Preregistration in Complex Contexts:
A Preregistration Template for the Application of Cognitive Models

Sophia Crüwell[1,2] and Nathan J. Evans[2,3]

[1] Meta-Research Innovation Center Berlin (METRIC-B), QUEST Center for

Transforming Biomedical Research, Berlin Institute of Health, Charité –

Universitätsmedizin Berlin, Germany

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

[3] School of Psychology, University of Newcastle, Australia

Word count: 8,283

Correspondence concerning this article may be addressed to: Sophia Crüwell (sophia.cruewell@charite.de)

Abstract

In recent years, open science practices have become increasingly popular in psychology and related sciences. These practices aim to increase rigour and transparency in science as a potential response to the challenges posed by the replication crisis. Many of these reforms – including the highly influential *preregistration* – have been designed for experimental work that tests simple hypotheses with standard statistical analyses, such as assessing whether an experimental manipulation has an effect on a variable of interest. However, psychology is a diverse field of research, and the somewhat narrow focus of the prevalent discussions surrounding and templates for preregistration has led to debates on how appropriate these reforms are for areas of research with more diverse hypotheses and more complex methods of analysis, such as cognitive modelling research within mathematical psychology. Our article attempts to bridge the gap between open science and mathematical psychology, focusing on the type of cognitive modelling that Crüwell, Stefan, & Evans (2019) labelled *model application*, where researchers apply a cognitive model as a *measurement tool* to test hypotheses about parameters of the cognitive model. Specifically, we (1) discuss several potential researcher degrees of freedom within model application, (2) provide the first preregistration template for model application, and (3) provide an example of a preregistered model application using our preregistration template. More broadly, we hope that our discussions and proposals constructively advance the debate surrounding preregistration in cognitive modelling, and provide a guide for how preregistration templates may be developed in other diverse or complex research contexts.

Keywords:

Cognitive modelling; Reproducibility; Open science; Preregistration; Transparency

The replication crisis has been an issue for psychology and related fields since at least 2011 (Pashler & Wagenmakers, 2012), though many of the associated problems are likely much older (cf. e.g., Sterling, 1959; Cohen, 1965; Meehl, 1967). These problems have led to the proposal of a variety of reforms – often termed *open science* practices – which emphasise rigour, specificity, the constraint of flexibility, and transparency. These practices include data sharing (Klein et al., 2018), preregistration (Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012), and the journal article format Registered Reports (Chambers et al., 2014). The term *open science* is commonly used to refer to these practices as they encourage openness in the sense of transparent and accessible research (Crüwell et al., 2018), which in combination with specificity and constraint are essential to counteract the effect of cognitive biases and other pressures that may influence scientific findings (Munafò et al., 2017). The current article focuses on the open science practice of preregistration, which we discuss in more detail below, and how it might be implemented within cognitive modelling studies, which we discuss in more detail in the following sections.

Preregistration involves the specification of a researcher's plans for a study, including hypotheses and analyses, typically *before* the study is conducted. This usually takes the form of a document that contains these plans, which is made available online. Preregistration can help constrain researcher degrees of freedom (i.e., undisclosed flexibility in study design, data collection, and/or data analysis; Simmons, Nelson, & Simonsohn, 2011), and alleviate the effects of questionable research practices (QRPs) such as hypothesising after results are known (HARKing; Kerr, 1998) or p-hacking. This is important as each of these practices can render the interpretations of results based on seemingly confirmatory analyses invalid (Wagenmakers et al., 2012). Thus, while most published studies in psychology claim to be confirmatory, it may be difficult to know whether these studies truly are confirmatory without the a-priori specification of hypotheses and analysis plans, particularly given the incredibly high incidence of findings falling in line with the "confirmatory" predictions of psychology studies (Fanelli, 2010).

Ideally, the constraint imposed by preregistration should clearly distinguish between the exploratory and confirmatory steps within a research project (i.e., separate prediction from "postdiction"; Wagenmakers et al., 2012). However, the fact that a study is preregistered should not be taken as a marker of quality, as the preregistration document may lack the specificity needed to effectively constrain potential researcher degrees of freedom (Veldkamp et al., 2018), and the decisions made in the preregistration may not be well justified or appropriate (Szollosi et al., 2019). The Registered Reports format allows for an assessment of the quality of the pre-specified plan through an initial round of peer-review before the study is conducted, meaning that researchers can alter their pre-specified plans based on reviewer feedback (Chambers et al., 2014). However, this is not the case for the standard practice of preregistration, and many psychology journals do not currently include a Registered Report article format, meaning that researchers may initially struggle to create preregistration documents that are appropriately detailed and justified (Nosek, Ebersole, DeHaven, & Mellor, 2018). Several preregistration templates have been developed to assist researchers in creating preregistration documents, such as those provided by the Open Science Framework (OSF; https://osf.io/prereg/) and AsPredicted (https://aspredicted.org/), as well as checklists to assess the quality and constraint of a preregistration document (Wicherts et al., 2016). These templates and checklists have been designed as general-purpose tools for experimental psychology. Therefore, they are applicable to studies where researchers are interested in testing simple hypotheses, such as whether an experimental manipulation has an effect on a variable of interest, with simple analysis tools, such as a null hypothesis significance test on an interaction term within an ANOVA.

A large proportion of psychology studies fall within the standard experimental framework that these general-purpose templates and checklists have been designed to accommodate, making these tools of broad use to many researchers in psychology. However, psychology is a diverse field of research, and several areas of psychology commonly involve studies with more diverse hypotheses and more complex methods of analysis. Importantly,

the central focus of preregistration endeavours on purely experimental research has led to debates on how appropriate preregistration is for psychological research that is not purely experimental, particularly in the area of *cognitive modelling*, where researchers use mathematical models that are formal representations of cognitive processes to better understand human cognition (Wagenmakers & Evans, 2018; Lewandowsky, 2019; Lee et al., 2019; Crüwell et al., 2019; MacEachern & Van Zandt, 2019; Szollosi et al., 2019; Vandekerckhove et al., 2019). Although some question the general usefulness of preregistration in areas of psychology research with more diverse hypotheses and more complex analyses (MacEachern & Van Zandt, 2019; Szollosi et al., 2019), others believe that preregistration could still serve an important purpose in constraining researcher degrees of freedom (Wagenmakers & Evans, 2018; Lee et al., 2019; Crüwell et al., 2019). However, the preregistration tools currently available to researchers may make achieving proper constraint practically infeasible, as the exact researcher degrees of freedom in these areas of research can differ greatly from those in purely experimental psychology (Wagenmakers & Evans 2018; Lee et al. 2019; Crüwell et al. 2019; Vandekerckhove et al. 2019; though also see Arnold et al. 2019 for a cognitive modelling study with a well-constrained preregistration using existing tools). Recent research has already begun to create more specific preregistration templates for more specific areas of research, such as in qualitative research (Haven & Grootel, 2019; Kern & Skrede Gleditsch, 2017), experience sampling methodology (Kirtley et al., 2019), secondary data analysis (Mertens & Krypotos, 2019; Weston et al., 2018; van den Akker et al., 2019), and fMRI studies (Flannery, 2018). Therefore, the further development of method- and field-specific preregistration templates and checklists may improve the applicability of preregistration to areas of psychology research with more diverse hypotheses and more complex analyses, similar to how the development of general-purpose preregistration templates and checklists have helped researchers to create well-constrained preregistration documents for purely experimental studies.

Our article aims to bridge the gap between previous preregistration endeavours and research in areas of psychology with diverse hypotheses and complex analyses. At a general

level, we wish to showcase that preregistration can be used to constrain potential researcher degrees of freedom in more complex research contexts, and that field specific preregistration templates are important for both understanding and constraining these degrees of freedom. More specifically, we develop an initial preregistration template for cognitive modelling research, building on the initial suggestions of Lee et al. (2019), who proposed the concept of "registered modelling reports", as well as Crüwell et al. (2019), who proposed four potential categories for different types of cognitive modelling and suggested that different categories are likely to have different researcher degrees of freedom (i.e., different categories require different templates). Our article focuses on the category of cognitive modelling that Crüwell et al. (2019) labelled *model application*, where researchers apply a cognitive model as a *measurement tool* to test hypotheses about parameters of the cognitive model, such as assessing whether an experimental manipulation has an effect on a parameter of interest.

The remainder of this article will take the following format. First, we explain the concept of cognitive modelling, as well as the category of model application and how it differs from other categories of cognitive modelling. Second, we detail several additional researcher degrees of freedom that we believe are relevant to model application, and how they can make the preregistration process more complicated than in purely experimental work. Third, we provide a preregistration template for model application, including an example implementation to showcase how it can help to constrain researcher degrees of freedom. Fourth, we discuss the potential limitations of our template – most notably, that our template only covers model application, and that cognitive modelling studies often involve an interplay between different categories of cognitive modelling – and future directions for developing templates for other categories of cognitive modelling. We hope that our discussions and proposals will constructively advance the debate surrounding preregistration in cognitive modelling, and provide a guide for how preregistration templates may be developed in other diverse or complex research contexts.

Cognitive Modelling

Cognitive modelling is the formal description of theories about psychological processes (Farrell & Lewandowsky, 2018). Unlike statistical models, cognitive models contain parameters that have psychological interpretations (Busemeyer & Diederich, 2010), and the assumptions of cognitive models are designed to align with actual human behaviour. Cognitive models do not intend to provide perfect representations of psychological processes, but rather useful formal instantiations of psychological theory (i.e., "All models are wrong, but some are useful."; Box, 1979, p.202), which can enable more precise insight than is possible using verbal theories (McClelland, 2009). Importantly, "cognitive modelling" serves as an umbrella term that comprises diverse hypotheses and methods of analysis, with cognitive models being capable of serving a variety of different functions to answer a range of different research questions, spanning the spectrum from purely exploratory to purely confirmatory.

Although the diversity within cognitive modelling research is undoubtedly a positive feature, as it allows for cognitive models to provide unique insights into psychological processes in a range of different contexts, this diversity can also lead to some potential pitfalls. One key example is the increasingly discussed abundance of "modeler's degrees of freedom" (Dutilh et al., 2018), where differences in approaches and interpretations – for example, in interpreting model success and model failure (McClelland, 2009; Roberts & Pashler, 2000) – can influence the results and conclusions of a study. Furthermore, the potential degrees of freedom in approaches and interpretations are not always transparent, which can impede the effectiveness of the research process, potentially leading to issues such as the rare reuse of existing models (Addyman & French, 2012). Importantly, these issues are precisely what many open science practices – particularly preregistration – have been designed to address in purely experimental areas of psychology research. Therefore, it seems possible that preregistration could be a useful tool for at least some parts of cognitive modelling research, which could reduce the potential modeler's degrees of freedom and allow for a more transparent research process.

Despite the intuitive appeal of preregistration as a possible solution to the abundance of modeler's degrees of freedom within cognitive modelling, the potential introduction of preregistration to cognitive modelling has been a contentious issue (Wagenmakers & Evans, 2018; Lewandowsky, 2019; Lee et al., 2019; Crüwell et al., 2019; MacEachern & Van Zandt, 2019; Szollosi et al., 2019; Vandekerckhove et al., 2019). One potential reason for this debate is the previously mentioned diversity within cognitive modelling, where cognitive models can be used in a range of different contexts to answer a range of different research questions. Importantly, putting a single set of constraints on cognitive modelling across all contexts would greatly reduce the modeller's degrees of freedom, but would also prevent much of the diversity that make cognitive models such a useful instrument within psychology research. Furthermore, different contexts are likely to have different relevant degrees of freedom, meaning that a single set of constraints for the entirety of cognitive modelling would likely be inappropriate and ineffective. As a potential first step to allowing the constraint of modeller's degrees of freedom, while also maintaining the diversity in cognitive modelling, Crüwell et al. (2019) proposed that it may be useful to separate cognitive modelling research into four discrete categories: *model development*, *model evaluation*, *model comparison*, and *model application*. Model development involves the initial development of a model, or the extension/reduction of an existing model to create a new model, which is often an iterative, exploratory process. Model evaluation involves assessing whether a model, or multiple models, can qualitatively capture certain trends in empirical data, with these trends often referred to as "qualitative benchmarks". Model comparison involves directly contrasting multiple models on their ability to account for a set of empirical data, which is usually performed quantitatively through model selection methods (e.g., AIC, Akaike 1974; BIC, Schwarz 1978; see Evans 2019a for a discussion).

