

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 The authors made the following contributions. Sven Panis: Conceptualization,
10 Writing - Original Draft Preparation, Writing - Review & Editing; Richard Ramsey:
11 Conceptualization, Writing - Review & Editing, Supervision.

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

14

Abstract

15 Time-to-event data such as response times, saccade latencies, and fixation durations form a
16 cornerstone of experimental psychology, and have had a widespread impact on our
17 understanding of human cognition. However, the orthodox method for analysing such data
18 – comparing means between conditions – is known to conceal valuable information about
19 the timeline of psychological effects, such as their onset time and duration. The ability to
20 reveal finer-grained, “temporal states” of cognitive processes can have important
21 consequences for theory development by qualitatively changing the key inferences that are
22 drawn from psychological data. Luckily, well-established analytical approaches, such as
23 event history analysis (EHA), are able to evaluate the detailed shape of time-to-event
24 distributions, and thus characterise the time course of psychological states. One barrier to
25 wider use of EHA, however, is that the analytical workflow is typically more
26 time-consuming and complex than orthodox approaches. To help achieve broader uptake,
27 in this paper we outline a set of tutorials that detail how to implement one distributional
28 method known as discrete-time EHA. We illustrate how to wrangle raw data files and
29 calculate descriptive statistics, as well as how to calculate inferential statistics via Bayesian
30 and frequentist multilevel regression modelling. Along the way, we touch upon several key
31 aspects of the workflow, such as how to specify regression models, the implications for
32 experimental design, as well as how to manage inter-individual differences. We finish the
33 article by considering the benefits of the approach for understanding psychological states,
34 as well as the limitations and future directions of this work. Finally, the project is written
35 in R and freely available, which means the general approach can easily be adapted to other
36 data sets, and all of the tutorials are available as .html files to widen access beyond R-users.

37 *Keywords:* response times, event history analysis, Bayesian multi-level regression
38 models, experimental psychology, cognitive psychology

39 Word count: X

40

1. Introduction

41 1.1 Motivation and background context: Comparing means versus 42 distributional shapes

43 In experimental psychology, it is standard practice to analyse reaction times (RTs),
44 saccade latencies, and fixation durations by calculating average performance across a series
45 of trials. Such mean-average comparisons have been the workhorse of experimental
46 psychology over the last century, and have had a substantial impact on theory development
47 as well as our understanding of the structure of cognition and brain function. However,
48 differences in mean RT conceal important pieces of information, such as when an
49 experimental effect starts, how long it lasts, how it evolves with increasing waiting time,
50 and whether its onset is time-locked to other events (Panis, 2020; Panis, Moran,
51 Wolkersdorfer, & Schmidt, 2020; Panis & Schmidt, 2016, 2022; Panis, Torfs, Gillebert,
52 Wagemans, & Humphreys, 2017; Panis & Wagemans, 2009; Wolkersdorfer, Panis, &
53 Schmidt, 2020). Such information is useful not only for the interpretation of experimental
54 effects under investigation, but also for cognitive psychophysiology and computational
55 model selection (Panis, Schmidt, Wolkersdorfer, & Schmidt, 2020).

56 As a simple illustration, Figure 1 shows the results of several simulated RT data sets,
57 which show how mean-average comparisons between two conditions can conceal the shape
58 of the underlying RT distributions. For instance, in examples 1-3, mean RT is always
59 comparable between two conditions, while the distributions differ (Figure 1, left). In
60 contrast, in examples 4-6, mean RT is lower in condition 2 compared to condition 1, but
61 the RT distributions differ in each case (Figure 1, right). Therefore, a comparison of means
62 would lead to a similar conclusion in examples 1-3, as well as examples 4-6, whereas a
63 comparison of the distributions would lead to a different conclusion in every case.

