

1 Event History Analysis for psychological time-to-event data: A tutorial in R with examples  
2 in Bayesian and frequentist workflows

3 Sven Panis<sup>1</sup> & Richard Ramsey<sup>1</sup>

4 <sup>1</sup> ETH Zürich

5 Author Note

6 Neural Control of Movement lab, Department of Health Sciences and Technology  
7 (D-HEST). Social Brain Sciences lab, Department of Humanities, Social and Political  
8 Sciences (D-GESS).

9 Correspondence concerning this article should be addressed to Sven Panis, ETH  
10 GLC, room G16.2, Gloriastrasse 37/39, 8006 Zürich. E-mail: sven.panis@hest.ethz.ch

11

## Abstract

12 Time-to-event data such as response times and saccade latencies form a cornerstone of  
13 experimental psychology, and have had a widespread impact on our understanding of  
14 human cognition. However, the orthodox method for analyzing such data – comparing  
15 means between conditions – is known to conceal valuable information about the timeline of  
16 psychological effects, such as their onset time and how they evolve with increasing waiting  
17 time. The ability to reveal finer-grained, “temporal states” of cognitive processes can have  
18 important consequences for theory development by qualitatively changing the key  
19 inferences that are drawn from psychological data. Luckily, well-established analytical  
20 approaches, such as event history analysis (EHA), are able to evaluate the detailed shape  
21 of time-to-event distributions, and thus characterize the time course of psychological states.  
22 One barrier to wider use of EHA, however, is that the analytical workflow is typically more  
23 time-consuming and complex than orthodox approaches. To help achieve broader uptake of  
24 EHA, in this paper we outline a set of tutorials that detail one distributional method  
25 known as discrete-time EHA. We touch upon several key aspects of the workflow, such as  
26 how to process raw data and specify regression models, and we also consider the  
27 implications for experimental design. We finish the article by considering the benefits of  
28 the approach for understanding psychological states, as well as the limitations and future  
29 directions of this work. Finally, the project is written in R and freely available, which  
30 means the approach can easily be adapted to other data sets.

31       *Keywords:* response times, event history analysis, Bayesian multilevel regression  
32 models, experimental psychology, cognitive psychology

33 Word count: 11664 (body) + 1593 (references) + 2394 (supplemental material)

34

## 1. Introduction

### 35 1.1 Motivation and background context: Comparing means versus 36 distributional shapes

37 In experimental psychology, it is standard practice to analyse response times (RTs),  
38 saccade latencies, and fixation durations by calculating average performance across a series  
39 of trials. Such comparisons between means have been the workhorse of experimental  
40 psychology over the last century, and have had a substantial impact on theory development  
41 as well as our understanding of the structure of cognition and brain function. Indeed, the  
42 view that mean values are truth and variations around the mean are error is deeply  
43 ingrained in experimental psychology (Bolger, Zee, Rossignac-Milon, & Hassin, 2019).

44 However, differences in mean RT conceal important pieces of information, such as when an  
45 experimental effect starts, how it evolves with increasing waiting time, and whether its  
46 onset is time-locked to other events (Panis, 2020; Panis, Moran, Wolkersdorfer, & Schmidt,  
47 2020; Panis & Schmidt, 2016, 2022; Panis, Torfs, Gillebert, Wagemans, & Humphreys,  
48 2017; Panis & Wagemans, 2009; Wolkersdorfer, Panis, & Schmidt, 2020). Such absolute  
49 timing information is useful not only for the interpretation of experimental effects under  
50 investigation, but also for cognitive psychophysiology and computational model selection  
51 (Panis, Schmidt, Wolkersdorfer, & Schmidt, 2020).

52 As a simple illustration, Figure 1 summarises simulated single-subject data (200 trials  
53 per condition) that shows how comparing means between two conditions can conceal the  
54 shapes of the underlying RT and accuracy distributions. Indeed, compared to the  
55 aggregation of data across trials (Figure 1A), a distributional approach offers the  
56 possibility to reveal the time course of psychological states (Figure 1B). For example,  
57 Figure 1B shows a first state (up to 400 ms after target onset) for which the early upswing  
58 in hazard is equal for both conditions, and the emitted responses are always correct in  
59 condition 1 and always incorrect in condition 2. In a second state (400 to 500 ms), hazard

- 60 is higher in condition 1, and conditional accuracies are close to .5 in both conditions. In a  
61 third state (>500 ms), the effect disappears in hazard, and all conditional accuracies are  
62 equal to 1 (see also Panis & Schmidt, 2016).

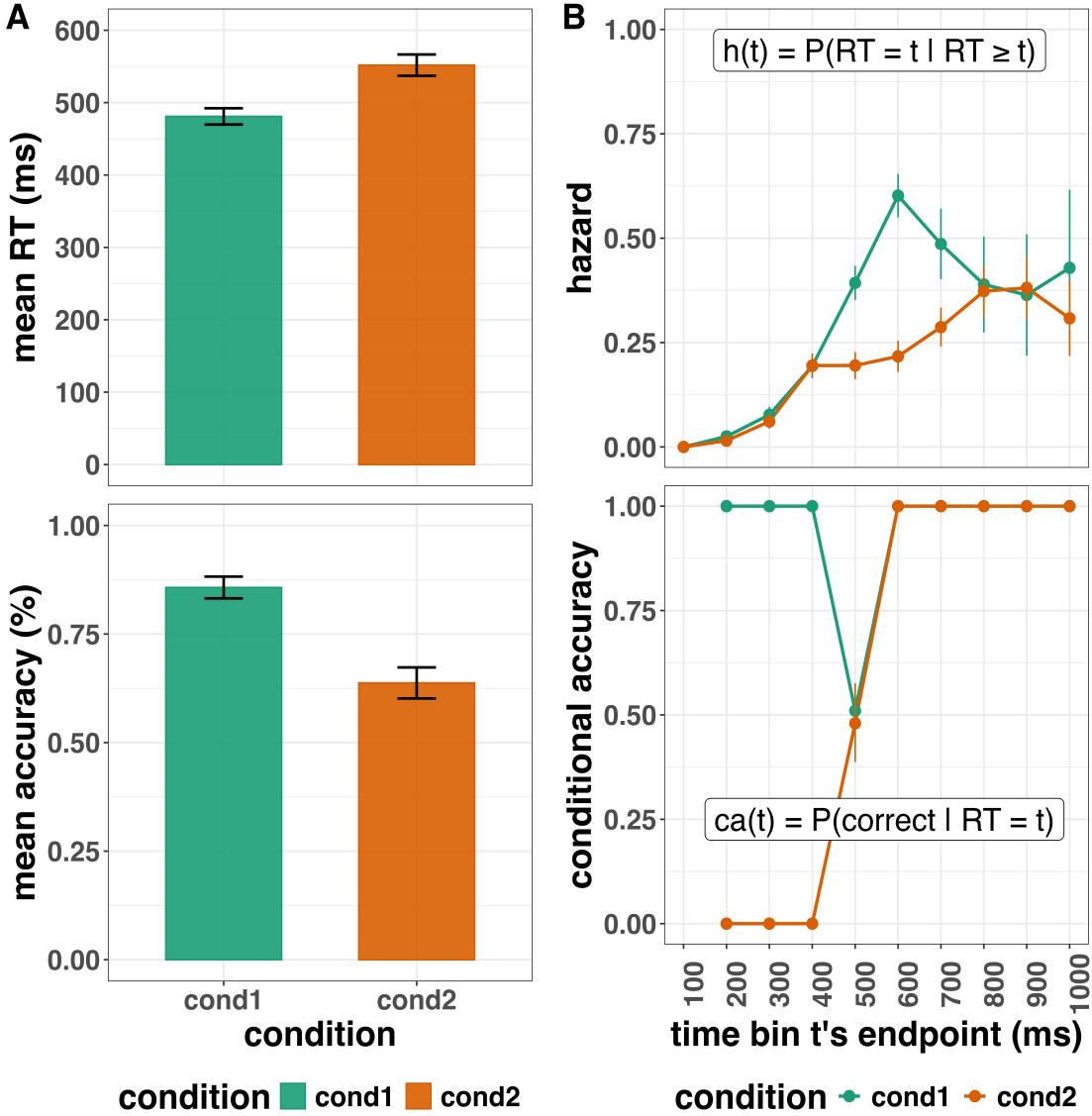


Figure 1. Simulated data showing mean performance versus distributional (EHA/SAT) analyses. (A) The mean RT (top) and overall accuracy (bottom) for two conditions are plotted. (B) The discrete-time hazard functions (top) and conditional accuracy functions (bottom) are plotted for the same data. The first second after target stimulus onset (time zero) is divided in ten bins of 100 ms. The first bin is (0,100], the last bin is (900,1000]. Note that the hazard and conditional accuracy estimates are plotted at the endpoint of each time bin. The definitions of discrete-time hazard and conditional accuracy are further explained in section 2.1.2. Error bars represent  $\pm 1$  standard error of the mean (A) or proportion (B).

63 Why does this matter for research in psychology? For many psychological questions,  
64 the estimation of such “temporal states” information can be theoretically meaningful by  
65 leading to more fine-grained understanding of psychological processes. Because EHA adds  
66 a relatively under-used but ever-present dimension – the passage of time – to the theory  
67 building toolkit, it provides one possible answer to the recent call for a temporal science of  
68 behavior (Abney, Fausey, Suarez-Rivera, & Tamis-LeMonda, 2025).

## 69 1.2 Aims

70 Our ultimate aim in this paper is twofold: first, we want to convince readers of the  
71 many benefits of using EHA when dealing with psychological RT data, and second, we  
72 want to provide a set of practical tutorials, which provide step-by-step instructions on how  
73 you actually perform a discrete-time EHA on RT data, as well as a complementary  
74 discrete-time speed-accuracy tradeoff (SAT) analysis on timed accuracy data in case of  
75 choice RT data (Figure 1B).

76 Even though EHA is a widely used statistical tool and there already exist many  
77 excellent reviews (Allison, 1982; Blossfeld & Rohwer, 2002; Box-Steffensmeier, 2004;  
78 Hosmer, Lemeshow, & May, 2011; Mills, 2011; Singer & Willett, 2003; Teachman, 1983)  
79 and tutorials (Allison, 2010; Elmer, Van Duijn, Ram, & Bringmann, 2023; Landes,  
80 Engelhardt, & Pelletier, 2020; Lougheed, Benson, Cole, & Ram, 2019; Stoolmiller, 2015;  
81 Stoolmiller & Snyder, 2006), we are not aware of any tutorials that are aimed specifically at  
82 psychological RT (+ accuracy) data, and which provide worked examples of the key data  
83 processing and Bayesian multilevel regression modelling steps. Set within this context, our  
84 overall aim is to introduce a set of tutorials, which explain **how** to do such analyses in the  
85 context of experimental psychology, rather than repeat in any detail **why** you may do  
86 them. Therefore, we hope that our tutorials will provide a pathway for research avenues in  
87 experimental psychology that have the potential to benefit from using EHA in the future.

### 88 1.3 Structure

89 In what follows, the paper is organised in three main sections. In Section 2, we  
90 provide a brief overview of EHA to orient the reader to the basic concepts that we will use  
91 throughout the paper and why such an approach might be relevant for research in  
92 experimental psychology. In Section 3, we outline a series of tutorials, which are written in  
93 the R programming language and publicly available on our Github page  
94 ([https://github.com/sven-panis/Tutorial\\_Event\\_History\\_Analysis](https://github.com/sven-panis/Tutorial_Event_History_Analysis)), along with all of the  
95 other code and material associated with the project. The tutorials provide hands-on,  
96 concrete examples of key parts of the analytical process, such as data wrangling, plotting  
97 descriptive statistics, model fitting and planning future studies, so that others can apply  
98 EHA to their own time-to-event data measured in RT tasks. In Section 4, we discuss the  
99 strengths and weaknesses of the approach for researchers in experimental psychology.

100 **2. What is event history analysis and why is it relevant to research in  
101 experimental psychology?**

102 **2.1 A brief introduction to event history analysis**

103 EHA is a class of statistical approaches to study the occurrence and timing of events,  
104 such as disease onset, marriages, arrests, and job terminations (Allison, 2010). In this  
105 section, we want to provide an intuition regarding how EHA works in general, as well as in  
106 the context of experimental psychology. For those who want more detailed treatment of  
107 EHA and/or regression equations, we refer the reader to several excellent textbooks on  
108 these topics (Allison, 2010; Gelman, Hill, & Vehtari, 2020; Singer & Willett, 2003; Winter,  
109 2019). We supply a few regression equations in section E of the Supplemental Material.

110 **2.1.1 Terminology and minimum requirements for EHA.** To avoid possible  
111 confusion in terminology used, it is worth noting that EHA is known by various labels,  
112 such as survival analysis, hazard analysis, duration analysis, failure-time analysis, and

113 transition analysis (Singer & Willett, 2003). In this paper, we choose to use the term EHA  
114 throughout.

115 In terms of minimum requirements to apply EHA, one must be able to:

- 116 1. define an event of interest that represents a qualitative change - a transition from one  
117 discrete state to another - that can be situated in time (e.g., a button press, a  
118 saccade onset, a fixation offset, etc.);
- 119 2. define time point zero in each trial (e.g., target stimulus onset, fixation onset, etc.);
- 120 3. measure the passage of time between time point zero and event occurrence in discrete  
121 or continuous time units in each trial.

122 These minimal requirements are fulfilled by the RT data obtained in single-button  
123 detection tasks, where the time-to-response is repeatedly measured in different trials in the  
124 same individual. In section A of the Supplemental Material we visualize this and other  
125 types of time-to-event data which are typically obtained in discrimination and bistable  
126 perception tasks.

127 **2.1.2 Types of EHA.** There are different types of modeling approaches in EHA.  
128 For example, the definition of hazard and the type of models employed depend on whether  
129 one is using continuous or discrete time units. As a lab, and mainly for practical reasons,  
130 we have much more experience using discrete-time EHA, and that is the approach that we  
131 describe and focus on in this paper. This choice may seem counter-intuitive, given that RT  
132 is typically treated as a continuous variable. However, continuous forms of EHA require  
133 much more data to estimate the continuous-time hazard (rate) function well (Bloxom,  
134 1984; Luce, 1991; Van Zandt, 2000). Thus, by trading a bit of temporal resolution for a  
135 lower number of trials, discrete-time methods seem ideal for dealing with typical  
136 psychological RT data sets for which there are less than ~200 trials per condition per  
137 participant (Panis, Schmidt, et al., 2020). Moreover, as indicated by Allison (2010),

138 learning discrete-time EHA methods first will help in learning continuous-time methods, so  
139 it seems like a good starting point.

