

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

3 Sven Panis¹ & Richard Ramsey¹

4 ¹ 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).

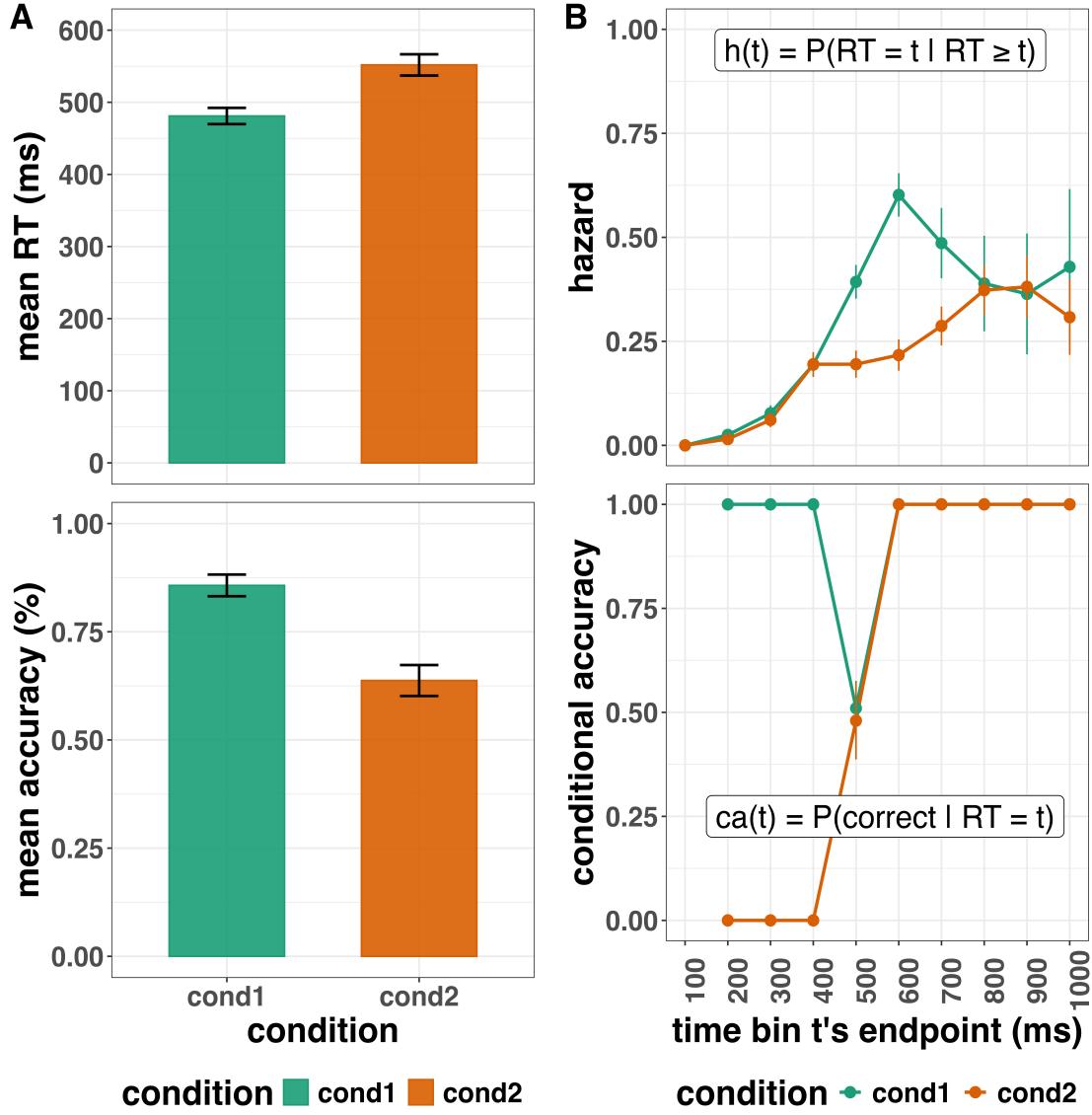


Figure 1. Simulated data showing mean performance versus distributional (EHA/SAT) analyses. (A) The mean RT (top) and overall accuracy (bottom) for two conditions are plotted. (B) The discrete-time hazard functions (top) and conditional accuracy functions (bottom) are plotted for the same data. The first 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. Error bars represent ± 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), 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 also visualize and discuss the types of time-to-event data that are obtained in
110 typical RT tasks and supply relevant regression equations in sections A and E of the
111 Supplemental Material, respectively.

112 **2.1.1 Terminology and minimum requirements for EHA.** To avoid possible

113 confusion in terminology used, it is worth noting that EHA is known by various labels,

114 such as survival analysis, hazard analysis, duration analysis, failure-time analysis, and

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

116 throughout.

117 In terms of minimum requirements to apply a single-event EHA, one must be able to:

118 1. define an event of interest that represents a qualitative change - a transition from one

119 discrete state to another - that can be situated in time (e.g., a button press, a

120 saccade onset, a fixation offset, etc.);

121 2. define time point zero in each trial (e.g., target stimulus onset, fixation onset, etc.);

122 3. measure the passage of time between time point zero and event occurrence in discrete

123 or continuous time units in each trial.

124 **2.1.2 Types of EHA.** There are different types of EHA. For example, the

125 definition of hazard and the type of models employed depend on whether one is using

126 continuous or discrete time units. As a lab, and mainly for practical reasons, we have much

127 more experience using discrete-time EHA, and that is the approach that we describe and

128 focus on in this paper. This choice may seem counter-intuitive, given that RT is typically

129 treated as a continuous variable. However, continuous forms of EHA require much more

130 data to estimate the continuous-time hazard (rate) function well (Bloxom, 1984; Luce,

131 1991; Van Zandt, 2000). Thus, by trading a bit of temporal resolution for a lower number

132 of trials, discrete-time methods seem ideal for dealing with typical psychological RT data

133 sets for which there are less than ~200 trials per condition per experiment (Panis, Schmidt,

134 et al., 2020). Moreover, as indicated by Allison (2010), learning discrete-time EHA methods

135 first will help in learning continuous-time methods, so it seems like a good starting point.

136 To apply discrete-time EHA, one divides the within-trial time in discrete, contiguous

137 time bins indexed by t (e.g., $t = 1:10$ time bins; Figure 1B). Then let RT be a discrete

138 random variable denoting the rank of the time bin in which a particular person's response

139 occurs in a particular trial (i.e., repeated measure). For example, a response in one trial

140 might occur at 546 ms and it would be in time bin 6 (any RTs from 501 ms to 600 ms).

141 One then calculates the sample-based estimate of the discrete-time hazard function of

142 event occurrence for each experimental condition (Figure 1B top). The discrete-time

143 hazard function gives you, for each time bin, the conditional probability that the event

144 occurs (sometime) in bin t , given that the event does not occur in previous bins. In other

145 words, it reflects the instantaneous risk that the event occurs in the current bin t , given

146 that it has not yet occurred in the past, i.e., in one of the prior bins ($t-1, t-2, \dots, 1$).

147 In the context of experimental psychology, it is often (but not always), the case that

148 responses can be classified as correct or incorrect. In those cases, one can also calculate the

149 conditional accuracy function (Figure 1B bottom). The conditional accuracy function gives

150 you for each time bin the conditional probability that a response is correct given that it is

151 emitted in time bin t (Allison, 2010; Kantowitz & Pachella, 2021; Wickelgren, 1977). The

152 $ca(t)$ function is also known as the micro-level speed-accuracy tradeoff (SAT) function. We

153 refer to this extended (hazard + conditional accuracy) analysis for choice RT data as

154 EHA/SAT. The definitions of these and other discrete-time functions are given in section B

155 of the Supplemental Material.

156 2.2 Benefits of event history analysis for research in experimental psychology

157 Statisticians and mathematical psychologists recommend focusing on the hazard

158 function when analyzing time-to-event data for various reasons (Holden, Van Orden, &

159 Turvey, 2009; Luce, 1991; Townsend, 1990). We do not cover these benefits in detail here,

160 as these are more general topics that have been covered elsewhere in textbooks (see also

161 section F of the Supplemental Material). Instead, here we focus on the benefits as we see

162 them for common research programmes in experimental psychology.

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

177 Second, compared to more conventional analytical approaches, EHA uses more of the
178 data because it deals with missing data differently. It is conventional with RT data to either
179 (a) use a response deadline and discard all trials without a response, or (b) wait in each
180 trial until a response occurs and then apply data trimming techniques, i.e., discarding too
181 short or too long RTs (and perhaps also erroneous responses) before calculating a mean RT
182 (Berger & Kiefer, 2021). Discarding data can introduce biases, however. Rather than treat
183 non-responses as missing data, EHA treats such trials as *right-censored* observations on the
184 variable RT, because all we know is that RT is greater than some value. Right-censoring is
185 a type of missing data problem and a nearly universal feature of survival data including RT
186 data. For example, if the censoring time was 1 second, then some trials result in observed
187 event times (those with a RT below 1 second), while the other trials result in response
188 times that are right-censored at 1 second. The fact that EHA can deal with

189 right-censoring, therefore, presents a analytical strength of the approach compared to many
190 common approaches in experimental psychology (ANOVA, linear regression, delta plots).

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

199 **2.3 Implications for research design in experimental psychology**

200 Performing EHA in experimental psychology has implications for how experiments
201 are designed. More specifically, we consider three implications that researchers will need to
202 consider when using discrete-time EHA.

203 First, one can use a response deadline in each trial because EHA deals with
204 right-censored observations.

205 Second, since the number of trials per condition are spread across bins, it is
206 important to have a relatively large number of trial repetitions per participant and per
207 condition. Accordingly, experimental designs using this approach typically focus on
208 factorial, within-subject designs, in which a large number of observations are made on a
209 relatively small number of participants (so-called small-*N* designs). This approach
210 emphasizes the precision and reproducibility of data patterns at the individual participant
211 level to increase the inferential validity of the design (Baker et al., 2021; Smith & Little,
212 2018). Note that because statistical power derives both from the number of participants
213 and from the number of repeated measures per participant and condition, small-*N* designs

²¹⁴ can still achieve what are generally considered acceptable levels of statistical power, if they
²¹⁵ have a sufficient amount of data overall (Baker et al., 2021; Smith & Little, 2018).

²¹⁶ Third, the width of each time bin will need to be determined. For instance, in Figure
²¹⁷ 1B we chose 100 ms in an arbitrary manner. In reality, however, bin width will need to be
²¹⁸ set by considering a number of factors simultaneously. The optimal bin width will depend
²¹⁹ on (a) the length of the observation period in each trial, (b) the rarity of event occurrence,
²²⁰ (c) the number of repeated measures (or trials) per condition per participant, and (d) the
²²¹ shape of the hazard function. Finding an appropriate bin width in a given user case before
²²² fitting models will require testing a number of options, when calculating and plotting the
²²³ descriptive statistics (see section 3.1). The goal is to find the smallest bin width that is
²²⁴ supported by the amount of data available. Based on our experience, a bin width of 50 ms
²²⁵ is a good starting value when the number of repeated measures is 100 or less. Too small
²²⁶ bin widths will result in erratic hazard functions as many bins will have no events, and
²²⁷ thus hazard estimates of zero. Interestingly, the time bins do not need to have the same
²²⁸ width. For example, Panis (2020) used larger bins towards the end of the observation
²²⁹ period, as fewer events occurred there.

