup>1</sup> · Corey N. White<sup>2</sup> · Jennifer S. Trueblood<sup>3</sup> · Jeffrey N. Rouder<sup>1</sup> · Dora Matzke<sup>4</sup> · Fábio P. Leite<sup>5</sup> · Alexander Etz<sup>1</sup> · Christopher Donkin<sup>6</sup> · Berna Devezer<sup>7</sup> · Amy H. Criss<sup>8</sup> · Michael D. Lee<sup>1</sup>

Published online: 9 October 2019

© Society for Mathematical Psychology 2019

## Abstract

The target article on robust modeling (Lee et al. [in review](#)) generated a lot of commentary. In this reply, we discuss some of the common themes in the commentaries; some are simple points of agreement while others are extensions of a practical or abstract nature. We also address a small number of disagreements or confusions.

**Keywords** Cognitive modeling · Reproducibility · Open science · Robustness · Model comparison

## Points of Broad Agreement

A number of commentaries reinforce arguments made in the target article, providing some reassuring robustness that other cognitive modelers have had similar experiences. Emmery et al. ([in review](#)) echo our warnings about over-reliance on benchmarks from their machine learning perspective. Kennedy et al. ([in review](#)) use prior predictive

checks to test model predictions in ways that we agree are principled and powerful, and join us in emphasizing generalization and prediction. Crüwell et al. ([in review](#)) develop categories of modeling goals, and map them onto approaches to robustness that we think are compatible with the target article.

## Conceptual Analysis and Extensions

Right off the bat, the commentaries raise interesting basic questions. De Boeck et al. ([in review](#)) ask whether all this means we think cognitive modeling is in crisis, and Gluck ([in review](#)) asks us to define what we mean by “robust.”

## Where is the Crisis?

To answer (De Boeck et al. [in review](#)): We do not believe cognitive modeling is *currently* in crisis, but we think it can always do better. If a neighbor’s house is found to rest on shaky ground, it is natural and probably wise to start thinking about one’s own foundations. We view recent open-science developments in psychology and other sciences as a useful starting point to explore whether and how we might similarly extend and improve the cognitive modeling enterprise.

## Functional Specification

We greatly appreciate Gluck’s ([in review](#)) comments about the term “robust” (see also Gluck et al. 2012). Indeed, robustness implies the presence of a *function* (something that is to

---

This article is the work of the authors and is reformatted from the original, which was published under a CC-BY Attribution 4.0 license and is available at <https://psyarxiv.com/6bu7s/>.

---

✉ Joachim Vandekerckhove  
joachim@uci.edu

<sup>1</sup> Department of Cognitive Sciences, University of California, Irvine, CA 92697, USA

<sup>2</sup> Department of Psychology, Missouri Western State University, Saint Joseph, MO 64507, USA

<sup>3</sup> Department of Psychology, Vanderbilt University, Nashville, TN 37235, USA

<sup>4</sup> Department of Psychology, University of Amsterdam, Amsterdam, Netherlands

<sup>5</sup> Department of Psychology, The Ohio State University at Lima, Lima, OH 45804, USA

<sup>6</sup> School of Psychology, University of New South Wales, Kensington, Australia

<sup>7</sup> Department of Business, University of Idaho, Moscow, ID 83844, USA

<sup>8</sup> Department of Psychology, Syracuse University, Syracuse, NY, USA

be left intact) and a *perturbation* (something that threatens the integrity of the function). The function of modeling is a topic of some discussion in these special issues—it may be to generalize, to explain, to translate to practice, to monetize, and so on. And, to differing degrees, these functions may be robust against certain perturbations, while being sensitive to others. Certainly, we want generalization and translation to be robust against, say, the identity of the lead researcher; but we want them to be sensitive to the data that are obtained. *Functional specification* would involve documenting the teleology of a model as well as what boundary conditions exist on its use. We appreciate that the interpretation of robustness can be diverse and that different functional specifications will naturally lead to different robust practices.

