

The importance of standards for sharing of computational models and data

Russell A Poldrack (Stanford University), Franklin Feingold (Stanford University), Michael J Frank (Brown University), Padraig Gleeson (University College London), Gilles de Hollander (Vrije Universiteit Amsterdam), Quentin JM Huys (University College London), Bradley C. Love (University College London), Christopher J. Markiewicz (Stanford University), Rosalyn Moran (King's College London), Petra Ritter (Charité Universitätsmedizin Berlin & Berlin Institute of Health), Brandon M. Turner (The Ohio State University), Tal Yarkoni (University of Texas at Austin), Ming Zhan (National Institute of Mental Health, NIH), Jonathan D. Cohen (Princeton University)

Abstract

The Target Article by Lee et al. (2019) highlights the ways in which ongoing concerns about research reproducibility extend to model-based approaches in cognitive science. Whereas Lee et al. focus primarily on the importance of research practices to improve model robustness, we propose that the transparent sharing of model specifications, including their inputs and outputs, is also essential to improving the reproducibility of model-based analyses. We outline an ongoing effort (within the context of the Brain Imaging Data Structure community) to develop standards for the sharing of the structure of computational models and their outputs.

Main text

The issues raised by Lee et al. (2019) regarding reproducibility and robustness in model-based approaches in cognitive science are well stated, and the proposed solutions would have important positive impacts on the field. In this commentary we focus on a related, complementary issue, that was not discussed in detail by Lee et al., yet that we believe to be equally important for improving reproducibility: the transparent sharing of model specifications, including their inputs and outputs. One strong motivation for our interest in the specification of model structure, inputs, and outputs comes from the increasing prevalence of model-based analyses in neuroimaging and cognitive neuroscience (Frank, 2015; Turner, Forstmann, Love, Palmeri, & Van Maanen, 2017).

The problem of reproducing computational models from the descriptions included in most published studies is comparable to that of reproducing statistical analyses: often, many of the details (parameter values and/or procedures) needed to replicate the results are not explicitly described, and are difficult to determine from the (often custom) code used to implement the model. This poses a problem not only for replicating results of models published by others, but often even when trying to re-implement a model developed earlier in the same laboratory using a software and/or hardware environment that is no longer accessible.

There are several different levels at which a modeling result might be reproducible (Benureau & Rougier, 2017), all of which have important implications for computational cognitive science. First, a researcher may wish to rerun the same code in exactly the same way as reported, in order to simply reproduce the results. Even this baseline level of reproducibility can be challenging across time within a single lab, as software interfaces and computational environments change. Second, a researcher may wish to rerun the same code while varying parameters or inputs (including different input datasets), in order to assess robustness, check for identifiability of parameters, or assess the ability of the model to recover parameters from simulated data. This requires an additional level of clarity regarding the structure of the inputs, outputs, and parameters. Finally, full reproducibility requires the ability to rerun the same model using a different implementation, in order to assess generalizability. The latter is necessary both to assess the reliance of any result on specific algorithmic or implementational choices, as well as to identify any potential software errors (that occur with increasing likelihood in any large software project, especially when software engineering best practices are not employed, as is often the case with scientific software). Achieving these ends requires a formal model specification, in terms of the equations, function definitions and/or algorithms that precisely and clearly communicate the theory from which any particular software implementation follows (Cooper & Guest, 2014), and that can be used to reproduce the published results (Fum, Del Missier, & Stocco, 2007).

Careful comparative analyses of different computational cognitive modeling approaches applied to the same data can even lead to important theoretical results. For example, Ratcliff and Childers (2015) performed a comparative analysis of different fitting routines for the diffusion decision model (DDM), a prominent model of response times. They showed that different fitting routines can lead to different parameter estimates. Similarly, Donkin et al. (2011) and Goldfarb et al. (2014) fitted both the DDM and a competing response time model, the linear ballistic accumulator model (LBA) to the same data sets. They showed that using different models, although conceptually similar, can lead to slightly different interpretations of response time patterns. Unfortunately, comparative studies such as these are currently very time-intensive, because different packages need to be installed and learned and these packages often have very different input/output specifications. An alternative approach would be to implement the models themselves, but this is potentially even more time-intensive and prone to errors. A common standard for the input/output structure of such packages and, better still, a standard format for specification of the models themselves would make important modeling comparison studies such as those mentioned above considerably less time-intensive and much easier to reproduce on independent data sets.