140 To apply discrete-time EHA, one divides the within-trial time in discrete, contiguous  
141 time bins indexed by  $t$  (e.g.,  $t = 1:10$  time bins; Figure 1B). Then let  $RT$  be a discrete  
142 random variable denoting the rank of the time bin in which a particular person's response  
143 occurs in a particular trial (i.e., repeated measure). For example, a response in one trial  
144 might occur at 546 ms and it would be in time bin 6 (any RTs from 501 ms to 600 ms).  
145 One then calculates the sample-based estimate of the discrete-time hazard function of  
146 event occurrence for each experimental condition (Figure 1B top). The discrete-time  
147 hazard function gives you, for each time bin, the conditional probability that the event  
148 occurs (sometime) in bin  $t$ , given that the event does not occur in previous bins. In other  
149 words, it reflects the instantaneous risk that the event occurs in the current bin  $t$ , given  
150 that it has not yet occurred in the past, i.e., in one of the prior bins ( $t-1, t-2, \dots, 1$ ).

151 In the context of experimental psychology, it is often (but not always), the case that  
152 responses can be classified as correct or incorrect. In those cases, one can also calculate the  
153 conditional accuracy function (Figure 1B bottom). The conditional accuracy function gives  
154 you for each time bin the conditional probability that a response is correct given that it is  
155 emitted in time bin  $t$  (Allison, 2010; Kantowitz & Pachella, 2021; Wickelgren, 1977). The  
156  $ca(t)$  function is also known as the micro-level speed-accuracy tradeoff (SAT) function. We  
157 refer to this extended (hazard + conditional accuracy) analysis for choice RT data as  
158 EHA/SAT. The definitions of these and other discrete-time functions are given in section B  
159 of the Supplemental Material.

## 160 **2.2 Benefits of event history analysis for research in experimental psychology**

161 Statisticians and mathematical psychologists recommend focusing on the hazard  
162 function when analyzing time-to-event data for various reasons (Holden, Van Orden, &

<sup>163</sup> Turvey, 2009; Luce, 1991; Townsend, 1990). We do not cover these benefits in detail here,  
<sup>164</sup> as these are more general topics that have been covered elsewhere in textbooks (see also  
<sup>165</sup> section G of the Supplemental Material). Instead, here we focus on the benefits as we see  
<sup>166</sup> them for common research programmes in experimental psychology.

<sup>167</sup> We highlight three benefits that we think are relevant to the domain of experimental  
<sup>168</sup> psychology. First, as illustrated in Figure 1, compared to averaging data across trials,  
<sup>169</sup> integrating results between hazard functions and their associated conditional accuracy  
<sup>170</sup> functions for choice RT data can be informative for understanding psychological processes,  
<sup>171</sup> in terms of inferences about the microgenesis and temporal organization of cognition and  
<sup>172</sup> theoretical development. As such, the approach permits different kinds of questions to be  
<sup>173</sup> asked, different inferences to be made, and it holds the potential to discriminate between  
<sup>174</sup> theoretical accounts of psychological and/or brain-based processes. For example, what kind  
<sup>175</sup> of theory or set of mechanisms could account for the shape of the functions and the  
<sup>176</sup> temporally localized effects reported in Figure 1B (Panis & Schmidt, 2016)? Are there new  
<sup>177</sup> auxiliary assumptions that computational models need to adopt (Panis, Moran, et al.,  
<sup>178</sup> 2020)? Will the temporal effect patterns align nicely with EEG findings (Panis & Schmidt,  
<sup>179</sup> 2022)? And are there new experiments that need to be performed to test the novel  
<sup>180</sup> predictions that follow from these analyses?

<sup>181</sup> Second, compared to more conventional analytical approaches, EHA uses more of the  
<sup>182</sup> data because it deals with missing data differently. It is conventional with RT data to either  
<sup>183</sup> (a) use a response deadline and discard all trials without a response, or (b) wait in each  
<sup>184</sup> trial until a response occurs and then apply data trimming techniques, i.e., discarding too  
<sup>185</sup> short or too long RTs (and perhaps also erroneous responses) before calculating a mean RT  
<sup>186</sup> (Berger & Kiefer, 2021). Discarding data can introduce biases, however. Rather than treat  
<sup>187</sup> non-responses as missing data, EHA treats such trials as *right-censored* observations on the  
<sup>188</sup> variable RT, because all we know is that RT is greater than some value. Right-censoring is  
<sup>189</sup> a type of missing data problem and a nearly universal feature of survival data including RT

190 data. For example, if the censoring time was 1 second, then some trials result in observed  
191 event times (those with a RT below 1 second), while the other trials result in response  
192 times that are right-censored at 1 second. The fact that EHA can deal with  
193 right-censoring, therefore, presents a analytical strength of the approach compared to many  
194 common approaches in experimental psychology (ANOVA, linear regression, delta plots).

195 Third, the approach is generalisable and applicable to many tasks that are commonly  
196 used in experimental psychology, such as detection, discrimination and bistable perception  
197 tasks, and to a range of common experimental manipulations, such as  
198 stimulus-onset-asynchrony (see section A of the Supplemental Material). The upshot is  
199 that one general analytical approach, which holds several potential advantages, is widely  
200 applicable to many substantive use-cases in the domain of experimental psychology,  
201 irrespective of the analyst's current view on the nature of cognition (Barack & Krakauer,  
202 2021).

### 203 2.3 Implications for research design in experimental psychology

204 Performing EHA in experimental psychology has implications for how experiments  
205 are designed. More specifically, we consider three implications that researchers will need to  
206 consider when using discrete-time EHA. First, one can use a response deadline in each trial  
207 because EHA deals with right-censored observations.

208 Second, since the number of trials per condition are spread across bins, it is  
209 important to have a relatively large number of trial repetitions per participant and per  
210 condition. Accordingly, experimental designs using this approach typically focus on  
211 factorial, within-subject designs, in which a large number of observations are made on a  
212 relatively small number of participants (so-called small- $N$  designs). This approach  
213 emphasizes the precision and reproducibility of data patterns at the individual participant  
214 level to increase the inferential validity of the design (Baker et al., 2021; Smith & Little,

215 2018). Note that because statistical power derives both from the number of participants  
216 and from the number of repeated measures per participant and condition, small- $N$  designs  
217 can still achieve what are generally considered acceptable levels of statistical power, if they  
218 have a sufficient amount of data overall (Baker et al., 2021; Smith & Little, 2018).

219 Third, the width of each time bin will need to be determined. For instance, in Figure  
220 1B we chose 100 ms in an arbitrary manner. In reality, however, bin width will need to be  
221 set by considering a number of factors simultaneously. The optimal bin width will depend  
222 on (a) the length of the observation period in each trial, (b) the rarity of event occurrence,  
223 (c) the number of repeated measures (or trials) per condition per participant, and (d) the  
224 shape of the hazard function. Finding an appropriate bin width in a given user case before  
225 fitting models will require testing a number of options, when calculating and plotting the  
226 descriptive statistics (see section 3.1). The goal is to find the smallest bin width that is  
227 supported by the amount of data available. Based on our experience, a bin width of 50 ms  
228 is a good starting value when the number of repeated measures is 100 or less. Too small  
229 bin widths will result in erratic hazard functions as many bins will have no events, and  
230 thus hazard estimates of zero. Interestingly, the time bins do not need to have the same  
231 width. For example, Panis (2020) used larger bins towards the end of the observation  
232 period, as fewer events occurred there.

### 233 3. Tutorials

234 Tutorials 1a and 1b show how to calculate and plot the descriptive statistics of  
235 EHA/SAT when there are one or two independent variables, respectively. Tutorials 2a and  
236 2b illustrate how to use Bayesian multilevel modeling to fit hazard and conditional  
237 accuracy models, respectively. Tutorials 3a and 3b show how to implement, respectively,  
238 multilevel models for hazard and conditional accuracy in the frequentist framework.  
239 Tutorial 4 shows how to use simulation and power analysis for planning experiments.  
240 Additionally, to further simplify the process for other users, the first two tutorials rely on a

241 set of our own custom functions that make sub-processes easier to automate, such as data  
242 wrangling and plotting functions (see section C of the Supplemental Material for a list of  
243 the custom functions).

244 The content of the tutorials, in terms of EHA and multilevel regression modelling, is  
245 mainly based on Allison (2010), Singer and Willett (2003), McElreath (2020), Heiss (2021),  
246 Kurz (2023a), and Kurz (2023b). We used R (Version 4.4.0; R Core Team, 2024) and the  
247 R-packages *bayesplot* (Version 1.11.1; Gabry, Simpson, Vehtari, Betancourt, & Gelman,  
248 2019), *brms* (Version 2.22.0; Bürkner, 2017, 2018, 2021), *citr* (Version 0.3.2; Aust, 2019),  
249 *cmdstanr* (Version 0.8.1.9000; Gabry, Češnovar, Johnson, & Brodner, 2024), *dplyr* (Version  
250 1.1.4; Wickham, François, Henry, Müller, & Vaughan, 2023), *forcats* (Version 1.0.0;  
251 Wickham, 2023a), *ggplot2* (Version 3.5.1; Wickham, 2016), *lme4* (Version 1.1.35.5; Bates,  
252 Mächler, Bolker, & Walker, 2015), *lubridate* (Version 1.9.3; Grolemund & Wickham, 2011),  
253 *Matrix* (Version 1.7.1; Bates, Maechler, & Jagan, 2024), *nlme* (Version 3.1.166; Pinheiro &  
254 Bates, 2000), *papaja* (Version 0.1.3; Aust & Barth, 2024), *patchwork* (Version 1.3.0;  
255 Pedersen, 2024), *purrr* (Version 1.0.2; Wickham & Henry, 2023), *RColorBrewer* (Version  
256 1.1.3; Neuwirth, 2022), *Rcpp* (Eddelbuettel & Balamuta, 2018; Version 1.0.13.1;  
257 Eddelbuettel & François, 2011), *readr* (Version 2.1.5; Wickham, Hester, & Bryan, 2024),  
258 *RJ-2021-048* (Bengtsson, 2021), *rstan* (Version 2.32.6; Stan Development Team, 2024),  
259 *standist* (Version 0.0.0.9000; Girard, 2024), *StanHeaders* (Version 2.32.10; Stan  
260 Development Team, 2020), *stringr* (Version 1.5.1; Wickham, 2023b), *tibble* (Version 3.2.1;  
261 Müller & Wickham, 2023), *tidybayes* (Version 3.0.7; Kay, 2024), *tidyverse* (Version 2.0.0; Wickham et al., 2019) and  
262 *tinylabels* (Version 0.2.4; Barth, 2023) for all reported analyses.

264 **3.1 Tutorial 1a: Calculating descriptive statistics using a life table**

265 **3.1.1 Data wrangling aims.** Our data wrangling procedures serve two related  
266 purposes. First, we want to summarise and visualise descriptive statistics using a life table.

267 A life table includes for each time bin, the risk set (i.e., the number of trials that are  
268 event-free at the start of the bin), the number of observed events, and the estimates of the  
269 discrete-time hazard probability  $h(t)$ , survival probability  $S(t)$ , probability mass  $P(t)$ ,  
270 possibly the conditional accuracy  $ca(t)$ , and their estimated standard errors (se). The  
271 definitions of these quantities are provided in section B of the Supplemental Material.

272 Second, we want to produce two different data sets that can each be submitted to  
273 different types of inferential modelling approaches. The two types of data structure we  
274 label as ‘person-trial’ data and ‘person-trial-bin’ data. The ‘person-trial’ data (Table 1)  
275 will be familiar to most researchers who record behavioural responses from participants, as  
276 it represents the measured RT and accuracy per trial within an experiment. This data set  
277 is used when fitting conditional accuracy models (Tutorials 2b and 3b).

Table 1  
*Data structure for ‘person-trial’ data*

pid	trial	condition	rt	accuracy
1	1	congruent	373.49	1
1	2	incongruent	431.31	1
1	3	congruent	455.43	0
1	4	incongruent	622.41	1
1	5	incongruent	535.98	1
1	6	incongruent	540.08	1
1	7	congruent	511.07	1
1	8	incongruent	444.42	1
1	9	congruent	678.69	1
1	10	congruent	549.79	1

*Note.* The first 10 trials for participant 1 are shown. These data are simulated and for illustrative purposes only.

278 In contrast, the ‘person-trial-bin’ data (Table 2) has a different, more extended  
 279 structure, which indicates in which bin a response occurred, if at all, in each trial.  
 280 Therefore, the ‘person-trial-bin’ data generates a 0 in each bin until an event occurs and  
 281 then it generates a 1 to signal an event has occurred in that bin. This data set is used  
 282 when fitting discrete-time hazard models (Tutorials 2a and 3a). It is worth pointing out  
 283 that there is no requirement for an event to occur at all (in any bin), as maybe there was  
 284 no response on that trial or the event occurred after the time window of interest. Likewise,  
 285 when the event occurs in bin 1 there would only be one row of data for that trial in the  
 286 person-trial-bin data set.

Table 2  
*Data structure for ‘person-trial-bin’ data*

pid	trial	condition	timebin	event
1	1	congruent	1	0
1	1	congruent	2	0
1	1	congruent	3	0
1	1	congruent	4	1
1	2	incongruent	1	0
1	2	incongruent	2	0
1	2	incongruent	3	0
1	2	incongruent	4	0
1	2	incongruent	5	1

*Note.* The first 2 trials for participant 1 from Table 1 are shown. The width of the time bins is 100 ms. These data are simulated and for illustrative purposes only.

287       **3.1.2 A real data wrangling example.** To illustrate how to quickly set up life  
 288       tables for calculating the descriptive statistics (functions of discrete time), we use a  
 289       published data set on masked response priming from Panis and Schmidt (2016), who were  
 290       interested in the temporal dynamics of the effect of prime-target congruency in RT and  
 291       accuracy data. In their first experiment, Panis and Schmidt (2016) presented a double  
 292       arrow for 94 ms that pointed left or right as the target stimulus with an onset at time  
 293       point zero in each trial. Participants had to indicate the direction in which the double  
 294       arrow pointed using their corresponding index finger, within 800 ms after target onset.  
 295       Response time and accuracy were recorded on each trial. Prime type (blank, congruent,

296 incongruent) and mask type were manipulated across trials (i.e., repeated measures of  
 297 time-to-response). Here we focus for each participant on the subset of 220 trials in which  
 298 no mask was presented. The 13-ms prime stimulus was a double arrow presented 187 ms  
 299 before target onset in the congruent (same direction as target) and incongruent (opposite  
 300 direction as target) prime conditions.

301 There are several data wrangling steps to be taken. First, we need to load the data  
 302 before we (a) supply required column names, and (b) specify the factor condition with the  
 303 correct levels and labels.

