Computational Brain & Behavior (2019) 2:255–265
https://doi.org/10.1007/s42113-019-00049-8

ORIGINAL PAPER

Robust Standards in Cognitive Science

Sophia Crüwell[1,2] · **Angelika M. Stefan**[1] · **Nathan J. Evans**[1]

Published online: 21 August 2019

Abstract

Recent discussions within the mathematical psychology community have focused on how Open Science practices may apply to cognitive modelling. Lee et al. (2019) sketched an initial approach for adapting Open Science practices that have been developed for experimental psychology research to the unique needs of cognitive modelling. While we welcome the general proposal of Lee et al. (2019), we believe a more fine-grained view is necessary to accommodate the adoption of Open Science practices in the diverse areas of cognitive modelling. Firstly, we suggest a categorization for the diverse types of cognitive modelling, which we argue will allow researchers to more clearly adapt Open Science practices to different types of cognitive modelling. Secondly, we consider the feasibility and usefulness of preregistration and lab notebooks for each of these categories and address potential objections to preregistration in cognitive modelling. Finally, we separate several cognitive modelling concepts that we believe Lee et al. (2019) conflated, which should allow for greater consistency and transparency in the modelling process. At a general level, we propose a framework that emphasizes local consistency in approaches while allowing for global diversity in modelling practices.

Keywords Cognitive modelling · Reproducibility · Open Science · Preregistration · Transparency

Introduction

Psychology and related fields have been gripped by the so-called "replication crisis" over the past several years (Pashler and Wagenmakers 2012), though discussions of the underlying problems date back several decades (cf. e.g., Sterling 1959; Cohen 1965; Meehl 1967). The continuing misuse of statistical methods has led to a substantial portion of the literature being unreplicable and most likely untenable (Ioannidis 2005; Open Science Collaboration 2015), which has led to a reform movement around "Open Science," where an increasing number of scientists have focused on improving the use and development of statistics and methodology within psychology. This movement has led to the development of a variety of Open Science practices, such as data sharing (Klein et al. 2018), preregistration (Wagenmakers et al. 2012), and preregistration's more rigorous sister Registered Reports (RRs; Chambers et al. 2014), with the goal of creating transparent and accessible psychological research (Nuijten 2018; Crüwell et al. 2018). Importantly, transparent scientific practices can counteract the effect of cognitive biases and other pressures that may affect the reproducibility of scientific findings (Munafò et al. 2017). Our article largely focuses on preregistration, a practice where researchers specify their research design, hypotheses, and analysis plan prior to the data collection (Nosek and Lindsey 2018).

Recent discussions within the mathematical psychology community have focused on how these Open Science practices may apply to cognitive modelling (see Wagenmakers and Evans 2018; Lewandowsky 2019, for reviews of the recent discussions on social media). Specifically, there are many instances in cognitive modelling where a lack of practical constraints (i.e., an abundance of "modeller's degrees of freedom," Dutilh et al. 2018) can lead to inconsistent and untransparent modelling approaches, which can affect interpretations of model success and failure (McClelland 2009;

✉ Sophia Crüwell
sophia.cruewell@charite.de

Angelika M. Stefan
a.m.stefan@uva.nl

Nathan J. Evans
nathan.j.evans@uon.edu.au

[1] Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands

[2] Meta-Research Innovation Center Berlin (METRIC-B), QUEST Center for Transforming Biomedical Research, Berlin Institute of Health, Charité – Universitätsmedizin Berlin, Berlin, Germany

Roberts and Pashler 2000), and prevent the reuse of existing models (e.g., Addyman and French 2012). Importantly, Open Science practices have been developed to counteract similar problems in experimental psychology, showing a clear value for similar practices within cognitive modelling, where they may lead to less biased modelling practices and broader reuse and refinement of existing models. However, there has been opposition to Open Science practices within mathematical psychology, with criticisms suggesting that Open Science practices can be inflexible and are poorly adapted to the needs of cognitive modelling (see Wagenmakers and Evans 2018, for a discussion).

Lee et al. (2019) sketched an initial approach for incorporating and adapting Open Science practices into cognitive modelling. Specifically, Lee et al. (2019) suggested several "good practices" within cognitive modelling, where studies should (1) preregister the models and evaluation criteria, (2) consider the full spectrum of models and evaluation criteria available, (3) ensure the robustness of findings, and (4) provide a distinction between confirmatory and exploratory analyses, with "postregistration"—in the form of a lab notebook—being an option for exploratory analyses. We welcome Lee et al.'s (2019) introduction of Open Science practices to cognitive modelling, as we believe that they will aid the consistency and transparency of research in cognitive modelling. However, we believe that a more fine-grained view is necessary to accommodate the adoption of Open Science practices in the diverse range of studies that fall under the "cognitive modelling" umbrella, where different types of cognitive modelling may have different specific requirements. Specifically, we believe that a lack of distinction between the diverse areas of cognitive modelling has led to overly polarized standpoints regarding Open Science practices, in particular preregistration, within the mathematical psychology community (see Wagenmakers and Evans 2018; Lewandowsky 2019). More separation and directive guidance is required to aid the adoption of Open Science practices in cognitive modelling.

Our article focuses on three key topics that we believe will aid the adoption of Open Science practices in cognitive modelling. Firstly, we provide a categorization for the diverse types of cognitive modelling, which we believe will aid the development of directive guidance for Open Science practices in different categories of cognitive modelling. Secondly, we address several potential objections to preregistration within cognitive modelling, discuss the categories of cognitive modelling that are best suited to preregistration, and suggest how preregistration may need to be adapted to suit the specific needs of each category. We also discuss the concept of "postregistration" and the use of lab notebooks, and the need for standards and consistency in post-hoc registration practices. Lastly, we attempt to separate several cognitive modelling concepts that we believe Lee et al. (2019) conflated, as their conflation may lead to inconsistent practices with misleading results within cognitive modelling.