²³⁰ 3. Tutorials

²³¹ Tutorials 1a and 1b show how to calculate and plot the descriptive statistics of
²³² EHA/SAT when there are one or two independent variables, respectively. Tutorials 2a and
²³³ 2b illustrate how to use Bayesian multilevel modeling to fit hazard and conditional
²³⁴ accuracy models, respectively. Tutorials 3a and 3b show how to implement, respectively,
²³⁵ multilevel models for hazard and conditional accuracy in the frequentist framework.
²³⁶ Tutorial 4 shows how to use simulation and power analysis for planning experiments.
²³⁷ Additionally, to further simplify the process for other users, the first two tutorials rely on a
²³⁸ set of our own custom functions that make sub-processes easier to automate, such as data
²³⁹ wrangling and plotting functions (see section C in the Supplemental Material for a list of

240 the custom functions).

241 Our list of tutorials is as follows:

- 242 • 1a. Wrangle raw data and calculate descriptive stats for one independent variable
- 243 • 1b. Wrangle raw data and calculate descriptive stats for two independent variables
- 244 • 2a. Bayesian multilevel modeling for $h(t)$
- 245 • 2b. Bayesian multilevel modeling for $ca(t)$
- 246 • 3a. Frequentist multilevel modeling for $h(t)$
- 247 • 3b. Frequentist multilevel modeling for $ca(t)$
- 248 • 4. Simulation and power analysis for planning experiments

249 We used R (Version 4.4.0; R Core Team, 2024) and the R-packages *bayesplot* (Version
250 1.11.1; Gabry, Simpson, Vehtari, Betancourt, & Gelman, 2019), *brms* (Version 2.22.0;
251 Bürkner, 2017, 2018, 2021), *citr* (Version 0.3.2; Aust, 2019), *cmdstanr* (Version 0.8.1.9000;
252 Gabry, Češnovar, Johnson, & Brønner, 2024), *dplyr* (Version 1.1.4; Wickham, François,
253 Henry, Müller, & Vaughan, 2023), *forcats* (Version 1.0.0; Wickham, 2023a), *ggplot2*
254 (Version 3.5.1; Wickham, 2016), *lme4* (Version 1.1.35.5; Bates, Mächler, Bolker, & Walker,
255 2015), *lubridate* (Version 1.9.3; Grolemund & Wickham, 2011), *Matrix* (Version 1.7.1;
256 Bates, Maechler, & Jagan, 2024), *nlme* (Version 3.1.166; Pinheiro & Bates, 2000), *papaja*
257 (Version 0.1.3; Aust & Barth, 2024), *patchwork* (Version 1.3.0; Pedersen, 2024), *purrr*
258 (Version 1.0.2; Wickham & Henry, 2023), *RColorBrewer* (Version 1.1.3; Neuwirth, 2022),
259 *Rcpp* (Eddelbuettel & Balamuta, 2018; Version 1.0.13.1; Eddelbuettel & François, 2011),
260 *readr* (Version 2.1.5; Wickham, Hester, & Bryan, 2024), *RJ-2021-048* (Bengtsson, 2021),
261 *rstan* (Version 2.32.6; Stan Development Team, 2024), *standist* (Version 0.0.0.9000; Girard,
262 2024), *StanHeaders* (Version 2.32.10; Stan Development Team, 2020), *stringr* (Version
263 1.5.1; Wickham, 2023b), *tibble* (Version 3.2.1; Müller & Wickham, 2023), *tidybayes*
264 (Version 3.0.7; Kay, 2024), *tidyverse* (Version 2.0.0; Wickham et al., 2019) and *tinylabels* (Version 0.2.4; Barth, 2023)

266 for all reported analyses. The content of the tutorials, in terms of EHA and multilevel
267 regression modelling, is mainly based on Allison (2010), Singer and Willett (2003),
268 McElreath (2020), Heiss (2021), Kurz (2023a), and Kurz (2023b).

269 **3.1 Tutorial 1a: Calculating descriptive statistics using a life table**

270 **3.1.1 Data wrangling aims.** Our data wrangling procedures serve two related
271 purposes. First, we want to summarise and visualise descriptive statistics using a life table.
272 A life table includes for each time bin, the risk set (i.e., the number of trials that are
273 event-free at the start of the bin), the number of observed events, and the estimates of the
274 discrete-time hazard function $h(t)$, survivor function $S(t)$, probability mass function $P(t)$,
275 possibly the conditional accuracy function $ca(t)$, and their estimated standard errors (se).
276 The definitions of these functions are provided in section A of the Supplemental Material.

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

283 `## Warning in attr(x, "align"): 'xfun::attr()' is deprecated.`
284 `## Use 'xfun::attr2()' instead.`
285 `## See help("Deprecated")`

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.

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

```

295 ## Warning in attr(x, "align"): 'xfun::attr()' is deprecated.
296 ## Use 'xfun::attr2()' instead.
297 ## See help("Deprecated")

```

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.

298 **3.1.2 A real data wrangling example.** To illustrate how to quickly set up life
 299 tables for calculating the descriptive statistics (functions of discrete time), we use a
 300 published data set on masked response priming from Panis and Schmidt (2016). In their
 301 first experiment, Panis and Schmidt (2016) presented a double arrow for 94 ms that
 302 pointed left or right as the target stimulus with an onset at time point zero in each trial.

303 Participants had to indicate the direction in which the double arrow pointed using their
 304 corresponding index finger, within 800 ms after target onset. Response time and accuracy
 305 were recorded on each trial. Prime type (blank, congruent, incongruent) and mask type
 306 were manipulated. Here we focus on the subset of trials in which no mask was presented.
 307 The 13-ms prime stimulus was a double arrow presented 187 ms before target onset in the
 308 congruent (same direction as target) and incongruent (opposite direction as target) prime
 309 conditions.

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

313 The required column names are as follows:

- 314 • “pid”, indicating unique participant IDs;
- 315 • “trial”, indicating each unique trial per participant;
- 316 • “condition”, a factor indicating the levels of the independent variable (1, 2, ...) and
 the corresponding labels;
- 318 • “rt”, indicating the response times in ms;
- 319 • “acc”, indicating the accuracies (1/0).

320 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")))
```

321 Next, we can set up the life tables and plots of the discrete-time functions $h(t)$, $S(t)$,
 322 $ca(t)$, and $P(t)$. To do so using a functional programming approach, one has to nest the
 323 data within participants using the `group_nest()` function, and supply a user-defined
 324 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))
```

325 Note that the censoring time (here: 600 ms) should be a multiple of the bin width
 326 (here: 40 ms). The censoring time should be a time point after which no informative
 327 responses are expected anymore. In experiments that implement a response deadline in
 328 each trial the censoring time can equal that deadline time point. Trials with a RT larger
 329 than the censoring time, or trials in which no response is emitted during the data collection
 330 period, are treated as right-censored observations in EHA. In other words, these trials are
 331 not discarded, because they contain the information that the event did not occur before the
 332 censoring time. Removing such trials before calculating the mean event time will result in
 333 underestimation of the true mean.

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

335 has one row for each time bin (of each trial) that is at risk for event occurrence. The
336 variable “event” in the person-trial-bin oriented data set indicates whether a response
337 occurs (1) or not (0) for each bin.

338 The next step is to set up the life table using our custom function setup_lt(),
339 calculate the conditional accuracies using our custom function calc_ca(), add the ca(t)
340 estimates to the life table using our custom function join_lt_ca(), and then plot the
341 descriptive statistics using our custom function plot_eha(). One can now inspect different
342 aspects, including the life table for a particular condition of a particular subject, and a plot
343 of the different functions for a particular participant.

344 In general, it is important to visually inspect the functions first for each participant,
345 in order to identify individuals that may be guessing (e.g., a flat conditional accuracy
346 function at .5 indicates that someone is just guessing), outlying individuals, and/or
347 different groups with qualitatively different behavior. Also, to select a bin width for fitting
348 models, one should test and compare various bin widths in the censor function, and select
349 the smallest one that is supported by the data. Too small bin widths will result in erratic
350 hazard functions because many bins will have estimates equal to zero.

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

```
353 ## Warning in attr(x, "align"): 'xfun::attr()' is deprecated.  
354 ## Use 'xfun::attr2()' instead.  
355 ## See help("Deprecated")
```

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.

357 probability mass functions for each prime condition for participant 6. By using
 358 discrete-time hazard functions of event occurrence – in combination with conditional
 359 accuracy functions for two-choice tasks – one can provide an unbiased, time-varying, and
 360 probabilistic description of the latency and accuracy of responses based on all trials of any
 361 data set.