304 The required column names are as follows:

- 305 • “pid”, indicating unique participant IDs;
- 306 • “trial”, indicating each unique trial per participant;
- 307 • “condition”, a factor indicating the levels of the independent variable (1, 2, ...) and  
     the corresponding labels;
- 309 • “rt”, indicating the response times in ms;
- 310 • “acc”, indicating the accuracies (1/0).

311 In the code of Tutorial 1a, this is accomplished as follows.

```
data_wr<-read_csv("../Tutorial_1_descriptive_stats/data/DataExp1_6subjects_wrangled.csv")
data_wr <- data_wr %>%
  rename(pid = vp, condition = prime_type, acc = respac, trial = TrialNr) %>%
  mutate(condition = condition + 1, # original levels were 0, 1, 2.
        condition = factor(condition,
                            levels=c(1,2,3),
                            labels=c("blank","congruent","incongruent")))
```

312 Next, we can set up the life tables and plot for each condition the discrete-time  
 313 hazard function  $h(t)$ , survivor function  $S(t)$ , probability mass function  $P(t)$ , and

314 conditional accuracy function `ca(t)`. To do so using a functional programming approach,  
 315 one has to nest the data within participants using the `group_nest()` function, and supply a  
 316 user-defined censoring time and bin width to our custom function “`censor()`”, as follows.

```
data_nested <- data_wr %>% group_nest(pid)

data_final <- data_nested %>%
  # ! user input: censoring time, and bin width
  mutate(censored = map(data, censor, 600, 40)) %>%
  # create person-trial-bin data set
  mutate(ptb_data = map(censored, ptb)) %>%
  # create life tables without ca(t)
  mutate(lifetable = map(ptb_data, setup_lt)) %>%
  # calculate ca(t)
  mutate(condacc = map(censored, calc_ca)) %>%
  # create life tables with ca(t)
  mutate(lifetable_ca = map2(lifetable, condacc, join_lt_ca)) %>%
  # create plots
  mutate(plot = map2(.x = lifetable_ca, .y = pid, plot_eha,1))
```

317 Note that the censoring time (here: 600 ms) should be a multiple of the bin width  
 318 (here: 40 ms). The censoring time should be a time point after which no informative  
 319 responses are expected anymore, in case one waits for a response in each trial. In  
 320 experiments that implement a response deadline in each trial the censoring time can equal  
 321 that deadline time point. Trials with a RT larger than the censoring time, or trials in  
 322 which no response is emitted during the data observation period, are treated as  
 323 right-censored observations in EHA. In other words, these trials are not discarded, because  
 324 they contain the information that the event did not occur before the censoring time.  
 325 Removing such trials before calculating the mean event time would result in  
 326 underestimation of the true mean.

327 The person-trial-bin oriented data set is created by our custom function `ptb()`, and it

328 has one row for each time bin (of each trial) that is at risk for event occurrence. The  
329 variable “event” in the person-trial-bin oriented data set indicates whether a response  
330 occurs (1) or not (0) for each bin. The next steps are to set up the life table using our  
331 custom function setup\_lt(), calculate the conditional accuracies using our custom function  
332 calc\_ca(), add the ca(t) estimates to the life table using our custom function join\_lt\_ca(),  
333 and then plot the descriptive statistics using our custom function plot\_eha(). One can now  
334 inspect different aspects, including the life table for a particular condition of a particular  
335 subject, and a plot of the different functions for a particular participant.

336 In general, it is important to visually inspect the functions first for each participant,  
337 in order to identify individuals that may not be following task instructions (e.g., a flat  
338 conditional accuracy function at .5 indicates that someone is just guessing), outlying  
339 individuals, and/or different groups with qualitatively different behavior. Also, to select a  
340 suited bin width for model fitting, one can test and compare various bin widths in the  
341 censor function, and select the smallest one that is supported by the data. Too small bin  
342 widths will result in erratic hazard functions because many bins will have estimates equal  
343 to zero.

344 Table 3 shows the life table for condition “blank” (no prime stimulus presented) for  
345 participant 6.

Table 3

*The life table for the blank prime condition of participant 6.*

bin	risk_set	events	hazard	se_haz	survival	se_surv	ca	se_ca
0	220	NA	NA	NA	1.00	0.00	NA	NA
40	220	0	0.00	0.00	1.00	0.00	NA	NA
80	220	0	0.00	0.00	1.00	0.00	NA	NA
120	220	0	0.00	0.00	1.00	0.00	NA	NA
160	220	0	0.00	0.00	1.00	0.00	NA	NA
200	220	0	0.00	0.00	1.00	0.00	NA	NA
240	220	0	0.00	0.00	1.00	0.00	NA	NA
280	220	7	0.03	0.01	0.97	0.01	0.29	0.17
320	213	13	0.06	0.02	0.91	0.02	0.77	0.12
360	200	26	0.13	0.02	0.79	0.03	0.92	0.05
400	174	40	0.23	0.03	0.61	0.03	1.00	0.00
440	134	48	0.36	0.04	0.39	0.03	0.98	0.02
480	86	37	0.43	0.05	0.22	0.03	1.00	0.00
520	49	32	0.65	0.07	0.08	0.02	1.00	0.00
560	17	9	0.53	0.12	0.04	0.01	1.00	0.00
600	8	4	0.50	0.18	0.02	0.01	1.00	0.00

*Note.* The column named “bin” indicates the endpoint of each time bin (in ms), and includes time point zero. For example the first bin is (0,40] with the starting point excluded and the endpoint included. At time point zero, no events can occur and therefore  $h(t=0)$  and  $ca(t=0)$  are undefined.  $se =$  standard error.  $ca =$  conditional accuracy.  $NA =$  undefined.

<sup>347</sup> probability mass functions for each prime condition for participant 6. By using  
<sup>348</sup> discrete-time hazard functions of event occurrence – in combination with conditional  
<sup>349</sup> accuracy functions for two-choice tasks – one can provide an unbiased, time-varying, and  
<sup>350</sup> probabilistic description of the latency and accuracy of responses based on all trials of any  
<sup>351</sup> RT data set.

## Descriptive stats for subject 6

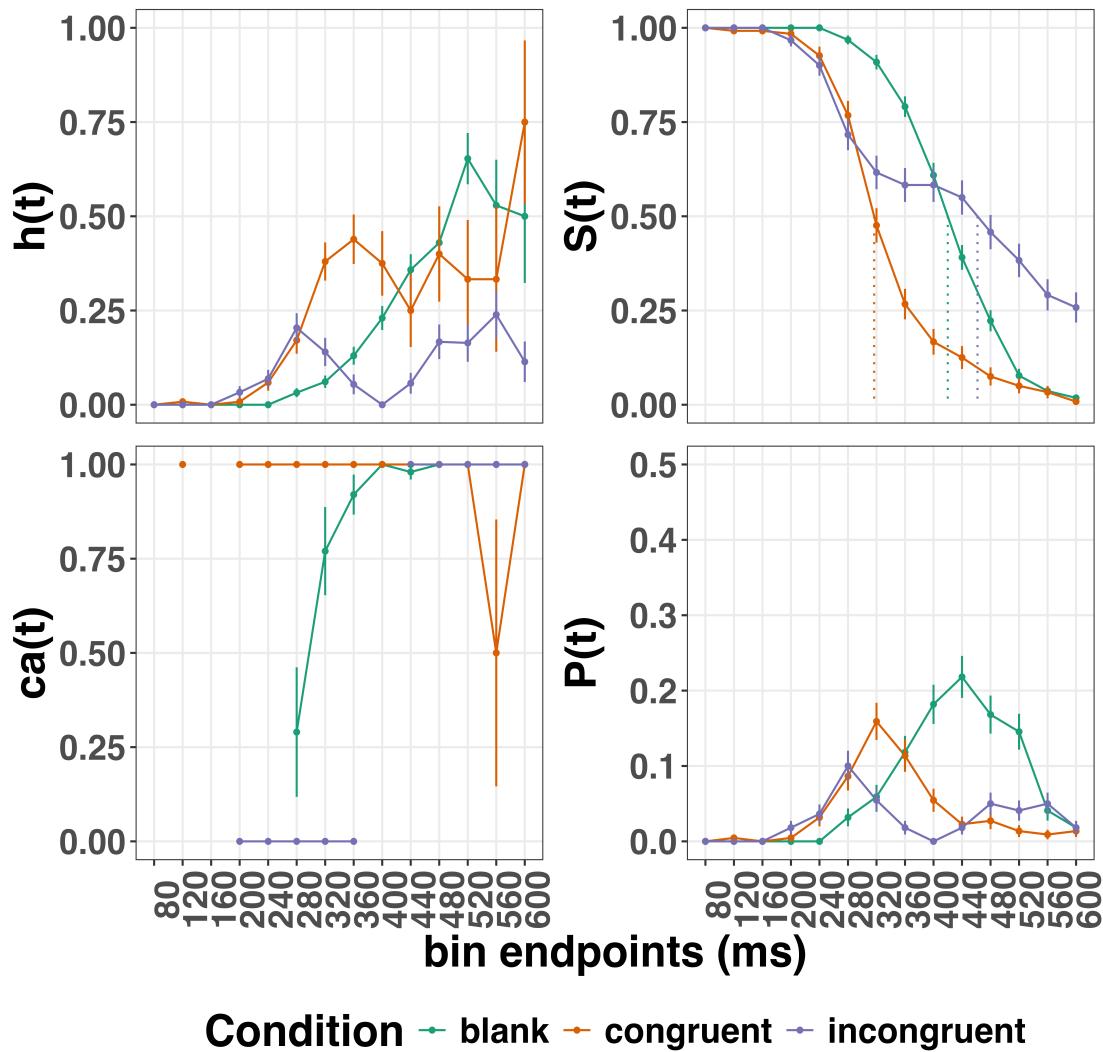


Figure 2. Estimated discrete-time hazard ( $h$ ), survivor ( $S$ ), conditional accuracy ( $ca$ ) and probability mass ( $P$ ) functions for participant 6. Vertical dotted lines indicate the estimated median RTs. Error bars represent  $\pm 1$  standard error of the respective proportion.

352 For example, for participant 6, the estimated hazard values in bin (240,280] are 0.03,

353 0.17, and 0.20 for the blank, congruent, and incongruent prime conditions, respectively. In

354 other words, when the waiting time has increased until *240 ms* after target onset, then the

355 conditional probability of response occurrence in the next 40 ms is more than five times

356 larger for both prime-present conditions, compared to the blank prime condition.

357 Furthermore, the estimated conditional accuracy values in bin (240,280] are 0.29, 1,

358 and 0 for the blank, congruent, and incongruent prime conditions, respectively. In other

359 words, if a response is emitted in bin (240,280], then the probability that it is correct is

360 estimated to be 0.29, 1, and 0 for the blank, congruent, and incongruent prime conditions,

361 respectively.

362 However, when the waiting time has increased until *400 ms* after target onset, then

363 the conditional probability of response occurrence in the next 40 ms is estimated to be

364 0.36, 0.25, and 0.06 for the blank, congruent, and incongruent prime conditions,

365 respectively. And when a response does occur in bin (400,440], then the probability that it

366 is correct is estimated to be 0.98, 1, and 1 for the blank, congruent, and incongruent prime

367 conditions, respectively.

368 These distributional results suggest that participant 6 is initially responding to the

369 prime even though (s)he was instructed to only respond to the target, that response

370 competition emerges in the incongruent prime condition around 300 ms, and that only

371 slower responses are fully controlled by the target stimulus. Qualitatively similar results

372 were obtained for the other five participants. When participants show qualitatively similar

373 distributional patterns, one might consider aggregating their data and plotting the

374 group-average distribution per condition (see Tutorial\_1a.Rmd). More generally, these

375 results go against the (often implicit) assumption in research on priming that all observed

376 responses are primed responses to the target stimulus. Instead, the distributional data

377 show that early responses are triggered exclusively by the prime stimulus, while only later

378 responses reflect primed responses to the target stimulus.

379 At this point, we have calculated and plotted the descriptive statistics for each type  
380 of prime stimulus in EHA/SAT. As we will show in later Tutorials, statistical models for  
381 hazard and conditional accuracy functions can be implemented as generalized linear mixed  
382 regression models predicting event occurrence (1/0) and conditional accuracy (1/0) in each  
383 bin of a selected time window for analysis. But first we consider calculating the descriptive  
384 statistics for within-subject designs with two independent variables.

385 **3.2 Tutorial 1b: Generalising to a more complex design**

386 So far in this paper, we have used a simple experimental design, which involved one  
387 condition with three levels. But psychological experiments are often more complex, with  
388 crossed factorial designs and/or conditions with more than three levels. The purpose of  
389 Tutorial 1b, therefore, is to provide a generalisation of the basic approach, which extends  
390 to a more complicated design. We feel that this might be useful for researchers in  
391 experimental psychology that typically use crossed factorial designs.

392 To this end, Tutorial 1b illustrates how to calculate and plot the descriptive statistics  
393 for the full data set of Experiment 1 of Panis and Schmidt (2016), which includes two  
394 independent variables: mask type and prime type. As we use the same functional  
395 programming approach as in Tutorial 1a, we simply present the sample-based functions for  
396 each participant as part of Tutorial\_1b.Rmd for those that are interested.

397 **3.3 Tutorial 2a: Fitting Bayesian hazard models to interval-censored RT data**

398 In this third tutorial, we illustrate how to fit Bayesian multilevel regression models to  
399 the RT data of the masked response priming data used in Tutorial 1a. Fitting (Bayesian or  
400 non-Bayesian) regression models to time-to-event data is important when you want to  
401 study how the shape of the hazard function depends on various predictors (Singer &

402 Willett, 2003).

403 In general, when fitting regression models, our lab adopts an estimation approach to  
404 multilevel regression (Kruschke & Liddell, 2018; Winter, 2019), which is heavily influenced  
405 by the Bayesian framework as suggested by Richard McElreath (Kurz, 2023b; McElreath,  
406 2020). We also use a “keep it maximal” approach to specifying varying (or random) effects  
407 (Barr, Levy, Scheepers, & Tily, 2013). This means that wherever possible we include  
408 varying intercepts and slopes per participant. To make inferences, we use two main  
409 approaches. We compare models of different complexity, using information criteria (e.g.,  
410 WAIC) and cross-validation (e.g., LOO), to evaluate out-of-sample predictive accuracy  
411 (McElreath, 2020). We also take the most complex model and evaluate key parameters of  
412 interest using point and interval estimates.