One Man's Meat is Another Man's Poison

"Cognitive modelling" is an umbrella term commonly used to describe a diverse range of studies that implement formalized models to better understand cognitive processes. These studies can range from purely confirmatory research, such as a direct comparison between two competing cognitive theories represented as formalized models (e.g., Evans et al. 2017a, 2019a, b; Palestro et al. 2018; Voskuilen et al. 2016; Evans and Hawkins 2019; Teodorescu and Usher 2013), to purely exploratory research, such as the initial development of a model for a novel experimental paradigm (e.g., Nosofsky 1986; Shiffrin and Steyvers 1997; Ratcliff 1978; Dougherty et al. 1999; Jones and Mewhort 2007; Vickers and Lee 2000). However, this diversity has been largely ignored within the recent discussions on how Open Science practices may apply to cognitive modelling, leaving the discussion open to potential misunderstandings about which Open Science practices are applicable to which types of cognitive modelling. We argue that the overly generic view of cognitive modelling may be the underlying reason for much of the opposition to Open Science practices within mathematical psychology. For example, in the context of the exploratory development of a model, a researcher might be horrified at the idea of "preregistration in cognitive modelling", and rightfully so, as preregistration is most useful for confirmatory research. There are other Open Science practices that are useful for increasing the transparency of exploratory research such as sharing your data, materials, and code (Klein et al. 2018), or prioritizing open access publishing (Tennant et al. 2016). We believe that the opposition to Open Science practices in cognitive modelling might be alleviated by a more specific categorization of different types of cognitive modelling, and a discussion of where specific Open Science practices are most applicable. Specifically, we propose that it may be useful to split cognitive modelling into the following four categories: *model application*, *model comparison*, *model evaluation*, and *model development*. Each of these different categories involves different research goals, uses different methods of assessment, and might differ in how well they are suited to different Open Science practices, particularly preregistration. Table 1 lists each of these different modelling categories, as well as some important factors that researchers should consider before implementing each of these categories of cognitive modelling, some of which we discuss in more detail throughout the remainder of our article. Importantly, we

believe that these considerations may be viewed as potential "researcher degrees of freedom," which may form the basis for future preregistration templates and documentation standards in cognitive modelling.

Model application consists of studies where an existing cognitive model, which is assumed to provide an adequate representation of the underlying cognitive process, is applied to empirical data to provide insight into how the cognitive process operates in that paradigm (e.g., Weigard and Huang-Pollock 2017; Ratcliff et al. 2001; Janczyk and Lerche 2019; Lerche et al. 2019; Wagenmakers et al. 2008; Ratcliff and Rouder 2000; Evans et al. 2018a, c). These applications often involve experimental studies with different groups and/or conditions, with researchers interested in how the cognitive process changes across these factors, measured by changes in the values of the model parameters. *Model application* is similar to the concept of "*measurement models*" discussed by Lee et al. (2019, p. 10), though we wish to distinguish our definition from their definition, as we believe that measurement is rarely the sole purpose of a model and that a single model can be used for both measurement and theory representation. For example, the diffusion model of decision-making (Ratcliff 1978) is often used for measurement in *model application* (e.g., Weigard and Huang-Pollock 2017; Ratcliff et al. 2001; Janczyk and Lerche 2019; Lerche et al. 2019; Wagenmakers et al. 2008; Ratcliff and Rouder 2000; Evans et al. 2018a), though it has also been used as a cognitive theory in *model comparison* (e.g., Voskuilen et al. 2016; Voss et al. 2019; Evans and Hawkins 2019; Evans et al. 2017a, 2019b), *model evaluation* (e.g., Ratcliff and Rouder 1998; Teodorescu and Usher 2013; Evans et al. 2019a; Cisek et al. 2009; Thura et al. 2012), and *model development* (e.g., Ratcliff 1978; Ratcliff and Rouder 1998; Ratcliff and Tuerlinckx 2002). *Model application* typically involves confirmatory research questions that are similar to those in traditional experimental research—with a cognitive model in place of a statistical model (i.e., a *t* test, ANOVA)—where a set of conflicting a priori hypotheses about how the parameters should vary over groups and/or conditions are assessed. Therefore, Open Science practices that have been developed for confirmatory experimental research—particularly preregistration—are clearly and readily applicable to the *model application* category, and existing preregistration templates could be adapted for *model application* with only minor amendments.

Model comparison consists of studies where multiple existing cognitive models are compared on their ability to account for empirical data, typically either based on their ability to provide an accurate explanation of the underlying process, or to predict future data in the same context (e.g., Voskuilen et al. 2016; Voss et al. 2019; Evans and Hawkins 2019; Evans et al. 2017a, 2019b). These assessments are usually made through quantitative model selection methods, which penalize models based on either their a priori flexibility (e.g., Kass and Raftery 1995; Evans and Brown 2018; Myung et al. 2006; Annis et al. 2019; Evans and Annis 2019; Gronau et al. 2017; Schwarz 1978) or their overfitting to the noise in samples of data (e.g., Spiegelhalter et al. 2002; Vehtari et al. 2017; Browne 2000; Akaike 1974). Importantly, models that are more flexible a priori will have an unfair advantage in accurately explaining the data than simpler models (Roberts and Pashler 2000; Myung and Pitt 1997; Evans et al. 2017b), and models that over-fit to a sample of data will predict future data more poorly than those that only capture the robust trends (Myung 2000). Although *model comparison* is less similar to confirmatory experimental research than *model application*, *model comparison* still typically involves confirmatory research questions about *which* models will be superior to others, making it well suited to preregistration. However, several additional factors need to be considered for the preregistration of *model comparison* beyond the factors in *model application*, such as the models to be compared (discussed by Lee et al. as "*the players in the game*," p.3) and the model selection methods for the comparison (discussed the Lee et al. as "*the rules of the game*." p.3). Furthermore, *model comparison* can often involve the use of existing data sets, meaning that researchers need to consider the scope of their comparison (i.e., the empirical contexts included in the comparison), and any potential preregistration template for *model comparison* would need to consider secondary data preregistration (see Mertens and Krypotos 2019; Weston et al. 2018).

Model evaluation consists of studies where one or multiple existing cognitive models are evaluated on their ability to account for specific patterns in empirical data (e.g., Ratcliff and Rouder 1998; Teodorescu and Usher 2013; Evans et al. 2019a; Cisek et al. 2009; Thura et al. 2012). These assessments are usually made through visual assessments of qualitative trends that can be plotted from the data, which are contrasted to the predictions that the model can make for these aspects of the data. It should be noted that *model evaluation* contains no correction for model flexibility and, therefore, should not be used to answer confirmatory research questions about *which* models are superior to others, as these comparisons will be biased towards more flexible models (see Roberts and Pashler 2000; Evans 2019b, for more detailed discussions). However, *model evaluation* is ideal for answering research questions about *why* certain models are found to be superior to others in *model comparison* and what further development may be required to create a better explanation of the underlying process, meaning that it is often used in combination with *model development*. *Model evaluation* can be used in a confirmatory manner when researchers have a

specific model, or models, that they wish to evaluate on a specific trend, or trends, and in these cases *model evaluation* requires a similar consideration to *model comparison* for preregistration. However, when *model evaluation* is combined with *model development*, *model evaluation* can become an iterative, exploratory process, which is less applicable to preregistration and related Open Science practices.