Descriptive stats for subject 6

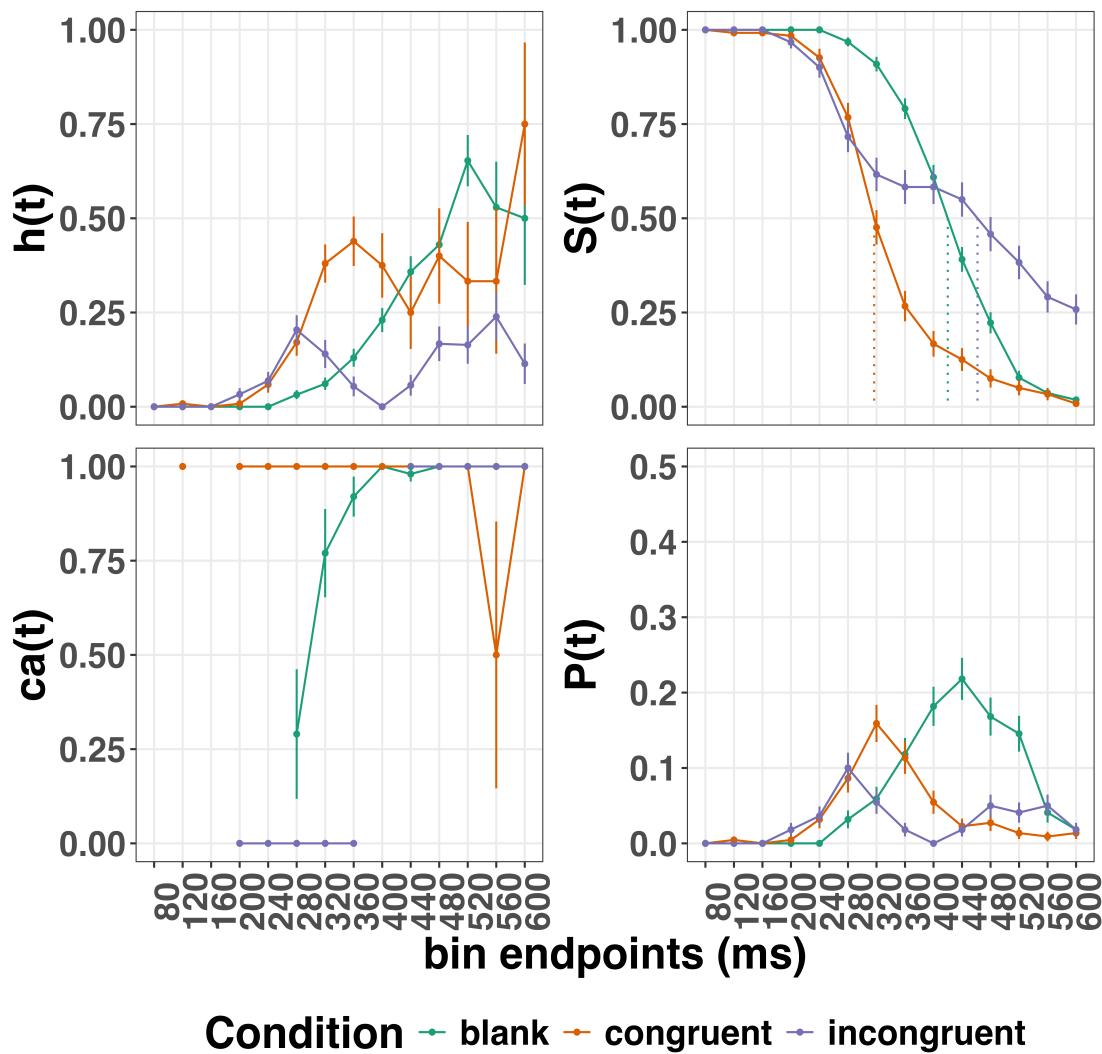


Figure 2. Estimated discrete-time hazard (h), survivor (S), conditional accuracy (ca) and probability mass (P) functions for participant 6. Vertical dotted lines indicate the estimated median RTs. Error bars represent ± 1 standard error of the respective proportion.

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

363 0.17, and 0.20 for the blank, congruent, and incongruent prime conditions, respectively. In

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

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

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

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

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

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

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

371 respectively.

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

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

374 0.36, 0.25, and 0.06 for the blank, congruent, and incongruent prime conditions,

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

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

377 conditions, respectively.

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

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

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

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

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

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

384 group-average distribution per condition (see Tutorial_1a.Rmd).

385 In general, these results go against the (often implicit) assumption in research on

386 priming that all observed responses are primed responses to the target stimulus. Instead,

387 the distributional data show that early responses are triggered exclusively by the prime

388 stimulus, while only later responses reflect primed responses to the target stimulus.

389 At this point, we have calculated, summarised and plotted descriptive statistics for
390 the key variables in EHA/SAT. As we will show in later Tutorials, statistical models for
391 $h(t)$ and $ca(t)$ can be implemented as generalized linear mixed regression models predicting
392 event occurrence (1/0) and conditional accuracy (1/0) in each bin of a selected time
393 window for analysis. But first we consider calculating the descriptive statistics for two
394 independent variables.

395 **3.2 Tutorial 1b: Generalising to a more complex design**

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

402 To this end, Tutorial 1b illustrates how to calculate and plot the descriptive statistics
403 for the full data set of Experiment 1 of Panis and Schmidt (2016), which includes two
404 independent variables: mask type and prime type. As we use the same functional
405 programming approach as in Tutorial 1a, we simply present the sample-based functions for
406 each participant as part of Tutorial_1b.Rmd for those that are interested.

407 **3.3 Tutorial 2a: Fitting Bayesian hazard models to discrete time-to-event data**

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

412 Willett, 2003).

413 In general, when fitting regression models, our lab adopts an estimation approach to
414 multilevel regression (Kruschke & Liddell, 2018; Winter, 2019), which is heavily influenced
415 by the Bayesian framework as suggested by Richard McElreath (Kurz, 2023b; McElreath,
416 2020). We also use a “keep it maximal” approach to specifying varying (or random) effects
417 (Barr, Levy, Scheepers, & Tily, 2013). This means that wherever possible we include
418 varying intercepts and slopes per participant. To make inferences, we use two main
419 approaches. We compare models of different complexity, using information criteria (e.g.,
420 WAIC) and cross-validation (e.g., LOO), to evaluate out-of-sample predictive accuracy
421 (McElreath, 2020). We also take the most complex model and evaluate key parameters of
422 interest using point and interval estimates.

423 **3.3.1 Hazard model considerations.** There are several analytic decisions one
424 has to make when fitting a discrete-time hazard model. First, one has to select an analysis
425 time window, i.e., a contiguous set of bins for which there is enough data for each
426 participant. Second, given that the dependent variable (event occurrence) is binary, one
427 has to select a link function (see section C in the Supplemental Material). The cloglog link
428 is preferred over the logit link when events can occur in principle at any time point within
429 a bin, which is the case for RT data (Singer & Willett, 2003). Third, one has to choose
430 whether to treat TIME (i.e., the time bin index t) as a categorical or continuous predictor.
431 And when you treat a variable as a categorical predictor, you can choose between reference
432 coding and index coding. With reference coding, one defines the variable as a factor and
433 selects one of the k categories as the reference level. Brm() will then construct $k-1$
434 indicator variables (see model M1d in Tutorial_2a.Rmd for an example). With index
435 coding, one constructs an index variable that contains integers that correspond to different
436 categories (see models M0i and M1i below). As explained by McElreath (2020), the
437 advantage of index coding is that the same prior can be assigned to each level of the index
438 variable, so that each category has the same prior uncertainty.

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

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

In tutorial_2a.Rmd we fit several Bayesian multilevel models (i.e., generalized linear mixed models) that differ in complexity to the person-trial-bin oriented data set that we created in Tutorial 1a. We decided to select the analysis time window (200,600] and the cloglog link. Below, we shortly discuss two of these models. The person-trial-bin data set is 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)))
```

456 3.3.2 Prior distributions. To get the posterior distribution of each model

457 parameter given the data, we need to specify prior distributions for the model parameters
458 which reflect our prior beliefs. In Tutorial_2a.Rmd we perform a few prior predictive
459 checks to make sure our selected prior distributions reflect our prior beliefs (Gelman,
460 Vehtari, et al., 2020).

461 The middle column of Supplementary Figure 2 (section E of the Supplemental
462 Material) shows six examples of prior distributions for an intercept on the logit and/or
463 cloglog scales. While a normal distribution with relatively large variance is often used as a
464 weakly informative prior for continuous dependent variables, rows A and B of
465 Supplementary Figure 2 show that specifying such distributions on the logit and cloglog
466 scales actually leads to rather informative distributions on the original probability scale, as
467 most mass is pushed to probabilities of 0 and 1.

468 3.3.3 Model M0i: A null model with index coding. When you do not want to
469 make assumptions about the shape of the hazard function, or its shape is not smooth but
470 irregular, then you can use a general specification of TIME, i.e., fit one grand intercept per
471 time bin. In this first model, we use a general specification of TIME using index coding,
472 and do not include experimental predictors. We call this model “M0i”.

473 Before we fit model M0i, we select the necessary columns from the data, and specify
474 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")

```

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

481 **3.3.4 Model M1i: Adding the effects of prime-target congruency.** Previous
 482 research has shown that psychological effects typically change over time (Panis, 2020;
 483 Panis, Moran, et al., 2020; Panis & Schmidt, 2022; Panis et al., 2017; Panis & Wagemans,
 484 2009). In the next model, therefore, we use index coding for both TIME (variable
 485 “timebin”) and the categorical predictor prime-target-congruency (variable “prime”), so
 486 that we get 30 grand intercepts, one for each combination of timebin level and prime level.
 487 Here is the model formula of this model that we call “M1i”.

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

488 Estimating model M1i took about 124 minutes.

489 **3.3.5 Compare the models.** We can compare the two models using the Widely
 490 Applicable Information Criterion (WAIC) and Leave-One-Out (LOO) cross-validation, and

491 look at model weights for both criteria (Kurz, 2023a; McElreath, 2020).

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

492 ## model_M0i model_M1i

493 ## 0 1

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

494 ## model_M0i model_M1i

495 ## 0 1

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

498 **3.3.6 Evaluating parameter estimates in model M1i.** To make inferences
499 from the parameter estimates in model M1i, we first plot the densities of the draws from
500 the posterior distributions of its population-level parameters in Figure 5, together with
501 point (median) and interval estimates (80% and 95% credible intervals).

Posterior distributions for population-level effects in Model M1i

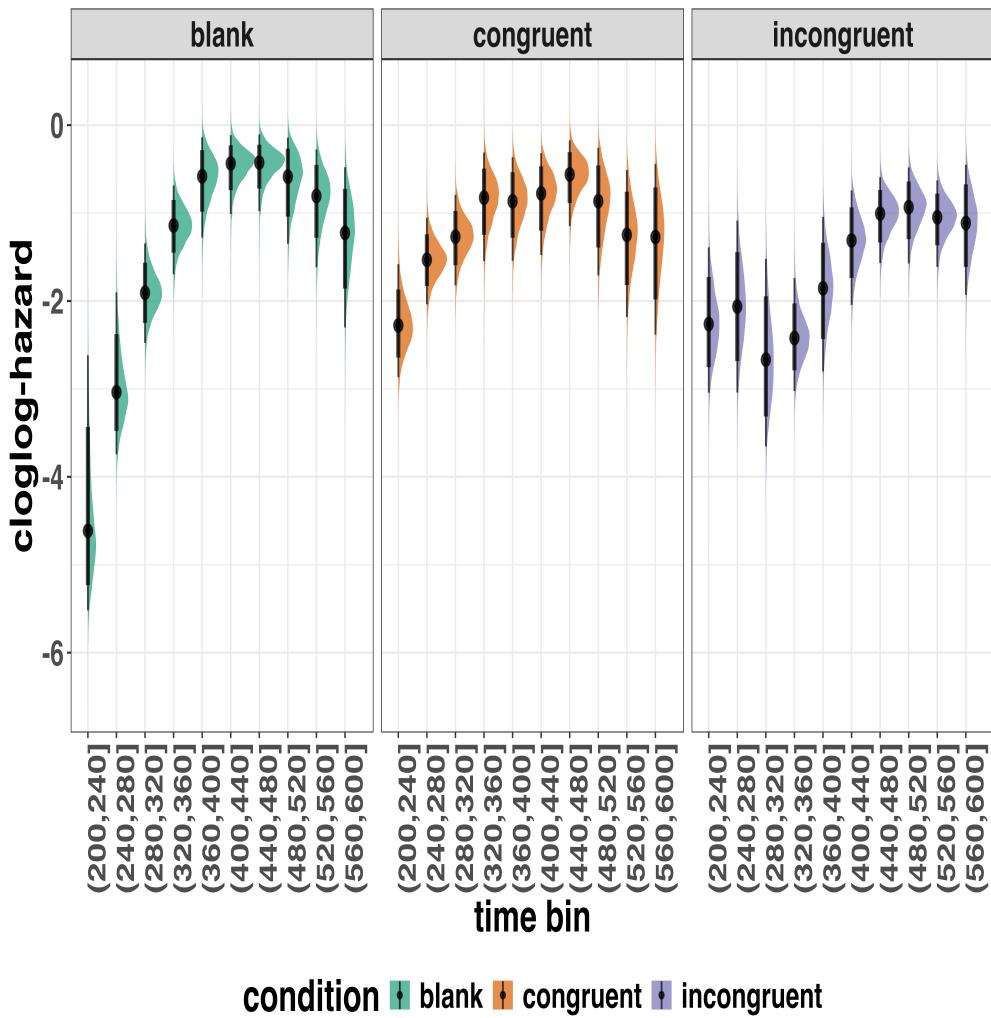


Figure 3. Medians and 80/95% credible intervals of the posterior distributions of the population-level parameters of model M1i.