As a potential narrow specification, a function of cognitive modeling could be to draw generalizable conclusions about the mind. Some common perturbations threatening that function are undocumented differences between labs, researchers' habits, locations, time, and other sources of variability that are not expected to moderate any conclusions but that nevertheless sometimes occur when researchers in experimental psychology try to replicate published results (Baribault et al. 2018; Dutilh et al. 2017; Silberzahn et al. 2018; Steegen et al. 2016). Robustness, in that narrow sense, is the desirable quality of a method or workflow that generates conclusions that are likely to replicate on repetition.

However, the field of cognitive modeling is broad and diverse and will certainly benefit from functional specifications that are similarly broad and diverse. For example, Gunzelmann (in review) points out that even pre- and post-registration will not solve some debates in cognitive modeling because the bounds on the phenomena to be explained are themselves not clear. Szollosi and Donkin (in review) raise the related problem of theoretical flexibility, which may similarly frustrate scientific progress even in the idealized case where all studies are preregistered. De Boeck et al. (in review) point out that robustness failures can be inherent in investigating non-robust phenomena. In all of these cases, functional specification—that is, being explicit about what one is trying to achieve—will undoubtedly help.

Kellen (in review) suggests a robustness whose function extends beyond statistical modeling and inference issues, to more complete accounts of the modeling process. Kellen (in review) draws upon seminal work by Suppes (1966) to suggest a richer framework that involves an interacting hierarchy of modeling practices that characterize the enterprise more completely. It is fair to note that our scope was not as ambitious as it could have been in that regard, and was focused on fairly concrete modeling practices.

## Solution-Oriented Modeling

Our target article discussed at some length the concept of *solution-oriented modeling* (Watts 2017). Not everyone agrees that solution-oriented science is worth pursuing. Indeed, some commenters are of the opinion that formal modeling, or even most of science, is by definition exploratory, creative, and difficult to plan (De Boeck et al. in review; Van Zandt and MacEachern in review; Lilburn et al. in review; Shiffrin in review).

On the extreme end of that spectrum, Shiffrin's "Misunderstanding the goal of modeling" seems rooted in the idea that the *only* valid goal for cognitive modeling is exploration and modeling is limited to the context of discovery. While we have no strong opinions on the precise ideal balance between the two modes of scientific research, we do believe that a critical component of science is the pursuit of confirmation and the context of justification.<sup>1</sup> It is of course every researcher's prerogative to be interested only in the *making* of claims, but we fundamentally disagree with Shiffrin (in review) when he writes that *confirming claims is not a goal of model-based analysis*. The ability to confirm a claim by successful prediction, and by demonstrating that it holds in repeated samples under predictable circumstances, seems a useful one when addressing practical issues.

Collectively, the rest of the commentaries make clear that there is no consensus about how important it is to develop and evaluate cognitive models in the context of real-world problems. On the one hand, Szollosi and Donkin (in review) appear to see the goal of cognitive modeling primarily in terms of improving theories, and leave little room for applications. On the other hand, Wilson et al. (in review) advocate for applications, working through what is needed for cognitive models to contribute to improving society and solving real-world problems, including the need for a full pipeline of reproducibility in the real-world situation. Neufeld and Cutler (in review) emphasize real-world possibilities afforded by clinical cognitive modeling, especially with regard to strong prediction tests about individual differences. Crüwell et al. (in review) agree that model application could benefit from porting over open-science practices. Starns et al. (in review) appear to be in wholehearted agreement that practical applications matter. Cox (in review) envisages a future repository of models, indexed by their function, and available for application by researchers, industry professionals, and others.

<sup>1</sup>"Discovery and justification are temporally distinct processes: At the beginning, something is discovered. Subsequently, it is justified" (Hoyningen-Huene 2006, p. 120).