Model development consists of all instances where models are altered in some way to create a new model and can be viewed as the preceding work required for all other categories of cognitive modelling (e.g., Brown and Heathcote 2005, 2008; Usher and McClelland 2001; Evans et al. 2018b; Nosofsky 1986; Shiffrin and Steyvers 1997; Ratcliff 1978; Dougherty et al. 1999; Jones and Mewhort 2007; Vickers and Lee 2000; Ratcliff and Rouder 1998; Ratcliff and Tuerlinckx 2002). The alterations can range from minor tweaks to an existing model (e.g., Brown and Heathcote 2008; Evans et al. 2018b; Ratcliff and Rouder 1998; Ratcliff and Tuerlinckx 2002), to constructing a novel model of a novel paradigm (e.g., Nosofsky 1986; Shiffrin and Steyvers 1997; Ratcliff 1978). The goal of the development process could be to create a simple model well suited to *model application*, a new explanation of the underlying process that will be compared with other models in *model comparison*, or a general model that provides a functional form capable of meeting a range of qualitative benchmarks in *model evaluation*. *Model development* is a broad category that encompasses many cognitive modelling studies and is often necessary to adapt an existing model to a new paradigm, even when development is not the focus of the study (Navarro 2019). In many of these cases, *model development* can be an iterative, exploratory process, meaning that Open Science practices designed for confirmatory experimental research, such as preregistration, have limited applicability. However, we believe that *model development* is the most crucial category of cognitive modelling, and therefore, future research should investigate how the process of *model development* can be made as transparent as possible, without restricting the iterative, exploratory process. The considerations in Table 1 may provide a valuable starting point for these investigations.

Note that our categorization of cognitive modelling might not be ideal for all purposes. Many researchers work with more than one category of cognitive modelling, and existing studies often combine these different categories. However, we believe that this categorization provides a useful tool for discussing the applicability of Open Science practices to

Table 1 Transparency considerations in different categories of cognitive modelling

Category	Important considerations for transparency
Model application[a]	Selecting a model to assume as the underlying process
	Creating a match between model parameters and application domain theory
	Deciding upon the method of parameter estimation (e.g., maximum likelihood)
	Choosing a method of statistical inference on parameters (e.g., Bayes factors)
Model comparison[a]	Selecting a subset of precisely defined models to be compared
	Selecting a suitable data set (or data sets) for the comparison
	Deciding upon the goal of the comparison (e.g., explanation or prediction)
	Deciding upon a comparison criterion that matches the goal (e.g., cross validation for prediction)
	Deciding upon the strength of evidence required for confidently selecting one model over another
	Deciding upon robustness checks to account for ancillary assumptions
Model evaluation[a]	Selecting the data trends or benchmarks of interest
	Deciding upon clear criteria for evaluation (e.g., directional, goodness-of-fit, visual)
	Defining the criteria for adequacy (e.g., when the model is seen to be descriptively accurate)
	Defining all a priori theoretically justifiable functional forms of the model
	Clearly separating confirmatory and exploratory aspects (e.g., data-driven changes to the model)
Model development[b]	Providing a clear and transparent documentation of the model exploration process
	Discussing existing theoretical justification for model components and functional form
	Distinguishing between theory-driven development and data-driven development
	Deciding which components of the model are core and which are ancillary
	Deciding upon the purpose of the model (e.g., tool for application, formalization of theory, both)
	Deciding upon evaluation criteria that will drive the model development (e.g., data trends, parameter identifiability)

[a]Considerations should be made before the modelling process

[b]Considerations can be made during any point of the exploratory process

cognitive modelling and determining where these practices may need to be adapted for the specific needs of different types of cognitive modelling.

The Devil Is in the Details

Lee et al. (2019) made an important contribution by suggesting several "good practices" within cognitive modelling. Here, we discuss two of these specific practices, which were intended as an initial approach for incorporating and adapting Open Science practices into cognitive modelling: "registered modelling reports" and "postregistration." The registered modelling reports format was proposed as an extension of the registered reports format used in confirmatory experimental research, where an article can receive an in-principle acceptance for publication after the first round of thorough reviews on a detailed research proposal and is then published after a second round of reviews when the study has been conducted (Chambers et al. 2015). Postregistration was proposed as a process of recording every detail of an exploratory modelling process in the form of a lab notebook, as an attempt to increase transparency in non-confirmatory settings. While we welcome the attempt to introduce practices that increase the consistency and transparency of modelling approaches, Lee et al. (2019) did not provide specific guidelines on how these practices should be implemented. As discussed above, there are many different types of cognitive modelling, ranging from the purely confirmatory to the purely exploratory, which each also differs in how Open Science practices would be best suited to their usage. Therefore, we argue that the current proposals of Lee et al. (2019) might be of limited value to mathematical psychology researchers and that some researchers may still object to registered modelling reports—and more generally, preregistration—as it is currently unclear which types of cognitive modelling they are most applicable to. Here, we use our proposed categorization to discuss the role of preregistration within cognitive modelling and address potential objections that researchers may have to preregistration in cognitive modelling. We also discuss Lee et al.'s proposal of "postregistration" in more detail, in what cases it is likely to be most applicable, and how it could be adapted to better ensure consistent and transparent research practices in cognitive modelling.

Preregistration

An important underlying focus of preregistration and registered reports is to highlight the difference between exploration (data dependent) and confirmation (data independent) in quantitative experimental research. The idea here is that so-called questionable research practices (QRPs)—such as hypothesizing after results are known (HARKing; Kerr 1998), p-hacking (Simmons et al. 2011; de Groot 2014), and other researcher degrees of freedom (Simmons et al. 2011)—may affect a study in ways which render the seemingly confirmatory results uninterpretable (Wagenmakers et al. 2012). Preregistrations or registered reports can help to counteract QRPs and unchecked researcher degrees of freedom by pre-specifying the hypotheses and analysis plans.