502 Because the parameter estimates are on the cloglog-hazard scale, we can ease our

503 interpretation by plotting the expected value of the posterior predictive distribution – the
 504 predicted hazard values – at the population level (Figure 6A), and for each participant in
 505 the data set (Figure 6B).

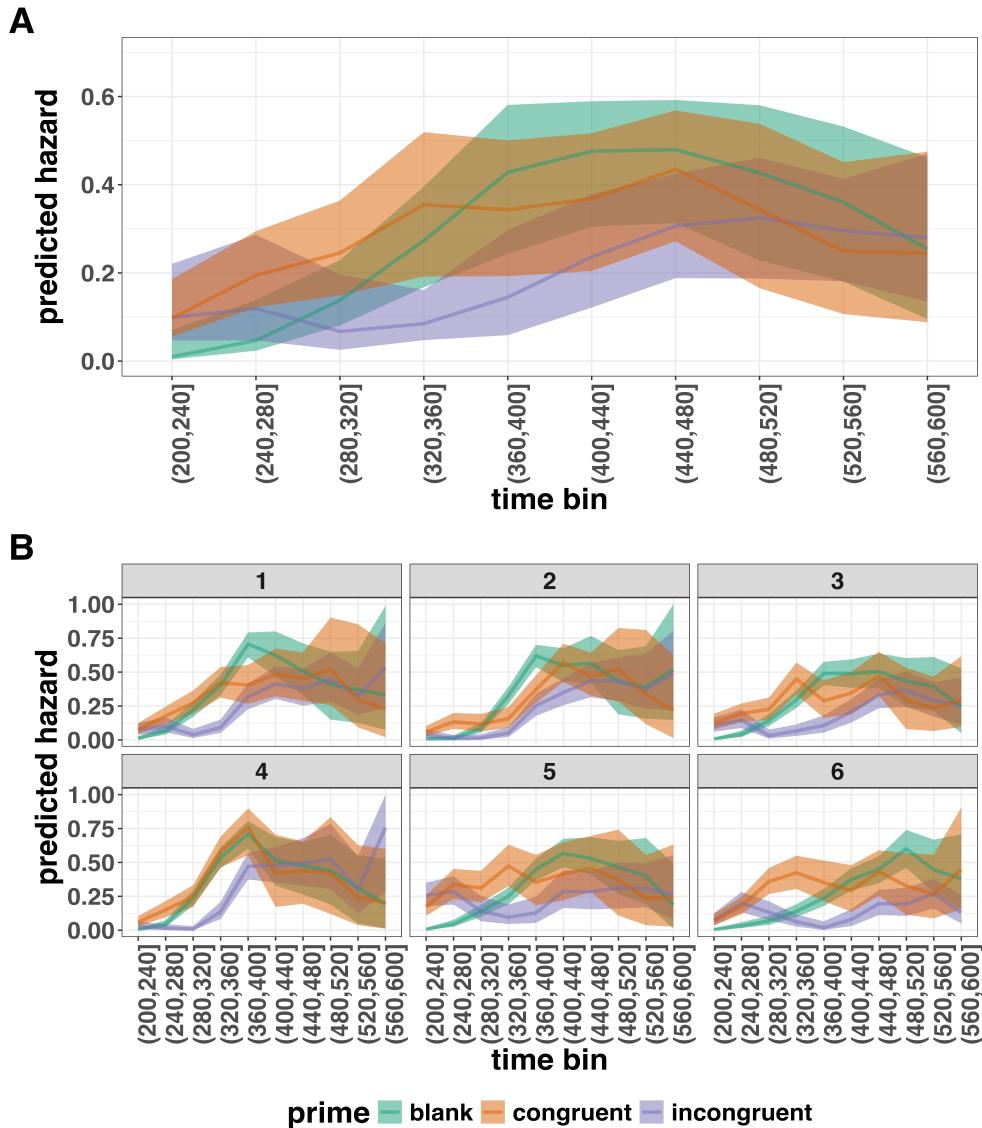


Figure 4. Point (median) and 80/95% credible interval summaries of the hazard estimates (expected values of the draws from the posterior predictive distributions) in each time bin at the population level (A), and for each participant (B).

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

507 relative to the blank prime condition, we can construct two contrasts (congruent-blank,

508 incongruent-blank), and plot the posterior distributions of these contrast effects, both at

509 the population level (Figure 7A; grand average marginal effect) and at the participant level

⁵¹⁰ (Figure 7B; subject-specific average marginal effect).

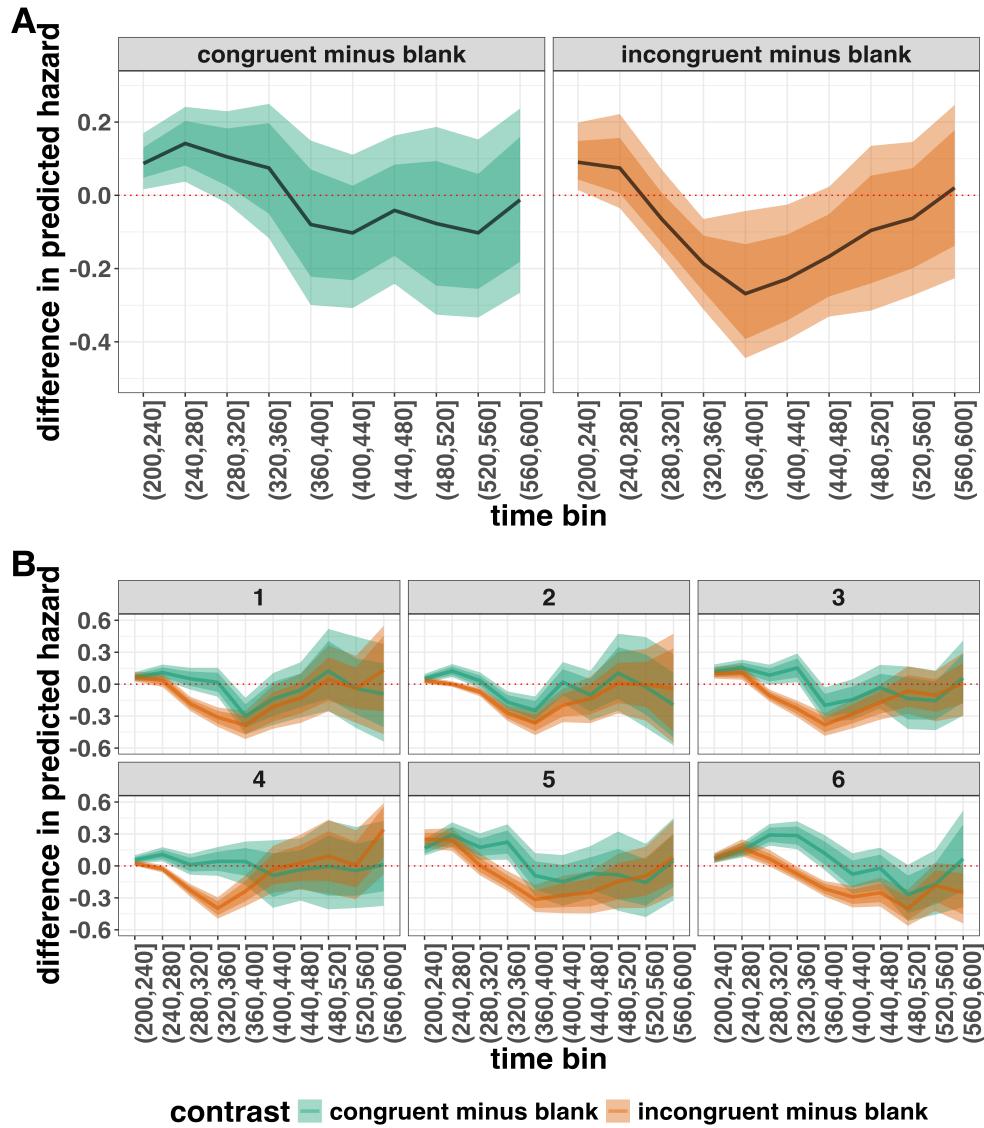


Figure 5. Point (mean) and 80/95% credible interval summaries of estimated differences in hazard in each time bin at the population level (A), and for each participant (B).

⁵¹¹ The point estimates and quantile intervals can be reported in a table (see
⁵¹² Tutorial_2a.Rmd for details).

⁵¹³ **Example conclusions for M1i.** What can we conclude from model M1i about
⁵¹⁴ our research question, i.e., the temporal dynamics of the effect of prime-target congruency

515 on RT? In other words, in which of the 40-ms time bins between 200 and 600 ms after
516 target onset does changing the prime from blank to congruent or incongruent affect the
517 hazard of response occurrence (for a prime-target stimulus-onset-asynchrony of 187 ms)?

518 If we want to estimate the population-level effect of prime type on hazard, we can
519 base our conclusion on Figure 7A. The contrast “congruent minus blank” was estimated to
520 be 0.09 hazard units in bin (200,240] (95% CrI = [0.02, 0.17]), and 0.14 hazard units in bin
521 (240,280]) (95% CrI = [0.04, 0.25]). For the other bins, the 95% credible interval contained
522 zero. The contrast “incongruent minus blank” was estimated to be 0.09 hazard units in bin
523 (200,240] (95% CrI = [0.01, 0.21]), -0.19 hazard units in bin (320,360] (95% CrI = [-0.31,
524 -0.06]), -0.27 hazard units in bin (360,400] (95% CrI = [-0.45, -0.04]), and -0.23 hazard
525 units in bin (400,440] (95% CrI = [-0.40, -0.03]). For the other bins, the 95% credible
526 interval contained zero.

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

534 If we want to focus more on inter-individual differences, we can study the
535 subject-specific hazard functions in Figure 7B. Note that three participants (1, 2, and 3)
536 show a negative difference for the contrast “congruent minus incongruent” in bin (360,400]
537 – subject 2 also in bin (320,360].