The sharing of code is clearly necessary for reproducibility (Eglen et al., 2017), but is rarely sufficient, as nicely stated by Buckheit and Donoho (1995): “An article about computational science ... is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures.” Sharing of the execution environment is now possible using containerized computing environments (such as Docker or Singularity), but

containerization alone only enables the first level of reproducibility outlined above. The higher levels of reproducibility require a clear way to specify the structure of the model so that it can be modified and/or replicated in a different environment, and compared with others.

The need for standards

The availability of standards for the organization of data and models has the potential to greatly enhance the utility of shared models and results; without such standards, inordinate effort is required on the part of one researcher simply to understand the organization of the information shared by another. Within the brain imaging community, the sharing of data has benefited substantially from community standards, first through the development of a common file format (NIfTI) supported by all major software packages, and more recently through the Brain Imaging Data Structure (BIDS) (Gorgolewski et al., 2016). The BIDS standard has been developed by a large number of researchers within the neuroimaging community, with the goal of providing a common organizational scheme for shared neuroimaging datasets. It defines a schema for the organization and naming of data files and folders, as well as a framework for the specification of metadata that are important for data analysis at multiple levels. Whereas BIDS first arose as a standard for raw magnetic resonance imaging data, it has been extended by the community to a number of additional data types including magnetoencephalography (MEG) and electroencephalography (EEG). In addition, ongoing work is extending the standard to derived data types, such as the outputs of neuroimaging analyses and the statistical models used to analyze these data.

Inspired by the increasing relevance of computational models in neuroimaging research and supported by the NIH BRAIN Initiative, a group of researchers (including the authors of this commentary) met in April 2019 to initiate the development of a standard framework for the description of computational models relevant to neuroscience and psychology. The breadth of modeling approaches represented by the attendees was intentionally wide, from cognitive models to dynamic neural models to spiking neural networks. At the same time, the group focused on a specific set of use cases, in order to ensure that the effort remained constrained to address common problems.

A first goal of the standards effort is to develop a scheme to organize the inputs to a computational model (such as behavioral or neuroscientific measurements) and the outputs of a modeling run. In many cases, this effort dovetailed with prior work in describing derived data types within BIDS, such as connectivity matrices that may be the result of structural or functional connectivity analysis, but which also may inform brain-network or statistical models. A further need was identified to standardize the representation of stimuli and training corpora used in empirical studies and/or computational simulations, which were left free-form in the original BIDS specification but would benefit from greater structure in the modeling context. Finally, several classes of derivatives specific to computational and statistical models were identified,

including descriptions of parameter distributions, traces of internal model variables and synthetic neuroimaging data.

A second (and arguably much more ambitious) goal of the effort was to develop a common framework for describing the structure of computational models themselves. The upside of a common specification for computational models is potentially enormous: researchers would be able to implement the same model in multiple environments; to easily inspect, evaluate, compare and extend one another's models; and to express even relatively complex models using a common generative syntax. However, the barriers to developing such a standard are also considerable. One challenge is the need to balance expressivity against simplicity: it is probably impossible for a relatively compact specification to capture the full breadth of computational models used in all of neuroscience and psychology, so difficult choices will have to be made in balancing the desire for comprehensiveness with the goal of practicality. The BIDS community heavily emphasizes the Pareto Principle or 80/20 Rule—a recognition that, with regard to adoption, the sweet spot for a new standard is a specification that is as simple as possible while capturing as wide a range of common use cases as possible. We expect that the same philosophy is likely to be fruitful in the case of computational models.

One potentially promising approach to developing a standard for model specification describing them in the form of a computational graph. This is an approach that is seeing increasing use in other domains (such as the data sciences and machine learning), and has already been adopted by at least two open source model specification projects that address very different levels of analysis in neuroscience: NeuroML (Cannon et al., 2014), focused on spiking network models containing elements ranging from integrate and fire neurons to biophysically detailed cells; and PsyNeuLink (psyneulink.org), focused on higher level models of system-level brain and cognitive function. Nevertheless, both share the idea that computational elements (whether neurons, neural populations, entire brain areas, or even abstract cognitive functions) can be expressed as a graph, the nodes of which describe the function carried out by each computational element (as well as its parameters), and the edges of which describe the exchange of information among them (i.e., the sources of inputs and destinations of outputs for the functions of the nodes they connect). Such a framework has the ability to encode a very broad class of models, ranging from neural networks to Bayesian models to large-scale models of neural dynamics. This could facilitate the exchange across the existing diversity of environments that focus on particular levels of analysis and/or theoretical approaches, such as NEURON (Carnevale & Hines, 2006), Emergent (Aisa et al. 2008), Nengo (Bekolay et al., 2014), The Virtual Brain (Sanz Leon et al., 2013), and ACT-R (Anderson et al., 2004). A key requirement of this approach is the ability to express the functions of nodes in a common form. This can be done either by defining an ontology of such functions, and/or referencing existing function libraries (such as MATLAB, Numpy, etc.). During the recent BIDS workshop, the group developed an outline for such an approach, that provides a promising starting point for a standard model description format that can help bridge between and integrate existing computational modeling efforts in neuroscience and psychology.