Model application, which will be the focus of our article, involves using an existing cognitive model to answer research questions about specific components of the underlying cognitive process. The chosen model is assumed to provide an accurate representation of the cognitive process underlying the data, and researchers are often interested in how

the components change over experimental conditions and/or groups, with these changes often having meaningful interpretations for related theories (e.g., task differences between older and younger adults; Ratcliff, Thapar, & McKoon, 2001). Model application involves using cognitive models in a similar manner to statistical models (e.g., ANOVA), though the assessments are performed on the theoretically meaningful parameters estimated within the cognitive model, rather than the variables directly observed within the data, creating several degrees of freedom that are not present in purely experimental research using only statistical models. However, model application is the category of cognitive modelling that is most closely related to the purely experimental research that previous preregistration efforts have been focused on, particularly as it is highly confirmatory in nature, making it the ideal category for initial preregistration efforts within cognitive modelling. In the next section we will provide a concrete discussion of potential modeller's degrees of freedom within model application – degrees of freedom that are not present in purely experimental research – which will form the basis for our preregistration template.

Researcher Degrees of Freedom in Model Application

What makes a good preregistration, and thus a good preregistration template? There are a variety of proposals for high-quality preregistrations, which differ in the amount of detail required from the researcher regarding their exact research plan (e.g., van 't Veer & Giner-Sorolla, 2016; Simmons et al., 2017; Wicherts et al., 2016). Similarly, different preregistration templates ask different questions, which differ in their level of detail and how much they prompt the researcher with specific instructions. Broadly, a good preregistration should make it easier to differentiate between exploratory and confirmatory research, by transparently constraining researcher degrees of freedom (i.e., flexibility in study design, data collection, analysis, and reporting; Wicherts et al., 2016; Simmons et al., 2011). Previous research has suggested that preregistration templates can have meaningful differences on the quality of preregistrations. For example, a comparison between preregistration documents that used one of the two commonly used

OSF preregistration templates showed that the less detailed OSF Standard Pre-Data Collection Registration template (since replaced by an improved standard form) led to less specific preregistrations than the more detailed OSF Prereg Challenge template, with the latter being better at restricting researcher degrees of freedom (Veldkamp et al., 2018).

Veldkamp et al. (2018) also proposed a stringent checklist for assessing the quality of preregistrations by checking whether all applicable researcher degrees of freedom are appropriately restricted by the preregistration. On a general level, it has been emphasised that a preregistration should lead to a plan that is specific (all steps to be taken are included), precise (each step is unambiguous), and exhaustive (there is no room left for other steps to be taken), so as to restrict the use of researcher degrees of freedom (Wicherts et al., 2016). Using standard templates, researchers in purely experimental areas of research may be better able to create a specific, precise, and exhaustive preregistration that restricts the researcher degrees of freedom present in these experimental contexts. However, restricting these commonly discussed generic researcher degrees of freedom is not sufficient for a successful preregistration in more complex contexts, such as model application. In fact, there have been cognitive modelling preregistrations submitted to the OSF that mentioned no more than the general modelling approach[1], which is understandable given that these preregistration templates were not designed for cognitive modelling studies.

In order to appropriately apply preregistration to research contexts that are not purely experimental, it is important to identify the unique researcher degrees of freedom within the area of research that the preregistration should ideally constrain. Below, we identify several degrees of freedom that we believe are present within model application, and then provide a more detailed discussion of these degrees of freedom and why they are important. Note that many previously identified researcher degrees of freedom – such as studying a vague hypothesis, ad-hoc exclusion of outliers, or specifying the pre-processing of data in an ad-hoc manner (Wicherts et al., 2016; Simmons et al., 2011) – are also applicable to model application, meaning that previous lists of researcher degrees of freedom should be *extended*

---

[1]Note that we do not cite specific example here, as we do not wish to "name and shame" researchers who have attempted to preregister their cognitive modelling studies.

for model application by adding the following category-appropriate modeller's degrees of freedom:

Cognitive Model

M1: Choosing a type of cognitive model.

M2: Specifying the exact parameterisation of the model(s).

M3: If applicable, specifying the (theoretical) motivation for these choices.

Parameter Estimation

E1: Deciding on the method of parameter estimation.

E2: Specifying settings/priors for parameter estimation.

E3: If the the data are going to be summarised into descriptive statistics, specifying which descriptive statistics will be used and how.

Statistical Inference

I1: Choosing a method of statistical inference on parameters (e.g., when comparing conditions). [This is largely covered by existing templates]

I2: Specifying what parameters will be assessed (e.g., allowed to vary across experimental conditions).

Robustness Checks

RC1: Performing and/or reporting robustness checks.

Issues and Peculiarities in Preregistering Model Application

The concrete list of modeller's degrees of freedom above provides an indication of what factors should be constrained by an exhaustive preregistration in model application, beyond the general researcher degrees of freedom. Here, we discuss these factors and some potential

pitfalls surrounding them, including the pre-specification of theoretical motivations, model parameterisations, estimation methods, preexisting data, outlier exclusions, parameter recovery studies, and robustness checks, as well as alternative minimal standards (i.e., when preregistration is not possible or not desired) for transparency in model application.

First and foremost, it is important that the choices made at all stages are justified and motivated, which is often not the case in cognitive modelling studies (Dutilh et al., 2018). The supplementary materials of Dutilh et al. (2018) contain a collection of analysis decisions and motivations of different teams analysing the same datasets using different methods[2]. While many of the conclusions made by different teams were similar, the discrepancies caused by each team's unique choices reveal how modeller's degrees of freedom can influence inferences, making well-motivated choices crucial. Considering that one of the main advantages of cognitive modelling is generalisability based on substantive explanations for psychological phenomena (Busemeyer & Diederich, 2010), one could argue that it is also a modeller's degree of freedom to not prespecify motivations or justifications for different decisions. For example, failing to specify which assumptions are central theoretical components of a model (i.e., the *core* assumptions) and which are only included out of practical necessity (i.e., the *ancillary* assumptions) can lead to situations where core assumptions can be written off as ancillary assumptions when they perform poorly, or do not produce the desired results (see Evans, 2019b, for a discussion). Similarly, when an assumption is not theoretically motivated, it should be clearly noted to avoid confusion or overinterpretation (Cooper & Guest, 2014). Therefore, preregistration seems to be useful both for the a-priori justification of choices, and for the clarification of which choices do not have clear justifications.

Furthermore, it is crucial to clearly pre-specify model(s) of interest – "the players of the game" (Lee et al., 2019) – as well as the specific parameterisation(s) of each model of interest, as there are a large number of possible variants of any model that may be applied (Heathcote et al., 2015). An ideal preregistration would include a complete plate diagram

---

[2]These supplementary materials can be found at `http://osf.io/egrnn`.

(or equivalent), such as Figure 1 below, where the model and parameterisation is specified, as well as any explicit dependencies between parameters, and any restrictions placed on the values of the parameters (see Appendix A). For example, if Bayesian hierarchical modelling is used for parameter estimation, the structure of the hierarchical model and the prior distributions over the parameters should be included within the parameterisation. Any post-hoc addition to, or modulation of, the model or the parameterisation should be clearly labelled as exploratory rather than confirmatory. Otherwise, the ability to adjust these factors can – at least in some cases – allow cognitive models to produce any possible result (cf Jones & Dzhafarov, 2014b,a), which would arguably be the ultimate degree of freedom. An effective preregistration of the model and the model parameterisation can constrain these potential researcher degrees of freedom, which would help ensure more rigorous model application work.

Following the model and its parameterisation, the next subsection of the preregistration template for model application is concerned with parameter estimation. In any situation when parameters are estimated, researchers should be clear on how they will estimate the parameters, and how the estimated parameters will be used as part of the analyses. However, in many cases researchers might only be interested in parameter estimation as a means of getting to the inference measure, as is the case for many Bayes factor approximation methods (Kass & Raftery, 1995; Annis et al., 2019; Evans & Annis, 2019; Gronau et al., 2017), which should be explicitly stated. Moreover, it is not always necessary to estimate the parameters to obtain an inference measure (see Evans & Brown, 2018), and in these cases the parameter estimation section may be replaced with a statement that explains why parameter estimation is not of interest here, meaning that any subsequent estimation of parameters would belong in an exploratory section.

One part of standard preregistration documents that will likely need to be adapted for many model application studies is information about original data collection, as many cognitive modelling studies reuse existing data. Preregistration of secondary data analyses is a challenge in itself, which has begun to be addressed with several proposed templates (see

Mertens & Krypotos, 2019; Weston et al., 2018; van den Akker et al., 2019). Although such a preregistration will not function in quite the same way as it is meant to with original data, the added transparency of openly stating pre-existing knowledge will hopefully provide similar credibility, and at the very least enable the reader to put the results into context (Nosek et al., 2018). Therefore, our preregistration template integrates parts of the preregistration template for secondary data proposed by Weston et al. (2018). In cases where existing data sets are used, the sections on conducting and designing the study – "Sampling Plan", "Design Plan", and "Variables" – should be replaced by the section "Data Description for Pre-existing Data".

Moreover, there are two aspects already included in general preregistration templates that still deserve a separate discussion in the context of cognitive modelling in general and model application in particular. Firstly, a necessary pre-specification that is already included in most preregistration templates is the decision on how outliers are excluded and data are pre-processed. In purely experimental studies, this is usually focused at the level of entire participants. This becomes more intricate in model application studies, as models are often estimated using the combined information from individual trials, meaning that researchers may want to exclude individual trials as well as entire participants from data analysis. Therefore, researchers conducting model application studies should clearly specify exclusion criteria for both individual trials and entire participants, and in which order these exclusions should be applied, as the exclusion of a subset of trials could influence whether or not a participant is excluded by the participant exclusion criteria. Secondly, in purely experimental studies, the aspect of power analysis is usually focused on the classic concept of statistical power (i.e., the probability of rejecting the null hypothesis given that it is false, within a null hypothesis significance testing framework). This also becomes more intricate within model application, as the analysis process is not simply a statistical test, and studies often use inference techniques other than null hypothesis significance testing. Other factors within model application are highly dependent on the experimental design, however, such as the precision of the parameters estimated from the cognitive model. Therefore, when

considering preregistration in model application, the classical power analysis used within purely experimental contexts could potentially be replaced with a parameter recovery study that matches the preregistered experimental design, model, and model parameterisation (Heathcote et al., 2015).

Although we believe that preregistration shows clear promise as a useful tool within model application, we can understand how some researchers within the field of mathematical psychology may still be skeptical regarding its utility. One previous argument against the utility of preregistration has been that cognitive modelling research has other, superior practices for ensuring the robustness of their findings, such as robustness checks (see e.g., a tweet by Morey, 2018). However, we disagree with the stance that robustness checks provide an alternative to preregistration, as the performing and reporting of robustness checks provide a large potential degree of freedom. Specifically, if researchers were to only perform or report the robustness checks that were successful in showing their results to be robust, then readers would likely become overconfident in the results of the study, as the findings were robust against all reported robustness checks. Therefore, although we agree that robustness checks are an important part of model application (and more generally, cognitive modelling research), we disagree that they are an alternative to preregistration, and instead are an aspect of the study that should be included within a preregistration document.

There may be other good arguments against the use of preregistration in model application, and researchers might be interested in alternative methods that also provide increased transparency. Any researcher looking for an alternative to preregistration could consider the provision of well-documented code as a minimal standard for making their study more transparent. In fact, writing and posting well-documented analysis code on an online repository before the analysis is conducted could even be considered a minimal preregistration, as it would function to constrain many of the degrees of freedom that we discussed above. Nevertheless, it should be noted that without sufficient documentation and comments the analysis code may be difficult to use, and thus, the results may be difficult

to reproduce. Therefore, researchers who choose this alternative to preregistration should invest a reasonable amount of time into developing well-documented code.