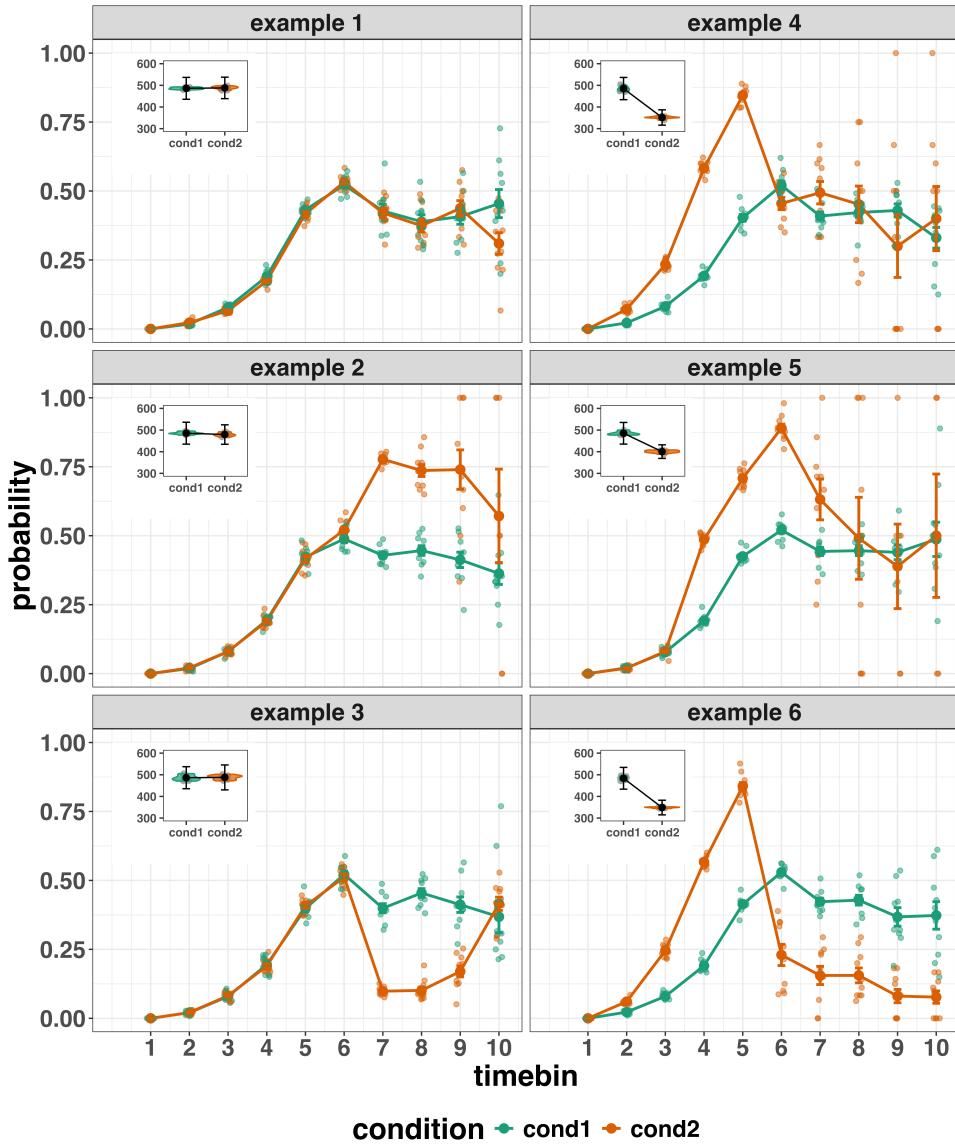


Figure 1. Means versus distributional shapes for six different simulated data set examples.

The first second after stimulus onset is divided in ten bins of 100 ms. Timebin indicates the bin rank. The first bin is (0,100], the last bin is (900,1000]. For our purposes here, it is enough to know that the distributions plotted represent the probability of an event occurring in that timebin, given that it has not yet occurred. Insets show mean reaction time per condition.