Moving forward, our group plans to further develop the standard for computational model description as part of the BIDS ecosystem, though its utility far beyond brain imaging may argue for its deployment as an independent standard (inspired by, but separate from, the BIDS initiative). This process will undoubtedly involve many iterations over an expanding set of concrete use cases. Researchers interested in contributing to this effort can find more information at the BIDS Github page, where all discussion regarding the standards takes place (<https://github.com/bids-standard/bids-specification/issues/230>).

Acknowledgements

Development of this article was supported by the BRAIN Initiative and National Institute of Mental Health (1R24MH114705).

References

- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–1060.
- Bekolay, T., Bergstra, J., Hunsberger, E., Dewolf, T., Stewart, T. C., Rasmussen, D., ... Eliasmith, C. (2014). Nengo: a Python tool for building large-scale functional brain models. *Frontiers in Neuroinformatics*, *7*, 48.
- Benureau, F. C. Y., & Rougier, N. P. (2017). Re-run, Repeat, Reproduce, Reuse, Replicate: Transforming Code into Scientific Contributions. *Frontiers in Neuroinformatics*, *11*, 69.
- Buckheit, J. B., & Donoho, D. L. (1995). WaveLab and Reproducible Research. In A. Antoniadis & G. Oppenheim (Eds.), *Wavelets and Statistics* (pp. 55–81). New York, NY: Springer New York.
- Cannon, R. C., Gleeson, P., Crook, S., Ganapathy, G., Marin, B., Piasini, E., & Silver, R. A. (2014). LEMS: a language for expressing complex biological models in concise and hierarchical form and its use in underpinning NeuroML 2. *Frontiers in Neuroinformatics*, *8*, 79.
- Carnevale, N. T., & Hines, M. L. (2006). *The NEURON Book*. Cambridge University Press.

- Donkin, C., Brown, S., Heathcote, A., & Wagenmakers, E.-J. (2011). Diffusion versus linear ballistic accumulation: different models but the same conclusions about psychological processes? *Psychonomic Bulletin & Review*, Vol. 18, pp. 61–69.
<https://doi.org/10.3758/s13423-010-0022-4>
- Eglen, S. J., Marwick, B., Halchenko, Y. O., Hanke, M., Sufi, S., Gleeson, P., ... Poline, J.-B. (2017). Toward standard practices for sharing computer code and programs in neuroscience. *Nature Neuroscience*, 20(6), 770–773.
- Frank, M. J. (2015). Linking Across Levels of Computation in Model-Based Cognitive Neuroscience. *An Introduction to Model-Based Cognitive Neuroscience*, pp. 159–177.
https://doi.org/10.1007/978-1-4939-2236-9_8
- Fum, D., Del Missier, F., & Stocco, A. (2007). *The cognitive modeling of human behavior: Why a model is (sometimes) better than 10,000 words*. Retrieved from
<https://pdfs.semanticscholar.org/0134/72f14486641aeab012242a3001c9f321df06.pdf>
- Goldfarb, S., Leonard, N. E., Simen, P., Caicedo-Núñez, C. H., & Holmes, P. (2014). A comparative study of drift diffusion and linear ballistic accumulator models in a reward maximization perceptual choice task. *Frontiers in Neuroscience*, 8, 148.
- Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., ... Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, 3, 160044.
- Ratcliff, R., & Childers, R. (2015). Individual differences and fitting methods for the two-choice diffusion model of decision making. *Decision*, Vol. 2, pp. 237–279.
<https://doi.org/10.1037/dec0000030>
- Sanz Leon, P., Knock, S. A., Woodman, M. M., Domide, L., Mersmann, J., McIntosh, A. R., & Jirsa, V. (2013). The Virtual Brain: a simulator of primate brain network dynamics. *Frontiers*

in Neuroinformatics, 7, 10.

Turner, B. M., Forstmann, B. U., Love, B. C., Palmeri, T. J., & Van Maanen, L. (2017).

Approaches to analysis in model-based cognitive neuroscience. *Journal of Mathematical Psychology*, 76, 65–79.