538 Future studies could (a) increase the number of participants to estimate the
539 proportion of “dippers” in the subject population, and/or (b) try to explain why this dip
540 occurs. For example, Panis and Schmidt (2016) concluded that active, top-down,

541 task-guided response inhibition effects emerge around 360 ms after the onset of the stimulus
542 following the prime (here: the target stimulus). Such a top-down inhibitory effect might
543 exist in our priming data set, because after some time participants will learn that the first
544 stimulus is not the one they have to respond to. To prevent a premature overt response to
545 the prime they thus might gradually increase a global response threshold during the
546 remainder of the experiment, which could result in a lower hazard in congruent trials
547 compared to blank trials, for bins after ~360 ms, and towards the end of the experiment.
548 This effect might be masked for incongruent primes by the response competition effect.

549 Interestingly, all subjects show a tendency in their mean difference (congruent minus
550 blank) to “dip” around that time (Figure 7B). Therefore, future modeling efforts could
551 incorporate the trial number into the model formula, in order to also study how the effects
552 of prime type on hazard change on the long experiment-wide time scale, next to the short
553 trial-wide time scale. In Tutorial_2a.Rmd we provide a number of model formulae that
554 should get you going.

555 3.4 Tutorial 2b: Fitting Bayesian conditional accuracy models

556 In this fourth tutorial, we illustrate how to fit a Bayesian multilevel regression model
557 to the timed accuracy data from the masked response priming data used in Tutorial 1a.
558 The general process is similar to Tutorial 2a, except that (a) we use the person-trial data,
559 (b) we use the logit link function, and (c) we change the priors. To keep the tutorial short,
560 we only fit one conditional accuracy model, which was based on model M1i from Tutorial
561 2a and labelled M1i_ca.

562 To make inferences from the parameter estimates in model M1i_ca, we first plot the
563 densities of the draws from the posterior distributions of its population-level parameters in
564 Figure 8, together with point (median) and interval estimates (80% and 95% credible
565 intervals).

Posterior distributions for population-level effects in Model M1i_ca

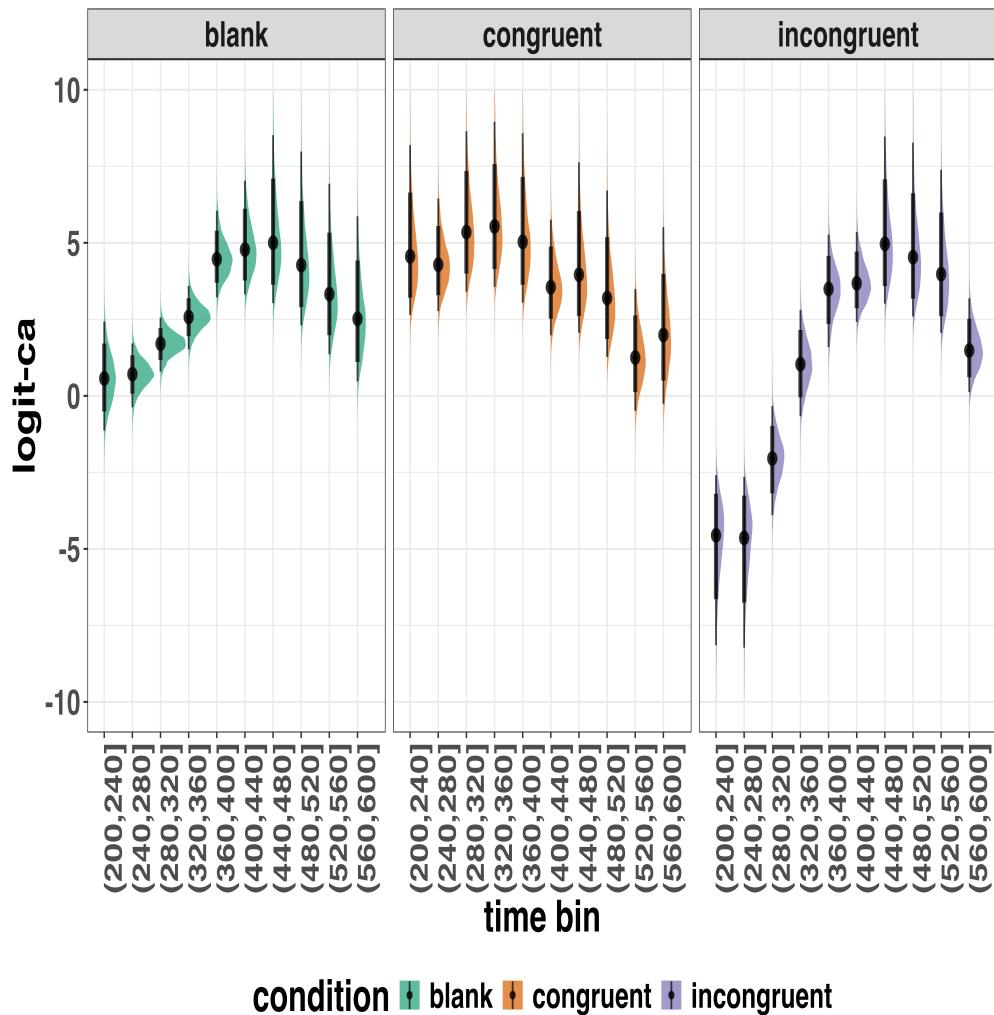


Figure 6. Medians and 80/95% credible intervals of the posterior distributions of the population-level parameters of model M1i_ca. ca = conditional accuracy.

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).

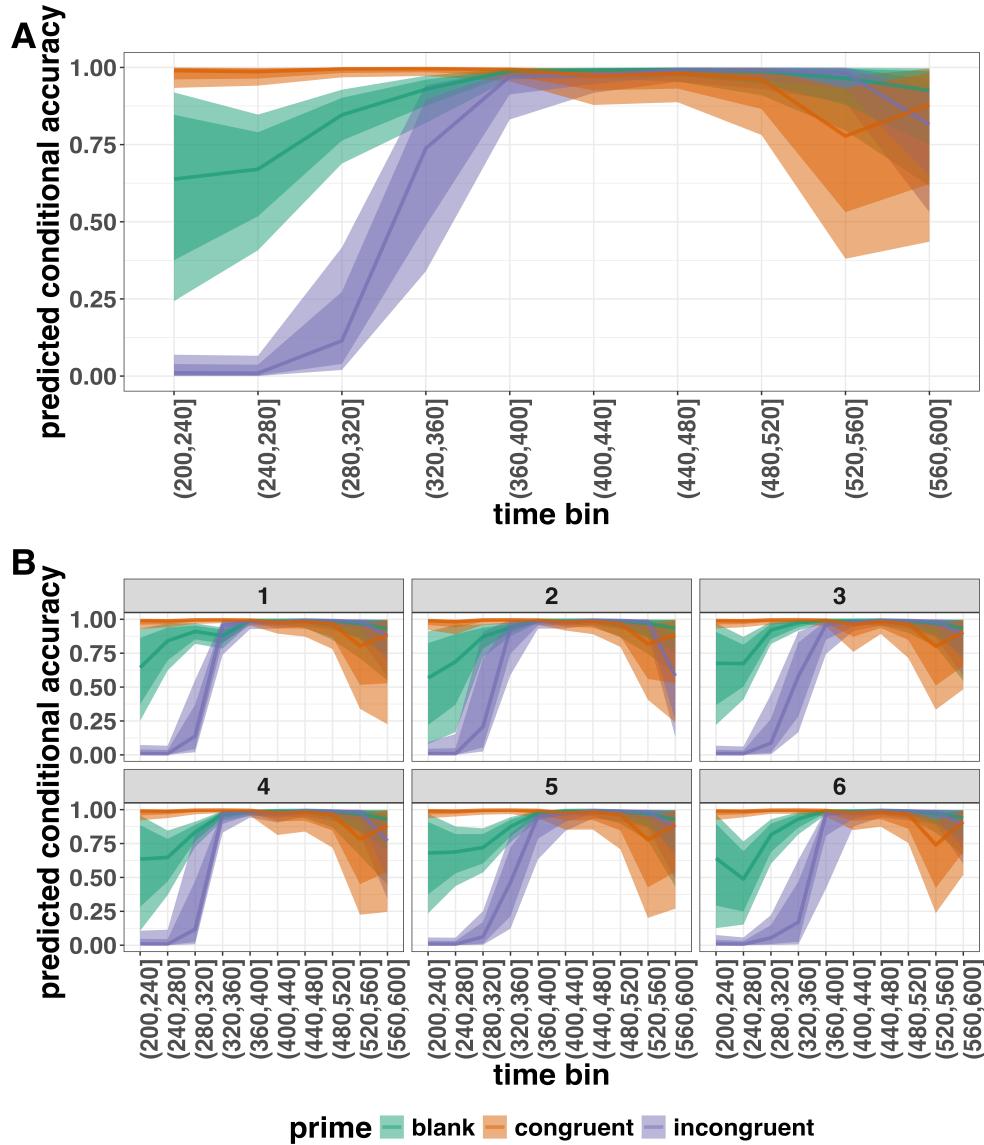


Figure 7. Point (median) and 80/95% credible interval summaries of the conditional accuracy estimates (expected values of the draws from the posterior predictive distributions) in each time bin at the population level (A), and for each participant (B).

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

relative to the blank prime condition, we can construct two contrasts (congruent-blank, incongruent-blank), and plot the posterior distributions of these contrast effects at the population level (Figure 10A) and for each participant (Figure 10B).