Importantly, the underlying strength of preregistration and registered reports are in their specificity: clearly specified research plans can prevent QRPs and create greater consistency and transparency in scientific findings. Although we agree with Lee et al. (2019) that registered reports could be a valuable tool in cognitive modelling studies, we do not believe that they provided adequate specificity to constrain researcher degrees of freedom and help prevent QRPs in cognitive modelling studies. Crucially, Lee et al. (2019) did not provide specific guidelines for how to implement these registered modelling reports, with their proposal limited to a general description that mirrors regular registered reports. However, different categories of cognitive modelling range from confirmatory to exploratory, which establishes the need for distinct implementation guidelines for preregistrations and registered reports in different categories, and possibly even the creation of new procedures that can increase transparency in a similar way. We argue that there cannot be a "one-size-fits-all" solution for preregistration or registered reports in cognitive modelling, as proposed by Lee et al. (2019). Instead, category-specific templates—or a flexible general registered report template with appropriate sub-templates—should be developed. Moreover, these category-specific templates need to be sufficiently detailed and actionable to allow researchers to implement registered modelling reports in a consistent manner. Creating these templates should be the goal of future research aiming to integrate Open Science practices into cognitive modelling. Moreover, as we argue below, while preregistration seems applicable to the largely confirmatory nature of *model application*, *model comparison*, and *model evaluation*, preregistration is less applicable to the largely exploratory nature of *model development*, meaning that other practices should be developed to increase transparency in *model development*.

Furthermore, we believe that Lee et al. (2019) did little to quell the past and potential objections to preregistration in cognitive modelling. We believe that many previously voiced objections against preregistration and Open Science practices in cognitive modelling (see Lewandowsky 2019) stem from overly general proposals being applied to category-specific challenges. Here, we use our proposed categorization to address four key potential objections to preregistration in cognitive modelling: objections that have been previously voiced by mathematical psychology researchers (see Lewandowsky 2019; Wagenmakers and

Evans 2018, for discussions) or are legitimate objections that we believe researchers may have.

Objection 1: "We cannot apply preregistration to cognitive modelling."

Given our categorization, this objection should be divided into four different possible objections: one for each category of cognitive modelling. We agree that preregistration is rarely applicable to the exploratory practice of *model development* and that Lee et al.'s proposal of lab notebooks—which we comment on below—is a more promising future avenue for this category. However, preregistration should be feasible in all other categories of cognitive modelling, and although implementing preregistration in *model comparison* and *model evaluation* may require major refinement to current preregistration guidelines and templates, preregistration in *model application* should be possible with minor amendments to existing guidelines and templates.

Objection 2: "Preregistration does not cover all specific needs of cognitive modelling."

We agree that this is currently the case and share the concern that overly general templates will not be useful for the diverse nature of cognitive modelling. However, we believe that this can be solved by creating specific templates for *model application*, *model comparison*, and *model evaluation*. In cases where multiple categories of cognitive modelling are used within a single study, a researcher can apply different templates for the different confirmatory analyses, which will separate the confirmatory and exploratory modelling efforts.

Objection 3: "Cognitive modelling often uses existing data, which cannot be preregistered."

We agree that creating consistent and transparent practices for the reuse of existing data is challenging, though it has been recently addressed within confirmatory experimental research. An existing preregistration template for studies using secondary data (https://osf.io/v4z3x/) proposes that researchers should create a detailed recording of their existing knowledge of the data set and any possible sources of bias. While bias from previous experience with the data cannot be ruled out, preregistering this information adds transparency to the modelling process, placing the findings into the context of previous knowledge for both the researcher and readers of the study.

Objection 4: "None of this applies to model development."

Although *model development* is rarely confirmatory, we caution against throwing the baby out with the bathwater. While preregistration may not be appropriate for *model development*, other Open Science practices may provide greater consistency and transparency to the process, such as the proposal of Lee et al. (2019) of "postregistration" in the form of lab notebooks, which we discuss below. Furthermore, as *model development* is often used in combination with other categories of cognitive modelling, such as *model evaluation*, preregistration may still be useful for the sections of the overall modelling process where the developed model is then applied, compared, and/or evaluated.

Lab Notebooks

Lee et al. (2019) proposed the idea of "postregistration," in the form of lab notebooks, for exploratory work. Based on our categorization, lab notebooks are most applicable to *model development* and could add transparency to the exploration of the different forking paths taken to reach the final model. Moreover, as noted by Lee et al. (2019), the documentation of the development processes may foster the publication of failed model development efforts, which could counteract file-drawer effects and add community knowledge about unfruitful development processes. Therefore, lab notebooks can be seen as an effective way to increase transparency in *model development*.

However, we argue that the overly general proposal of Lee et al. (2019) lacks the necessary detail to create consistent and transparent practices. Firstly, it should be clearly noted that documenting choices in notebooks is in no way comparable to preregistration. Lab notebooks are an important step towards greater transparency, but the term "postregistration" can be misleading and may result in the retroactive framing of exploratory processes as "registered" confirmatory analyses. Secondly, a lab notebook must be constrained and accessible to ensure consistency and transparency. Lee et al. (2019) suggested that "*Modeling notebooks can be created using existing software tools such as Jupyter or Rmarkdown*" (p.8). However, no framework, guidelines, or standards were provided for what these notebooks should consist of and how they should be structured, meaning that the current proposal may result in thousands of lab notebooks that are overly detailed, missing important information, poorly documented, and/or not reproducible. Given that cognitive modelling work already suffers from accessibility problems resulting from the use of a multitude of different programming languages and a poor adherence to good coding practices (Addyman and French 2012), it does not seem unreasonable to believe that similar issues will be present within lab notebooks.

We believe that specific standards are crucial for the success of post-hoc study registration. The standards for post-hoc study registration could be based on the format of exploratory reports, which aim to promote transparency in exploratory work (McIntosh 2017). However, while exploratory reports help promote the value of exploratory work, the precise advantages of the specific proposal of

McIntosh (2017) are still unclear. Another possibility would be adapting the idea of living preregistration documents (Haven and Grootel 2019) to *model development*. At a minimum, lab notebooks should adhere to a set of basic standards for coding and data sharing (Addyman and French 2012), such as the Google Style Guides[1] or the psych data standard project,[2] and researchers must agree on specific standards for what aspects need to be recorded and how they should be detailed. An agreed set of standards would also provide the opportunity for more rigorous review processes (e.g., code review), which could increase transparency while also decreasing errors. We also believe that these standards should extend beyond empirical studies to simulation studies, as the selective reporting of specific simulations results can provide an incomplete and inaccurate picture of the properties of a model, though the specific standards for simulation studies may have different requirements than those of empirical studies, and we leave a more detailed discussion of this topic to future research. Ideally, if a researcher were to leave a project, a lab notebook should allow their successor to immediately fill their role based on the notes: a lofty gold-standard, but one worth attempting to accomplish.