## Conceptual Transparency

The issues of functional specification and solution-oriented modeling are, to our minds, tightly related. Not only Gluck's (in review) commentary but also those by Cox (in review), Szollosi and Donkin (in review), Gunzelmann (in review), De Boeck et al. (in review), and Kellen (in review), and Heathcote (in review) strengthen our belief that functional specification (e.g., that a modeling exercise is addressing a practical problem) has a role to play in the pre- and post-registration of cognitive models. Many of the critiques may be addressed by the conceptual clarity that functional specification provides. We think that functional specification is a potentially important addition to the ideas in the target article.

## Conceptual Extensions

Another line of conceptual extension is to understand the interplay between confirmatory and exploratory approaches. Buzbas (in review) suggests working toward a more formal framework for understanding pre- and post-data model adjustment, and makes the case that it can be valid to do some sorts of confirmatory inference when the analysis is data-dependent. More generally, Buzbas (in review) makes a welcome appeal for close ties with formal statistics. Heathcote (in review) also sees models as having inherently interacting exploratory and confirmatory aspects. Lilburn et al. (in review) argue for the very real possibility that the distinction between confirmatory and exploratory approaches will become over-simplified, and make a compelling case for preserving nuance.

## Practical Refinements and Extensions

Many commentaries seem in broad agreement with our major points, but raise challenges or suggest solutions when it comes to the details. For example, several commentaries discuss the details of pre- and post-registration. Crüwell et al. (in review) and Lilburn et al. (in review) point out that there is work to be done fleshing out the details of exactly how lab book and similar ideas be implemented. Palmeri (in review) and De Boeck et al. (in review), and Heathcote (in review), among others, raise the issue of the level of detail that is appropriate for post-registration in particular. They are right that there is always a balance between effort and reward (and, e.g., the didactic value of a post-registration will be difficult to measure), and it will be interesting to see where the field finds the appropriate balance. Heathcote (in review) puts it well: “any scheme to prescribe modeling research practices needs to be mindful of the compliance (and hence opportunity) cost” (p. xx).

Despite these reservations, there seems to be broad agreement that cognitive modeling will benefit from the recent upsurge in sharing, documenting, and curating data and code. But there will certainly be limits: Emmerly et al. (in review) give sobering lessons from the neighboring field of machine learning relating how hard it can be to put desirable developments like reproducible and interpretable code into pervasive practice (see also Stodden et al. 2016).

## Even Better Practices Before Data are Collected

Vanpaemel (in review) extends our emphasis on pre-registering predictions to pre-registering risky predictions. We completely agree and think the Bayesian framework of data priors, which essentially functions to find critical tests of theories and models, is especially apposite for our purposes. Kennedy et al. (in review) discuss in detail the use of prior predictive checks, which are a principled tool for expressing the exact predictions of a model and a powerful method for pre-registering model comparison plans.

Heck and Ertfelder (in review) and Pitt and Myung (in review) thoroughly discuss methods that optimize the expected information gain from data collection, which we believe are powerful and useful for model comparison, especially when data are not cheap to collect.

Starns et al. (in review) expand on the practice of blinded analysis (see also Dutilh et al. 2019), of which there are now a few worked examples in the literature (e.g., Dutilh et al. 2017; Morey et al. 2019; Starns et al. in press).

## Even Better Practices After Data are Collected

As argued by Blaha (in review), data visualization and visual contact between data and models is a central component of model-based data analysis as well as model development. Especially as access to data becomes more democratic—and not everyone who is charged to interpret data is necessarily trained in quantitative methods—graphical methods will become more influential. The rise of data visualization as an area of research in its own right is exciting.

## Openness, Transparency, and Reproducibility

Poldrack et al. (in review) emphasize openness and transparency with a worked example coming from the Brain Imaging Data Structure community in cognitive neuroscience. They argue that “the transparent sharing of model specifications, including their inputs and outputs, is also essential to improving the reproducibility of model-based analyses.” In a similar vein, Broomell et al. (in review) encourage registering details of stimuli, and Heathcote encourages broader registered documentation

beyond the modeling code and behavioral data needed for the associated paper.