413 **3.3.1 Hazard model considerations.** There are several analytic decisions one  
414 has to make when fitting a discrete-time hazard model. First, because the first few bins  
415 typically contain no responses, one has to select an analysis time window, i.e., a contiguous  
416 set of bins for which there is data for each participant. Second, given that the dependent  
417 variable (event occurrence) is binary, one has to select a link function (see section D of the  
418 Supplemental Material). The cloglog link is preferred over the logit link when events can  
419 occur in principle at any time point within a bin, which is the case for RT data (Singer &  
420 Willett, 2003). Third, one has to choose whether to treat TIME (i.e., the time bin index t)  
421 as a categorical or continuous predictor (see also section E of the Supplemental Material).  
422 For example, when you want to know if cloglog-hazard is changing linearly or quadratically  
423 over time, you should treat TIME as a continuous predictor. When you are only interested  
424 in the effect of covariates on hazard, you can treat TIME as a categorical predictor (i.e., fit  
425 an intercept for each bin), in which case you can choose between reference coding and  
426 index coding. With reference coding, one defines the variable as a factor and selects one of  
427 the k categories as the reference level. Brm() will then construct k-1 indicator variables  
428 (see model M1d in Tutorial\_2a.Rmd for an example). With index coding, one constructs

429 an index variable that contains integers that correspond to different categories (see models  
 430 M0i and M1i below). As explained by McElreath (2020), the advantage of index coding is  
 431 that the same prior can be assigned to each level of the index variable, so that each  
 432 category has the same prior uncertainty.

433 In the case of a large- $N$  design without repeated measurements, the parameters of a  
 434 discrete-time hazard model can be estimated using standard logistic regression software  
 435 after expanding the typical person-trial data set into a person-trial-bin data set (Allison,  
 436 2010). When there is clustering in the data, as in the case of a small- $N$  design with  
 437 repeated measurements, the parameters of a discrete-time hazard model can be estimated  
 438 using population-averaged methods (e.g., Generalized Estimating Equations), and Bayesian  
 439 or frequentist generalized linear mixed models (Allison, 2010).

440 In general, there are three assumptions one can make or relax when adding  
 441 experimental predictor variables and other covariates: The linearity assumption for  
 442 continuous predictors (the effect of a 1 unit change is the same anywhere on the scale), the  
 443 additivity assumption (predictors do not interact), and the proportionality assumption  
 444 (predictors do not interact with TIME).

445 In tutorial\_2a.Rmd we fit several Bayesian multilevel models (i.e., generalized linear  
 446 mixed models) that differ in complexity to the person-trial-bin oriented data set that we  
 447 created in Tutorial 1a. We decided to select the analysis time window (200,600] and the  
 448 cloglog link. Below, we shortly discuss two of these models. The person-trial-bin data set is  
 449 prepared as follows.

```
# read in the file we saved in tutorial 1a
ptb_data <- read_csv("Tutorial_1_descriptive_stats/data/inputfile_hazard_modeling.csv")

ptb_data <- ptb_data %>%
  # select analysis time range: (200,600] with 10 bins (time bin ranks 6 to 15)
  filter(period > 5) %>%
```

```

# define categorical predictor TIME as index variable named timebin
mutate(timebin = factor(period, levels = c(6:15)),
       # factor "condition" using reference coding, with "blank" as the reference level
       condition = factor(condition, labels = c("blank", "congruent", "incongruent")),
       # categorical predictor "prime" with index coding
       prime = ifelse(condition=="blank", 1, ifelse(condition=="congruent", 2, 3)),
       prime = factor(prime, levels = c(1,2,3)))

```

450       **3.3.2 Prior distributions.** To get the posterior distribution of each model

451 parameter given the data, we need to specify prior distributions for the model parameters  
 452 which reflect our prior beliefs. In Tutorial\_2a.Rmd we perform a few prior predictive  
 453 checks to make sure our selected prior distributions reflect our prior beliefs (Gelman,  
 454 Vehtari, et al., 2020).

455       The middle column of Supplementary Figure 3 (section F of the Supplemental  
 456 Material) shows six examples of prior distributions for an intercept on the logit and/or  
 457 cloglog scales. While a normal distribution with relatively large variance is often used as a  
 458 weakly informative prior for continuous dependent variables, rows A and B of  
 459 Supplementary Figure 3 show that specifying such distributions on the logit and cloglog  
 460 scales actually leads to rather informative distributions on the original probability scale, as  
 461 most mass is pushed to probabilities of 0 and 1.

462       **3.3.3 Model M0i: A null model with index coding.** When you do not want to  
 463 make assumptions about the shape of the hazard function, or its shape is not smooth but  
 464 irregular, then you can use a general specification of TIME, i.e., fit one grand intercept per  
 465 time bin. In this first baseline or reference model, we use a general specification of TIME  
 466 using index coding, and do not include experimental predictors. We call this model “M0i”.  
 467 The other model (see section 3.3.4) extends model M0i by including our experimental  
 468 predictor prime type.

469       Before we fit model M0i, we select the necessary columns from the data, and specify

470 our priors. In the code of Tutorial 2a, model M0i is specified as follows.

```
model_M0i <-
  brm(data = data_M0i,
       family = bernoulli(link="cloglog"),
       formula = event ~ 0 + timebin + (0 + timebin | pid),
       prior = priors_M0i,
       chains = 4, cores = 4,
       iter = 3000, warmup = 1000,
       control = list(adapt_delta = 0.999,
                      step_size = 0.04,
                      max_treedepth = 12),
       seed = 12, init = "0",
       file = "Tutorial_2_Bayesian/models/model_M0i")
```

471 After selecting the bernoulli family and the cloglog link, the model formula is  
 472 specified. The specification “0 + …” removes the default intercept in brm(). The fixed  
 473 effects include an intercept for each level of timebin. Each of these intercepts is allowed to  
 474 vary across individuals (variable pid). We request 2000 samples from the posterior  
 475 distribution for each of four chains. Estimating model M0i took about 30 minutes on a  
 476 MacBook Pro (Sonoma 14.6.1 OS, 18GB Memory, M3 Pro Chip).

477 **3.3.4 Model M1i: Adding the effects of prime-target congruency.** Previous  
 478 research has shown that psychological effects typically change over time (Panis, 2020;  
 479 Panis, Moran, et al., 2020; Panis & Schmidt, 2022; Panis et al., 2017; Panis & Wagemans,  
 480 2009). In the next model, therefore, we use index coding for both TIME (variable  
 481 “timebin”) and the categorical predictor prime-target-congruency (variable “prime”), so  
 482 that we get 30 grand intercepts, one for each combination of timebin level and prime level.  
 483 Here is the model formula of this model that we call “M1i”.

```
event ~ 0 + timebin:prime + (0 + timebin:prime | pid)
```

484      Estimating model M1i took about 124 minutes.

485      **3.3.5 Compare the models.** We can compare the two models using the Widely  
 486      Applicable Information Criterion (WAIC) and Leave-One-Out (LOO) cross-validation, and  
 487      look at model weights for both criteria (Kurz, 2023a; McElreath, 2020).

```
model_weights(model_M0i, model_M1i, weights = "loo") %>% round(digits = 2)
```

488    ## model\_M0i model\_M1i  
 489    ##            0            1

```
model_weights(model_M0i, model_M1i, weights = "waic") %>% round(digits = 2)
```

490    ## model\_M0i model\_M1i  
 491    ##            0            1

492      Clearly, both the loo and waic weighting schemes assign a weight of 1 to model M1i,  
 493      and a weight of 0 to the other simpler model.

494      **3.3.6 Evaluating parameter estimates in model M1i.** To make inferences  
 495      from the parameter estimates in model M1i, we first plot the densities of the draws from  
 496      the posterior distributions of its population-level parameters in Figure 5, together with  
 497      point (median) and interval estimates (80% and 95% credible intervals).

## Posterior distributions for population-level effects in Model M1i

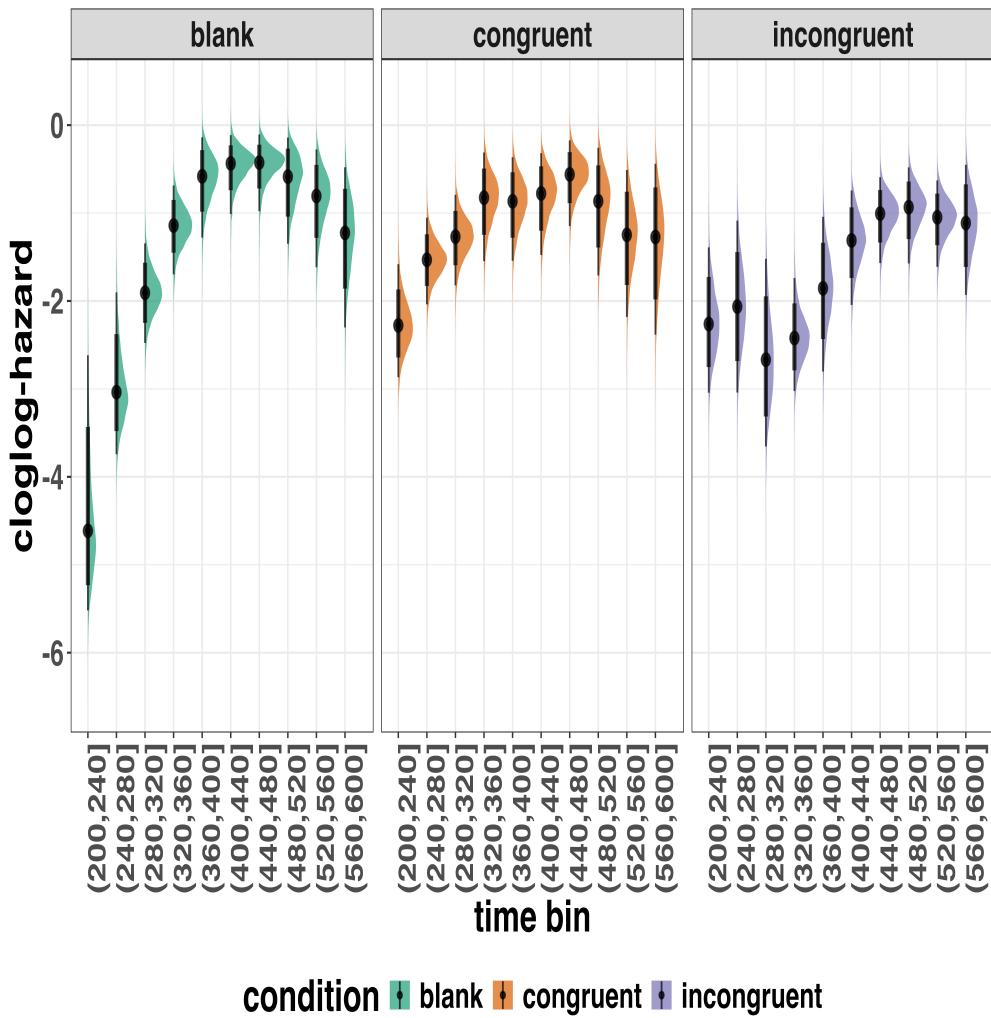


Figure 3. Medians and 80/95% credible intervals of the posterior distributions of the population-level parameters of model M1i.

498 Because the parameter estimates are on the cloglog-hazard scale, we can ease our  
 499 interpretation by plotting the expected value of the posterior predictive distribution – the  
 500 predicted hazard values – at the population level (Figure 6A), and for each participant in  
 501 the data set (Figure 6B).

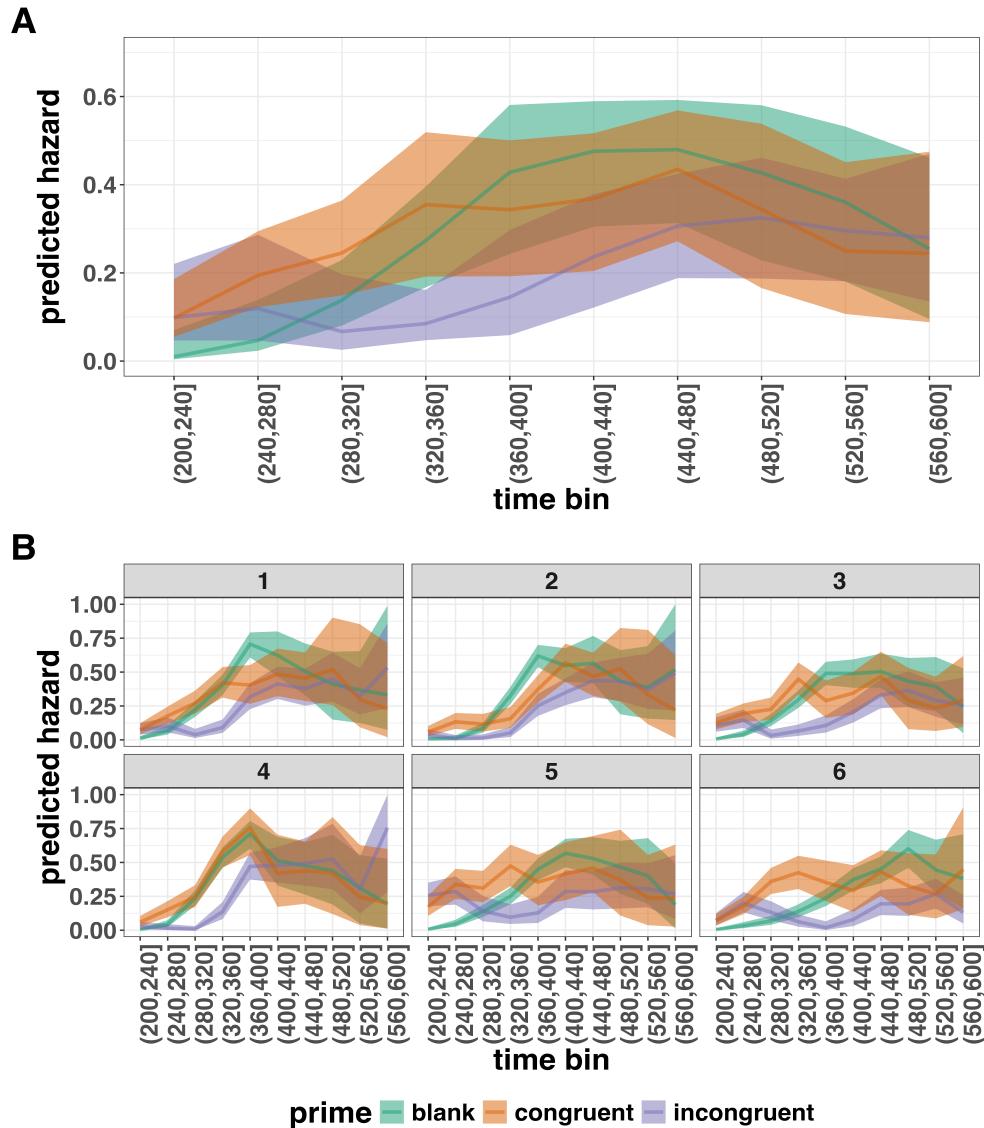


Figure 4. Point (median) and 80/95% credible interval summaries of the hazard estimates (expected values of the draws from the posterior predictive distributions) in each time bin at the population level (A), and for each participant (B).