A Template for Preregistration in Model Application

Taking the previous considerations into account, we developed a template for preregistration in model application, which can either be used in the context of standard preregistration, or as a basis for the Registered Modelling Reports journal format suggested by Lee et al. (2019), which builds on the conventional Registered Reports journal format (Chambers et al., 2014). Note that we are *not* attempting to claim that this is the only way that researchers should preregister studies involving model application, or that researchers must use preregistration in model application studies. Rather, we believe that our preregistration template may be a useful tool to help ensure robustness and transparency in model application studies. It should also be noted that there is already at least one example of a modelling study using a detailed preregistration (Arnold, Heck, Bröder, Meiser, & Boywitt, 2019), meaning that our template is not the only method for creating a highly constraining preregistration document in model application. However, this study by Arnold et al. (2019) seems to be the exception to the rule: preregistration in model application appears to be quite rare, and most other preregistration documents in model application do not sufficiently constrain the types of modeller's degrees of freedom that we discuss above. This suggests that a preregistration template specific to model application and constraining the appropriate modeller's degrees of freedom could be of great use.

Note that to make our preregistration template as concrete as possible, each part of our preregistration template will be accompanied by an example related to our example application. The details of our example application can be found in the next section.

Preregistration Template

Taking into account the modeller's degrees of freedom and the potential issues that we discussed earlier, we combined the preregistration template "OSF Prereg"[3] with parts of the secondary data analysis template (Weston et al., 2018), and used this as a basis for our model application preregistration template. Based on our discussion of the unique modeller's degrees of freedom in model application, our template also involved adapting, removing, and adding some sections. We discuss the sections that we have added below, and the full template can be seen in Appendix B.

It should also be noted that in order for a preregistration to be exhaustive, it is important to repeatedly use clarifying words such as "only" to constrain the choices made to *only* those mentioned in the preregistration (Wicherts et al., 2016). A template for preregistration can be helpful in emphasising this, as it asks specific questions and hopefully encourages exhaustiveness. Therefore, the additions proposed here prompt answers that are as specific as possible. In particular, we added the following sections:

A. Cognitive Model (Required)

A.1 Please include the type of model used (e.g., diffusion model, linear ballistic accumulator model), and a specific parameterisation/parameterisations.

A.2 **Example** *As in Evans, Bennett, & Brown (2018), the parameters of a simple diffusion model will be estimated, namely only: drift rate (v), starting point (z), threshold (a), non-decision time (ter). This differs from Evans & Brown (2017), where the full diffusion model was estimated, i.e., including between-trial variability parameters for drift rate, starting point, and non-decision time. These between-trial variability parameters were not relevant for Evans & Brown (2017), and without them, the simple diffusion model has better parameter recovery results (Lerche & Voss, 2016). Figure 1 shows a plate diagram of the hierarchical structure used for the qualitative*

---

[3]See: `http://docs.google.com/document/d/1DaNmJEtBy04bq1l5OxS4JAscdZEkUGATURWwnBKLYxk/edit`. This is a detailed preregistration template, and seems to be based on the detailed Prereg challenge template which was positively evaluated earlier.

model-based analysis assessing 1) whether groups appear to get closer to optimality over time, 2) whether each group differs from optimality, and 3) whether there appears to be a difference between the groups (see Analysis Plan for more information); i indexes participants, and j indexes blocks. Only the threshold parameter varies between blocks, to estimate changes in the speed accuracy trade-off.

A.3 **More information** The architecture of the model should be pre-specified in a way that is specific, precise, and exhaustive. To this end, you should ideally include a plate diagram (see e.g., Figure 1) and specify the relevant equations. Motivate your choices. Note: If you are using e.g. Bayesian hierarchical modelling for parameter estimation, the structure of the hierarchical model and the prior distribution over the parameters belong into this parameterisation as well.

B. Method of Parameter Estimation (Required)

B.1 Please specify and motivate your method of parameter estimation.

B.2 **Example** *Only Bayesian hierarchical modeling will be used to estimate the parameters of the diffusion model, constraining individual-level parameters to follow group-level truncated normal distributions. For the estimation model (see Figure 1), the two groups (fixed-trial and fixed-time) are given a separate hierarchical structure, and the group-level parameters are not constrained between groups. Following Evans & Brown (2017) and Evans et al. (2018), we will use likelihood functions taken from the "fast-dm" toolbox (Voss & Voss, 2007) for the calculation of the density function of the simple diffusion model. For the first model, for sampling from the posterior distributions over parameters, we will use Markov-chain Monte Carlo with differential evolution proposals (Turner et al., 2013), using 66 chains, drawing 3,000 samples from each, and discarding the first 1,500 samples (as in Evans & Brown, 2017, see supplementary materials).*

B.3 **More information** If you are not interested in the parameters and are going straight to statistical inference without estimating the parameters, please state this clearly and motivate this choice. If you are using Bayesian methods, specify and motivate priors. In general, specify as much as possible, including for example the starting point (distribution) for estimation. If the data are going to be summarised into descriptive statistics, state which descriptive statistics will be used, and how.

C. Robustness Checks and Sensitivity Analyses

C.1 Please specify any planned robustness checks and/or sensitivity analyses, if any.

C.2 **Example** *The key analysis will be replicated a) including participants/trials that were initially excluded in line with our exclusion criteria, and b) using a model in which the threshold parameter and the drift rate parameter vary across blocks. Their results will be mentioned alongside the key results, and interpreted accordingly. If these results show a lack of robustness, this will be an interesting outcome worthy of further investigation.*

C.3 **More information** This section ensures that robustness checks are not performed and/or reported selectively. It is important to note that, given the preregistration of modelling and analyses, it should be clear that any lack of robustness is at least not due to post-hoc, data-driven choices.

By adding these sections to a combination of the new standard, detailed OSF preregistration template and the template for secondary data use, and by adapting relevant sections to the needs of model application research, we provide a template for limiting researcher and modeler's degrees of freedom in model application. The section on cognitive models requires a choice of cognitive model and the specification of the exact parameterisation(s) used, and encourages the description of specific motivations for each choice. Any post-hoc addition or modulation is then clearly exploratory rather than confirmatory research. In the parameter section, the researchers are asked to provide
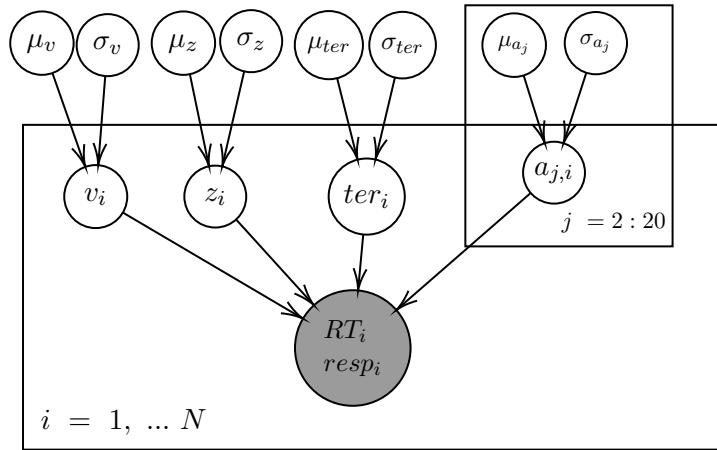
Figure 1. : Example of a plate diagram used to preregister the parameterisation of the model, taken from our example application; corresponding distributions can be found in the full preregistration in Appendix A.

information on the method of parameter estimation, including specific settings and/or priors used. The robustness section asks researchers to specify which robustness checks and sensitivity analyses will be performed, and how these will be interpreted. In the fully adapted and combined preregistration template (see Appendix B), researchers are further asked to specify for example how they are going to do statistical inference on which parameters.

Example Application

Example Application Background

Our example application focuses on applying evidence accumulation models (also commonly referred to as sequential sampling models), which have been hugely useful and influential in the (cognitive) psychology literature (Evans & Wagenmakers, 2019; Forstmann, Ratcliff, & Wagenmakers, 2016), and are thus an ideal focus for our discussions of preregistration in model application. Evidence accumulation models describe the fundamental process of making a decision between alternatives in the presence of noise (e.g., Ratcliff, 1978; Ratcliff & Smith, 2004; Brown & Heathcote, 2008; Usher & McClelland, 2001), where evidence accumulates for the different decision alternatives until the evidence

for one reaches a threshold, and a decision is made. These models provide an account of the well-known speed-accuracy trade-off, where decision-makers must strike a balance between speed and accuracy, as collecting more information leads to more accurate results at the cost of more time, whereas deciding more quickly might lead to more errors. The key parameters of evidence accumulation models are the *drift rate*, which reflects how good people are at the task, the *decision threshold*, which reflects how cautious people are at the task, and the *non-decision time*, which reflects how long is spent on perceptual and motor processes.

Our example application aimed to replicate and analytically extend a surprising result on the speed-accuracy trade-off in Evans & Brown (2017) using a previously collected data set. Evans & Brown (2017) used a random dot motion task with a fixed difficulty level for all decisions (in order to ensure that a single fixed boundary could produce an optimal threshold setting; Bogacz et al. 2006) to investigate whether participants can optimise their speed-accuracy trade-off – in terms of reward rate – in different settings. This involved experimentally varying the amount of feedback participants received and whether participants were given a fixed amount of time or a fixed number of trials in each block, and then estimating the parameters of the diffusion model using Bayesian hierarchical modelling. The main finding of Evans & Brown (2017) was that with enough practice and feedback, people were able to optimise the speed-accuracy trade-off, and that they do so faster with increasing amounts of feedback. Interestingly, they also found that participants who completed a fixed number of trials per block were closer to optimality than participants who completed trials for a fixed amount of time in each block, which is in conflict to previous research (Starns & Ratcliff, 2012). This surprising result is the focus of our example application; specifically, we aim to replicate the superiority of "fixed-trial" conditions over "fixed-time" conditions in leading participants closer to reward rate optimality, using a different data set and updated analysis methods.

As all feedback groups in Evans & Brown (2017) showed the same general effect, the focus in this example dataset is on the effects of a "medium" amount of feedback for fixed

trials and fixed time groups, replicating the middle row of their Figure 3. The assessment in Evans & Brown (2017) was rather qualitative and not very rigorously defined, making this a good opportunity to show how preregistration in cognitive modelling can add rigor and transparency in situations with many potential researcher degrees of freedom. Furthermore, we analytically extended this replication by calculating Bayes factors to more robustly establish whether the groups differ in their distance from optimality, and if so, determine the strength of evidence in favour of the effect (which is not provided by the method in Evans & Brown 2017).

Example Application Sample & Materials

We used a subset of an existing dataset consisting of 70 participants who were recruited at the University of Newcastle and received course credit for their participation. The task used was the random dot kinetogram, following Evans & Brown (2017) and Evans et al. (2018). The participants were randomly (and equally) divided into the two groups of fixed trials and fixed time. Note that power analyses are not currently possible for the application of complex cognitive models, and more generally, the concept of power is only applicable within a significance testing framework with meaningful cut-off points between an effect being present and not being present. Instead, we planned to use Bayes factors for a more continuous, strength of evidence approach.

For all participants, the first block of trials was excluded to allow for participants to become adequately practiced at the task. Trials with response times below 150ms or above 10000ms were excluded as anticipatory responses and trials where participants lost attention, respectively[4]. Participants with task accuracy below 60% (following Evans et al., 2018) or less than 200 eligible trials (based on the number of trials required for accurate parameter estimation) were excluded. Applying these criteria resulted in the exclusion of 9 participants from the fixed-time condition (8 due to accuracy, 1 due to too few eligible trials), and 10 participants from the fixed-trial condition (all due to accuracy).

---

[4]Note that this is a particularly important aspect to preregister in model application, as a wide variety of "rules" for RT outlier exclusion can be found in the literature, and even within a single paper.

Example Application Preregistration

Our example application was preregistered at https://osf.io/39t5x/ (or see Appendix A for the full preregistration document). We include parts of the preregistraton here. For example, we preregistered the following hypotheses:

H1 With suitable practice and medium feedback (cf Evans & Brown, 2017), participants get closer to optimality with each block of trials. (directional)

H2 After suitable practice and medium feedback (cf Evans & Brown, 2017), participants will have an approximately optimal speed-accuracy trade-off. (directional)

H3 Participants who complete a fixed number of trials are closer to optimality than participants who complete trials in a fixed amount of time. (directional)

To test these hypotheses, we preregistered the following:

Testing H1/2/3: In addition to the statistical analyses, we will qualitatively compare the posterior distributions of the decision threshold parameters (actual thresholds for each block as estimated using the model in Figure 1) against the posterior predictive distributions for the optimal threshold calculated as above.