Here, too, these practices can serve different functions (error checking and reproducibility, improve public confidence, encourage broad research participation) and fortify against different threats.

## Disagreements and Confusions

In a few cases, we simply disagree with the commenters in limited ways.

### Confirmatory does not prevent Exploratory

A broad concern engendered by the target article related to the relative merits of confirmatory and exploratory research. We emphasized new ideas that involved confirmatory approaches, especially in the form of pre-registration.

Many commentaries provided excellent general arguments (and specific examples) of the foundational role exploratory research plays in theory and model development. Van Zandt and MacEachern ([in review](#)) emphasize “the subjective and exploratory nature of model development” (p. xx). Neufeld and Cutler ([in review](#)) and Shiffrin ([in review](#)) emphasize the merits of data-driven research based on abductive reasoning, in which models are built to be descriptively adequate accounts of available data. Palmeri ([in review](#)) relates a clear and instructive case study regarding the predictive failure of one model of category learning leading to the exploratory adequacy and subsequent robustness of another.

Lilburn et al. ([in review](#)) argue that the dichotomy between exploration and confirmation implies that confirmation is to be preferred, but good practice is to mix exploration and confirmation. We believe the latter opinion is well in line with our proposed practice of “complete modeling,” but we do not share Lilburn et al.’s ([in review](#)) concern that exploratory modeling would evaporate if a new branch of confirmatory modeling should arise. All of the authors of the target article engage in exploratory modeling all of the time, and recognize its fundamental value. We have no intent to criticize exploratory approaches to modeling; we certainly agree that exploratory research should never be considered “merely exploratory” (an expression used by both Lilburn et al. [in review](#), p. xx, and Palmeri [in review](#), p. xx); and we believe that anyone who wants to engage exclusively in exploratory research should be free to do so confidently. We think it is possible to introduce ideas like pre-registration that bolster confirmatory research practices in cognitive modeling without diluting the exploratory approach to research on which the field has been built and continues to thrive. Likewise, post-registration can bolster

exploratory model development by reporting paths taken during the exploration.

However, we also believe some commenters underestimate the practical feasibility of confirmatory modeling. To give an example, Van Zandt and MacEachern ([in review](#)) extensively refer to Yu et al. (2011) as exemplary of the fundamentally creative, exploratory, even unpredictable nature of model development. But consider the steps taken in that article from the point of view of the three analysts. First, they are handed a small data set, to which they ply their art and generate, in a creative fashion, an appropriate model. *Then they provide to a neutral third party a detailed description of their modeling strategy and the resulting model.* Finally, their model is subjected to a larger data set and evaluated on its out-of-sample performance. This does not strike us as incompatible with confirmatory research so much as an excellent example of a worked Registered Modeling Report *avant la lettre*.

### Power vs. Planning

Perhaps more than any other of our comments, our dismissal of power analysis drew explicit criticism. Our dismissal should not be construed too broadly; elsewhere in the target article, we recommend prospective analyses such as parameter recovery simulation to establish sufficient resolution and optimal experimental design analysis to determine appropriate sample sizes. Narrowly defined, in the context of null hypothesis testing and prior to obtaining the data, power analysis is a fine tool (Gluth and Jarecki [in review](#)). More broadly, several articles in these special issues discuss design optimization (Heck and Ertfelder [in review](#); Pitt and Myung [in review](#)) and other prospective analyses (Kennedy et al. [in review](#); Vanpaemel [in review](#)) that encompass or supersede classical power analysis.

However, as far as (narrow) power analysis goes, we felt it necessary to emphasize the select areas of application because its importance is often overstated beyond its appropriate niche role. For example, “post hoc” power analyses are now sometimes requested by reviewers or required by journal policies (e.g., *Journal of Experimental Psychology: Human Perception and Performance* 2019) in order to determine, retrospectively, whether a completed study would have had the prospective power to detect the effect it did, in fact, detect. Such fallacious calculations of power for data analysis are at best uninformative and more likely misleading to readers and reviewers and we do not recommend them (Gelman 2018; Hoenig 2001).