65 data across trials, a distributional approach offers the possibility to reveal the time course
66 of psychological states. As such, the approach permits different kinds of questions to be
67 asked, different inferences to be made, and it holds the potential to discriminate between
68 different theoretical accounts of psychological and/or brain-based processes. For example,
69 the distributions in Example 4 show that the effect starts between 100 and 200 ms (in
70 timebin 2) and is gone when the waiting time reaches 500 ms or more. In contrast, in
71 Example 5, the effect starts around 300 ms and is gone by 700 ms. And in the Example 6,
72 the effect reverses between 500 and 600 ms. What kind of theory or theories could account
73 for such effects? Are there new auxiliary assumptions that theories need to adopt? And are
74 there new experiments that need to be performed to test the novel predictions that follow
75 from these analyses? As we show later using published examples, for many psychological
76 questions, such “temporal states” information can be theoretically meaningful by leading to
77 more fine-grained understanding of psychological processes, as well as adding a relatively
78 under-used dimension – the passage of time – to the theory building toolkit.

79 From a historical perspective, it is worth noting that the development of analytical
80 tools that can estimate or predict whether and when events will occur is not a new
81 endeavour. Indeed, hundreds of years ago, analytical methods were developed to predict
82 the duration of time until people died (e.g., Makeham, William M., 1860). The same logic
83 has been applied to psychological time-to-event data, as previously demonstrated (Panis,
84 Schmidt, et al., 2020). Here, in the current paper, we focus on a distributional method for
85 time-to-event data known as discrete-time Event History Analysis (EHA), a.k.a. survival
86 analysis, hazard analysis, duration analysis, failure-time analysis, and transition analysis
87 (Singer & Willett, 2003). We hope to show the value of EHA for knowledge and theory
88 building in cognitive psychology and related areas of research, such as cognitive
89 neuroscience. Moreover, we provide tutorials that provide step-by-step code and
90 instructions in the hope that we can enable others to use EHA in a more routine, efficient
91 and effective manner.

92 1.2 Aims and structure of the paper

93 In this paper, we focus on discrete-time EHA. We first provide a brief overview of
94 EHA to orient the reader to the basic concepts that we will use throughout the paper.
95 However, this will remain relatively short, as this has been covered in detail before (Allison,
96 1982, 2010; Singer & Willett, 2003). Indeed, our primary aim here is to introduce a set of
97 tutorials, which explain **how** to do such analyses, rather than repeat in any detail **why** you
98 may do them.

99 We provide six different tutorials, which are written in the R programming language
100 and publicly available on our Github and the Open Science Framework (OSF) pages, along
101 with all of the other code and material associated with the project. The tutorials provide
102 hands-on, concrete examples of key parts of the analytical process, so that others can apply
103 EHA to their own time-to-event data sets. Each tutorial is provided as an RMarkdown file,
104 so that others can download and adapt the code to fit their own purposes. Additionally,
105 each tutorial is made available as a .html file, so that it can be viewed by any web browser,
106 and thus available to those that do not use R. Finally, the manuscript itself is written in R
107 using the `papaja()` package (Aust & Barth, 2024), which makes it computationally
108 reproducible, in terms of the underlying data and figures.

109 In Tutorial 1a, we illustrate how to process or “wrangle” a previously published RT +
110 accuracy data set to calculate descriptive statistics when there is one independent variable.
111 The descriptive statistics are plotted, and we comment on their interpretation. In Tutorial
112 1b we provide a generalisation of this approach to illustrate how one can calculate the
113 descriptive statistics when using a more complex design, such as when there are two
114 independent variables.

115 In Tutorial 2a, we illustrate how one can fit Bayesian multi-level regression models to
116 RT data using the R package `brms`. We discuss possible link functions, and plot the
117 model-based effects of our predictors of interest. In Tutorial 2b we fit Bayesian multi-level

regression models to *timed* accuracy data to perform a micro-level speed-accuracy tradeoff (SAT) analysis, which complements the EHA of RT data for choice RT data. In Tutorial 3a, we illustrate how to fit the same type of multilevel regression model for RT data in a frequentist framework using the R package lme4. We then briefly compare and contrast these inferential frameworks when applied to EHA. In Tutorial 3b, we illustrate how to perform the SAT analysis in a frequentist framework.