502 As we are actually interested in the effects of congruent and incongruent primes,

503 relative to the blank prime condition, we can construct two contrasts (congruent-blank,  
504 incongruent-blank), and plot the posterior distributions of these contrast effects, both at  
505 the population level (Figure 7A) and at the participant level (Figure 7B).

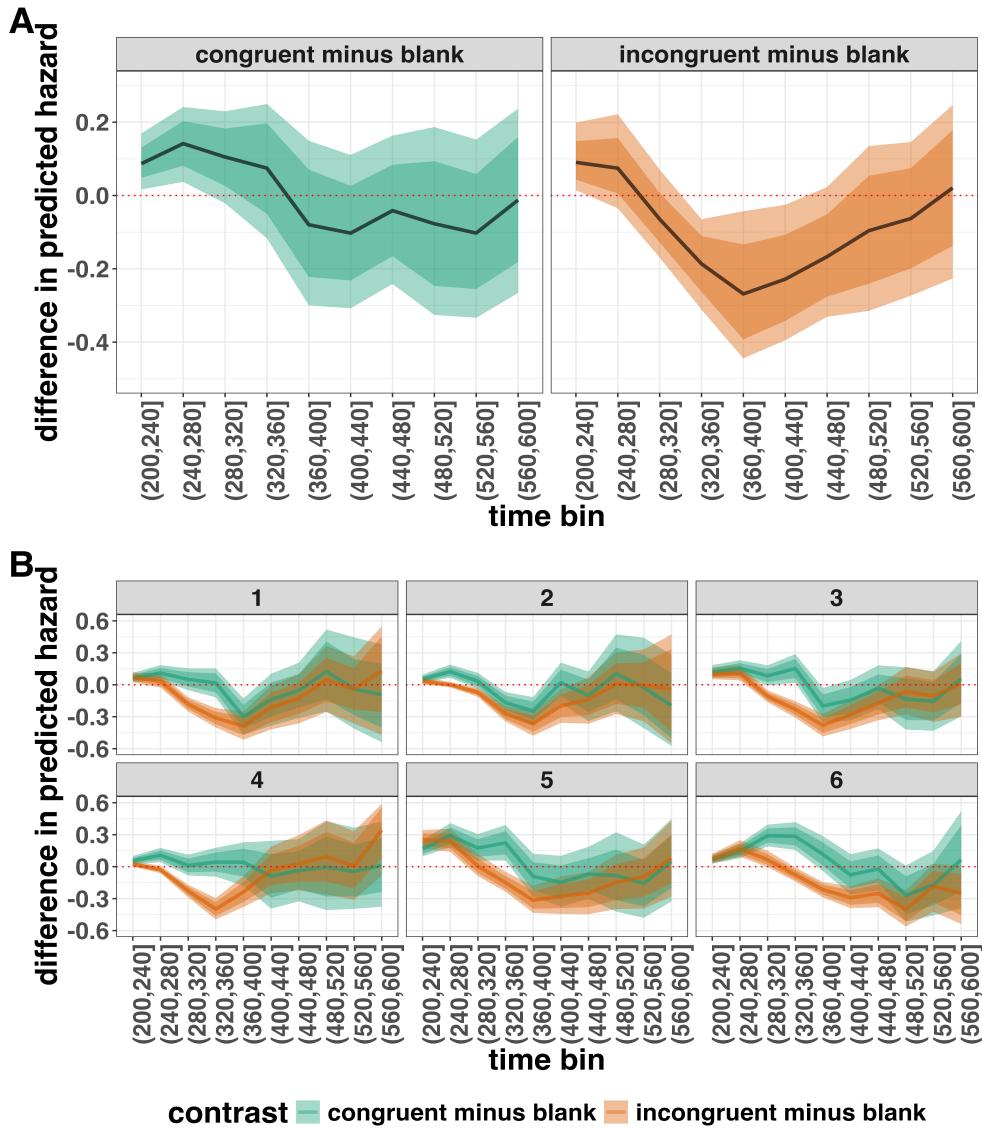


Figure 5. Point (mean) and 80/95% credible interval summaries of estimated differences in hazard in each time bin at the population level (A), and for each participant (B).

506 The point estimates and quantile intervals can also be reported in a table (see

507 Tutorial\_2a.Rmd for details).

508 **Example conclusions for M1i.** What can we conclude from model M1i about

509 our research question, i.e., the temporal dynamics of the effect of prime-target congruency

510 on RT? In other words, in which of the 40-ms time bins between 200 and 600 ms after

511 target onset does changing the prime from blank to congruent or incongruent affect the  
512 hazard of response occurrence (for a prime-target stimulus-onset-asynchrony of 187 ms)?

513 If we want to estimate the population-level effect of prime type on hazard, we can  
514 base our conclusion on Figure 4A. The contrast “congruent minus blank” was estimated to  
515 be 0.09 hazard units in bin (200,240] (95% CrI = [0.02, 0.17]), and 0.14 hazard units in bin  
516 (240,280]) (95% CrI = [0.04, 0.25]). For the other bins, the 95% credible interval contained  
517 zero. The contrast “incongruent minus blank” was estimated to be 0.09 hazard units in bin  
518 (200,240] (95% CrI = [0.01, 0.21]), -0.19 hazard units in bin (320,360] (95% CrI = [-0.31,  
519 -0.06]), -0.27 hazard units in bin (360,400] (95% CrI = [-0.45, -0.04]), and -0.23 hazard  
520 units in bin (400,440] (95% CrI = [-0.40, -0.03]). For the other bins, the 95% credible  
521 interval contained zero.

522 There are thus two phases of performance for the average person between 200 and  
523 600 ms after target onset. In the first phase, the addition of a congruent or incongruent  
524 prime stimulus increases the hazard of response occurrence compared to blank prime trials  
525 in the time period (200, 240]. In the second phase, only the incongruent prime decreases  
526 the hazard of response occurrence compared to blank primes, in the time period (320,440].  
527 The sign of the effect of incongruent primes on the hazard of response occurrence thus  
528 depends on how much waiting time has passed since target onset.

529 If we want to focus more on inter-individual differences, we can study the  
530 subject-specific hazard functions in Figure 4B. Note that three participants (1, 2, and 3)  
531 show a negative difference for the contrast “congruent minus incongruent” in bin (360,400]  
532 – subject 2 also in bin (320,360].

533 Interestingly, all subjects show a tendency in their mean difference (congruent minus  
534 blank) to “dip” around that time (Figure 4B). Therefore, future modeling efforts could  
535 incorporate the trial number into the model formula, in order to also study how the effects  
536 of prime type on hazard change on the long experiment-wide time scale, next to the short

537 trial-wide time scale. In Tutorial\_2a.Rmd we provide a number of model formulae that  
538 should get you going.

539 **3.4 Tutorial 2b: Fitting Bayesian conditional accuracy models**

540 In this fourth tutorial, we illustrate how to fit a Bayesian multilevel regression model  
541 to the timed accuracy data from the masked response priming data used in Tutorial 1a.  
542 The general process is similar to Tutorial 2a, except that (a) we use the person-trial data,  
543 (b) we use the logit link function, and (c) we change the priors. To keep the tutorial short,  
544 we only fit one conditional accuracy model, which was based on model M1i from Tutorial  
545 2a and labelled M1i\_ca.

546 To make inferences from the parameter estimates in model M1i\_ca, we first plot the  
547 densities of the draws from the posterior distributions of its population-level parameters in  
548 Figure 8, together with point (median) and interval estimates (80% and 95% credible  
549 intervals).

## Posterior distributions for population-level effects in Model M1i\_ca

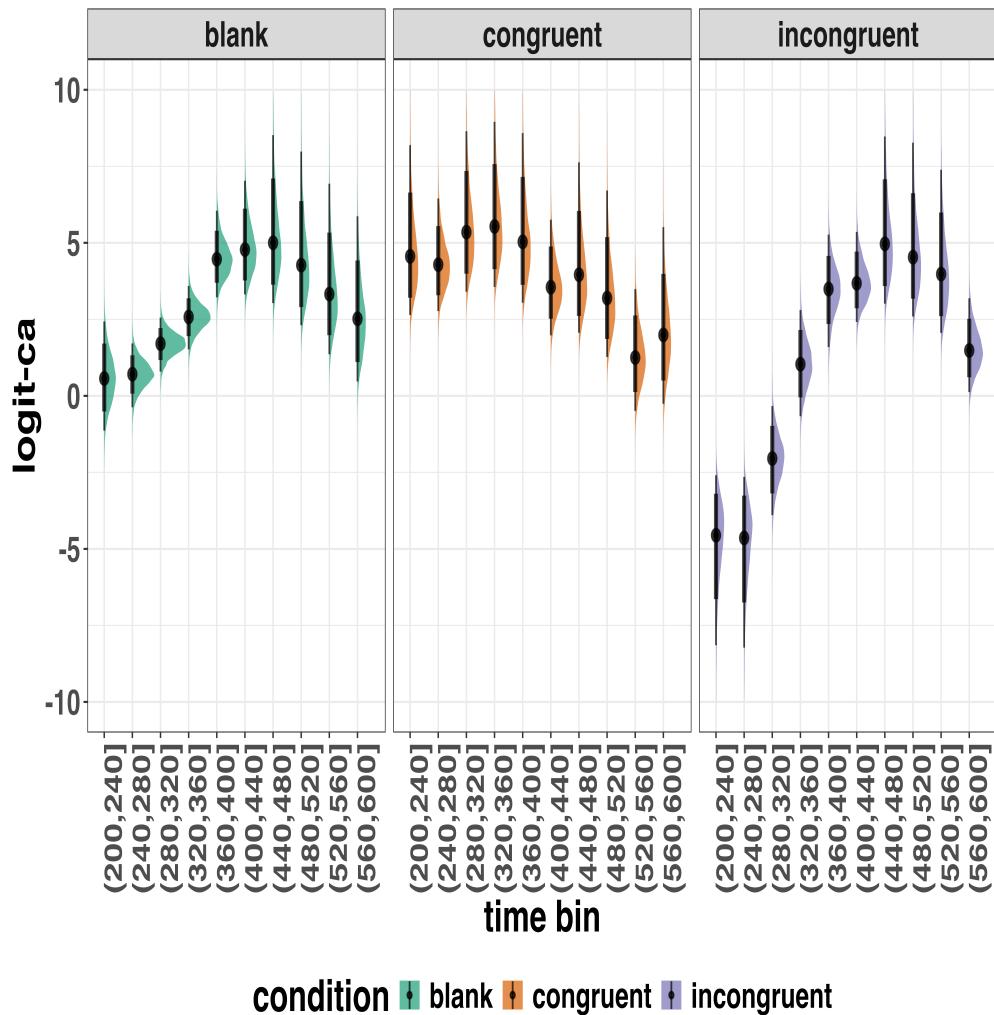
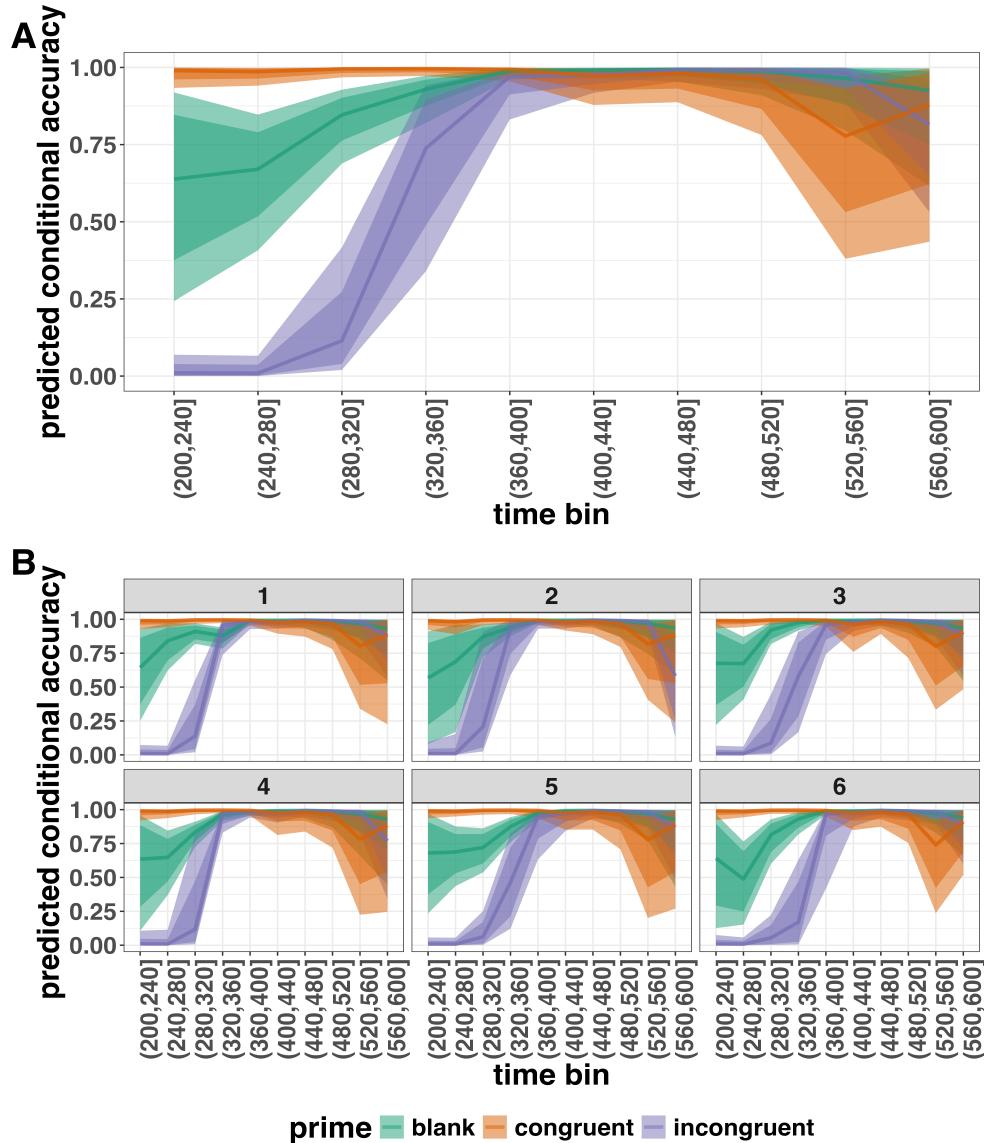


Figure 6. Medians and 80/95% credible intervals of the posterior distributions of the population-level parameters of model M1i\_ca. ca = conditional accuracy.

Because the parameter estimates are on the logit-ca scale, we can ease our

interpretation by plotting the expected value of the posterior predictive distribution – the predicted conditional accuracies – at the population level (Figure 9A), and for each participant in the data set (Figure 9B).