Testing H2: Using only the second half of all 20 blocks (11-20, so as to account for participants adjusting to the task), we will test whether each group, separately, differs from optimality using Bayes factors, approximated with the Savage-Dickey Ratio on μ_c (Wagenmakers et al., 2010, see Figure 2 for the corresponding plate diagram).

Testing H3: Again using only blocks 11-20, we will use the Savage-Dickey method on Δ_c ($\Delta_c = \mu_{c_1} - \mu_{c_2}$) to test whether the groups differ in their distance from optimality.

The preregistered models can be seen in Figures 1, 2, and 3, and further preregistered information such as on the distributions of the parameters and the relations between them, and on the parameter estimation method can be found in the full preregistration document (see Appendix A).
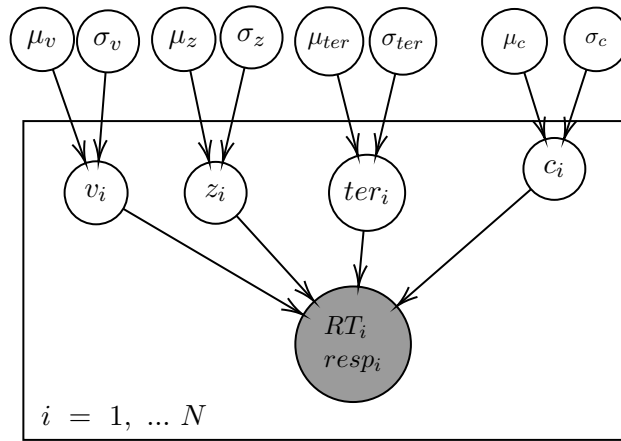
Figure 2. : Plate diagram of the hierarchical structure used for the quantitative model-based analysis to test whether each group differs from optimality using Bayes factors, approximated with the Savage-Dickey Ratio on μ_c. RT stands for reaction time, resp stands for response accuracy.
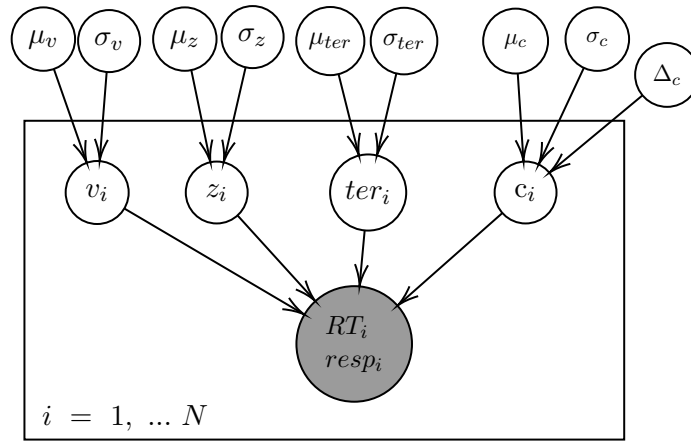

Figure 3. : Plate diagram of the hierarchical structure used for the quantitative model-based analysis to test the hypothesis whether the groups differ from each other in their difference from optimality. RT stands for reaction time, resp stands for response accuracy.
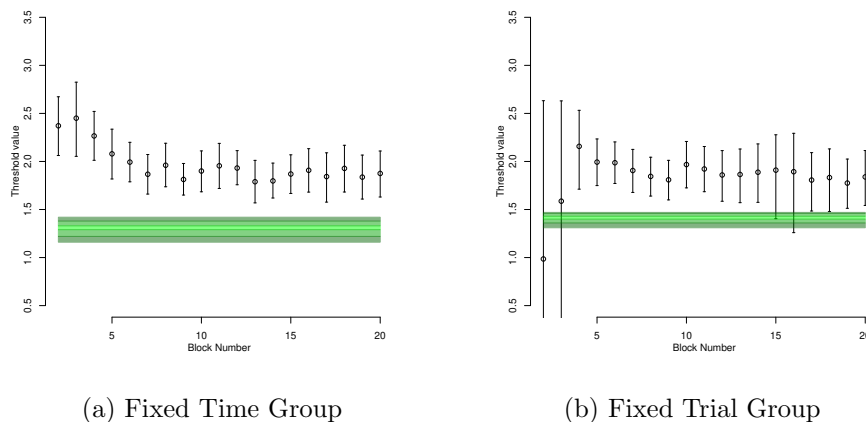
(a) Fixed Time Group   (b) Fixed Trial Group

Figure 4. : Plots comparing the posterior distributions of the decision threshold parameters against the posterior predictive distributions for the optimal threshold, for the fixed time group (left) and the fixed trial group (right).

Example Application Results

Figure 4 shows the posterior distributions of the decision threshold parameters against the posterior predictive distributions for the optimal threshold. We can qualitatively observe the same trend to optimality over time as in Evans & Brown (2017) and Evans et al. (2018). However, from this qualitative analysis, it is unclear whether the groups differ in the extent to which they move towards optimality. Testing each group separately reveals strong evidence for the participants of both groups being too cautious in their decision making ($BF_{Time} = 149$, $BF_{Trial} = 31.664$). Using the Savage-Dickey method on Δ_c leads to weak evidence for the groups not differing in their distance from optimality ($BF = 1.182$), albeit in the direction of the fixed trial group being closer to optimality than the fixed time group. Fixed-time participants completed an average of 25.66 trials per block, and fixed-trial participants took an average of 92.93 seconds per block.

Example Application Discussion

Our results do not fully replicate Evans & Brown (2017). In our replication, participants did move towards optimality given practice and feedback, but there was no

clear difference between the fixed-time and fixed-trial groups. This may be a result of our updated statistical analysis methods, as our qualitative pattern of results look similar to those from Evans & Brown (2017), and it is our new quantitative analyses that suggest that there is no evidence for a difference between groups. Regardless, our findings indicate that there is not necessarily a difference between fixed-trial blocks and fixed-time blocks in how close people are able to come to reward rate optimality, and therefore, this perceived difference in Evans & Brown (2017) should be interpreted with caution. It should also be noted that our findings showed strong evidence for participants being suboptimally cautious, which again is somewhat against the conclusions of Evans & Brown (2017), but makes sense in the context of Evans et al. (2018) who suggested that only specific experimental designs (e.g., slower trial-to-trial timing) will result in people achieving reward rate optimality.

General Discussion

In this article, we have proposed a concrete list of modeller's degrees of freedom, developed a preregistration template for model application, and showcased the possibility of preregistering model application studies with an example application. Our overarching goal was to display how preregistration templates can be developed within areas of psychology that have diverse hypotheses and complex analyses, such as cognitive modelling, with a more specific goal of making preregistration in model application more feasible. We also believe that our concrete list of modeller's degrees of freedom may be a useful resource for researchers when they consider which aspects of a model application study may be prone to flexibility, regardless of whether or not they choose to perform preregistration. However, as we argued previously, we believe that preregistration is the best tool currently available for constraining researcher degrees of freedom, and we believe that model application studies may benefit from the use of preregistration and our template. Importantly, we believe that our article provides a first step towards bridging the gap between open science and mathematical psychology, and we hope that our work will provide new insights to the debate on whether preregistration has utility within cognitive modelling research.

Nevertheless, deviations from a preregistration should always be possible, for example if the researcher gained important knowledge in between writing the preregistration and analysing the data, or if the data violate assumptions of the planned analyses – "the preregistration of a model and the way it will be used may not survive contact with the data" (Lee et al., 2019, p. 4). Furthermore, minor accidental omissions which likely do not affect the outcome can happen with any preregistration, and should be transparently reported. Although we believe that our preregistration template allowed us to create a highly constrained preregistration document for our example application, we wish to note that the amount of freedom available – and therefore, the amount of detail required in the preregistration – was at times surprising. Even after carefully going over the example application preregistration several times, we still accidentally omitted a minor detail about the process of parameter estimation: that we used a migration algorithm to assist convergence (as did Evans & Brown, 2017; Evans et al., 2018). We hope that the clear imperfection in our own example application demonstrates several important aspect of preregistration: that (1) a preregistration is not a prison, (2) it is always possible to make changes given that they are well motivated and reported in a transparent manner (DeHaven 2017; though see Claesen et al. 2019 for a critical perspective on the lack of transparent reporting of deviations from preregistered plans), and (3) creating preregistration documents is a learning process that improves with time. Similarly, exploratory work is important and should not suffer as a result of an increased move towards preregistered confirmatory work, and we believe that the interplay of unrestricted creativity and constrained verification remains important (Wagenmakers et al., 2018).

One potential critique of our preregistration template could be that the prompts are too open-ended. However, it has previously been found that a format with specific, open-ended questions is better at restricting researcher degrees of freedom than a purely open ended template (Veldkamp et al., 2018). Supporting this, it has been argued that some increase in transparency and constraint is better than none (Nuijten, 2019). Furthermore, templates that are too long or ask too many questions that are too specific might deter

researchers from using the template. Therefore, we believe that our preregistration template strikes a sensible balance between a completely open-ended template and an overly long template, as our template prompts researchers to disclose key information that relates to the modeller's degrees of freedom that we discuss within our article, but does not ask researchers hundreds of small questions regarding every potential detail of their study.

Limitations

First and foremost, it should be noted that our proposed preregistration template is an initial proposal of what a preregistration might look like in cognitive modelling, and specifically for the category of model application. We do not intend for our template to be a definitive answer to preregistration in cognitive modelling, or even in model application. Instead, our aim is to create an initial tool for researchers who are interested in preregistering their model application study, but are unsure of how to do so. Furthermore, although we attempted to make our template as generalizable as possible, it should be noted that many of our modeller's degrees of freedom were inspired by our previous experience working with evidence accumulation models – the class of models that we applied within our example application. Researchers more familiar with other classes of models may have different opinions on which degrees of freedom should be constrained within a model application preregistration document. It is also possible that researchers may experience unique issues that are not covered by our preregistration template when attempting to preregister studies that differ greatly from our example application, such as when using a different class of model. Therefore, we hope that others will critique and build upon our initial work, leading to the further development of a range of preregistration templates for cognitive modelling.

A second limitation of our preregistration template is that it only covers the cognitive modelling category of model application. As stated earlier, we believe that model application is an ideal category for initial preregistration efforts within cognitive modelling, as it is highly confirmatory in nature and is the most closely related to purely experimental research, meaning that it best falls in line with the focus of most previous

preregistration efforts. However, cognitive modelling studies often involve the interplay between several categories of cognitive modelling, meaning that some may question the utility of a preregistration template for a single category of cognitive modelling. For example, a researcher might want to assess whether a certain parameter estimated from the cognitive model differs over experimental conditions, but only if the model provides an adequate account of the data that the parameters are being estimated from, which would constitute an interplay of model application with model evaluation. We agree that the potential interplay between different categories of cognitive modelling limits the possible scope of our preregistration template, though as discussed earlier, we believe that our template is only an initial step towards preregistration in cognitive modelling, and that future research may be able to find solutions for the potential interplay between categories. In the above example, the researcher could combine our preregistration template for model application with a preregistration template for model evaluation. This would allow the researcher to evaluate the fit of the model in a constrained way based on a-priori defined criteria. If the fit is found to be inadequate, the researcher is able to make alterations to the model based on a-priori defined rules, thereby preventing any uneccessary deviations from the preregistration for the model application. Although the above scenario is not currently possible, as currently there is no preregistration template for model evaluation or guidelines for how to combine these templates, we believe that (1) the preregistration of the model application aspect of the study would be better than no preregistration at all, and (2) many cognitive models used in model application are already sufficiently developed to provide an accurate account of a range of data, meaning that the model evaluation step may not always be necessary.