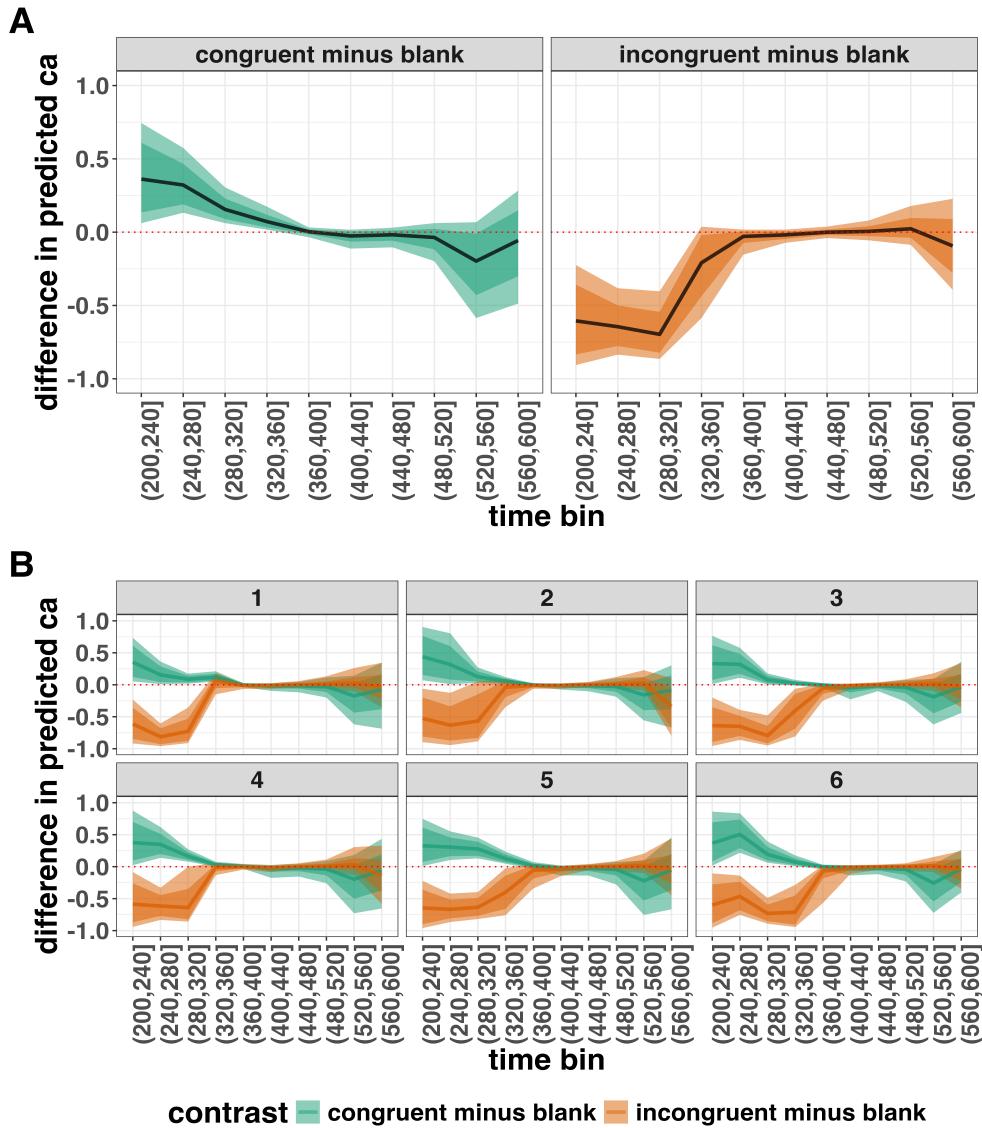


Figure 8. Point (mean) and 80/95% credible interval summaries of estimated differences in conditional accuracy in each time bin at the population level (A), and for each participant (B).

574 Based on Figure 10A we see that on the population level congruent primes have a
 575 positive effect on the conditional accuracy of emitted responses in time bins (200,240],
 576 (240,280], (280,320], and (320,360], relative to the estimates in the baseline condition
 577 (blank prime; red dashed lines in Figure 10A). Incongruent primes have a negative effect on

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

580 **3.5 Tutorial 4: Planning**

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

582 **3.5.1 Background.** The general approach to planning that we adopt here involves
583 simulating reasonably structured data to help guide what you might be able to expect from
584 your data once you collect it (Gelman, Vehtari, et al., 2020). The basic structure and code
585 follows the examples outlined by Solomon Kurz in his ‘power’ blog posts
586 (<https://solomonkurz.netlify.app/blog/bayesian-power-analysis-part-i/>) and Lisa
587 DeBruine’s R package `faux{}` (<https://debruine.github.io/faux/>) as well as these related
588 papers (DeBruine & Barr, 2021; Pargent, Koch, Kleine, Lermer, & Gaube, 2024).

589 **3.5.2 Basic workflow.** The basic workflow is as follows:

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

596 Ideally, in the above workflow, we would also fit a model to each dataset and
597 summarise the model output, rather than the raw data. However, when each model takes
598 several hours to build, and we may want to simulate many 1000s of datasets, it can be
599 computationally demanding for desktop machines. So, for ease, here we just use the raw
600 simulated datasets to guide future expectations.

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

3.5.3 Fit a regression model and simulate one dataset.

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

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

We then took parameters from this model and used them to create a single dataset

with 200 trials per condition for 10 individual participants. The raw data and the simulated data are plotted in Figure 12 and show quite close correspondence, which is re-assuring. But, this is only one dataset. What we really want to do is simulate many datasets and vary parameters of interest, which is what we turn to in the next section.

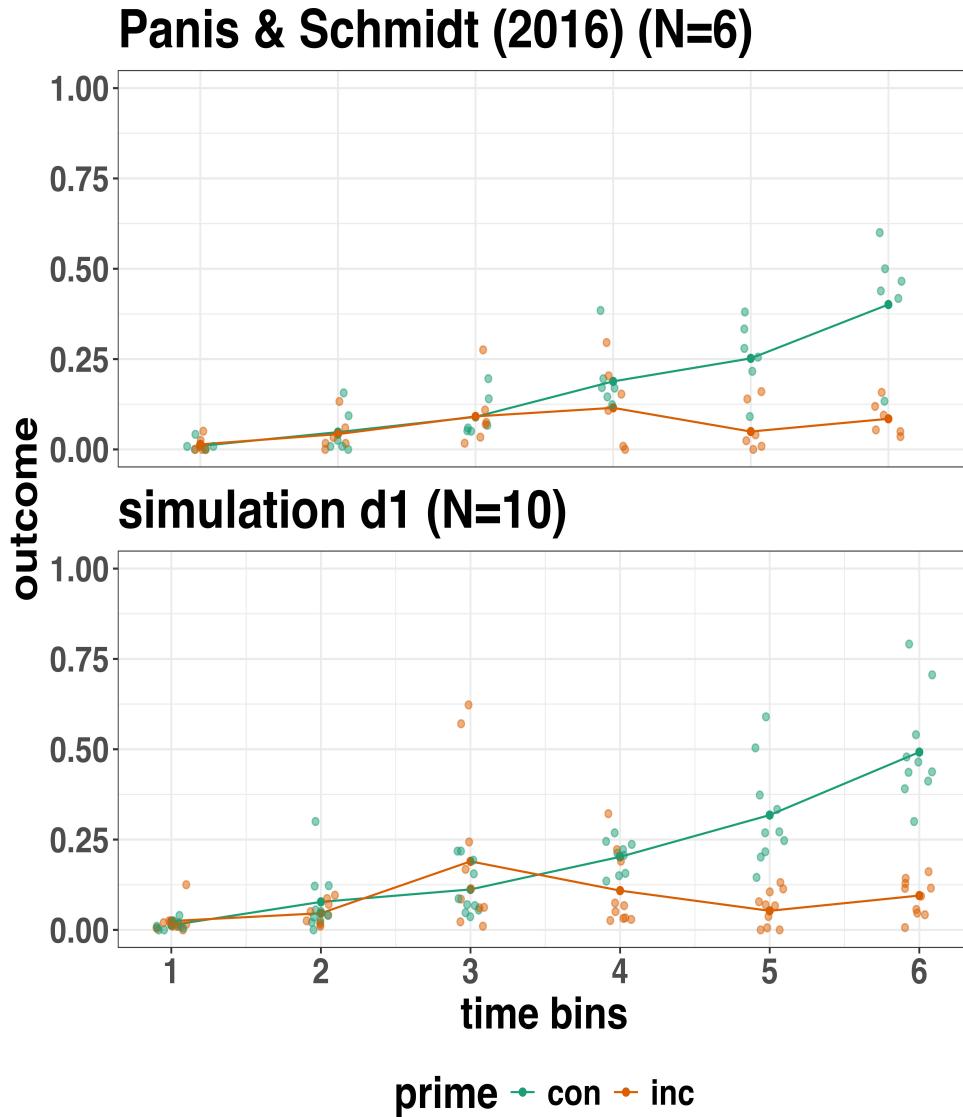


Figure 9. Raw data from Panis and Schmidt (2016) and simulated data from 10 participants.

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

614 3.5.4 Simulate and summarise data across a range of parameter values.

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

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

⁶⁴⁶ calculations might be as follows (trial count = 200 per condition in all cases):

- ⁶⁴⁷ • For a 70% reduction (0.3 hazard ratio), N=10 would give nearly 100% power.
- ⁶⁴⁸ • For a 60% reduction (0.4 hazard ratio), N=10 would give nearly 90% power.
- ⁶⁴⁹ • For a 50% reduction (0.5 hazard ratio), N=15 would give over 80% power.

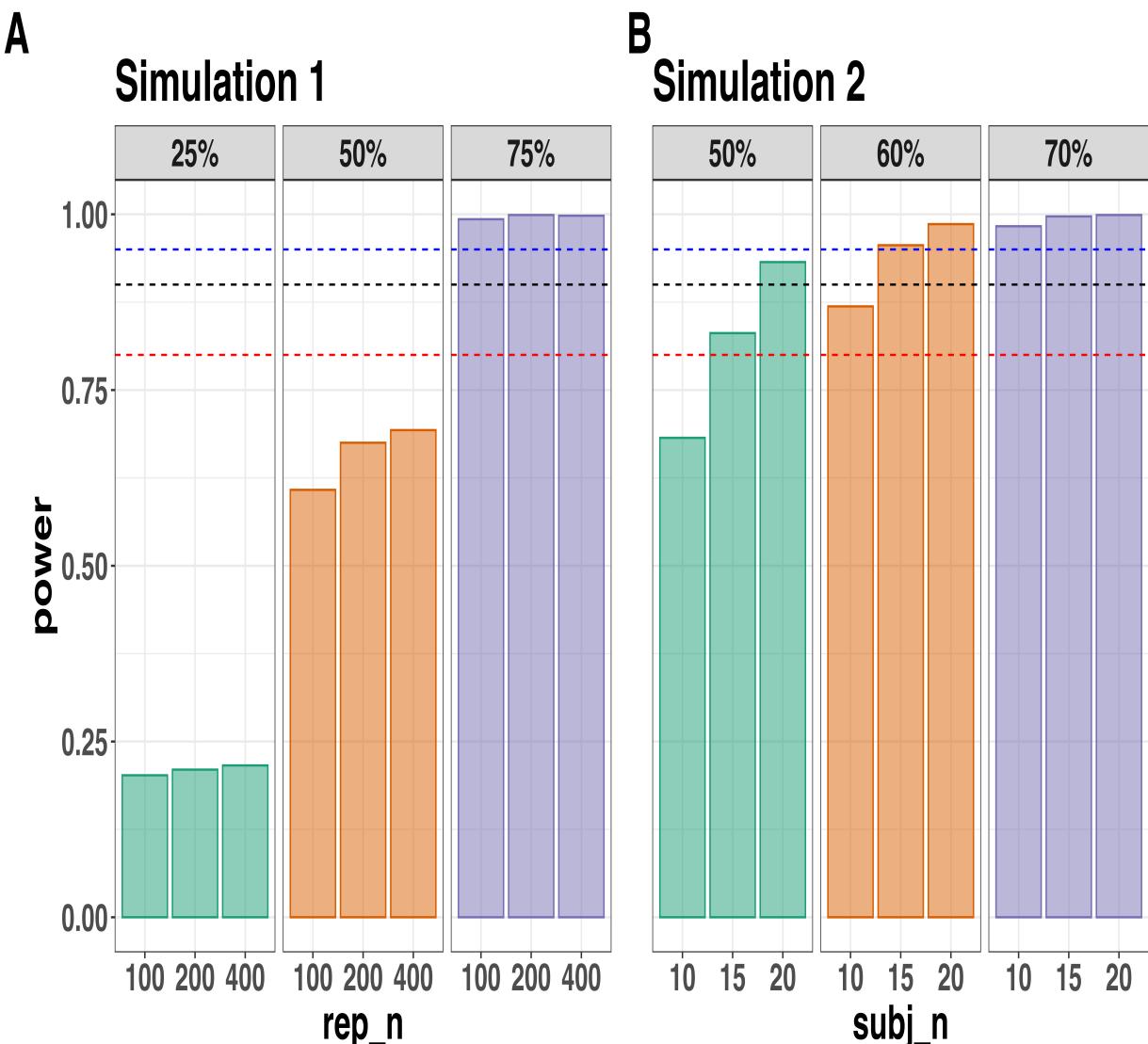


Figure 10. Statistical power across data Simulation 1 (A) and Simulation 2 (B). Power was calculated as the percentage of lower-bound 95% confidence intervals that exclude zero when the difference between prime condition is calculated (congruent - incongruent). In Simulation 1, the effect size was varied 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.