Our recommendations regarding good practices before data are collected notwithstanding, even if a researcher did not use *any* method to determine their sample size (because they did not think to, or because they had no control over the size of the sample; and because their

data collection plan did not undergo peer review), their data are just as valid and their conclusions should be considered equally. After the data are collected, all that matters are the data and the models. The misconception that the expected informativeness of a study (sometimes expressed in a prospective power analysis) is indicative of the actual informativeness of the study as realized, that is, the “misconception that what holds on average—across an ensemble of hypothetical experiments—also holds for each case individually” (Wagenmakers et al. 2015, p. 913) is sometimes termed the *power fallacy*.

### Qualitative and Quantitative Fit

A point on which we think there is surface confusion, but underlying agreement, relates to incorporating qualitative features into model evaluation and comparison. For example, Shiffrin’s (in review) commentary argues that we over-emphasized quantitative fitting of models to data at the expense of more qualitative evaluations that are relevant to identifying the causal factors at the heart of effective theories and models. In the target article, we discussed the use of more general utility functions to compare models with data, allowing a capability to reward or penalize models in terms of the connection between key theoretical predictions and qualitative properties of the data. But where we see qualitative and quantitative evaluations as two ends on a continuum of utilities for scoring models to data, Crüwell et al. (in review) argue that “these different forms of assessment reflect fundamentally different goals of implementing models” (p. xx).

Kellen (in review) advocates for more flexible evaluation criteria of this sort, capable of answering questions like “What are the diverging predictions and how do they connect with the theoretical claims made in each model?” (p. xx). He also emphasizes that how to measure depends on goals and that goals may shift the desired balance between fit and parsimony. A similar argument is made by Cox (in review), who places an emphasis on functions such as measurement models and substantive models.

In general, we believe there is underlying agreement that assessing models based on qualitative patterns is both sensible and worthwhile above and beyond quantitative model fitting (see also Navarro 2019).

### Conclusion, for Now

Our target article aimed to use the crisis of confidence in experimental psychology as a catalyst to think about ways in which cognitive modeling might be more robust. The discussion surrounding the article has played out in many

different venues—conference symposia, social media, and now in two back-to-back special issues of *Computational Brain & Behavior*—and has become more vibrant and more broadly based than we had ever imagined.

Several of the commentaries propose more detailed, or better, practical recommendations. Others provide thoughtful conceptual analysis and extensions. Yet others pose interesting questions about the goals and application areas of model-based analysis. It is clear that this rich discussion stems from the large variation in modeling approaches, the differing goals of modeling exercises, and the wide diversity of researcher backgrounds that exist in the field and that are represented among the many contributors. Such diversity in goals and approaches—balancing discovery with justification, model building with model testing, and encompassing models with measurement models—is a major strength of cognitive modeling (Devezer et al. 2019).

Our main takeaway from this debate is set up nicely by Gluck’s (in review) observation that a determination of robustness of a practice requires a function of that practice. It is clear that cognitive modeling, like much of science in general, has multiple potential functions. Lucid specification of one’s goals is hence a prerequisite for determining whether certain methodological practices are robust or rather brittle. In that vein, we hope this pair of special issues is only the start of a growing compendium of good practices in cognitive modeling.

**Funding Information** This article is the product of the Workshop on Robust Social Science held in St. Petersburg, FL, in June 2018. The workshop was made possible by generous funding from the National Science Foundation (grant #BCS-1754205) to Joachim Vandekerckhove and Michael Lee of the University of California, Irvine. Alexander Etz was supported by NSF GRFP #DGE-1321846. Berna Devezer was supported by NIGMS of the NIH under award #P20GM104420. Dora Matzke was supported by a Veni grant (#451-15-010) from the Netherlands Organization of Scientific Research (NWO). Jennifer Trueblood was supported by NSF #SES-1556325. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the National Science Foundation.