Future Directions

When considering the future of preregistration in cognitive modelling, we believe that mathematical psychology is one of the fields of psychology best suited to creating constrained preregistration documents. Cognitive modelling is a highly theory-driven field of research

(van Rooij, 2019), and the formal nature of cognitive models means that they make precise predictions about empirical data, which when compared to the widespread lack of theory in other parts of psychology (Muthukrishna & Henrich, 2019) suggests that certain categories of cognitive modelling – model application, model comparison, and model evaluation – may lend themselves well to preregistration. However, we believe that this will only be the case when appropriate tools – such as preregistration templates – are available and tailored to cognitive modelling research. Therefore, we believe that future research efforts should focus on developing preregistration templates for the other categories of cognitive modelling proposed by Crüwell et al. (2019), either through extending our preregistration template for model application or by creating new preregistration templates. With specific templates for each category, a project including more than one modelling category could use the templates for each category to constrain the degrees of freedom in each part of the process, either simultaneously or sequentially. For example, a researcher might preregister a model comparison between a series of candidate models to decide which best accounts for the data, and after conducting this comparison might preregister a model evaluation for each of the models on a series of data trends to determine why the best model performed better than the other models. We believe that this system of iterative preregistrations for different categories of cognitive modelling within a single study provides the ideal balance between constraint and diversity, as researchers are free to investigate the data in as much detail as they wish, but each analysis performed is constrained.

At a general level, we believe that there should be an increased focus on the development of preregistration templates for specific fields and/or methods of research. Although excessively general templates might seem appealing, as they appear to unite different fields of research within a single framework, these general templates will often lack the specificity to properly constrain the researcher degrees of freedom for each field. No preregistration is perfect, and using an overly general template – or one designed for a different field of research – is likely better than using no template or eschewing preregistration entirely. Nevertheles, we believe that future preregistration efforts should

focus on developing specific preregistration templates for more fields of research, so that researchers can avoid unnecessarily decreasing their ability to effectively constrain researcher degrees of freedom.

References

Addyman, C., & French, R. M. (2012). Computational modeling in cognitive science: A manifesto for change. *Topics in Cognitive Science*, *4*(3), 332-341.

Akaike, H. (1974). A new look at the statistical model identification. In *Selected papers of hirotugu akaike* (pp. 215–222). 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.

Arnold, N. R., Heck, D. W., Bröder, A., Meiser, T., & Boywitt, C. D. (2019). Testing hypotheses about binding in context memory with a hierarchical multinomial modeling approach. *Experimental Psychology*.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Eeview*, *113*(4), 700.

Box, G. E. (1979). Robustness in the strategy of scientific model building. In *Robustness in statistics* (pp. 201–236). Elsevier.

Brown, S., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*(3), 153–178.

Busemeyer, J. R., & Diederich, A. (2010). *Cognitive modeling.* Sage.

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.

Claesen, A., Gomes, S. L. B. T., Tuerlinckx, F., & vanpaemel, w. (2019). *Preregistration: Comparing dream to reality.* Retrieved from `psyarxiv.com/d8wex`

Cohen, J. (1965). Some statistical issues in psychological research. *Handbook of Clinical Psychology*, 95–121.

Cooper, R. P., & Guest, O. (2014). Implementations are not specifications: Specification, replication and experimentation in computational cognitive modeling. *Cognitive Systems Research*, *27*, 42–49.

Crüwell, S., Stefan, A. M., & Evans, N. J. (2019). Robust standards in cognitive science. *Computational Brain & Behavior*, *2*, 255–265.

Crüwell, S., van Doorn, J., Etz, A., Makel, M. C., Moshontz, H., Niebaum, J., ... Schulte-Mecklenbeck, M. (2018). 8 easy steps to open science: An annotated reading list. *PsyArXiv*.

DeHaven, A. (2017, 05). *Preregistration: A plan, not a prison [Blog Post].* Retrieved from `https://cos.io/blog/preregistration-plan-not-prison/`

Dutilh, G., Annis, J., Brown, S. D., Cassey, P., Evans, N. J., Grasman, R. P., ... others (2018). The quality of response time data inference: A blinded, collaborative assessment of the validity of cognitive models. *Psychonomic Bulletin & Review*, 1–19.

Evans, N. J. (2019a). Assessing the practical differences between model selection methods in inferences about choice response time tasks. *Psychonomic Bulletin & Review*, 1–29.

Evans, N. J. (2019b). *What factors are most important in finding the best model of a psychological process? comment on navarro (2018).* Retrieved from `psyarxiv.com/rkyp3`

Evans, N. J., & Annis, J. (2019). Thermodynamic integration via differential evolution: A method for estimating marginal likelihoods. *Behavior Research Methods*, *51*(2), 930–947.

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

Evans, N. J., & Brown, S. D. (2017). People adopt optimal policies in simple decision-making, after practice and guidance. *Psychonomic Bulletin & Review*, *24*(2), 597–606.

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.

Evans, N. J., & Wagenmakers, E.-J. (2019). Evidence accumulation models: Current limitations and future directions. Retrieved from `https://psyarxiv.com/74df9`

Fanelli, D. (2010). "Positive" results increase down the hierarchy of the sciences. *PloS One*, *5*(4), e10068.

Farrell, S., & Lewandowsky, S. (2018). *Computational Modeling of Cognition and Behavior*. Cambridge University Press.

Flannery, J. (2018, 05). *fMRI Preregistration Template.* Retrieved from `https://osf.io/dvb2e/`

Forstmann, B., Ratcliff, R., & Wagenmakers, E.-J. (2016, January). Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions. *Annual Review of Psychology*, *67*(1), 641–666.

Gronau, Q. F., Sarafoglou, A., Matzke, D., Ly, A., Boehm, U., Marsman, M., ... Steingroever, H. (2017). A tutorial on bridge sampling. *Journal of mathematical psychology*, *81*, 80–97.

Haven, T. L., & Grootel, D. L. V. (2019). Preregistering qualitative research. *Accountability in Research*, *0*(0), 1-16.

Heathcote, A., Brown, S. D., & Wagenmakers, E.-J. (2015). An introduction to good practices in cognitive modeling. In *An Introduction to Model-based Cognitive Neuroscience* (pp. 25–48). Springer.

Jones, M., & Dzhafarov, E. N. (2014a). Analyzability, ad hoc restrictions, and excessive flexibility of evidence-accumulation models: Reply to two critical commentaries. *Psychological Review*.

Jones, M., & Dzhafarov, E. N. (2014b). Unfalsifiability and mutual translatability of major modeling schemes for choice reaction time. *Psychological Review*, *121*(1), 1.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795.

Kern, F., & Skrede Gleditsch, K. (2017). *Exploring pre-registration and pre-analysis plans for qualitative inference.* Retrieved from `https://www.researchgate.net/publication/319141144_Exploring_Pre-registration_and_Pre-analysis_Plans_for_Qualitative_Inference`

Kerr, N. L. (1998). Harking: Hypothesizing after the results are known. *Personality and Social Psychology Review*, *2*(3), 196–217.

Kirtley, O., Lafit, G., Hiekkaranta, A., Achterhof, R., & Myin-Germeys, I. (2019). A pre-registration template for studies using Experience Sampling Methods (ESM). Retrieved from `https://osf.io/2chmu/`

Klein, O., Hardwicke, T. E., Aust, F., Breuer, J., Danielsson, H., Mohr, A. H., ... Frank, M. C. (2018). A practical guide for transparency in psychological science. *Collabra: Psychology*, *4*(1).

Lee, M. D., Criss, A. H., Devezer, B., Donkin, C., Etz, A., Leite, F. P., ... others (2019). Robust modeling in cognitive science. *Computational Brain & Behavior*, 1–13.

Lerche, V., & Voss, A. (2016). Model complexity in diffusion modeling: Benefits of making the model more parsimonious. *Frontiers in Psychology*, *7*, 1324.

Lewandowsky, S. (2019, 01). *# PSprereg: A permanent home and homage to the Twitterati [Blog Post]*. Retrieved from `https://featuredcontent.psychonomic.org/psprereg-a-permanent-home-and-homage-tothe-twitterati/`

MacEachern, S. N., & Van Zandt, T. (2019). Preregistration of modeling exercises may not be useful. *Computational Brain & Behavior*, *2*(3-4), 179–182.

McClelland, J. L. (2009). The place of modeling in cognitive science. *Topics in Cognitive Science*, *1*(1), 11–38.

Meehl, P. E. (1967). Theory-testing in psychology and physics: A methodological paradox. *Philosophy of Science*, *34*(2), 103–115.

Mertens, G., & Krypotos, A.-M. (2019). *Preregistration of secondary analyses.* Retrieved from `psyarxiv.com/ph4q7`

Morey, R. D. (2018, 11). *"Also, robustness checks, model comparison, cross validation/holdout, and blinding (where appropriate) can help. Pre-reg is a blunt tool; modellers themselves have developed many other options that are more closely matched to the problem" [Tweet].* Retrieved from `https://twitter.com/richarddmorey/status/1064208099840540673`

Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., Du Sert, N. P., ... Ioannidis, J. P. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), 0021.

Muthukrishna, M., & Henrich, J. (2019). A problem in theory. *Nature Human Behaviour*, 1.

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, *115*(11), 2600–2606.

Nuijten, M. B. (2019). Practical tools and strategies for researchers to increase replicability. *Developmental Medicine & Child Neurology*, *61*(5), 535–539.

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.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, *111*(2), 333.

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.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? a comment on theory testing. *Psychological Review*, *107*(2), 358–367.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, *6*(2), 461–464.

Simmons, J., Nelson, L., & Simonsohn, U. (2017, 11). *How to properly preregister a study [Blog Post]*. Retrieved from `http://datacolada.org/64`

Simmons, J., Nelson, L. D., & Simonsohn, U. (2011, October). False-positive psychology. *Psychological Science*, *22*(11), 1359–1366.

Starns, J. J., & Ratcliff, R. (2012, February). Age-related differences in diffusion model boundary optimality with both trial-limited and time-limited tasks. *Psychonomic Bulletin & Review*, *19*(1), 139–145.

Sterling, T. D. (1959, March). 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.

Szollosi, A., Kellen, D., Navarro, D., Shiffrin, R., van Rooij, I., Van Zandt, T., & Donkin, C. (2019). Preregistration is redundant, at best. *PsyArXiv*.

Turner, B. M., Sederberg, P. B., Brown, S. D., & Steyvers, M. (2013). A method for efficiently sampling from distributions with correlated dimensions. *Psychological Methods*, *18*(3), 368.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*(3), 550–592.

van 't Veer, A. E., & Giner-Sorolla, R. (2016, November). Pre-registration in social psychology—a discussion and suggested template. *Journal of Experimental Social Psychology*, *67*, 2–12.

Vandekerckhove, J., White, C. N., Trueblood, J. S., Rouder, J. N., Matzke, D., Leite, F. P., . . . others (2019). Robust diversity in cognitive science. *Computational Brain & Behavior*, *2*(3-4), 271–276.

van den Akker, O., Weston, S. J., Campbell, L., Chopik, W. J., Damian, R. I., Davis-Kean, P., . . . others (2019). Preregistration of secondary data analysis: A template and tutorial.

van Rooij, I. (2019, 01). *Psychological science needs theory development before preregistration [Blog Post]*. Retrieved from `https://featuredcontent.psychonomic.org/psychological-science-needs-theory-development-before-preregistration/`

Veldkamp, C. L. S., Bakker, M., van Assen, M. A. L. M., Crompvoets, E. A. V., Ong, H. H., Nosek, B. A., . . . Wicherts, J. M. (2018, Sep). *Ensuring the quality and specificity of preregistrations*. Retrieved from `psyarxiv.com/cdgyh`

Voss, A., & Voss, J. (2007). Fast-dm: A free program for efficient diffusion model analysis. *Behavior Research Methods*, *39*(4), 767–775.

Wagenmakers, E.-J., Dutilh, G., & Sarafoglou, A. (2018, July). The creativity-verification cycle in psychological science: New methods to combat old idols. *Perspectives on Psychological Science*, *13*(4), 418–427.

Wagenmakers, E.-J., & Evans, N. (2018, 11). *"Don't Interfere with my Art": On the Disputed Role of Preregistration in Exploratory Model Building [Blog Post]*. Retrieved from `https://www.bayesianspectacles.org/dont-interfere-with-my-art-on-the-disputedrole-of-preregistration-in-exploratory-model-building/`

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.

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

Weston, S. J., Mellor, D., Bakker, M., van den Akker, O., Campbell, L., Ritchie, S. J., ... Nguyen, T. (2018). *Preregistration of secondary data analysis template.* Retrieved from `https://osf.io/v4z3x/`

Wicherts, J. M., Veldkamp, C. L. S., Augusteijn, H. E. M., Bakker, M., van Aert, R. C. M., & van Assen, M. A. L. M. (2016, November). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. *Frontiers in Psychology*, *7*.