Chalk and Cheese

One final point of brief debate is the cognitive modelling concepts that Lee et al. (2019) mentioned within their "good practices." Specifically, we believe that Lee et al.'s discussion conflated several distinct theoretical goals for implementing cognitive models and that the chosen cognitive modelling assessment should match the specific theoretical goals of the implementation, as different assessments can potentially lead to different conclusions. In this section, we discuss the difference between answering *which* and *why* questions, the difference between *explanation* and *prediction*, and the difference between updating *prior model odds* and updating *prior parameter distributions*.

The Difference Between Which and Why

Lee et al. (2019) proposed a "*continuum of utilities (i.e., cost functions) for 'scoring' a model against data*" (p. 6). Specifically, Lee et al. (2019) discussed two different ways in which researchers commonly assess the ability of models to account for empirical data: assessing how well different models can account for different qualitative benchmarks, as in our category of *model evaluation*, or comparing models on a flexibility-corrected goodness-of-fit metric, as in our category of *model comparison*. Lee et al. (2019) suggested that these different assessments could be treated as "*two end-points on a continuum of utilities*," where researchers "*balance between giving weight to qualitatively important data patterns, while still measuring overall quantitative agreement*" (p. 6). In practice, this would involve researchers deciding upon the relative weight given to the qualitative benchmarks and quantitative fit and then selecting the model that provides the best-weighted performance across both of these factors.

Although we agree that both *model evaluation* and *model comparison* are important for understanding psychological processes, we disagree that these two categories should form a continuum for selecting between models. Rather, we argue that these different forms of assessment reflect fundamentally different goals of implementing models that answer fundamentally different questions (see Evans 2019b, for a more in-depth discussion). Specifically, *model comparison* provides the most appropriate answer to *which* model provides the best account of a sample of data, as *model comparison* methods have been specifically designed to achieve this goal, taking into account all aspects of the data providing corrections for model flexibility. In contrast, *model evaluation* provides the most appropriate answer to *why* some models perform better than others according to model selection, as *model evaluation* provides visual insights into how a model succeeds or fails to account for specific parts of the data, and does not attempt to correct for flexibility.

The Difference Between Explanation and Prediction

Lee et al. (2019) also proposed that researchers should preregister the assessment criteria (i.e., the "*rules of the game*", pp. 3f) when performing *model comparison* (according to our categorization) in a registered modelling report. Specifically, Lee et al. (2019) provided a token example where different evaluation criteria lead to different models being selected and, therefore, opposite theoretical conclusions. While we agree that researchers should preregister the model selection method used for *model comparison* to limit researcher degrees of freedom, making a sensible choice for this preregistration requires knowledge of what method best suits the research question: an issue that Lee et al. (2019) wrote off as "*a challenging statistical and methodological question that remains an active area of debate and research throughout the empirical sciences and statistics*" (p. 4). Although we agree that understanding and developing methods of *model comparison* is an ongoing area of research, we believe that there is a clear rationale for why researchers should prefer specific methods in specific

[1] http://google.github.io/styleguide/

[2] https://github.com/psych-ds/psych-DS

situations and that this rationale is required for principled preregistration of assessment criteria (Evans 2019a; Gronau and Wagenmakers 2019).

In most cases, mathematical psychology researchers use models to either explain or predict a cognitive process (Yarkoni and Westfall 2017). As mentioned previously, selecting a model that maximizes one of these goals requires a method that provides some penalty for flexibility, as models with greater a priori flexibility provide less constrained (and hence poorer) explanations of a process, and models that over-fit to a sample of data provide poorer predictions about future data. Importantly, specific methods have been designed to correct for each type of flexibility: Bayesian model selection methods (e.g., Kass and Raftery 1995; Evans and Brown 2018; Annis et al. 2019; Evans and Annis 2019; Gronau et al. 2017; Schwarz 1978) punish models for their a priori flexibility, reflected in their integration of the unnormalized posterior probability over the parameter space, and out-of-sample prediction methods (e.g., Spiegelhalter et al. 2002; Vehtari et al. 2017; Browne 2000; Akaike 1974) punish models for overfitting to samples of data, reflected in their assessment of unseen samples of data.[3] Therefore, some simple and clear guidelines already exist for how researchers should choose a method for *model comparison*: when researchers are interested in providing the best explanation of a cognitive process, they should use a *model comparison* method that penalizes for a priori flexibility, such as Bayesian model selection (though also see the minimum description length principle; Myung et al. 2006); when researchers are interested in best predicting future data, they should use a *model comparison* method that penalizes for overfitting, such as out-of-sample prediction (see Evans 2019a, for a more in-depth discussion). Furthermore, we believe that *model comparison* methods without corrections for a priori flexibility or overfitting (e.g., the RSME, MAD, correlations, and LL methods mentioned by Lee et al. 2019) can be ignored in most situations, as these methods only provide insight into which model maximizes the fit for a sample of data, and in the context of explanation or prediction will provide conclusions that are biased towards more flexible models.

The Difference Between Updating Prior Model Odds and Updating Prior Parameter Distributions

Lee et al. (2019) also proposed that the distinction between confirmatory and exploratory analysis can be posed in terms of Bayesian updating. Specifically, Lee et al. (2019) suggested that in confirmatory analyses "*claims are sought about the relative probability of models, based on the data*" and require prior odds, and in exploratory analyses "*it is difficult to make claims about prior probabilities of models*" (p. 10). While we agree that the prior model odds are an important part of Bayesian model selection that is often overlooked by researchers and may be the difference between confirmatory and exploratory analyses in some case, we disagree that this is a "*useful way to think of the distinction*" (p. 10). We argue that this distinction is extremely limited, as it is only applicable to a single method (i.e., Bayesian model selection) within a single category of cognitive modelling (i.e., *model comparison*) and ignores the process of theory updating.