*Figure 7.* Point (median) and 80/95% credible interval summaries of the conditional accuracy estimates (expected values of the draws from the posterior predictive distributions) in each time bin at the population level (A), and for each participant (B).

554 As we are actually interested in the effects of congruent and incongruent primes,  
 555 relative to the blank prime condition, we can construct two contrasts (congruent-blank,  
 556 incongruent-blank), and plot the posterior distributions of these contrast effects at the  
 557 population level (Figure 10A) and for each participant (Figure 10B).

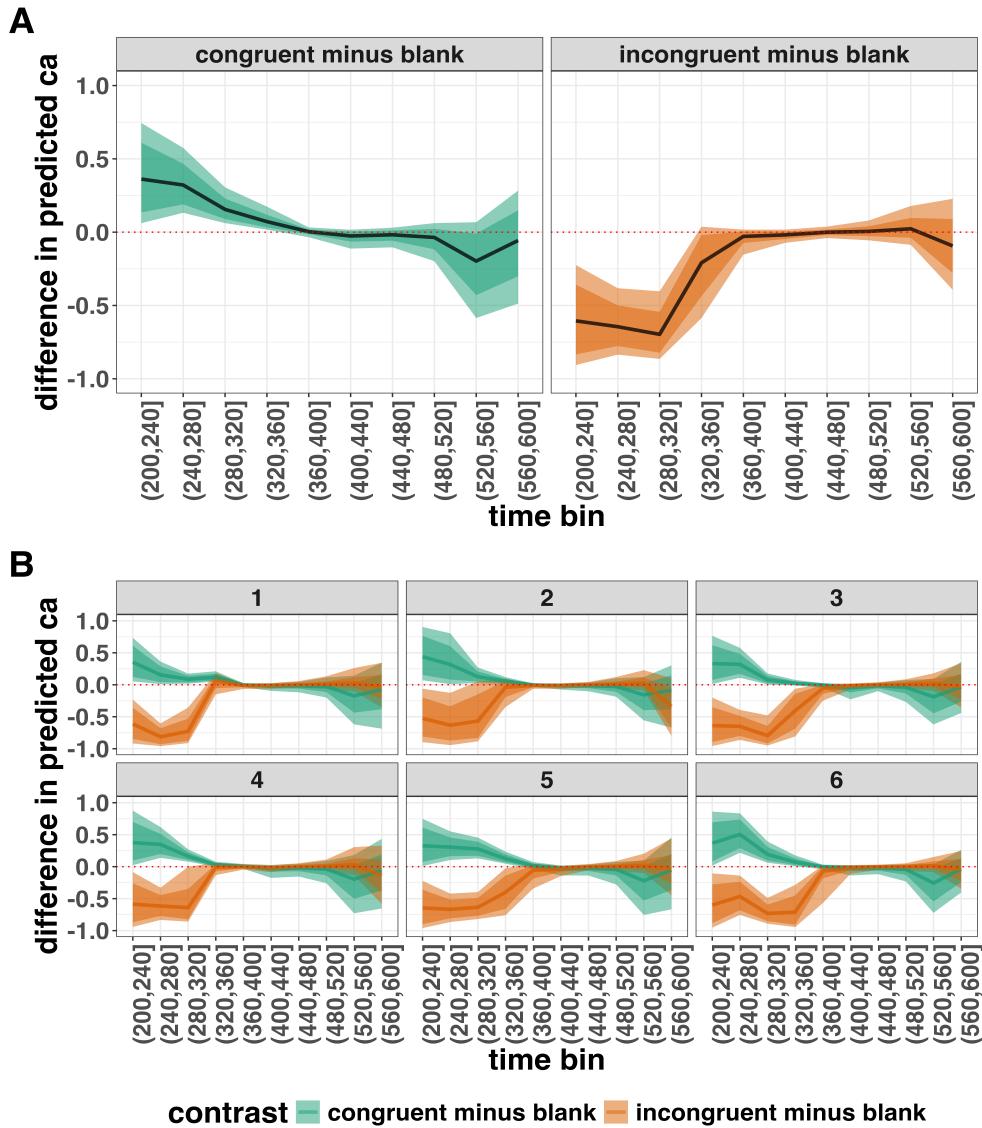


Figure 8. Point (mean) and 80/95% credible interval summaries of estimated differences in conditional accuracy in each time bin at the population level (A), and for each participant (B).

558 Based on Figure 10A we see that on the population level congruent primes have a  
 559 positive effect on the conditional accuracy of emitted responses in time bins (200,240],  
 560 (240,280], (280,320], and (320,360], relative to the estimates in the baseline condition  
 561 (blank prime; red dashed lines in Figure 10A). Incongruent primes have a negative effect on

562 the conditional accuracy of emitted responses in the first time bins, relative to the  
563 estimates in the baseline condition.

564 Before we move to our Tutorial on planning experiments, we also provide code to fit  
565 hazard and conditional accuracy models in the frequentist framework (see  
566 Tutorial\_3a.Rmd and Tutorial\_3b.Rmd). However, because multilevel generalized linear  
567 regression models often do not converge with complex random-coefficient structures, we do  
568 not discuss them here.

### 569 3.5 Tutorial 4: Planning

570 In the final tutorial, we look at planning a future experiment, which uses EHA.

571 **3.5.1 Background.** The general approach to planning that we adopt here involves  
572 simulating reasonably structured data to help guide what you might be able to expect from  
573 your data once you collect it (Gelman, Vehtari, et al., 2020). The basic structure and code  
574 follows the examples outlined by Solomon Kurz in his ‘power’ blog posts  
575 (<https://solomonkurz.netlify.app/blog/bayesian-power-analysis-part-i/>) and Lisa  
576 DeBruine’s R package faux{} (<https://debruine.github.io/faux/>) as well as these related  
577 papers (DeBruine & Barr, 2021; Pargent, Koch, Kleine, Lermer, & Gaube, 2024).

578 **3.5.2 Basic workflow.** The basic workflow is as follows:

- 579 1. Fit a regression model to existing data.
- 580 2. Use the regression model parameters to simulate new data.
- 581 3. Write a function to create 1000s of datasets and vary parameters of interest (e.g.,  
582 sample size, trial count, effect size).
- 583 4. Summarise the simulated data to estimate likely power or precision of the research  
584 design options.

585 Ideally, in the above workflow, we would also fit a model to each dataset and  
586 summarise the model output, rather than the raw data. However, when each model takes

587 several hours to build, and we may want to simulate many 1000s of datasets, it can be  
588 computationally demanding for desktop machines. So, for ease, here we just use the raw  
589 simulated datasets to guide future expectations.

590 In the below, we only provide a high-level summary of the process and let readers  
591 dive into the details within the tutorial should they feel so inclined.

592 **3.5.3 Fit a regression model and simulate one dataset.** We again use the  
593 data from Panis and Schmidt (2016) to provide a worked example. We fit an index coding  
594 model on a subset of time bins (six time bins in total) and for two prime conditions  
595 (congruent and incongruent). We chose to focus on a subsample of the data to ease the  
596 computational burden. We also used a full varying effects structure, with the model  
597 formula as follows:

```
event ~ 0 + timebin:prime + (0 + timebin:prime | pid)
```

598 We then took parameters from this model and used them to create a single dataset  
599 with 200 trials per condition for 10 individual participants. The raw data and the  
600 simulated data are plotted in Figure 12 and show quite close correspondence, which is  
601 re-assuring. But, this is only one dataset. What we really want to do is simulate many  
602 datasets and vary parameters of interest, which is what we turn to in the next section.

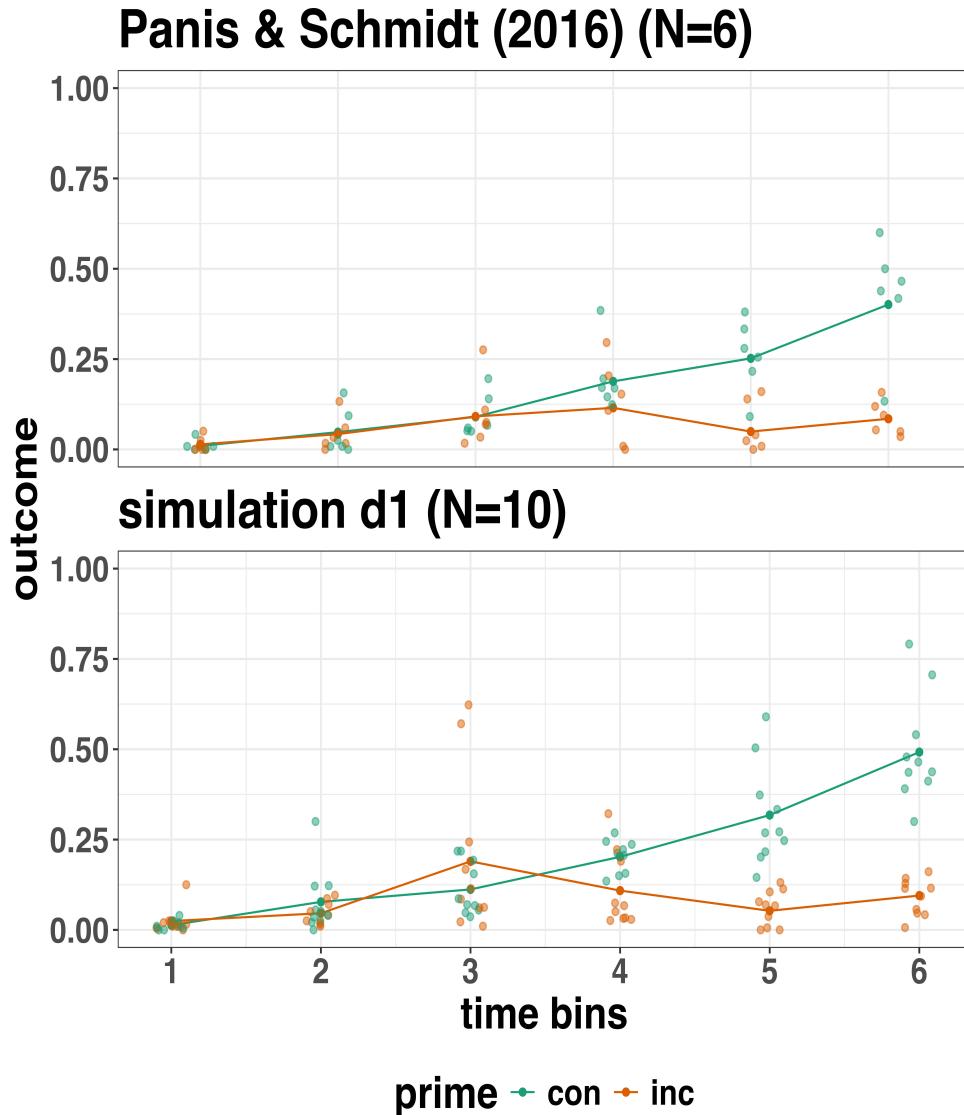


Figure 9. Raw data from Panis and Schmidt (2016) and simulated data from 10 participants.

### 3.5.4 Simulate and summarise data across a range of parameter values.

603     Here we use the same data simulation process as used above, but instead of simulating one  
 604     dataset, we simulate 1000 datasets per variation in parameter values. Specifically, in  
 605     Simulation 1, we vary the number of trials per condition (100, 200, and 400), as well as the  
 606     effect size in bin 6. We focus on bin 6 only, in terms of varying the effect size, just to make  
 607     things simpler and easier to understand. The effect size observed in bin 6 in this subsample  
 608

of data was a 79% reduction in hazard value from the congruent prime (0.401 hazard value) to the incongruent prime condition (0.085 hazard value). In other words, a hazard ratio of 0.21 (e.g.,  $0.085/0.401 = 0.21$ ). As a starting point, we chose three effect sizes, which covered a fairly broad range of hazard ratios (0.25, 0.5, 0.75), which correspond to a 75%, 50% and 25% reduction in hazard value as a function of prime condition.

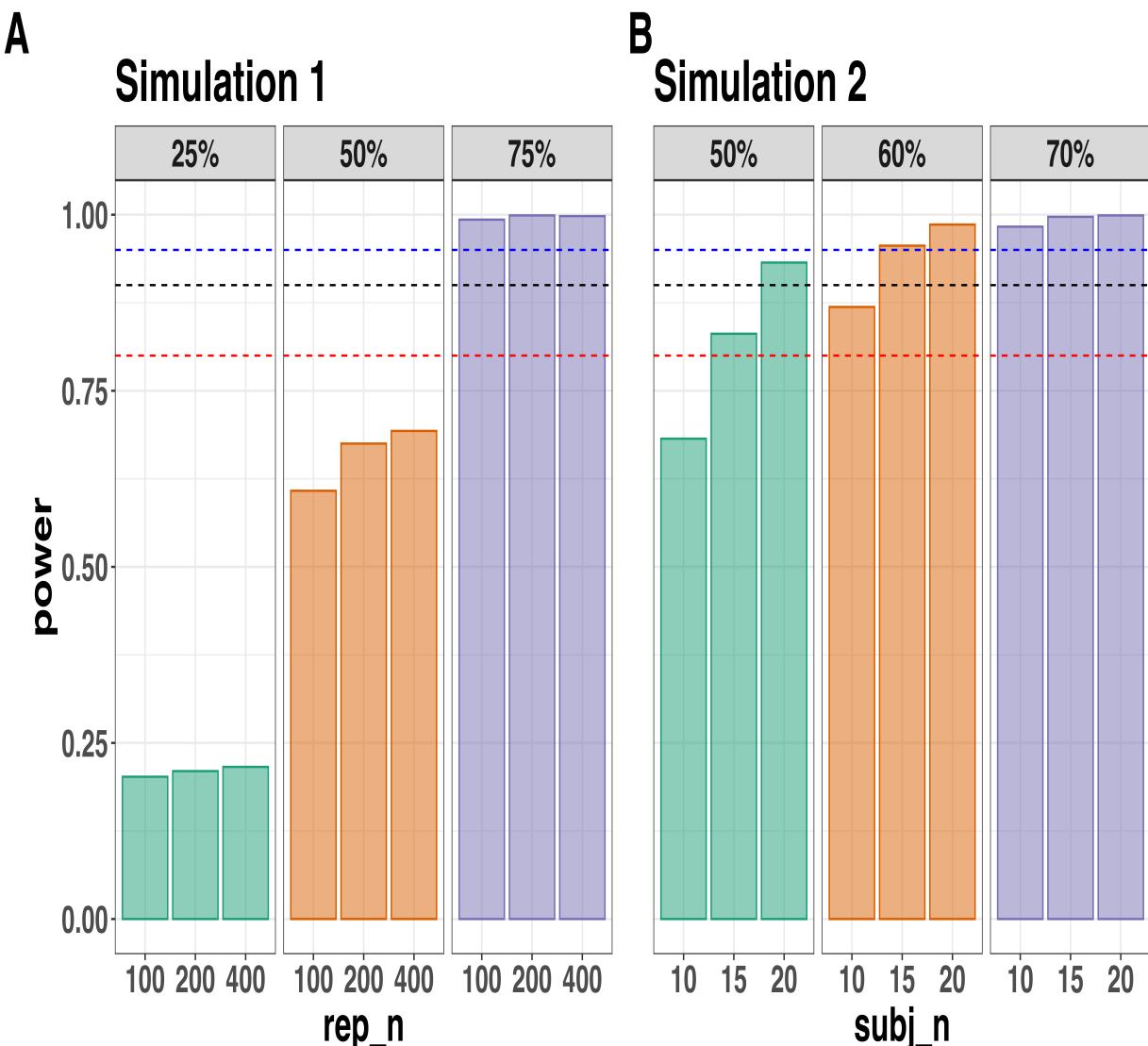
Summary results from Simulation 1 are shown in Figure 13A. Figure 13A depicts statistical “power” as calculated by the percentage of lower-bound 95% confidence intervals that exclude zero when the difference between prime condition is calculated (congruent - incongruent). In other words, what fraction of the simulated datasets generated an effect of prime that excludes the criterion mark of zero. We are aware that “power” is not part of a Bayesian analytical workflow, but we choose to include it here, as it is familiar to most researchers in experimental psychology.