Appendix

Content:

A Preregistration Example Application

B Preregistration Template for Model Application

Preregistration for
Replicating Evans & Brown (2017)

Sophia Crüwell and Nathan Evans

23 May 2019

1 Study Information

1.1 Title

Replicating Evans & Brown (2017): Do fixed trial sessions lead participants closer to optimality than fixed time sessions?

1.2 Authors

Sophia Crüwell, Nathan J. Evans

1.3 Description

This study replicates Evans & Brown (2017), focussing on what they termed the medium feedback condition to investigate if participants in the fixed trials condition turn out to be closer to optimality than participants in fixed time sessions, given practice and feedback.

1.4 Hypotheses

H1 With suitable practice and medium feedback (cf Evans & Brown, 2017), participants get closer to optimality with each block of trials. (directional)

H2 After suitable practice and medium feedback (cf Evans & Brown, 2017),participants will have an approximately optimal speed-accuracy trade-off. (directional)

H3 Participants who complete a fixed number of trials are closer to optimality than participants who complete trials in a fixed amount of time. (directional)

2 Data Description for Pre-existing Data

2.1 Name or brief description of dataset(s)

Motion discrimination task with a random dot kinematogram in 70 participants. The data used for this study is only a relevant subset of the full data set, which includes more groups and 133 participants in total.

2.2 Is this data open or publicly available?

The data are currently not openly available. They were collected using JATOS and are stored on a University of Newcastle server.

2.3 How can the data be accessed? Provide link if available online:

N/A

2.4 Date of download or access:

N/A

2.5 Data Source

Own Lab Collection – the data were collected in 2018 by Tayler Porteiro, Honours student at the University of Newcastle (Australia).

2.6 Codebook

N/A

2.7 Sampling and data collection procedures

70 participants were recruited at the University of Newcastle and completed the experiment online, for which they received course credit. The participants were randomly (and equally) divided into the two groups of fixed trials and fixed time. The task used the random dot kinetogram, following Evans & Brown (2017) and Evans et al. (2018).

2.8 Prior work based on the dataset

There does not exist prior work based on this data set.

2.9 Prior Research Activity

We have used all of these variables & some of these models before on this (sub)set of data.

2.10 Prior Knowledge current dataset

SC has not analysed these data before. NJE has fit the model from Figure 1 to the data and checked the results from one of the groups that are not relevant for the current study, and has run the model-based analyses using the model in Figure 2 and here also checked the consistency of the Bayes factor approximation for one of the irrelevant groups.

3 Sampling Plan

N/A as data preexisting

4 Design Plan

N/A as data preexisting

5 Variables

N/A as data preexisting

6 Data Cleaning

6.1 Data exclusion

For all participants, the first block of trials will be excluded to allow for participants to become adequately practiced at the task. Trials with response times below 150ms or above 10000ms will be excluded as anticipatory responses and trials where participants lost attention, respectively. Participants with task accuracy below 60% or less than 200 eligible trials are excluded. The number of eligible trials was decided after examining the data but not the dependent variable, based on the number of trials required for accurate parameter estimation. The original study used 70 % as the accuracy cut-off, but we are using 60 % following Evans et al. (2018), as this was found to be a more realistic point of exclusion. We argue that participants performing below 70% may just be urgent, in which case their exclusion might bias our analyses, but those performing below 60 % would be too close to chance.

6.2 Missing data

N/A – There is no missing data in this data set.

7   Cognitive Modelling

7.1   Choice of Cognitive Model

As in Evans et al. (2018) the parameters of a simple diffusion model will be estimated, namely only: drift rate (v), starting point (z), threshold (a), non-decision time (ter). This differs from Evans & Brown (2017), where the full diffusion model was estimated, i.e., including between-trial variability parameters for drift rate, starting point, and non-decision time. These between-trial variability parameters were not relevant for Evans & Brown (2017), and without them, the simple diffusion model has better parameter recovery results (Lerche & Voss, 2016). Figure 1 shows a plate diagram of the hierarchical structure used for the qualitative model-based analysis assessing 1) whether groups appear to get closer to optimality over time, 2) whether each group differs from optimality, and 3) whether there appears to be a difference between the groups (see Analysis Plan for more information); i indexes participants, and j indexes blocks. Only the threshold parameter varies between blocks, to estimate changes in the speed accuracy trade-off.
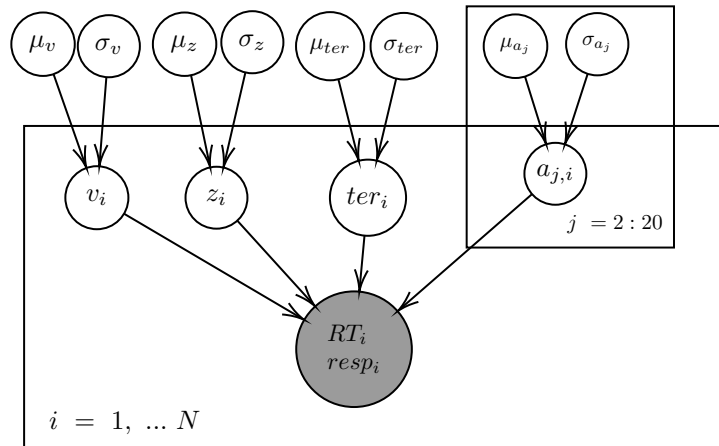

Figure 1: Plate diagram of the hierarchical structure used to qualitatively assess 1) whether groups appear to move closer to optimality over time, 2) whether each group differs from optimality, and 3) whether there appears to be a difference between the groups. RT stands for reaction time, resp stands for response accuracy.

The parameters of the estimation model shown in the plate diagram of Figure 1 are distributed as:

Data level:
$$(\text{RT}_i, \text{resp}_i) \sim \text{Diffusion}\left(v_i, z_i, \text{ter}_i, a_{2:20,i}\right)$$

Group level:
$$v_i \sim N\left(\mu_v, \sigma_v\right)$$
$$\frac{z_i}{a_{2:20,i}} \sim TN\left(\mu_{\frac{z}{a}}, \sigma_{\frac{z}{a}}, 0, 1\right)$$
$$ter_i \sim TN\left(\mu_{ter}, \sigma_{ter}, 0, \infty\right)$$
$$a_{2:20,i} \sim TN\left(\mu_{a_{2:20}}, \sigma_{a_{2:20}}, 0, \infty\right)$$

Prior distributions:
$$\mu_v \sim N(3, 3)$$
$$\mu_{\frac{z}{a}} \sim TN(0.5, 0.5, 0, 1)$$
$$\mu_{ter} \sim TN(0.3, 1, 0, \infty)$$
$$\mu_{a_{2:20}} \sim TN(2, 2, 0, \infty)$$
$$\sigma_v, \sigma_{a_{2:20}} \sim \Gamma(1, 1)$$
$$\sigma_{\frac{z}{a}}, \sigma_{ter} \sim \Gamma(0.5, 0.5)$$

Figure 2 and 3 show the hierarchical structures used for the quantitative model-based analyses. For the first quantitative model-based analysis, the threshold parameter a is replaced by parameter c ($c_i = a_i - b_i$, where b_i is the optimal threshold for person i), which does not vary across blocks, as we are here interested in comparing the mean threshold of each group to the optimal threshold using Bayes factors, approximated with the Savage-Dickey Ratio on mu_c (see Analysis Plan for more information). The parameters shown in the plate diagram of Figure 2 are thus distributed as:

Data level:
$$(\text{RT}_i\,,\text{resp}_i) \sim \text{Diffusion}\,(v_i, z_i, \text{ter}_i, a_i)$$

Group level:
$$v_i \sim N\,(\mu_v, \sigma_v)$$
$$\frac{z_i}{a_i} \sim TN\,\left(\mu_{\frac{z}{a}}, \sigma_{\frac{z}{a}}, 0, 1\right)$$
$$\text{ter}_i \sim TN\,(\mu_{ter}, \sigma_{ter}, 0, \infty)$$
$$c_i \sim TN\,(\mu_c, \sigma_c, 0, \infty)$$

Prior distributions:
$$\mu_v \sim N(3, 1)$$
$$\mu_{\frac{z}{a}} \sim TN(0.5, 0.1, 0, 1)$$
$$\mu_{ter} \sim TN(0.3, 1, 0, \infty)$$
$$\mu_c \sim N(0, 0.2)$$
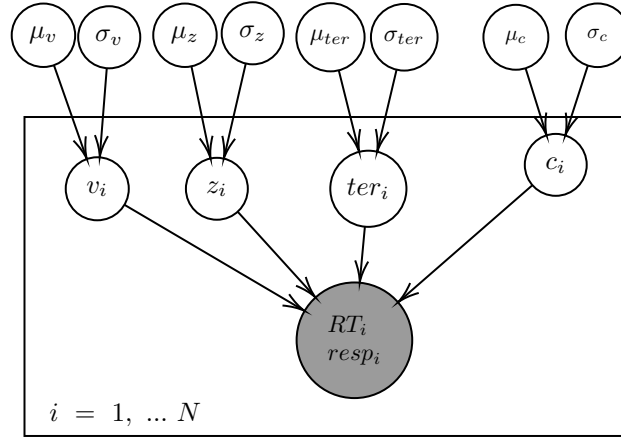$$\sigma_{\frac{z}{a}}, \sigma_{ter}, \sigma_v, \sigma_c \sim \Gamma(1, 1)$$

Figure 2: Plate diagram of the hierarchical structure used for the quantitative model-based analysis to test whether each group differs from optimality using Bayes factors, approximated with the Savage-Dickey Ratio on mu_c. RT stands for reaction time, resp stands for response accuracy.

In the second model for the quantitative model-based analyses (Figure 3), we add a third hyperparameter for c, namely the difference parameter Δ_c between the two groups' means for c. Using this model, we will quantitatively test whether the groups differ from each other in their distance from optimality using Bayes factors, approximated with the Savage-Dickey Ratio on Δ_c. The parameters shown in the plate diagram of Figure 3 are thus distributed as:

Data level:
$$(\text{RT}_i, \text{resp}_i) \sim \text{Diffusion}(v_i, z_i, \text{ter}_i, a_i)$$

Group level Fixed Time:
$$v_i \sim N(\mu_v, \sigma_v)$$
$$\frac{z_i}{a_i} \sim TN(\mu_{\frac{z}{a}}, \sigma_{\frac{z}{a}}, 0, 1)$$
$$\text{ter}_i \sim TN(\mu_{ter}, \sigma_{ter}, 0, \infty)$$
$$c_i \sim N(\mu_c - \Delta_c, \sigma_c)$$

Group level Fixed Trial:
$$v_i \sim N(\mu_v, \sigma_v)$$
$$\frac{z_i}{a_i} \sim TN(\mu_z, \sigma_z, 0, 1)$$
$$\text{ter}_i \sim TN(\mu_{ter}, \sigma_{ter}, 0, \infty)$$
$$c_i \sim N(\mu_c + \Delta_c, \sigma_c)$$

Prior distributions:
$$\mu_v \sim N(3, 1)$$
$$\mu_{\frac{z}{a}} \sim TN(0.5, 0.1, 0, 1)$$
$$\mu_{ter} \sim TN(0.3, 1, 0, \infty)$$
$$\mu_c \sim N(0, 0.2)$$
$$\mu_{\Delta_c} \sim N(0, 0.1)$$
$$\sigma_{\frac{z}{a}}, \sigma_{ter}, \sigma_v, \sigma_a, \sigma_{\Delta_c} \sim \Gamma(1, 1)$$

7.2   Method of Parameter Estimation

Only Bayesian hierarchical modeling will be used to estimate the parameters of the diffusion model, constraining individual-level parameters to follow group-level truncated normal distributions. For the estimation model (see Figure 1), the two groups (fixed-trial and fixed-time) are given a separate hierarchical structure, and the group-level parameters are not constrained between groups. Following Evans & Brown (2017) and Evans et al. (2018), we will use likelihood functions taken from the "fast-dm" toolbox Voss & Voss (2007) for the calculation of the density function of the simple diffusion model. For the first model, for sampling from the posterior distributions over parameters, we will use Markov-chain Monte Carlo with differential evolution proposals (Turner et al., 2013), using 66 chains, drawing 3,000 samples from each, and discarding the first 1,500 samples (as in Evans & Brown, 2017, see supplementary materials). The number of chains differs from that in Evans & Brown (2017), as it is based on the number of free parameters in the model. For the second and third models, the same procedure is used with 12 chains (fewer free parameters), 21,000
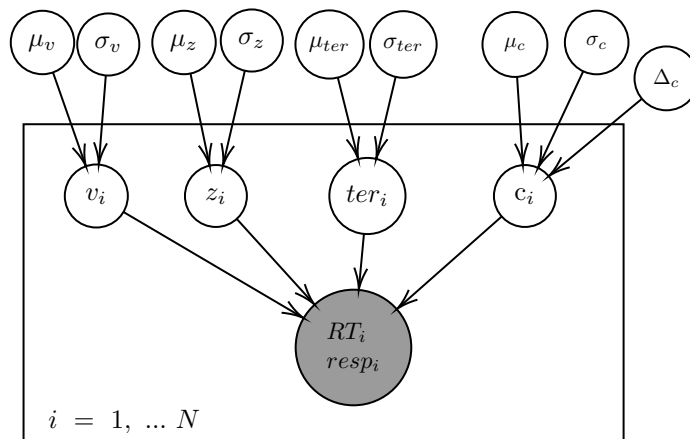
Figure 3: Plate diagram of the hierarchical structure used for the quantitative model-based analysis to test the hypothesis whether the groups differ from each other in their difference from optimality. RT stands for reaction time, resp stands for response accuracy.

samples (to ensure greater precision for the more quantitatively precise Bayes factors comparison), with the first 1,500 discarded as burn-in.