650 **3.5.5 Planning decisions.** Now that we have summarised our simulated data,

651 what planning decisions could we make about a future study? More concretely, how many
652 trials per condition should we collect and how many participants should we test? Like
653 almost always when planning future studies, the answer depends on your objectives, as well
654 as the available resources (Lakens, 2022). There is no straightforward and clear-cut answer.

655 Some considerations might be as follows:

- 656 • How much power or precision are you looking to obtain in this particular study?
- 657 • Are you running multiple studies that have some form of replication built in?
- 658 • What level of resources do you have at your disposal, such as time, money and
659 personnel?
- 660 • How easy or difficult is it to obtain the specific type of sample?

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

662 hazard ratio of 0.4 or 0.5 as a target effect size since this is much smaller than that
663 observed previously (Panis & Schmidt, 2016). Then we might pick the corresponding
664 combination of trial count per condition (e.g., 200) and participant sample size (e.g., N=10
665 or N=15) that takes you over the 80% power mark. If we wanted to maximise power based
666 on these simulations, and we had the time and resources available, then we would test
667 N=20 participants, which would provide >90% power for an effect size of 0.5.

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

669 what planning choices we made based on these data simulations, we would not solely rely
670 on data collected from one single study. Instead, we would run a follow-up experiment that
671 replicates and extends the initial result. By doing so, we would aim to avoid the Cult of
672 the Isolated Single Study (Nelder, 1999; Tong, 2019), and thus reduce the reliance on any
673 one type of planning tool, such as a power analysis. Then, we would look for common
674 patterns across two or more experiments, rather than trying to make the case that a single
675 study on its own has sufficient evidential value to hit some criterion mark.

676

4. Discussion

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

692 **4.1 What are the main use-cases of EHA for understanding cognition and brain
693 function?**

694 For those researchers, like ourselves, who are primarily interested in understanding
695 human cognitive and brain systems, we consider two broadly-defined, main use-cases of
696 EHA. First, as we hope to have made clear by this point, EHA is one way to investigating
697 a “temporal states” approach to cognitive processes, by tracking behavior as a function of
698 step-wise increases in absolute waiting time. EHA thus provides a way to uncover the
699 microgenesis of cognitive effects, by revealing when cognitive states may start and stop,
700 how states are replaced with others, as well as what they may be tied to or interact with.

701 Therefore, if your research questions concern **when psychological states occur, and**
702 **how they are temporally organized**, our EHA tutorials could be useful tools for you to
703 use.

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

714 **4.2 Model complexity versus interpretability**

715 Hazard and conditional accuracy models can quickly become very complex when
716 adding more than one time scale, due to the many possible higher-order interactions. For
717 example, some of the models discussed in Tutorial 2a, which we did not focus on in the
718 main text, contain two time scales as covariates: the passage of time on the within-trial
719 time scale, and the passage of time on the across-trial (or within-experiment) time scale.
720 However, when trials are presented in blocks, and blocks of trials within sessions, and when
721 the experiment comprises a number of sessions, then four time scales can be defined
722 (within-trial, within-block, within-session, and within-experiment). From a theoretical
723 perspective, adding more than one time scale – and their interactions – can be important
724 to capture plasticity and other learning effects that may play out on such longer time
725 scales, and that are probably present in each experiment in general (REF DFT). From a
726 practical perspective, therefore, some choices need to be made to balance the amount of

727 data that is being collected per participant, condition and across the varying timescales.
728 As one example, if there are several times cales of relevance, then it might be prudent for
729 interpretational purposes to limit the number of experimental predictor variables
730 (conditions). This is of course where planning and data simulation efforts would be
731 important to provide a guide to experimental design choices (see Tutorial 4 and section
732 2.3).

733 **4.3 Limitations**

734 Compared to the orthodox method – comparing means between conditions – the
735 most important limitation of multilevel hazard and conditional accuracy modeling is that it
736 might take a long time to estimate the parameters using Bayesian methods or the model
737 might have to be simplified significantly to use frequentist methods (ref ??).

738 Another issue is that you need a relatively large number of trials per condition to
739 estimate the discrete-time hazard function with relatively high temporal resolution (e.g., 20
740 ms), which is required when testing predictions of process models of cognition. Indeed, in
741 general, there is a trade-off between the number of trials per condition and the temporal
742 resolution (i.e., bin width) of the discrete-time hazard function. Therefore, we recommend
743 researchers to collect as many trials as possible per experimental condition, given the
744 available resources and considering the participant experience (e.g., fatigue and boredom).

745 For instance, if the maximum session length deemed reasonable is between 1 and 2 hours,
746 what is the maximum number of trials per condition that you could reasonably collect?
747 After consideration, it might be worth conducting multiple testing sessions per participant
748 and/or reducing the number of experimental conditions. There is a user-friendly online tool
749 for calculating statistical power as a function of the number of trials as well as the number
750 of participants, and this might be worth consulting to guide the research design process
751 (Baker et al., 2021). Finally, if you have a lot of repeated measurements per condition per
752 participant, you can of course also try continuous-time methods (ref).

753

5. Conclusions

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

763

Author contributions

764 Conceptualization: S. Panis and R. Ramsey; Software: S. Panis and R. Ramsey;
765 Writing - Original Draft Preparation: S. Panis; Writing - Review & Editing: S. Panis and
766 R. Ramsey; Supervision: R. Ramsey.

767

Conflicts of Interest

768 The author(s) declare that there were no conflicts of interest with respect to the
769 authorship or the publication of this article.

770

Prior versions

771 All of the submitted manuscript and Supplemental Material was previously posted to
772 a preprint archive: <https://doi.org/10.31234/osf.io/57bh6>

773

Supplemental Material

774

Disclosures

775 **Data, materials, and online resources**

776 Link to public archive:
777 https://github.com/sven-panis/Tutorial_Event_History_Analysis
778 Supplemental Material: Panis_Ramsey_suppl_material.pdf

779 **Ethical approval**

780 Ethical approval was not required for this tutorial in which we reanalyze existing
781 data sets.

782

References

- 783 Abney, D. H., Fausey, C. M., Suarez-Rivera, C., & Tamis-LeMonda, C. S. (2025).
784 Advancing a temporal science of behavior. *Trends in Cognitive Sciences*.
785 <https://doi.org/10.1016/j.tics.2025.05.010>
- 786 Allison, P. D. (1982). Discrete-Time Methods for the Analysis of Event Histories.
787 *Sociological Methodology*, 13, 61. <https://doi.org/10.2307/270718>
- 788 Allison, P. D. (2010). *Survival analysis using SAS: A practical guide* (2. ed.). Cary, NC:
789 SAS Press.
- 790 Aust, F. (2019). *Citr: 'RStudio' add-in to insert markdown citations*. Retrieved from
791 <https://github.com/crsh/citr>
- 792 Aust, F., & Barth, M. (2024). *papaja: Prepare reproducible APA journal articles with R*
793 *Markdown*. <https://doi.org/10.32614/CRAN.package.papaja>
- 794 Baker, D. H., Vilidaite, G., Lygo, F. A., Smith, A. K., Flack, T. R., Gouws, A. D., &
795 Andrews, T. J. (2021). Power contours: Optimising sample size and precision in
796 experimental psychology and human neuroscience. *Psychological Methods*, 26(3),
797 295–314. <https://doi.org/10.1037/met0000337>
- 798 Barack, D. L., & Krakauer, J. W. (2021). Two views on the cognitive brain. *Nature*
799 *Reviews Neuroscience*, 22(6), 359–371. <https://doi.org/10.1038/s41583-021-00448-6>
- 800 Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for
801 confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*,
802 68(3), 10.1016/j.jml.2012.11.001. <https://doi.org/10.1016/j.jml.2012.11.001>
- 803 Barth, M. (2023). *tinylabes: Lightweight variable labels*. Retrieved from
804 <https://cran.r-project.org/package=tinylabes>
- 805 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects
806 models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
807 <https://doi.org/10.18637/jss.v067.i01>
- 808 Bates, D., Maechler, M., & Jagan, M. (2024). *Matrix: Sparse and dense matrix classes and*

- 809 *methods*. Retrieved from <https://Matrix.R-forge.R-project.org>
- 810 Bengtsson, H. (2021). A unifying framework for parallel and distributed processing in r
811 using futures. *The R Journal*, 13(2), 208–227. <https://doi.org/10.32614/RJ-2021-048>
- 812 Berger, A., & Kiefer, M. (2021). Comparison of Different Response Time Outlier Exclusion
813 Methods: A Simulation Study. *Frontiers in Psychology*, 12, 675558.
814 <https://doi.org/10.3389/fpsyg.2021.675558>
- 815 Blossfeld, H.-P., & Rohwer, G. (2002). *Techniques of event history modeling: New*
816 *approaches to causal analysis*, 2nd ed (pp. x, 310). Mahwah, NJ, US: Lawrence
817 Erlbaum Associates Publishers.
- 818 Bloxom, B. (1984). Estimating response time hazard functions: An exposition and
819 extension. *Journal of Mathematical Psychology*, 28(4), 401–420.
820 [https://doi.org/10.1016/0022-2496\(84\)90008-7](https://doi.org/10.1016/0022-2496(84)90008-7)
- 821 Bolger, N., Zee, K. S., Rossignac-Milon, M., & Hassin, R. R. (2019). Causal processes in
822 psychology are heterogeneous. *Journal of Experimental Psychology: General*, 148(4),
823 601–618. <https://doi.org/10.1037/xge0000558>
- 824 Box-Steffensmeier, J. M. (2004). Event history modeling: A guide for social scientists.
825 Cambridge: University Press.
- 826 Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan.
827 *Journal of Statistical Software*, 80(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>
- 828 Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms.
829 *The R Journal*, 10(1), 395–411. <https://doi.org/10.32614/RJ-2018-017>
- 830 Bürkner, P.-C. (2021). Bayesian item response modeling in R with brms and Stan. *Journal*
831 *of Statistical Software*, 100(5), 1–54. <https://doi.org/10.18637/jss.v100.i05>
- 832 DeBruine, L. M., & Barr, D. J. (2021). Understanding Mixed-Effects Models Through
833 Data Simulation. *Advances in Methods and Practices in Psychological Science*, 4(1),
834 2515245920965119. <https://doi.org/10.1177/2515245920965119>
- 835 Eddelbuettel, D., & Balamuta, J. J. (2018). Extending R with C++: A Brief Introduction