The results of Simulation 1 show that if we were targeting an effect size similar to the one reported in the original study, then testing 10 participants and collecting 100 trials per condition would be enough to provide over 95% power. However, we could not be as confident about smaller effects, such as a hazard ratio of 50% or 25%. From this simulation, we can see that somewhere between an effect size of a 50% and 75% reduction in hazard value, power increases to a range that most researchers would consider acceptable (i.e., >95% power). To probe this space a little further, we decided to run a second simulation, which varied different parameters.

In Simulation 2, we varied the effect size between a different range of values (0.5, 0.4, 0.3), which correspond to a 50%, 60% and 70% reduction in hazard value as a function of prime condition. In addition, we varied the number of participants per experiment between 10, 15, and 20 participants. Given that trial count per condition made little difference to power in Simulation 1, we fixed trial count at 200 trials per condition in Simulation 2. Summary results from Simulation 2 are shown in Figure 13B. A summary of these power

635 calculations might be as follows (trial count = 200 per condition in all cases):

- 636 • For a 70% reduction (0.3 hazard ratio), N=10 would give nearly 100% power.
- 637 • For a 60% reduction (0.4 hazard ratio), N=10 would give nearly 90% power.
- 638 • For a 50% reduction (0.5 hazard ratio), N=15 would give over 80% power.



*Figure 10.* Statistical power across data Simulation 1 (A) and Simulation 2 (B). Power was calculated as the percentage of lower-bound 95% confidence intervals that exclude zero when the difference between prime condition is calculated (congruent - incongruent). In Simulation 1, the effect size was varied between a 25%, 50% and 75% reduction in hazard value, whereas the trial count was varied between 100, 200 and 400 trials per condition (the number of participants was fixed at N=10). In Simulation 2, the effect size was varied between a 50%, 60% and 70% reduction in hazard value, whereas the number of participants was varied between N=10, 15 and 20 (the number of trials per condition was fixed at 200). The dashed lines represent 80% (red), 90% (black) and 95% (blue) power. Abbreviations: rep\_n = the number of trials per experimental condition; subj\_n = the number of participants per simulated experiment.

639       **3.5.5 Planning decisions.** Now that we have summarised our simulated data,

640 what planning decisions could we make about a future study? More concretely, how many

641 trials per condition should we collect and how many participants should we test? Like

642 almost always when planning future studies, the answer depends on your objectives, as well

643 as the available resources (Lakens, 2022). There is no straightforward and clear-cut answer.

644 Some considerations might be as follows:

- 645     • How much power or precision are you looking to obtain in this particular study?

- 646     • Are you running multiple studies that have some form of replication built in?

- 647     • What level of resources do you have at your disposal, such as time, money and

648 personnel?

- 649     • How easy or difficult is it to obtain the specific type of sample?

650       If we were running this kind of study in our lab, what would we do? We might pick a

651 hazard ratio of 0.4 or 0.5 as a target effect size since this is much smaller than that

652 observed previously (Panis & Schmidt, 2016). Then we might pick the corresponding

653 combination of trial count per condition (e.g., 200) and participant sample size (e.g., N=10

654 or N=15) that takes you over the 80% power mark. If we wanted to maximise power based

655 on these simulations, and we had the time and resources available, then we would test

656 N=20 participants, which would provide >90% power for an effect size of 0.5.

657       **But**, and this is an important “but”, unless there are unavoidable reasons, no matter

658 what planning choices we made based on these data simulations, we would not solely rely

659 on data collected from one single study. Instead, we would run a follow-up experiment that

660 replicates and extends the initial result. By doing so, we would aim to avoid the Cult of

661 the Isolated Single Study (Nelder, 1999; Tong, 2019), and thus reduce the reliance on any

662 one type of planning tool, such as a power analysis. Then, we would look for common

663 patterns across two or more experiments, rather than trying to make the case that a single

664 study on its own has sufficient evidential value to hit some criterion mark.

665

#### 4. Discussion

666 This main motivation for writing this paper is the observation that EHA and SAT  
667 analysis remain under-used in psychological research. As a consequence, the field of  
668 psychological research is not taking full advantage of the many benefits EHA/SAT provides  
669 compared to more conventional analyses. By providing a freely available set of tutorials,  
670 which provide step-by-step guidelines and ready-to-use R code, we hope that researchers  
671 will feel more comfortable using EHA/SAT in the future. Indeed, we hope that our  
672 tutorials may help to overcome a barrier to entry with EHA/SAT, which is that such  
673 approaches require more analytical complexity compared to standard approaches. While  
674 we have focused here on within-subject, factorial, small- $N$  designs, it is important to realize  
675 that EHA/SAT can be applied to other designs as well (large- $N$  designs with only one  
676 measurement per subject, between-subject designs, etc.). As such, the general workflow  
677 and associated code can be modified and applied more broadly to other contexts and  
678 research questions. In the following, we discuss the main use-cases, issues relating to model  
679 complexity and interpretability, as well as limitations of the approach.

680 **4.1 What are the main use-cases of EHA for understanding cognition and brain  
681 function?**

682 For those researchers, like ourselves, who are primarily interested in understanding  
683 human cognitive and brain systems, we consider two broadly-defined, main use-cases of  
684 EHA. First, as we hope to have made clear by this point, EHA is one way to investigating  
685 a “temporal states” approach to cognitive processes, by tracking behavior as a function of  
686 step-wise increases in absolute waiting time. EHA thus provides a way to uncover the  
687 microgenesis of cognitive effects, by revealing when cognitive states may start and stop,  
688 how states are replaced with others, as well as what they may be tied to or interact with.  
689 Therefore, if your research questions concern **when psychological states occur, and**

690 how they are temporally organized, our EHA tutorials could be useful tools for you to  
691 use.

692 Second, even if you are not primarily interested in studying the temporal  
693 organization of cognitive states, EHA could still be a useful tool to consider using, in order  
694 to qualify inferences that are being made based on comparisons between means. Given that  
695 distinctly different inferences can be made from the same data based on whether one  
696 computes a mean across trials or a RT distribution of events (Figure 1), it may be  
697 important for researchers to supplement comparisons between means with EHA. Therefore,  
698 if you have a lot of right-censored observations in your RT data set, and/or your research  
699 question concerns whether and when responses occur, and whether and when experimental  
700 manipulations affect the instantaneous risk of response occurrence, then EHA should be  
701 your method of choice.

## 702 4.2 Model complexity versus interpretability

703 Hazard and conditional accuracy models can quickly become very complex when  
704 adding more than one time scale, due to the many possible higher-order interactions. For  
705 example, some of the models discussed in Tutorial 2a, which we did not focus on in the  
706 main text, contain two time scales as covariates: the passage of time on the within-trial  
707 time scale, and the passage of time on the across-trial (or within-experiment) time scale.  
708 However, when trials are presented in blocks, and blocks of trials within sessions, and when  
709 the experiment comprises a number of sessions, then four time scales can be defined  
710 (within-trial, within-block, within-session, and within-experiment). From a theoretical  
711 perspective, adding more than one time scale – and their interactions – can be important  
712 to capture plasticity and other learning effects that may play out on such longer time  
713 scales, and that are probably present in each experiment in general (Schöner & Spencer,  
714 2016). From a practical perspective, therefore, some choices need to be made to balance  
715 the amount of data that is being collected per participant, condition and across the varying

716 timescales. As one example, if there are several times cales of relevance, then it might be  
717 prudent for interpretational purposes to limit the number of experimental predictor  
718 variables (conditions). This is of course where planning and data simulation efforts would  
719 be important to provide a guide to experimental design choices (see Tutorial 4 and section  
720 2.3).

721 **4.3 Limitations**

722 Compared to the orthodox method – comparing means between conditions – the  
723 most important limitation of multilevel hazard and conditional accuracy modeling is that it  
724 might take a long time to estimate the parameters using Bayesian methods or the model  
725 might have to be simplified significantly to use frequentist methods.

726 Another issue is that you need a relatively large number of trials per condition to  
727 estimate the discrete-time hazard function with relatively high temporal resolution (e.g., 20  
728 ms), which is required when testing predictions of process models of cognition. Indeed, in  
729 general, there is a trade-off between the number of trials per condition and the temporal  
730 resolution (i.e., bin width) of the discrete-time hazard function. Therefore, we recommend  
731 researchers to collect as many trials as possible per experimental condition, given the  
732 available resources and considering the participant experience (e.g., fatigue and boredom).  
733 For instance, if the maximum session length deemed reasonable is between 1 and 2 hours,  
734 what is the maximum number of trials per condition that you could reasonably collect?  
735 After consideration, it might be worth conducting multiple testing sessions per participant  
736 and/or reducing the number of experimental conditions. There is a user-friendly online tool  
737 for calculating statistical power as a function of the number of trials as well as the number  
738 of participants, and this might be worth consulting to guide the research design process  
739 (Baker et al., 2021). Finally, if you have a lot of repeated measurements per condition per  
740 participant, you can of course also try continuous-time methods (Allison, 2010; Elmer et  
741 al., 2023).

742

## 5. Conclusions

743       Estimating the temporal distributions of RT and accuracy provide a rich source of  
744      information on the time course of cognitive processing, which have been largely  
745      undervalued in the history of experimental psychology and cognitive neuroscience. We  
746      hope that by providing a set of hands-on, step-by-step tutorials, which come with  
747      custom-built and freely available code, researchers will feel more comfortable embracing  
748      EHA and investigating the shape of empirical hazard functions and the temporal profile of  
749      cognitive states. On a broader level, we think that wider adoption of such approaches will  
750      have a meaningful impact on the inferences drawn from data, as well as the development of  
751      theories regarding the structure of cognition.

752

**Author contributions**

753

Conceptualization: S. Panis and R. Ramsey; Software: S. Panis and R. Ramsey;

754

Writing - Original Draft Preparation: S. Panis; Writing - Review & Editing: S. Panis and

755

R. Ramsey; Supervision: R. Ramsey.

756

**Conflicts of Interest**

757

The author(s) declare that there were no conflicts of interest with respect to the

758

authorship or the publication of this article.

759

**Prior versions**

760

All of the submitted manuscript and Supplemental Material was previously posted to

761

a preprint archive: <https://doi.org/10.31234/osf.io/57bh6>

762

**Supplemental Material**

763

**Disclosures**

764

**Data, materials, and online resources**

765

Link to public archive:

766

[https://github.com/sven-panis/Tutorial\\_Event\\_History\\_Analysis](https://github.com/sven-panis/Tutorial_Event_History_Analysis)

767

Supplemental Material: Panis\_Ramsey\_suppl\_material.pdf

768

**Ethical approval**

769

Ethical approval was not required for this tutorial in which we reanalyze existing

770

data sets.

**References**

- 771
- 772 Abney, D. H., Fausey, C. M., Suarez-Rivera, C., & Tamis-LeMonda, C. S. (2025).  
773 Advancing a temporal science of behavior. *Trends in Cognitive Sciences*.  
774 <https://doi.org/10.1016/j.tics.2025.05.010>
- 775 Allison, P. D. (1982). Discrete-Time Methods for the Analysis of Event Histories.  
776 *Sociological Methodology*, 13, 61. <https://doi.org/10.2307/270718>
- 777 Allison, P. D. (2010). *Survival analysis using SAS: A practical guide* (2. ed.). Cary, NC:  
778 SAS Press.
- 779 Aust, F. (2019). *Citr: 'RStudio' add-in to insert markdown citations*. Retrieved from  
780 <https://github.com/crsh/citr>
- 781 Aust, F., & Barth, M. (2024). *papaja: Prepare reproducible APA journal articles with R*  
782 *Markdown*. <https://doi.org/10.32614/CRAN.package.papaja>
- 783 Baker, D. H., Vilidaite, G., Lygo, F. A., Smith, A. K., Flack, T. R., Gouws, A. D., &  
784 Andrews, T. J. (2021). Power contours: Optimising sample size and precision in  
785 experimental psychology and human neuroscience. *Psychological Methods*, 26(3),  
786 295–314. <https://doi.org/10.1037/met0000337>
- 787 Barack, D. L., & Krakauer, J. W. (2021). Two views on the cognitive brain. *Nature*  
788 *Reviews Neuroscience*, 22(6), 359–371. <https://doi.org/10.1038/s41583-021-00448-6>
- 789 Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for  
790 confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*,  
791 68(3), 10.1016/j.jml.2012.11.001. <https://doi.org/10.1016/j.jml.2012.11.001>
- 792 Barth, M. (2023). *tinylabes: Lightweight variable labels*. Retrieved from  
793 <https://cran.r-project.org/package=tinylabes>
- 794 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects  
795 models using lme4. *Journal of Statistical Software*, 67(1), 1–48.  
796 <https://doi.org/10.18637/jss.v067.i01>
- 797 Bates, D., Maechler, M., & Jagan, M. (2024). *Matrix: Sparse and dense matrix classes and*