8   Analysis Plan

8.1   Statistical models

We previously specified three hypotheses:

> H1 With suitable practice and medium feedback (cf Evans & Brown, 2017), participants get closer to optimality with each block of trials. (directional)

> H2 After suitable practice and medium feedback (cf Evans & Brown, 2017),participants will have an approximately optimal speed-accuracy trade-off. (directional)

> H3 Participants who complete a fixed number of trials are closer to optimality than participants who complete trials in a fixed amount of time. (directional)

In order to test these hypotheses, we will first need to calculate mean response time and accuracy for each participant and both groups. The reward rate will be calculated using:

$$\frac{PC}{MRT + ITI + FDT + (1 - PC) * ET}$$

Here, PC is the probability of a correct response, MRT is the mean correct response time, ITI refers to the inter-trial interval, FDT refers to feedback display time, and ET refers to the error time-out.

The optimal threshold setting maximises reward rate given the estimated values of all other parameters, and is identified by calculating the expected accuracy and mean response time for each setting of the threshold parameter (Bogacz et al., 2006).

Our hypotheses will be quantitatively tested by means of only the following tests:

Testing H2: Using only the second half of all 20 blocks (11-20, so as to account for participants adjusting to the task), we will test whether each group, separately, differs from optimality using Bayes factors, approximated with the Savage-Dickey Ratio on mu_c (Wagenmakers et al., 2010, see Figure 2 above for the corresponding plate diagram).

Testing H3: Again using only blocks 11-20, we will use the Savage-Dickey method on Δ_c ($\Delta_c = \mu_{c1} - \mu_{c2}$) to test whether the groups differ in their distance from optimality.

8.2   Other analyses

Testing H1/2/3: In addition to the statistical analyses listed in 8.1, we will qualitatively compare the posterior distributions of the decision threshold parameters (actual thresholds for each block as estimated using the model in Figure 1) against the posterior predictive distributions for the optimal threshold calculated as above.

8.3   Inference criteria

These are the inference criteria for the analyses in 8.1 and 8.2:

1. Criterion for BF testing H2: Following Jeffreys (1961) for the interpretation of strength of evidence given a Bayes Factor. If $c = 0$, the behaviour of the group is optimal, if $c < 0$ it is urgent, and if $c > 0$ it is cautious.

2. Criterion for BF testing H3: Following Jeffreys (1961) for the interpretation of strength of evidence given a Bayes Factor. If $\Delta_c = 0$, the groups are the same regarding their difference from optimality, and if $\Delta_c < 0$ or $\Delta_c > 0$, the groups differ in their distance from optimality.

3. Criterion to test H1, qualitatively evaluating the plots comparing posterior to posterior predictive distribution: If the actual thresholds of both (or one) group(s) clearly show a trend towards optimality, we will conclude that participants move towards optimising the speed-accuracy trade-off. Otherwise, our conclusion will be suitably less strong and discuss any lack of clarity.

8.4 Exploratory Analyses

We are currently not planning any exploratory analyses.

9 Robustness Checks/Sensitivity Analyses

There are no planned robustness checks or sensitivity analyses.

References

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological review*, *113*(4), 700.

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

Evans, N. J., & Brown, S. D. (2017). People adopt optimal policies in simple decision-making, after practice and guidance. *Psychonomic Bulletin & Review*, *24*(2), 597–606.

Jeffreys, H. (1961). *Theory of probability* (Third ed.). Oxford, UK: Clarendon Press.

Lerche, V., & Voss, A. (2016). Model complexity in diffusion modeling: Benefits of making the model more parsimonious. *Frontiers in psychology*, *7*, 1324.

Turner, B. M., Sederberg, P. B., Brown, S. D., & Steyvers, M. (2013). A method for efficiently sampling from distributions with correlated dimensions. *Psychological Methods*, *18*(3), 368.

Voss, A., & Voss, J. (2007). Fast-dm: A free program for efficient diffusion model analysis. *Behavior Research Methods*, *39*(4), 767–775.

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.

Preregistration Template for Model Application

Study Information

1. Title (required)
   Provide an informative working title for your project. This title may change before publication.
   Example: *The effect of practice and feedback on optimality in the speed-accuracy tradeoff.*
   More info: This title needs to be descriptive and specific to your project. Avoid exceedingly vague titles such as 'Decision making preregistration plan'.

2. Authors (required)
   Provide the names and order of the authors involved in this project. This can be changed later.
   Example: *Cora Dubrey, Fernando Crabtree, & Annette Bopp*

3. Description (optional)
   Briefly describe your study, giving some information on its background and purpose.
   Example: *While animals have been found to be able to reach optimality in the speed-accuracy tradeoff, the evidence for humans is mixed. It is unclear whether humans only do not act optimally by default, or whether they are unable to do so. Here, we want to investigate this issue by giving participants time to practice, and varying the amount of feedback they receive, to see if these affect whether they reach optimality in the speed-accuracy tradeoff.*
   More info: Keep this description short – while it is helpful to give some context for your project, you do not have to include a whole introductory section for the purposes of this preregistration.

4. Hypotheses (required)
   List the hypotheses to be tested. Ensure specificity, precisenesss, and exhaustiveness by stating all your hypotheses/predictions as specifically and unambiguously as possible, ideally also emphasising that you will *only* be testing these hypotheses in your confirmatory analyses. Do specify if the hypotheses are directional or non-directional, and state the direction if appropriate.
   Example: *We will test the following hypotheses only :*
   H1 With suitable practice and medium feedback, participants get closer to optimality with each block of trials. (directional)
   H2 After suitable practice and medium feedback, participants will have an approximately optimal speed-accuracy trade-off. (directional)

H3 Participants who complete a fixed number of trials are closer to optimality than participants who complete trials in a fixed amount of time. (directional)

Data Description for Preexisting Data (delete as appropriate)

If you are using pre-existing data, please fill in this section and ignore the sections Sampling Plan, Design Plan, and Variables. If you are collecting the data yourself, please ignore this section and fill in the sections Sampling Plan, Design Plan, and Variables.

5. Name or brief description of dataset(s) (required):
   Example: *Motion discrimination task with a random dot kinematogram in 70 participants. The data used for this study is only a relevant subset of the full data set, which includes more groups and 133 participants in total.*

6. Is this data open or publically available? (required) [Yes/No]

7. How can the data be accessed? Provide link if available online. (required)
   Example: *The data are currently not openly available.*

8. Date of download or access: (required)
   Example: *Accessed on 01/05/2019.*

9. Data Source (required)
   Please select and describe what entity originally collected this data. (required)
   Examples/Options:
   National Data Set – a nationally representative sample collected by another team of researchers.
   Private Organizational Data – Internally collected data by an organization made available for academic purposes.
   Own Lab Collection-Data were connected by one of the analysts' lab.
   Other Lab Collection – Data were collected by another researcher's lab (analysts were not involved in data collection).
   Meta-Analysis – A systematic review of published studies.
   Multi-lab collaboration – Data were collected at several sites using the same procedure.
   Other – please explain

10. Codebook (required)
    Some studies (usually publically available) offer codebooks to describe their data. If such a codebook is available, please link to it here or upload the document.

11. Sampling and data collection procedures (required)

If the data collection procedure is already well documented, please provide a link to the information. If the data collection procedure is not yet well documented, please describe, to the best of your ability, how data were collected. What populations were sampled from, what were the recruitment efforts, what was the procedure for running participants through the study, were researchers blind to the research question, hypotheses or conditions, was randomization of any kind used, etc?

Example: *70 participants were recruited at the University of Newcastle and completed the experiment online, for which they received course credit. The participants were randomly (and equally) divided into the two groups of fixed trials and fixed time. The task used the random dot kinetogram, following Evans & Brown (2017) and Evans et al. (2018).*

12. Prior work based on the dataset (required)

> Have you published/presented any previous work based on this dataset? Include any publications, conference presentations (papers, posters), or working papers (in-prep, unpublished, preprints) based on this data set that you have worked on.
> **Example:** *There does not exist prior work based on this data set.*

13. Prior Research Activity (required)

> Have you worked with these data before? Describe any prior research activity using these data in a specific and transparent way.
> **Example:** *We have used all of these variables & some of these models before on this (sub)set of data.*

14. Prior Knowledge current dataset (required)

> Describe any prior knowledge of the dataset. Be specific and transparent.
> **Example:** *We have already fit the model from Figure 1 [not included here] to the data and checked the results from one of the groups that are not relevant for the current study, and has run the model-based analyses using the model in Figure 2 [not included in example] and here also checked the consistency of the Bayes factor approximation for one of the irrelevant groups.*

Sampling Plan (if own data collection)

If you are using pre-existing data, you may delete this section.

15. Data collection procedures (required)

Please describe your data collection process. If you are using human subjects, this should include the population from which you obtain subjects, recruitment efforts, payment for participation, how subjects will be selected for eligibility from the initial pool (e.g. inclusion and exclusion rules), and your study timeline. For studies that do not

include human subjects, include information about how and for how long samples will be collected, the source or location of samples, and/or the batch numbers you will use.

Example: *Participants will be students recruited through advertisements at a university, where they will receive course credit for their participation.*

More information: The answer to this question should be specific enough that another person could repeat the data collection procedures and recreate the study population. Alternatively, if the study population would be unable to be reproduced because it relies on a specific set of circumstances unlikely to be recreated (e.g., a community of people from a specific time and location), the criteria and methods for creating the group and the rationale for this unique set of subjects should be clear.

16. Sample size (required)
Describe the sample size of your study. How many units will be analyzed in the study? If the units are not individuals, then describe the size requirements for each unit. Eiter state exact numbers or an expected range.

Example: *Our target sample size is 70 participants. We will attempt to recruit up to 90, to allow for exclusions.*

17. Sample size rationale (optional)
This could include an arbitrary constraint such as time, money, or personnel, power analysis (if applicable), or an analysis based on a parameter recovery study.

Example: *The number of subjects is based on a parameter recovery study.*

More information: This gives you an opportunity to specifically state how the sample size will be determined. A wide range of possible answers is acceptable; remember that transparency is more important than principled justifications. Any pre-specified reasoning behind the sample size is preferable to ambiguity and potential confusion for the reader.

18. Stopping rule (optional)
If you cannot pre-specify your sample size, specify a stopping rule, i.e., how you will decide when to terminate your data collection.

Example: *We will post participant sign-up slots by week on the preceding Friday night, with 20 spots posted per week. We will post 20 new slots each week if, on that Friday night, we are below 320 participants.*

More information: Unacceptable rationales include stopping based on e.g. p-values if checkpoints and stopping rules are not specified. If you have control over your sample size, then including a stopping rule is not necessary, though it must be clear in this question or a previous question how an exact sample size is attained.