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

- 863 <https://doi.org/10.59350/wbn93-edb02>
- 864 Holden, J. G., Van Orden, G. C., & Turvey, M. T. (2009). Dispersion of response times
865 reveals cognitive dynamics. *Psychological Review*, 116(2), 318–342.
- 866 <https://doi.org/10.1037/a0014849>
- 867 Hosmer, D. W., Lemeshow, S., & May, S. (2011). *Applied Survival Analysis: Regression*
868 *Modeling of Time to Event Data* (2nd ed). Hoboken: John Wiley & Sons.
- 869 Kantowitz, B. H., & Pachella, R. G. (2021). The Interpretation of Reaction Time in
870 Information-Processing Research 1. *Human Information Processing*, 41–82.
- 871 <https://doi.org/10.4324/9781003176688-2>
- 872 Kay, M. (2024). *tidybayes: Tidy data and geoms for Bayesian models*.
- 873 <https://doi.org/10.5281/zenodo.1308151>
- 874 Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing,
875 estimation, meta-analysis, and power analysis from a Bayesian perspective.
876 *Psychonomic Bulletin & Review*, 25(1), 178–206.
- 877 <https://doi.org/10.3758/s13423-016-1221-4>
- 878 Kurz, A. S. (2023a). *Applied longitudinal data analysis in brms and the tidyverse* (version
879 0.0.3). Retrieved from <https://bookdown.org/content/4253/>
- 880 Kurz, A. S. (2023b). *Statistical rethinking with brms, ggplot2, and the tidyverse: Second*
881 *edition* (version 0.4.0). Retrieved from <https://bookdown.org/content/4857/>
- 882 Lakens, D. (2022). Sample Size Justification. *Collabra: Psychology*, 8(1), 33267.
- 883 <https://doi.org/10.1525/collabra.33267>
- 884 Landes, J., Engelhardt, S. C., & Pelletier, F. (2020). An introduction to event history
885 analyses for ecologists. *Ecosphere*, 11(10), e03238. <https://doi.org/10.1002/ecs2.3238>
- 886 Lougheed, J. P., Benson, L., Cole, P. M., & Ram, N. (2019). Multilevel survival analysis:
887 Studying the timing of children's recurring behaviors. *Developmental Psychology*,
888 55(1), 53–65. <https://doi.org/10.1037/dev0000619>
- 889 Luce, R. D. (1991). *Response times: Their role in inferring elementary mental organization*

- 890 (1. issued as paperback). Oxford: Univ. Press.
- 891 McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and*
892 *STAN* (2nd ed.). New York: Chapman and Hall/CRC.
- 893 <https://doi.org/10.1201/9780429029608>
- 894 Mills, M. (2011). *Introducing Survival and Event History Analysis*. 1 Oliver's Yard, 55 City
895 Road, London EC1Y 1SP United Kingdom: SAGE Publications Ltd.
- 896 <https://doi.org/10.4135/9781446268360>
- 897 Müller, K., & Wickham, H. (2023). *Tibble: Simple data frames*. Retrieved from
898 <https://CRAN.R-project.org/package=tibble>
- 899 Nelder, J. A. (1999). From Statistics to Statistical Science. *Journal of the Royal Statistical
900 Society. Series D (The Statistician)*, 48(2), 257–269. Retrieved from
901 <https://www.jstor.org/stable/2681191>
- 902 Neuwirth, E. (2022). *RColorBrewer: ColorBrewer palettes*. Retrieved from
903 <https://CRAN.R-project.org/package=RColorBrewer>
- 904 Panis, S. (2020). How can we learn what attention is? Response gating via multiple direct
905 routes kept in check by inhibitory control processes. *Open Psychology*, 2(1), 238–279.
906 <https://doi.org/10.1515/psych-2020-0107>
- 907 Panis, S., Moran, R., Wolkersdorfer, M. P., & Schmidt, T. (2020). Studying the dynamics
908 of visual search behavior using RT hazard and micro-level speed–accuracy tradeoff
909 functions: A role for recurrent object recognition and cognitive control processes.
910 *Attention, Perception, & Psychophysics*, 82(2), 689–714.
911 <https://doi.org/10.3758/s13414-019-01897-z>
- 912 Panis, S., Schmidt, F., Wolkersdorfer, M. P., & Schmidt, T. (2020). Analyzing Response
913 Times and Other Types of Time-to-Event Data Using Event History Analysis: A Tool
914 for Mental Chronometry and Cognitive Psychophysiology. *I-Perception*, 11(6),
915 2041669520978673. <https://doi.org/10.1177/2041669520978673>
- 916 Panis, S., & Schmidt, T. (2016). What Is Shaping RT and Accuracy Distributions? Active

- 917 and Selective Response Inhibition Causes the Negative Compatibility Effect. *Journal of*
918 *Cognitive Neuroscience*, 28(11), 1651–1671. https://doi.org/10.1162/jocn_a_00998
- 919 Panis, S., & Schmidt, T. (2022). When does “inhibition of return” occur in spatial cueing
920 tasks? Temporally disentangling multiple cue-triggered effects using response history
921 and conditional accuracy analyses. *Open Psychology*, 4(1), 84–114.
922 <https://doi.org/10.1515/psych-2022-0005>
- 923 Panis, S., Torfs, K., Gillebert, C. R., Wagemans, J., & Humphreys, G. W. (2017).
924 Neuropsychological evidence for the temporal dynamics of category-specific naming.
925 *Visual Cognition*, 25(1-3), 79–99. <https://doi.org/10.1080/13506285.2017.1330790>
- 926 Panis, S., & Wagemans, J. (2009). Time-course contingencies in perceptual organization
927 and identification of fragmented object outlines. *Journal of Experimental Psychology:
928 Human Perception and Performance*, 35(3), 661–687.
929 <https://doi.org/10.1037/a0013547>
- 930 Pargent, F., Koch, T. K., Kleine, A.-K., Lermer, E., & Gaube, S. (2024). A Tutorial on
931 Tailored Simulation-Based Sample-Size Planning for Experimental Designs With
932 Generalized Linear Mixed Models. *Advances in Methods and Practices in Psychological
933 Science*, 7(4), 25152459241287132. <https://doi.org/10.1177/25152459241287132>
- 934 Pedersen, T. L. (2024). *Patchwork: The composer of plots*. Retrieved from
935 <https://patchwork.data-imaginist.com/>
- 936 Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in s and s-PLUS*. New York:
937 Springer. <https://doi.org/10.1007/b98882>
- 938 R Core Team. (2024). *R: A language and environment for statistical computing*. Vienna,
939 Austria: R Foundation for Statistical Computing. Retrieved from
940 <https://www.R-project.org/>
- 941 Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling
942 Change and Event Occurrence*. Oxford, New York: Oxford University Press.
- 943 Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-N design.

- 944 *Psychonomic Bulletin & Review*, 25(6), 2083–2101.
945 <https://doi.org/10.3758/s13423-018-1451-8>
- 946 Stan Development Team. (2020). *StanHeaders: Headers for the R interface to Stan*.
947 Retrieved from <https://mc-stan.org/>
- 948 Stan Development Team. (2024). *RStan: The R interface to Stan*. Retrieved from
949 <https://mc-stan.org/>
- 950 Stoolmiller, M. (2015). *An Introduction to Using Multivariate Multilevel Survival Analysis*
951 to Study Coercive Family Process (Vol. 1; T. J. Dishion & J. Snyder, Eds.). Oxford
952 University Press. <https://doi.org/10.1093/oxfordhb/9780199324552.013.27>
- 953 Stoolmiller, M., & Snyder, J. (2006). Modeling heterogeneity in social interaction processes
954 using multilevel survival analysis. *Psychological Methods*, 11(2), 164–177.
955 <https://doi.org/10.1037/1082-989X.11.2.164>
- 956 Teachman, J. D. (1983). Analyzing social processes: Life tables and proportional hazards
957 models. *Social Science Research*, 12(3), 263–301.
958 [https://doi.org/10.1016/0049-089X\(83\)90015-7](https://doi.org/10.1016/0049-089X(83)90015-7)
- 959 Tong, C. (2019). Statistical Inference Enables Bad Science; Statistical Thinking Enables
960 Good Science. *The American Statistician*, 73(sup1), 246–261.
961 <https://doi.org/10.1080/00031305.2018.1518264>
- 962 Townsend, J. T. (1990). Truth and consequences of ordinal differences in statistical
963 distributions: Toward a theory of hierarchical inference. *Psychological Bulletin*, 108(3),
964 551–567. <https://doi.org/10.1037/0033-2909.108.3.551>
- 965 Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin &*
966 *Review*, 7(3), 424–465. <https://doi.org/10.3758/BF03214357>
- 967 Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics.
968 *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)
- 969 Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New
970 York. Retrieved from <https://ggplot2.tidyverse.org>

- 971 Wickham, H. (2023a). *Forcats: Tools for working with categorical variables (factors)*.
972 Retrieved from <https://forcats.tidyverse.org/>
- 973 Wickham, H. (2023b). *Stringr: Simple, consistent wrappers for common string operations*.
974 Retrieved from <https://stringr.tidyverse.org>
- 975 Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., ...
976 Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43),
977 1686. <https://doi.org/10.21105/joss.01686>
- 978 Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *Dplyr: A
979 grammar of data manipulation*. Retrieved from <https://dplyr.tidyverse.org>
- 980 Wickham, H., & Henry, L. (2023). *Purrr: Functional programming tools*. Retrieved from
981 [https://purrr.tidyverse.org/](https://purrr.tidyverse.org)
- 982 Wickham, H., Hester, J., & Bryan, J. (2024). *Readr: Read rectangular text data*. Retrieved
983 from <https://readr.tidyverse.org>
- 984 Wickham, H., Vaughan, D., & Girlich, M. (2024). *Tidyr: Tidy messy data*. Retrieved from
985 <https://tidyr.tidyverse.org>
- 986 Winter, B. (2019). *Statistics for Linguists: An Introduction Using R*. New York:
987 Routledge. <https://doi.org/10.4324/9781315165547>
- 988 Wolkersdorfer, M. P., Panis, S., & Schmidt, T. (2020). Temporal dynamics of sequential
989 motor activation in a dual-prime paradigm: Insights from conditional accuracy and
990 hazard functions. *Attention, Perception, & Psychophysics*, 82(5), 2581–2602.
991 <https://doi.org/10.3758/s13414-020-02010-5>