In tutorial 4, we illustrate one approach to planning how much data to collect in an experiment using EHA. We use data simulation techniques to vary sample size and trial count per condition until a certain degree of statistical power or precision is reached. [[more to come here, once we have written the tutorial]].

In summary, even though EHA is a widely used statistical tool and there already exist many excellent reviews (e.g., Blossfeld & Rohwer, 2002; Box-Steffensmeier, 2004; Hosmer, Lemeshow, & May, 2011; Teachman, 1983) and tutorials (e.g., Allison, 2010; Landes, Engelhardt, & Pelletier, 2020) on its general use-cases, we are not aware of any tutorials that are aimed at psychological time-to-event data, and which provide worked examples of the key data processing and multi-level regression modelling steps. Therefore, our ultimate goal is twofold: first, we want to convince readers of the many benefits of using EHA when dealing with time-to-event data with a focus on psychological time-to-event data, and second, we want to provide a set of practical tutorials, which provide step-by-step instructions on how you actually perform a discrete-time EHA on time-to-event data such as RT data, as well as a complementary discrete-time SAT analysis on timed accuracy data.

2. A brief introduction to event history analysis

For a comprehensive background context to EHA, we recommend several excellent textbooks (Allison, 2010; Singer & Willett, 2003). Likewise, for a general introduction to understanding regression equations, we recommend several excellent textbooks (Gelman,

¹⁴³ Hill, & Vehtari, 2020; Winter, 2019). Our focus here is not on providing a detailed account
¹⁴⁴ of the underlying regression equations, since this topic has been comprehensively covered
¹⁴⁵ many times before. Instead, we want to provide an intuition regarding how EHA works in
¹⁴⁶ general, as well as in the context of experimental psychology. As such, we only supply
¹⁴⁷ regression equations in the supplementary material.

¹⁴⁸ **2.1 Basic features of event history analysis**

¹⁴⁹ To apply EHA, one must be able to:

- ¹⁵⁰ 1. define an event of interest that represents a qualitative change that can be situated in
¹⁵¹ time (e.g., a button press, a saccade onset, a fixation offset, etc.);
- ¹⁵² 2. define time point zero (e.g., target stimulus onset, fixation onset);
- ¹⁵³ 3. measure the passage of time between time point zero and event occurrence in discrete
¹⁵⁴ or continuous time units.

¹⁵⁵ In EHA, the definition of hazard and the type of models employed depend on
¹⁵⁶ whether one is using continuous or discrete time units. Since our focus here is on hazard
¹⁵⁷ models that use discrete time units, we describe that approach. After dividing time in
¹⁵⁸ discrete, contiguous time bins indexed by t (e.g., $t = 1:10$ timebins), let RT be a discrete
¹⁵⁹ random variable denoting the rank of the time bin in which a particular person's response
¹⁶⁰ occurs in a particular experimental condition. For example, the first response might occur
¹⁶¹ at 546 ms and it would be in timebin 6 (any RTs from 501 ms to 600 ms).

¹⁶² Discrete-time EHA focuses on the discrete-time hazard function of event occurrence
¹⁶³ and the discrete-time survivor function (Figure 2). The equations that define both of these
¹⁶⁴ functions are reported in part A of the supplementary material. The discrete-time hazard
¹⁶⁵ function gives you, for each time bin, the probability that the event occurs (sometime) in

¹⁶⁶ bin t, given that the event does not occur in previous bins. In other words, it reflects the
¹⁶⁷ instantaneous likelihood that the event occurs in the current bin, given that it has not yet
¹⁶⁸ occurred in the past, i.e., in one of the prior bins. In contrast, the discrete-time survivor
¹⁶⁹ function cumulates the bin-by-bin risks of event *nonoccurrence* to obtain the survival
¹⁷⁰ probability, the probability that the event occurs after bin t. In other words, the survivor
¹⁷¹ function gives you for each time bin the likelihood that the event occurs in the future, i.e.,
¹⁷² in one of the subsequent timebins.