In Bayesian terms, the process of theory updating can often be represented as updating the prior distributions of the parameters: the plausibility assigned to different parameter values before having observed some sample of data. After having observed the sample of data, a researcher may be interested in refining their model based on the information contained within the sample of data for inferences on future samples of data. This could involve updating the prior distributions to the estimated posterior distributions (i.e., today's posterior is tomorrow's prior; Lindley 1972; Wagenmakers et al. 2010), or altering the functional form of the model by adding or removing parameters (i.e., adding/removing prior distributions). In contrast, updating the prior model odds suggests that researchers believe that these exact models (including the prior distributions) will be the comparison of interest in future samples of data, and serves as an update to the running tally of the relative probability of the two models. Therefore, refining a model by updating the prior distributions changes the precise form of the model, meaning that researchers must choose between refining their theories and updating the relative probability of the two models. Importantly, we argue that confirmatory settings exist where it is difficult to make claims about prior odds and exploratory settings exist where reasonable prior odds can be derived. For instance, we believe that in some situations, researchers may wish to perform confirmatory analyses after refining theories: a case of a confirmatory analysis without updated prior model odds. Furthermore, we believe that in other situations, researchers may wish to add a model to the comparison that they have previous knowledge about (e.g., the unrefined model), in a secondary, exploratory step: a case of an exploratory analysis that could involve prior model odds. Therefore, we disagree with the proposed distinction between confirmatory and exploratory based on prior model odds and instead suggest that after observing a sample of data, researchers should carefully consider whether they wish to refine their theory through updating the prior distributions or adjust the relative

[3]Note that the "information criteria" do not actually assess unseen data, but instead provide an approximation to leave-one-out cross validation under certain assumptions.

probability of the two models through updating the prior model odds.

Conclusion

Our article provided a discussion of several important issues that we believe were not fully addressed by Lee et al. (2019). Firstly, we proposed a clear categorization for the diverse types of cognitive modelling, and through this categorization proposed a framework for Open Science practices in cognitive modelling that emphasizes local consistency in approaches, while allowing for global diversity in modelling practices. This is an important step towards both category-appropriate guidance and more fruitful discussions regarding Open Science practices in cognitive modelling. Secondly, we addressed potential objections to preregistration in cognitive modelling and argued that preregistration and lab notebooks need further, more detailed development to be useful tools of transparency and consistency for cognitive modelling. We also provided several suggestions for how these practices could be further developed, and how categorizations, such as the one that we proposed, may help this process. More generally, Open Science practices, and especially preregistration, should be adapted more specifically for individual fields to create the greatest potential benefits. Lastly, we addressed several cognitive modelling concepts that are closely related to Open Science practices but that we believe Lee et al. (2019) conflated and provided a detailed discussion of how these concepts differ. We also discussed when each of these concepts is likely to be relevant for researchers, which should allow greater consistency in cognitive modelling practices. We hope that the discussions within our article will help advance the field of mathematical psychology from robust discussions to robust standards for Open Science practices in cognitive modelling.

References

Addyman, C., & French, R.M. (2012). Computational modeling in cognitive science: a manifesto for change. *Topics in Cognitive Science*, *4*(3), 332–341. https://doi.org/10.1111/j.1756-8765.2012.01206.x.

Akaike, H. (1974). A new look at the statistical model identification. In Parzen, E., Tanabe, K., Kitagawa, G. (Eds.) *Selected papers of Hirotugu Akaike* (pp. 215–222). New York: Springer.

Annis, J., Evans, N.J., Miller, B.J., Palmeri, T.J. (2019). Thermodynamic integration and steppingstone sampling methods for estimating Bayes factors: a tutorial. *Journal of Mathematical Psychology*, *89*, 67–86. https://doi.org/10.1016/j.jmp.2019.01.005.

Brown, S.D., & Heathcote, A. (2005). A ballistic model of choice response time. *Psychological Review*, *112*(1), 117. https://doi.org/10.1037/0033-295x.112.1.117.

Brown, S.D., & Heathcote, A. (2008). The simplest complete model of choice response time: linear ballistic accumulation. *Cognitive Psychology*, *57*(3), 153–178. https://doi.org/10.1016/j.cogpsych.2007.12.002.

Browne, M.W. (2000). Cross-validation methods. *Journal of Mathematical Psychology*, *44*(1), 108–132. https://doi.org/10.1006/jmps.1999.1279.

Chambers, C.D., Feredoes, E., Muthukumaraswamy, S.D., Etchells, P. (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.

Chambers, C.D., Dienes, Z., McIntosh, R.D., Rotshtein, P., Willmes, K. (2015). Registered reports: realigning incentives in scientific publishing. *Cortex*, *66*, A1–A2. https://doi.org/10.1016/j.cortex.2015.03.022.

Cisek, P., Puskas, G.A., El-Murr, S. (2009). Decisions in changing conditions: the urgency-gating model. *Journal of Neuroscience*, *29*(37), 11560–11571. https://doi.org/10.1523/jneurosci.1844-09.2009.

Cohen, J. (1965). Some statistical issues in psychological research. In Wolman, B.B. (Ed.) *Handbook of clinical psychology* (pp. 95–121). New York: McGraw-Hill.

Crüwell, S., van Doorn, J., Etz, A., Makel, M.C., Moshontz, H., Niebaum, J., Orben, A., Parsons, S., Schulte-Mecklenbeck, M. (2018). 7 easy steps to open science: An annotated reading list. PsyArXiv. https://doi.org/10.31234/osf.io/cfzyx.

de Groot, A.D. (2014). The meaning of "significance" for different types of research [translated and annotated by Eric-Jan Wagenmakers, Denny Borsboom, Josine Verhagen, Rogier Kievit, Marjan Bakker, Angelique Cramer, Dora Matzke, Don Mellenbergh, and Han L J. van der Maas]. *Acta Psychologica*, *148*, 188–194. https://doi.org/10.1016/j.actpsy.2014.02.001.

Dougherty, M.R., Gettys, C.F., Ogden, E.E. (1999). MINERVA-DM: a memory processes model for judgments of likelihood. *Psychological Review*, *106*(1), 180. https://doi.org/10.1037/0033-295x.106.1.180.

Dutilh, G., Annis, J., Brown, S.D., Cassey, P., Evans, N.J., Grasman, R.P.P.P., Hawkins, G.E., Heathcote, A., Holmes, W.R., Krypotos, A.-M., Kupitz, C.N., Leite, F.P., Lerche, V., Lin, Y.-S., Logan, G.D., Palmeri, T.J., Starns, J.J., Trueblood, J.S., van Maanen, L., van Ravenzwaaij, D., Vandekerckhove, J., Visser, I., Voss, A., White, C.N., Wiecki, T.V., Rieskamp, J., Donkin, C. (2018). The quality of response time data inference: a blinded, collaborative assessment of the validity of cognitive models. *Psychonomic Bulletin & Review*, 1–19. https://doi.org/10.3758/s13423-017-1417-2.

Evans, N.J. (2019a). Assessing the practical differences between model selection methods in inferences about choice response time tasks. *Psychonomic Bulletin & Review*, 1–29. https://doi.org/10.3758/s13423-018-01563-9.