- 798       methods. Retrieved from <https://Matrix.R-forge.R-project.org>
- 799   Bengtsson, H. (2021). A unifying framework for parallel and distributed processing in r  
800       using futures. *The R Journal*, 13(2), 208–227. <https://doi.org/10.32614/RJ-2021-048>
- 801   Berger, A., & Kiefer, M. (2021). Comparison of Different Response Time Outlier Exclusion  
802       Methods: A Simulation Study. *Frontiers in Psychology*, 12, 675558.  
803       <https://doi.org/10.3389/fpsyg.2021.675558>
- 804   Blossfeld, H.-P., & Rohwer, G. (2002). *Techniques of event history modeling: New  
805       approaches to causal analysis*, 2nd ed (pp. x, 310). Mahwah, NJ, US: Lawrence  
806       Erlbaum Associates Publishers.
- 807   Bloxom, B. (1984). Estimating response time hazard functions: An exposition and  
808       extension. *Journal of Mathematical Psychology*, 28(4), 401–420.  
809       [https://doi.org/10.1016/0022-2496\(84\)90008-7](https://doi.org/10.1016/0022-2496(84)90008-7)
- 810   Bolger, N., Zee, K. S., Rossignac-Milon, M., & Hassin, R. R. (2019). Causal processes in  
811       psychology are heterogeneous. *Journal of Experimental Psychology: General*, 148(4),  
812       601–618. <https://doi.org/10.1037/xge0000558>
- 813   Box-Steffensmeier, J. M. (2004). Event history modeling: A guide for social scientists.  
814       Cambridge: University Press.
- 815   Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan.  
816       *Journal of Statistical Software*, 80(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>
- 817   Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms.  
818       *The R Journal*, 10(1), 395–411. <https://doi.org/10.32614/RJ-2018-017>
- 819   Bürkner, P.-C. (2021). Bayesian item response modeling in R with brms and Stan. *Journal  
820       of Statistical Software*, 100(5), 1–54. <https://doi.org/10.18637/jss.v100.i05>
- 821   DeBruine, L. M., & Barr, D. J. (2021). Understanding Mixed-Effects Models Through  
822       Data Simulation. *Advances in Methods and Practices in Psychological Science*, 4(1),  
823       2515245920965119. <https://doi.org/10.1177/2515245920965119>
- 824   Eddelbuettel, D., & Balamuta, J. J. (2018). Extending R with C++: A Brief Introduction

- 825 to Rcpp. *The American Statistician*, 72(1), 28–36.
- 826 <https://doi.org/10.1080/00031305.2017.1375990>
- 827 Eddelbuettel, D., & François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal  
828 of Statistical Software*, 40(8), 1–18. <https://doi.org/10.18637/jss.v040.i08>
- 829 Elmer, T., Van Duijn, M. A. J., Ram, N., & Bringmann, L. F. (2023). Modeling  
830 categorical time-to-event data: The example of social interaction dynamics captured  
831 with event-contingent experience sampling methods. *Psychological Methods*.  
832 <https://doi.org/10.1037/met0000598>
- 833 Gabry, J., Češnovar, R., Johnson, A., & Broder, S. (2024). *Cmdstanr: R interface to  
834 'CmdStan'*. Retrieved from <https://github.com/stan-dev/cmdstanr>
- 835 Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization  
836 in bayesian workflow. *J. R. Stat. Soc. A*, 182, 389–402.  
837 <https://doi.org/10.1111/rssa.12378>
- 838 Gelman, A., Hill, J., & Vehtari, A. (2020). Regression and Other Stories.  
839 [https://www.cambridge.org/highereducation/books/regression-and-other-  
841 stories/DD20DD6C9057118581076E54E40C372C](https://www.cambridge.org/highereducation/books/regression-and-other-<br/>840 stories/DD20DD6C9057118581076E54E40C372C); Cambridge University Press.  
<https://doi.org/10.1017/9781139161879>
- 842 Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y., ...  
843 Modrák, M. (2020). *Bayesian Workflow*. arXiv.  
844 <https://doi.org/10.48550/arXiv.2011.01808>
- 845 Girard, J. (2024). *Standist: What the package does (one line, title case)*. Retrieved from  
846 <https://github.com/jmgirard/standist>
- 847 Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate.  
848 *Journal of Statistical Software*, 40(3), 1–25. Retrieved from  
849 <https://www.jstatsoft.org/v40/i03/>
- 850 Heiss, A. (2021, November 10). A Guide to Correctly Calculating Posterior Predictions  
851 and Average Marginal Effects with Multilevel Bayesian Models.

- 852 https://doi.org/10.59350/wbn93-edb02
- 853 Holden, J. G., Van Orden, G. C., & Turvey, M. T. (2009). Dispersion of response times  
854 reveals cognitive dynamics. *Psychological Review*, 116(2), 318–342.
- 855 https://doi.org/10.1037/a0014849
- 856 Hosmer, D. W., Lemeshow, S., & May, S. (2011). *Applied Survival Analysis: Regression*  
857 *Modeling of Time to Event Data* (2nd ed). Hoboken: John Wiley & Sons.
- 858 Kantowitz, B. H., & Pachella, R. G. (2021). The Interpretation of Reaction Time in  
859 Information-Processing Research 1. *Human Information Processing*, 41–82.
- 860 https://doi.org/10.4324/9781003176688-2
- 861 Kay, M. (2024). *tidybayes: Tidy data and geoms for Bayesian models*.  
862 https://doi.org/10.5281/zenodo.1308151
- 863 Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing,  
864 estimation, meta-analysis, and power analysis from a Bayesian perspective.  
865 *Psychonomic Bulletin & Review*, 25(1), 178–206.
- 866 https://doi.org/10.3758/s13423-016-1221-4
- 867 Kurz, A. S. (2023a). *Applied longitudinal data analysis in brms and the tidyverse* (version  
868 0.0.3). Retrieved from https://bookdown.org/content/4253/
- 869 Kurz, A. S. (2023b). *Statistical rethinking with brms, ggplot2, and the tidyverse: Second*  
870 *edition* (version 0.4.0). Retrieved from https://bookdown.org/content/4857/
- 871 Lakens, D. (2022). Sample Size Justification. *Collabra: Psychology*, 8(1), 33267.  
872 https://doi.org/10.1525/collabra.33267
- 873 Landes, J., Engelhardt, S. C., & Pelletier, F. (2020). An introduction to event history  
874 analyses for ecologists. *Ecosphere*, 11(10), e03238. https://doi.org/10.1002/ecs2.3238
- 875 Lougheed, J. P., Benson, L., Cole, P. M., & Ram, N. (2019). Multilevel survival analysis:  
876 Studying the timing of children's recurring behaviors. *Developmental Psychology*,  
877 55(1), 53–65. https://doi.org/10.1037/dev0000619
- 878 Luce, R. D. (1991). *Response times: Their role in inferring elementary mental organization*

- 879 (1. issued as paperback). Oxford: Univ. Press.
- 880 McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and*  
881 *STAN* (2nd ed.). New York: Chapman and Hall/CRC.  
882 <https://doi.org/10.1201/9780429029608>
- 883 Mills, M. (2011). *Introducing Survival and Event History Analysis*. 1 Oliver's Yard, 55 City  
884 Road, London EC1Y 1SP United Kingdom: SAGE Publications Ltd.  
885 <https://doi.org/10.4135/9781446268360>
- 886 Müller, K., & Wickham, H. (2023). *Tibble: Simple data frames*. Retrieved from  
887 <https://CRAN.R-project.org/package=tibble>
- 888 Nelder, J. A. (1999). From Statistics to Statistical Science. *Journal of the Royal Statistical*  
889 *Society. Series D (The Statistician)*, 48(2), 257–269. Retrieved from  
890 <https://www.jstor.org/stable/2681191>
- 891 Neuwirth, E. (2022). *RColorBrewer: ColorBrewer palettes*. Retrieved from  
892 <https://CRAN.R-project.org/package=RColorBrewer>
- 893 Panis, S. (2020). How can we learn what attention is? Response gating via multiple direct  
894 routes kept in check by inhibitory control processes. *Open Psychology*, 2(1), 238–279.  
895 <https://doi.org/10.1515/psych-2020-0107>
- 896 Panis, S., Moran, R., Wolkersdorfer, M. P., & Schmidt, T. (2020). Studying the dynamics  
897 of visual search behavior using RT hazard and micro-level speed–accuracy tradeoff  
898 functions: A role for recurrent object recognition and cognitive control processes.  
899 *Attention, Perception, & Psychophysics*, 82(2), 689–714.  
900 <https://doi.org/10.3758/s13414-019-01897-z>
- 901 Panis, S., Schmidt, F., Wolkersdorfer, M. P., & Schmidt, T. (2020). Analyzing Response  
902 Times and Other Types of Time-to-Event Data Using Event History Analysis: A Tool  
903 for Mental Chronometry and Cognitive Psychophysiology. *I-Perception*, 11(6),  
904 2041669520978673. <https://doi.org/10.1177/2041669520978673>
- 905 Panis, S., & Schmidt, T. (2016). What Is Shaping RT and Accuracy Distributions? Active

- 906 and Selective Response Inhibition Causes the Negative Compatibility Effect. *Journal of*  
907 *Cognitive Neuroscience*, 28(11), 1651–1671. [https://doi.org/10.1162/jocn\\_a\\_00998](https://doi.org/10.1162/jocn_a_00998)
- 908 Panis, S., & Schmidt, T. (2022). When does “inhibition of return” occur in spatial cueing  
909 tasks? Temporally disentangling multiple cue-triggered effects using response history  
910 and conditional accuracy analyses. *Open Psychology*, 4(1), 84–114.  
911 <https://doi.org/10.1515/psych-2022-0005>
- 912 Panis, S., Torfs, K., Gillebert, C. R., Wagemans, J., & Humphreys, G. W. (2017).  
913 Neuropsychological evidence for the temporal dynamics of category-specific naming.  
914 *Visual Cognition*, 25(1-3), 79–99. <https://doi.org/10.1080/13506285.2017.1330790>
- 915 Panis, S., & Wagemans, J. (2009). Time-course contingencies in perceptual organization  
916 and identification of fragmented object outlines. *Journal of Experimental Psychology:  
917 Human Perception and Performance*, 35(3), 661–687.  
918 <https://doi.org/10.1037/a0013547>
- 919 Pargent, F., Koch, T. K., Kleine, A.-K., Lermer, E., & Gaube, S. (2024). A Tutorial on  
920 Tailored Simulation-Based Sample-Size Planning for Experimental Designs With  
921 Generalized Linear Mixed Models. *Advances in Methods and Practices in Psychological  
922 Science*, 7(4), 25152459241287132. <https://doi.org/10.1177/25152459241287132>
- 923 Pedersen, T. L. (2024). *Patchwork: The composer of plots*. Retrieved from  
924 <https://patchwork.data-imaginist.com/>
- 925 Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in s and s-PLUS*. New York:  
926 Springer. <https://doi.org/10.1007/b98882>
- 927 R Core Team. (2024). *R: A language and environment for statistical computing*. Vienna,  
928 Austria: R Foundation for Statistical Computing. Retrieved from  
929 <https://www.R-project.org/>
- 930 Schöner, G., & Spencer, J. P. (2016). *Dynamic thinking: A primer on dynamic field theory*.  
931 New York, NY: Oxford University Press.  
932 <https://doi.org/10.1093/acprof:oso/9780199300563.001.0001>

- 933 Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling  
934 Change and Event Occurrence*. Oxford, New York: Oxford University Press.
- 935 Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-N design.  
936 *Psychonomic Bulletin & Review*, 25(6), 2083–2101.  
937 <https://doi.org/10.3758/s13423-018-1451-8>
- 938 Stan Development Team. (2020). *StanHeaders: Headers for the R interface to Stan*.  
939 Retrieved from <https://mc-stan.org/>
- 940 Stan Development Team. (2024). *RStan: The R interface to Stan*. Retrieved from  
941 <https://mc-stan.org/>
- 942 Stoolmiller, M. (2015). *An Introduction to Using Multivariate Multilevel Survival Analysis  
943 to Study Coercive Family Process* (Vol. 1; T. J. Dishion & J. Snyder, Eds.). Oxford  
944 University Press. <https://doi.org/10.1093/oxfordhb/9780199324552.013.27>
- 945 Stoolmiller, M., & Snyder, J. (2006). Modeling heterogeneity in social interaction processes  
946 using multilevel survival analysis. *Psychological Methods*, 11(2), 164–177.  
947 <https://doi.org/10.1037/1082-989X.11.2.164>
- 948 Teachman, J. D. (1983). Analyzing social processes: Life tables and proportional hazards  
949 models. *Social Science Research*, 12(3), 263–301.  
950 [https://doi.org/10.1016/0049-089X\(83\)90015-7](https://doi.org/10.1016/0049-089X(83)90015-7)
- 951 Tong, C. (2019). Statistical Inference Enables Bad Science; Statistical Thinking Enables  
952 Good Science. *The American Statistician*, 73(sup1), 246–261.  
953 <https://doi.org/10.1080/00031305.2018.1518264>
- 954 Townsend, J. T. (1990). Truth and consequences of ordinal differences in statistical  
955 distributions: Toward a theory of hierarchical inference. *Psychological Bulletin*, 108(3),  
956 551–567. <https://doi.org/10.1037/0033-2909.108.3.551>
- 957 Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin &  
958 Review*, 7(3), 424–465. <https://doi.org/10.3758/BF03214357>
- 959 Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics.

- 960        *Acta Psychologica*, 41(1), 67–85. [https://doi.org/10.1016/0001-6918\(77\)90012-9](https://doi.org/10.1016/0001-6918(77)90012-9)
- 961    Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New  
962        York. Retrieved from <https://ggplot2.tidyverse.org>
- 963    Wickham, H. (2023a). *Forcats: Tools for working with categorical variables (factors)*.  
964        Retrieved from <https://forcats.tidyverse.org/>
- 965    Wickham, H. (2023b). *Stringr: Simple, consistent wrappers for common string operations*.  
966        Retrieved from <https://stringr.tidyverse.org>
- 967    Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., ...  
968        Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43),  
969        1686. <https://doi.org/10.21105/joss.01686>
- 970    Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *Dplyr: A  
971        grammar of data manipulation*. Retrieved from <https://dplyr.tidyverse.org>
- 972    Wickham, H., & Henry, L. (2023). *Purrr: Functional programming tools*. Retrieved from  
973        <https://purrr.tidyverse.org/>
- 974    Wickham, H., Hester, J., & Bryan, J. (2024). *Readr: Read rectangular text data*. Retrieved  
975        from <https://readr.tidyverse.org>
- 976    Wickham, H., Vaughan, D., & Girlich, M. (2024). *Tidyr: Tidy messy data*. Retrieved from  
977        <https://tidyr.tidyverse.org>
- 978    Winter, B. (2019). *Statistics for Linguists: An Introduction Using R*. New York:  
979        Routledge. <https://doi.org/10.4324/9781315165547>
- 980    Wolkersdorfer, M. P., Panis, S., & Schmidt, T. (2020). Temporal dynamics of sequential  
981        motor activation in a dual-prime paradigm: Insights from conditional accuracy and  
982        hazard functions. *Attention, Perception, & Psychophysics*, 82(5), 2581–2602.  
983        <https://doi.org/10.3758/s13414-020-02010-5>