Design Plan (if own data collection)

If you are using pre-existing data, you may delete this section. In this section, you will be asked to describe the experimental design of your study. Remember that this research plan is designed to register a single study, so if you have multiple experimental designs, please complete a separate preregistration. Note that this is about experimental design; your modelling design choices will be registered in a later section.

19.   Study type (required)
   A. Experiment - A researcher randomly assigns treatments to study subjects, this includes field or lab experiments. This is also known as an intervention experiment and includes randomized controlled trials.
   B. Observational Study - Data is collected from study subjects that are not randomly assigned to a treatment. This includes surveys, natural experiments, and regression discontinuity designs.
   C. Other

20.   Blinding (required)
   Blinding describes who is aware of the experimental manipulations within a study. Mark all that apply.
   A. No blinding is involved in this study.
   B. For studies that involve human subjects, they will not know the treatment group to which they have been assigned.
   C. Personnel who interact directly with the study subjects (either human or non-human subjects) will not be aware of the assigned treatments. (Commonly known as "double blind")
   D. Personnel who analyze the data collected from the study are not aware of the treatment applied to any given group.

21.   Is there any additional blinding in this study?

22.   Experimental design (required)
   Describe your experimental design.
   Example: *We have a 3 (feedback: low, medium, high) by 2 (fixed time, fixed trial) by 23 (practice block) design.*
   More info: This question has a variety of possible answers. The key is for a researcher to be as detailed as is necessary given the specifics of their design.

23.   Randomization (optional)
   If you are doing a randomized study, how will you randomize, and at what level?

Example: *We will use block randomization, where each participant will be randomly assigned to one of the four equally sized, predetermined blocks. The random number list used to create these four blocks will be created using the web applications available at http://random.org.*
More info: If randomization is required for the study, the method should be specified here, not simply the source of random numbers.

Variables (if own data collection)

In this section you can describe all variables that are manipulated and measured in your experiment. If you are using pre-existing data, you may delete this section.

24. Manipulated variables (optional)
    Describe all variables you plan to manipulate and the levels or treatment arms of each variable. This is not applicable to any observational study.
    Example: *We manipulated the level of feedback given to participants, and whether they completed trials in a fixed amount of time, or a fixed number of trials.*
    More information: For any experimental manipulation, you should give a precise definition of each manipulated variable.

25. Measured variables (required)
    Describe each variable that you will measure.
    Example: *The outcome variables measured are reaction time and response accuracy.*
    More information: Observational studies will include only measured variables. As with the previous questions, the answers here must be precise.

26. Indices (optional)
    If any measurements are going to be combined into an index (or even a mean), what measures will you use and how will they be combined? Include either a formula or a precise description of your method.
    Example: *For each cell and each individual, we will calculate mean response time and accuracy.*
    More information: If you are using multiple pieces of data to construct a single variable, how will this occur? Both the data that are included and the formula or weights for each measure must be specified. Standard summary statistics, such as "means" do not require a formula, though more complicated indices require either the exact formula or, if it is an established index in the field, the index must be unambiguously defined. For example, "biodiversity index" is too broad, whereas "Shannon's biodiversity index" is appropriate.

Data Cleaning

27.   Data exclusion (required)
How will you determine what data (e.g., participants or trials), if any, will be excluded from your analyses? How will outliers be handled? Will you use any awareness checks?
Example: *For all participants, the first block of trials will be excluded to allow for participants to become adequately practiced at the task. Trials with response times below 150ms or above 10000ms will be excluded as anticipatory responses and trials where participants lost attention, respectively. Participants with task accuracy below 60% or less than 200 eligible trials (based on the number of trials required for accurate parameter estimation) will be excluded.*
More information: Any rule for excluding a particular set of data is acceptable. You may describe rules for excluding a participant and/or for identifying outlier data.

28.   Missing data (optional)
How will you deal with incomplete or missing data?
Example: *If a subject does not complete the entire duration of the experiment, that subject will not be included in the analysis.*
More information: Any relevant explanation is acceptable. As a final reminder, remember that the final analysis must follow the specified plan, and deviations must be either strongly justified or included as a separate, exploratory analysis.

Cognitive Modelling

29.   Cognitive Model (Required)
Please include the type of model used (e.g. diffusion model, Linear ballistic accumulator model), and a specific parameterisation/parameterisations.
Example: *The parameters of a simple diffusion model will be estimated, namely only: drift rate (v), starting point (z), threshold (a), non-decision time (ter). This differs from [previous paper], where the full diffusion model was estimated, i.e., including between-trial variability parameters for drift rate, starting point, and non-decision time. These between-trial variability parameters are not relevant here, and without them, the simple diffusion model has better parameter recovery results (Lerche & Voss, 2016). Figure 1 shows a plate diagram of the hierarchical structure used for the qualitative model-based analysis assessing 1) whether groups appear to get closer to optimality over time, 2) whether each group differs from optimality, and 3) whether there appears to be a difference between the groups (see Analysis Plan for more information); i indexes participants, and j indexes blocks. Only the threshold parameter varies between blocks,*

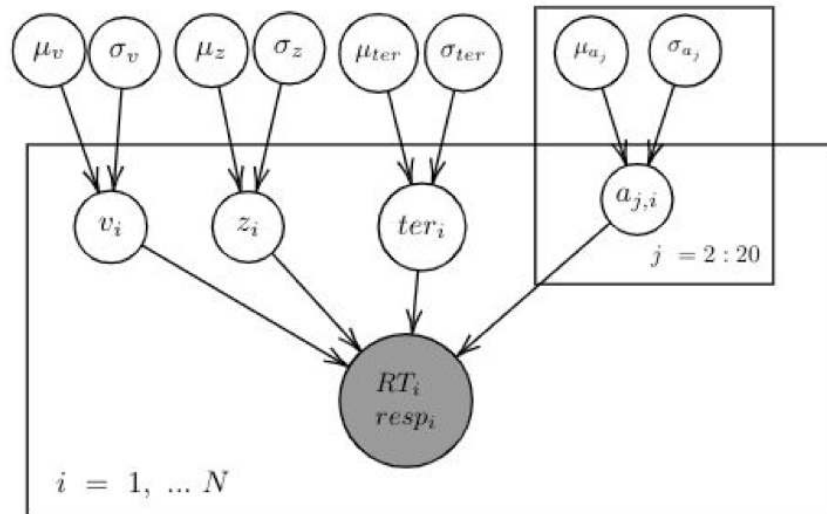*to estimate changes in the speed accuracy trade-off.*

Figure 1. Plate diagram of the hierarchical structure used. RT stands for reaction time, resp stands for response accuracy.

The parameters of the estimation model shown in the plate diagram of Figure 1 are distributed as: [include specific distributions, e.g. at data level and group level, and the relevant prior distributions]

More information: The architecture of the model should be pre-specified in a way that is specific, precise, and exhaustive. To this end, you should ideally include a plate diagram and specify the relevant equations. Motivate your choices. Note: If you are using e.g. Bayesian hierarchical modelling for parameter estimation, the structure of the hierarchical model and the prior distribution over the parameters belong into this parameterisation as well.

30. Method of Parameter Estimation (Required)
    Please specify and motivate your method of parameter estimation.
    Example: *Only Bayesian hierarchical modeling will be used to estimate the parameters of the diffusion model, constraining individual-level parameters to follow group-level truncated normal distributions. For the estimation model (see Figure 1, the two groups (fixed-trial and fixed-time) are given a separate hierarchical structure, and the group-level parameters are not constrained between groups. We will use likelihood functions taken from the ``fast-dm'' toolbox (Voss & Voss, 2007) for the calculation of the density function of the simple diffusion model. For the first model, for sampling from the posterior distributions over parameters, we will use Markov-chain Monte Carlo with differential evolution proposals (Turner et al., 2013), using 66 chains, drawing 3,000 samples from each, and discarding the first 1,500 samples.*

More information: If you are not interested in the parameters and are going straight to statistical inference without estimating the parameters, please state this clearly and motivate this choice. If you are using Bayesian methods, specify and motivate priors. In general, specify as much as possible, including e.g. the starting point (distribution) for estimation. If the data are going to be summarised into descriptive statistics, state which descriptive statistics will be used, and how.

Robustness checks

31. Robustness Checks and Sensitivity Analyses (Required)
     Please specify any planned robustness checks and/or sensitivity analyses, if any.
     Example: The key analysis will be replicated a) including participants/trials that were initially excluded in line with our exclusion criteria, and b) using a model in which the threshold parameter and the drift rate parameter vary across blocks. Their results will be mentioned alongside the key results, and interpreted accordingly. If these results show a lack of robustness, this will be an interesting outcome worthy of further investigation.
     More information: This section ensures that robustness checks are not performed and/or reported selectively. It is important to note that, given the preregistration of modelling and analyses, it should be clear that any lack of robustness is at least not due to post-hoc, data-driven choices.

Analysis Plan

Please specify at least one confirmatory analysis. The confirmatory analyses described here have to be included in the final article, with a clear distinction from any additional exploratory analyses. This preregistration has to state up front which parameters are assessed, for example to vary across conditions. Only then is it a confirmatory analysis, otherwise it is exploratory. You may describe exploratory plans here, but a clear confirmatory analysis is needed.

32. Statistical Analyses
     Specify the methods that will be used to test each hypothesis as precisely as possible. In particular, specify the method(s) and process(es) of (statistical) inference and on which parameters of interest you will be applying them. In the case of model-based analyses, make sure to also include the relevant models in the Cognitive Modelling section above. Keep in mind that any analyses not mentioned and specified in these confirmatory sections have to be clearly labeled as exploratory in the final output.
     Example: *Our hypotheses will be quantitatively tested by means of only the following tests:*
     Testing H2: Using only the second half of all 20 blocks (11-20, so as to account for participants adjusting to the task), we will test whether each group, separately, differs from optimality using Bayes factors, approximated with the Savage-Dickey Ratio on μc.

Testing H3: Again using only blocks 11-20, we will use the Savage-Dickey method on Δc ($\Delta c = \mu c1 - \mu c2$) to test whether the groups differ in their distance from optimality.

More information: As with all of the other questions, the key is to provide a specific recipe for analyzing the collected data. Ask yourself: is enough detail provided to run the same analysis again with the information provided by the user?

33.   Other analyses
Specify any other confirmatory analyses you intend to perform.
Example: *Testing H1/2/3: In addition to the statistical analyses listed above, we will qualitatively compare the posterior distributions of the decision threshold parameters (actual thresholds for each block as estimated using the model in Figure 1) against the posterior predictive distributions for the optimal threshold calculated as above.*

34.   Inference criteria (optional)
What criteria will you use to make inferences? Please describe the information you will use (e.g. p-values, Bayes Factors, etc.), as well as a cut-off criterion, where appropriate. Will you be using one or two tailed tests for each of your analyses? If you are comparing multiple conditions or testing multiple hypotheses, will you account for this?
Example: *These are the inference criteria for our analyses listed above:*
1. Criterion for BF testing H2: Following Jeffreys (1961) for the interpretation of strength of evidence given a Bayes Factor. If c = 0, the behaviour of the group is optimal, if c < 0 it is urgent, and if c > 0 it is cautious.
2. Criterion for BF testing H3: Following Jeffreys (1961) for the interpretation of strength of evidence given a Bayes Factor. If Δc = 0, the groups are the same regarding their difference from optimality, and if Δc < 0 or Δc > 0, the groups differ in their distance from optimality.
3. Criterion to test H1, qualitatively evaluating the plots comparing posterior to posterior predictive distribution: If the actual thresholds of both (or one) group(s) clearly show a trend towards optimality, we will conclude that participants move towards optimising the speed-accuracy trade-off. Otherwise, our conclusion will be suitably less strong and discuss any lack of clarity.

35.   Exploratory analysis (optional)
Describe any exploratory analyses you expect to do.
Example: *We are currently not planning any exploratory analyses.*
More information: An exploratory test is any test where a prediction is not made up front, or there are multiple possible tests that you are going to use. A statistically significant finding in an exploratory test is a great way to form a new confirmatory hypothesis, which could be registered at a later time. It is crucial to clearly distinguish confirmatory from exploratory results in your final article.

Other

36.   Other (Optional)
      Any other information not included in the above. You may include e.g., literature cited,
      disclosures of any related work such as replications or work that uses the same data, or
      other context that will be helpful for future readers and/or yourself.