Evans, N.J. (2019b). What factors are most important in finding the best model of a psychological process? Comment on Navarro 2018. PsyArXiv. https://doi.org/10.31234/osf.io/rkyp3.

Evans, N.J., & Annis, J. (2019). Thermodynamic integration via differential evolution: a method for estimating marginal likelihoods. *Behavior Research Methods*, 1–18. https://doi.org/10.3758/s13428-018-1172-y.

Evans, N.J., & Brown, S.D. (2018). Bayes factors for the linear ballistic accumulator model of decision-making. *Behavior Research Methods*, *50*(2), 589–603. https://doi.org/10.3758/s13428-017-0887-5.

Evans, N.J., & Hawkins, G.E. (2019). When humans behave like monkeys: feedback delays and extensive practice increase the efficiency of speeded decisions. *Cognition*, *184*, 11–18. https://doi.org/10.1016/j.cognition.2018.11.014.

Evans, N.J., Hawkins, G.E., Boehm, U., Wagenmakers, E.-J., Brown, S.D. (2017a). The computations that support simple decision-making: a comparison between the diffusion and urgency-gating models. *Scientific Reports, 7*(1). https://doi.org/10.1038/s41598-017-16694-7.

Evans, N.J., Howard, Z.L., Heathcote, A., Brown, S.D. (2017b). Model flexibility analysis does not measure the persuasiveness of a fit. *Psychological Review*, *124*(3), 339–345. https://doi.org/10.1037/rev0000057.

Evans, N.J., Bennett, A.J., Brown, S.D. (2018a). Optimal or not; depends on the task. *Psychonomic Bulletin & Review*, 1–8.

Evans, N.J., Brown, S.D., Mewhort, D.J., Heathcote, A. (2018b). Refining the law of practice. *Psychological Review*, *125*(4), 592–605. https://doi.org/10.1037/rev0000105.

Evans, N.J., Steyvers, M., Brown, S.D. (2018c). Modeling the covariance structure of complex datasets using cognitive models: an application to individual differences and the heritability of cognitive ability. *Cognitive Science*, *42*(6), 1925–1944. https://doi.org/10.1111/cogs.12627.

Evans, N.J., Dutilh, G., Wagenmakers, E.-J., van der Maas, H.L. (2019a). Double responding: a new constraint for models of speeded decision making. PsyArXiv. https://doi.org/10.31234/osf.io/nf2q9.

Evans, N.J., Hawkins, G.E., Brown, S.D. (2019b). The role of passing time in decision-making. *Journal of Experimental Psychology, Learning, Memory, & Cognition*. https://doi.org/10.31234/osf.io/3wq6g.

Gronau, Q.F., Sarafoglou, A., Matzke, D., Ly, A., Boehm, U., Marsman, M., Leslie, D.S., Forster, J.J., Wagenmakers, E.-J., Steingroever, H. (2017). A tutorial on bridge sampling. *Journal of Mathematical Psychology*, *81*, 80–97. https://doi.org/10.31222/osf.io/m8ujg.

Gronau, Q.F., & Wagenmakers, E.-J. (2019). Limitations of Bayesian leave-one-out cross-validation for model selection. *Computational Brain & Behavior*, *2*(1), 1–11. https://doi.org/10.1007/s42113-018-0011-7.

Haven, T.L., & Grootel, D.L.V. (2019). Preregistering qualitative research. *Accountability in Research*, 1–16. https://doi.org/10.1080/08989621.2019.1580147.

Ioannidis, J.P. (2005). Why most published research findings are false? *PLoS Medicine*, *2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124.

Janczyk, M., & Lerche, V. (2019). A diffusion model analysis of the response-effect compatibility effect. *Journal of Experimental Psychology: General*, *148*(2), 237–251. https://doi.org/10.1037/xge0000430.

Jones, M.N., & Mewhort, D.J. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, *114*(1), 1–37. https://doi.org/10.1037/0033-295x.114.1.1.

Kass, R.E., & Raftery, A.E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572.

Kerr, N.L. (1998). HARKing: hypothesizing after the results are known. *Personality and Social Psychology Review*, *2*(3), 196–217. https://doi.org/10.1207/s15327957pspr0203_4.

Klein, O., Hardwicke, T.E., Aust, F., Breuer, J., Danielsson, H., Mohr, A.H., IJzerman, H., Nilsonne, G., Vanpaemel, W., Frank, M.C. (2018). A practical guide for transparency in psychological science. *Collabra: Psychology*, *4*, 1. https://doi.org/10.31234/osf.io/rtygm.

Lee, M.D., Criss, A.H., Devezer, B., Donkin, C., Etz, A., Leite, F.P., Matzke, D., Rouder, J.N., Trueblood, J.S., White, C.N., Vandekerckhove, J. (2019). Robust modeling in cognitive science. Computational Brain & Behavior. https://doi.org/10.1007/s42113-019-00029-y.

Lerche, V., Christmann, U., Voss, A. (2019). Impact of context information on metaphor elaboration. *Experimental Psychology*, *65*(6), 370–384. https://doi.org/10.1027/1618-3169/a000422.

Lewandowsky, S. (2019). PSprereg: a permanent home and homage to the twitterati. Blog Post.

Lindley, D.V. (1972). *Bayesian statistics, a review* Vol. 2. Philadelphia: Society for Industrial and Applied Mathematics.

McClelland, J.L. (2009). The place of modeling in cognitive science. *Topics in Cognitive Science*, *1*(1), 11–38. https://doi.org/10.1111/j.1756-8765.2008.01003.x.

McIntosh, R.D. (2017). Exploratory reports: a new article type for Cortex. *Cortex*, *96*, A1–A4. https://doi.org/10.1016/j.cortex.2017.07.014.

Meehl, P.E. (1967). Theory-testing in psychology and physics: a methodological paradox. *Philosophy of Science*, *34*(2), 103–115. https://doi.org/10.1086/288135.

Mertens, G., & Krypotos, A.-M. (2019). Preregistration of secondary analyses. PsyArXiv. https://doi.org/10.31234/osf.io/ph4q7.

Munafò, M.R., Nosek, B.A., Bishop, D.V., Button, K.S., Chambers, C.D., Du Sert, N.P., Simonsohn, U., Wagenmakers, E.-J., Ware, J.J., Ioannidis, J.P. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), 0021. https://doi.org/10.1038/s41562-016-0021.

Myung, I.J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology*, *44*(1), 190–204. https://doi.org/10.1006/jmps.1999.1283.