### References

- Baribault, B., Donkin, C., Little, D.R., Trueblood, J.S., Oravecz, Z., van Ravenzwaaij, D. (2018). Metastudies for robust tests of theory. *Proceedings of the National Academy of Sciences*, *115*, 2607–2612.
- Blaha, L. (in review). We have not looked at our results until we have displayed them effectively. *Computational Brain & Behavior*.
- Broomell, S.B., Sloman, S., Blaha, L.M., Chelen, J. (in review). Interpreting model comparison requires understanding model-stimulus relationships. *Computational Brain & Behavior*.
- Buzbas, E. (in review). Need of mathematical formalism in proposals for robust modeling. *Computational Brain & Behavior*.

- Cox, D.J. (in review). The many functions of quantitative modeling. *Computational Brain & Behavior*.
- Crüwell, S., Stefan, A.M., Evans, N.J. (in review). Robust standards in cognitive science. *Computational Brain & Behavior*.
- De Boeck, P., Jeon, M., Gore, L. (in review). Beyond registration pre and post. *Computational Brain & Behavior*.
- Devezer, B., Nardin, L.G., Baumgaertner, B., Buzbas, E.O. (2019). Scientific discovery in a model-centric framework: reproducibility, innovation, and epistemic diversity. *PLOS ONE*, *14*(5), e0216125.
- Dutilh, G., Sarafoglou, A., Wagenmakers, E.-J. (2019). Flexible yet fair: blinding analyses in experimental psychology. (via osf.io/d79r8).
- Dutilh, G., Vandekerckhove, J., Ly, A., Matzke, D., Pedroni, A., Frey, R. (2017). A test of the diffusion model explanation for the worst performance rule using preregistration and blinding. *Attention, Perception, & Performance*, *79*, 713–725.
- Emmery, C., Kádár, A., Wiltshire, T.J., Hendrickson, A.T. (in review). Towards replication in computational cognitive modeling: a machine learning perspective. *Computational Brain & Behavior*.
- Gelman, A. (2018). Don't calculate post-hoc power using observed estimate of effect size. *Annals of Surgery*, *20*, 1.
- Gluck, K.A. (in review). What does it mean for psychological modeling to be more robust? *Computational Brain & Behavior*.
- Gluck, K.A., McNamara, J.M., Brighton, H., Dayan, P., Kareev, Y., Krause, J., et al. (2012). Robustness in a variable environment. In J.R. Stevens, & P. Hammerstein (Eds.) *Evolution and the mechanisms of decision making*. (p. 195–214). *Stringmann Forum Report, vol. 11, J. Lupp, series ed.* Cambridge: MIT Press.
- Gluth, S., & Jarecki, J.B. (in review). On the importance of power analyses for cognitive modeling. *Computational Brain & Behavior*.
- Gunzelmann, G. (in review). Promoting cumulation in models of the human mind. *Computational Brain & Behavior*.
- Heathcote, A. (in review). What do the rules for the wrong game tell us about how to play the right game? *Computational Brain & Behavior*.
- Heck, D.W., & Ertfelder, E. (in review). Maximizing the expected information gain of cognitive modeling via design optimization. *Computational Brain & Behavior*.
- Hoening, J.M. (2001). The abuse of power: the pervasive fallacy of power calculations for data analysis. *The American Statistician*, *55*(1), 19–24.
- Hoyningen-Huene, P. (2006). Context of discovery versus context of justification and Thomas Kuhn. *Revisiting discovery and justification* (p. 119–131). Dordrecht.
- Journal of experimental psychology: human perception and performance. Submission questionnaire (2019).
- Kellen, K. (in review). A model hierarchy for psychological science. *Computational Brain & Behavior*.
- Kennedy, L., Simpson, D., Gelman, A. (in review). The experiment is just as important as the likelihood in understanding the prior: a cautionary note on robust cognitive modelling. *Computational Brain & Behavior*.