Myung, I.J., & Pitt, M.A. (1997). Applying occam's razor in modeling cognition: a Bayesian approach. *Psychonomic Bulletin & Review*, *4*(1), 79–95. https://doi.org/10.3758/bf03210778.

Myung, I.J., Navarro, D.J., Pitt, M.A. (2006). Model selection by normalized maximum likelihood. *Journal of Mathematical Psychology*, *50*(2), 167–179. https://doi.org/10.1016/j.jmp.2005.06.008.

Navarro, D.J. (2019). Between the devil and the deep blue sea: tensions between scientific judgement and statistical model selection. *Computational Brain & Behavior*, *2*(1), 28–34. https://doi.org/10.31234/osf.io/39q8y.

Nosek, B.A.N., & Lindsey, S.E. (2018). Preregistration becoming the norm in psychological science. *APS Observer*, *31*, 3.

Nosofsky, R.M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39. https://doi.org/10.1037/0096-3445.115.1.39.

Nuijten, M.B. (2018). Practical tools and strategies for researchers to increase replicability. *Developmental Medicine & Child Neurology*, *61*(5), 535–539. https://doi.org/10.1111/dmcn.14054.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. https://doi.org/10.1126/science.aac4716.

Palestro, J.J., Weichart, E., Sederberg, P.B., Turner, B.M. (2018). Some task demands induce collapsing bounds: evidence from a behavioral analysis. *Psychonomic Bulletin & Review*, *25*(4), 1225–1248. https://doi.org/10.3758/s13423-018-1479-9.

Pashler, H., & Wagenmakers, E.-J. (2012). Editors' introduction to the special section on replicability in psychological science: a

crisis of confidence? *Perspectives on Psychological Science*, *7*(6), 528–530. https://doi.org/10.1177/1745691612465253.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. https://doi.org/10.1037/0033-295x.85.2.59.

Ratcliff, R., & Rouder, J.N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*(5), 347–356. https://doi.org/10.1111/1467-9280.00067.

Ratcliff, R., & Rouder, J.N. (2000). A diffusion model account of masking in two–choice letter identification. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 127–140. https://doi.org/10.1037/0096-1523.26.1.127.

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review*, *9*(3), 438–481. https://doi.org/10.3758/bf03196302.

Ratcliff, R., Thapar, A., McKoon, G. (2001). The effects of aging on reaction time in a signal detection task. *Psychology and Aging*, *16*(2), 323–341. https://doi.org/10.1037/0882-7974.16.2.323.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, *107*(2), 358–367. https://doi.org/10.1037/0033-295x.107.2.358.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464. https://doi.org/10.1214/aos/1176344136.

Shiffrin, R.M., & Steyvers, M. (1997). A model for recognition memory: REM–retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4*(2), 145–166. https://doi.org/10.3758/bf03209391.

Simmons, J.P., Nelson, L.D., Simonsohn, U. (2011). False-positive psychology. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(4), 583–639. https://doi.org/10.1111/1467-9868.00353.

Sterling, T.D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance–or vice versa. *Journal of the American Statistical Association*, *54*(285), 30. https://doi.org/10.2307/2282137.

Tennant, J.P., Waldner, F., Jacques, D.C., Masuzzo, P., Collister, L.B., Hartgerink, C.H. (2016). The academic, economic and societal impacts of open access: an evidence-based review. *F1000Research*, *5*, 632. https://doi.org/10.12688/f1000research.8460.2.

Teodorescu, A.R., & Usher, M. (2013). Disentangling decision models: from independence to competition. *Psychological Review*, *120*(1), 1–38. https://doi.org/10.1037/a0030776.

Thura, D., Beauregard-Racine, J., Fradet, C.-W., Cisek, P. (2012). Decision making by urgency gating: theory and experimental support. *Journal of Neurophysiology*, *108*(11), 2912–2930. https://doi.org/10.1152/jn.01071.2011.

Usher, M., & McClelland, J.L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological Review*, *108*(3), 550–592. https://doi.org/10.1037/0033-295x.108.3.550.

Vehtari, A., Gelman, A., Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4.

Vickers, D., & Lee, M.D. (2000). Dynamic models of simple judgments: II. Properties of a self-organizing PAGAN (Parallel, Adaptive, Generalized Accumulator Network) model for multi-choice tasks. *Nonlinear Dynamics, Psychology, and Life Sciences*, *4*(1), 1–31. https://doi.org/10.1023/A:1009571011764.

Voskuilen, C., Ratcliff, R., Smith, P.L. (2016). Comparing fixed and collapsing boundary versions of the diffusion model. *Journal of Mathematical Psychology*, *73*, 59–79. https://doi.org/10.1016/j.jmp.2016.04.008.

Voss, A., Lerche, V., Mertens, U., Voss, J. (2019). Sequential sampling models with variable boundaries and non-normal noise: a comparison of six models. *Psychonomic Bulletin & Review*, 1–20. https://doi.org/10.3758/s13423-018-1560-4.

Wagenmakers, E.-J., & Evans, N. (2018). "Don't interfere with my art": on the disputed role of preregistration in exploratory model building. Blog Post.

Wagenmakers, E.-J., Ratcliff, R., Gomez, P., McKoon, G. (2008). A diffusion model account of criterion shifts in the lexical decision task. *Journal of Memory and Language*, *58*(1), 140–159. https://doi.org/10.1016/j.jml.2007.04.006.

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., Grasman, R. (2010). Bayesian hypothesis testing for psychologists: a tutorial on the Savage–Dickey method. *Cognitive Psychology*, *60*(3), 158–189. https://doi.org/10.1016/j.cogpsych.2009.12.001.

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H.L.J., Kievit, R.A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, *7*(6), 632–638. https://doi.org/10.1177/1745691612463078.

Weigard, A., & Huang-Pollock, C. (2017). The role of speed in ADHD-related working memory deficits: a time-based resource-sharing and diffusion model account. *Clinical Psychological Science*, *5*(2), 195–211. https://doi.org/10.1177/2167702616668320.

Weston, S.J., Mellor, D., Bakker, M., van den Akker, O., Campbell, L., Ritchie, S.J., Chopik, W.J., Damian, R.I., Kosie, J., Soderberg, C.K., Ebersole, C.R., Brown, B., Davis-Kean, P., Hall, A., Kruse, E., Olsen, J., Valentine, K., Nguyen, T. (2018). Preregistration of secondary data analysis template. https://osf.io/v4z3x/.

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: lessons from machine learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122. https://doi.org/10.1177/1745691617693393.