- Lee, M.D., Criss, A.H., Devezer, B., Donkin, C., Etz, A., Leite, F.P. (in review). Robust modeling in cognitive science. *Computational Brain & Behavior*.
- Lilburn, S.D., Little, D.R., Osth, A.F., Smith, P.L. (in review). Cultural problems cannot be solved with technical solutions alone. *Computational Brain & Behavior*.
- Morey, R., Kaschak, M.P., Díez-Álamo, A.M., Glenberg, A.M., Zwaan, R.A., Lakens, D. (2019). A pre-registered, multi-lab non-replication of the action-sentence compatibility effect (ACE). (Article submitted for publication).
- Navarro, D.J. (2019). Between the devil and the deep blue sea: Tensions between scientific judgement and statistical model selection. *Computational Brain & Behavior*, *2*, 28–34.
- Neufeld, R.W.J., & Cutler, C.D. (in review). Potential contributions of clinical mathematical psychology to robust modeling in cognitive science. *Computational Brain & Behavior*.
- Palmeri, T.J. (in review). On testing and developing cognitive models. *Computational Brain & Behavior*.
- Pitt, M., & Myung, J.I. (in review). Robust modeling through design optimization. *Computational Brain & Behavior*.
- Poldrack, R., Feingold, F., Frank, M.J., Gleeson, P., de Hollander, G., Huys, Q.J.M. (in review). The importance of standards for sharing of computational models and data. *Computational Brain & Behavior*.
- Shiffrin, R. (in review). Misunderstanding the goal of modeling. *Computational Brain & Behavior*.
- Silberzahn, R., Uhlmann, E.L., Martin, D.P., Anselmi, P., Aust, F., Awtry, E.A. (2018). Many analysts, one data set: making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science*, *1*(3), 337–356.
- Starns, J.J., Cataldo, A.M., Rotello, C.M. (in review). Blinded inference: an opportunity for mathematical modelers to lead the way in research reform. *Computational Brain & Behavior*.
- Starns, J.J., Cataldo, A.M., Rotello, C.M., Annis, J., Aschenbrenner, A., Broder, A. (in press). Assessing theoretical conclusions with blinded inference to investigate a potential inference crisis. *Advances in Methods and Practices in Psychological Science*.
- Steege, S., Tuerlinckx, F., Gelman, A. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*(5), 702–712.
- Stodden, V., McNutt, M., Bailey, D.H., Deelman, E., Gil, Y., Hanson, B. (2016). Enhancing reproducibility for computational methods. *Science*, *354*(6317), 1240–1241.
- Suppes, P. (1966). Models of data. In *Studies in logic and the foundations of mathematics* (Vol. 44, pp. 252–261). Elsevier.
- Szollosi, A., & Donkin, C. (in review). Neglected sources of flexibility in psychological theories: from replicability to good explanations. *Computational Brain & Behavior*.
- Van Zandt, T., & MacEachern, S.N. (in review). Preregistration of modeling exercises may not be useful. *Computational Brain & Behavior*.
- Vanpaemel, W. (in review). The really risky registered modeling report for incentivizing strong tests. *Computational Brain & Behavior*.
- Wagenmakers, E.-J., Verhagen, J., Ly, A., Bakker, M., Lee, M.D., Matzke, D. (2015). A power fallacy. *Behavior Research Methods*, *47*(4), 913–917.
- Watts, D.J. (2017). Should social science be more solution-oriented? *Nature Human Behaviour*, *1*(1), 0015.
- Wilson, M.D., Boag, R.J., Strickland, L. (in review). All models are wrong, some are useful, but are they reproducible? *Computational Brain & Behavior*.
- Yu, Q., MacEachern, S.N., Peruggia, M. (2011). Bayesian synthesis: Combining subjective analyses, with an application to ozone data. *The Annals of Applied Statistics*, *5*(2B), 1678–1698.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.