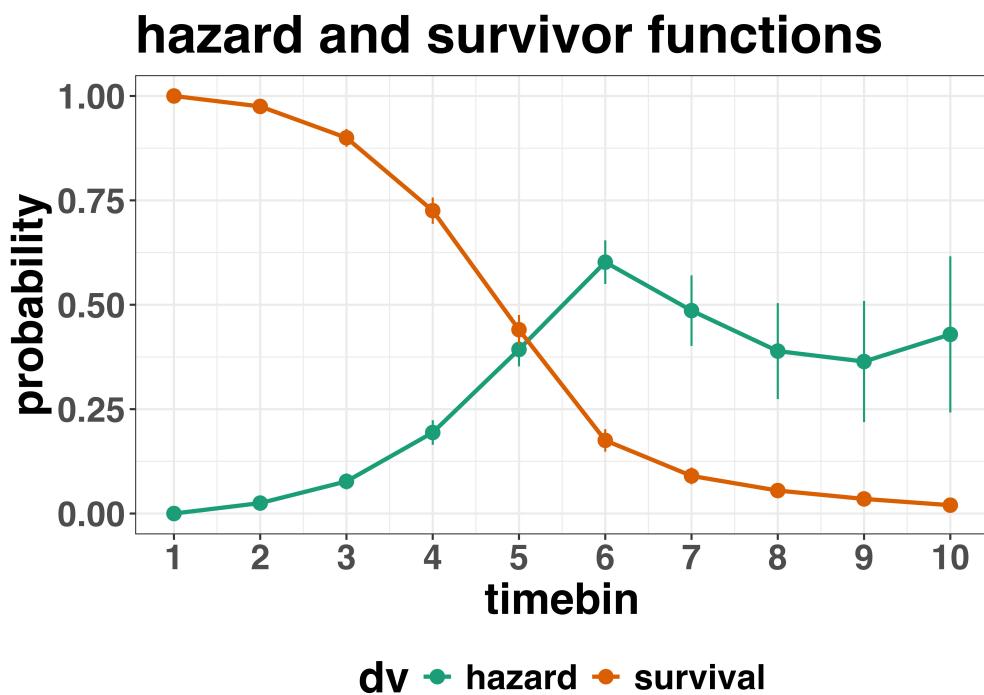


Figure 2. Discrete-time hazard and survivor functions. Discrete time-to-event data were simulated for 200 trials of 1 experimental condition. While the hazard function is the vehicle for inferring the time course of cognitive processes, the survival probability $S(t-1)$ can help to qualify or provide context to the interpretation of the hazard probability $h(t)$. For example, the high hazard of $.60 = h(t=6)$ is experienced only by 44 percent of the trials, as $S(t-1=5) = .44$. Because the survivor function is a decreasing function of time, the error bars in later parts of the hazard function will always be wider and less precise compared to earlier parts.

173 2.2 Benefits of event history analysis

174 Statisticians and mathematical psychologists recommend focusing on the hazard
175 function when analyzing time-to-event data for various reasons. We do not cover these
176 benefits in detail here, as these are more general topics that have been covered elsewhere in
177 textbooks. Instead, we briefly summarise list the benefits below, and refer the reader to
178 section F of Supplementary Materials for more detailed coverage of the benefits. A
179 summary of the benefits are as follows:

- 180 1. Hazard functions are more diagnostic than density functions when one is interested in
181 studying the detailed shape of a RT distribution (Holden et al., 2009).
- 182 2. RT distributions may differ from each other in multiple ways, and hazard functions
183 allow one to capture these differences that mean-average comparisons may conceal
184 (Townsend, 1990).
- 185 3. EHA takes account of more of the data collected in a typical speeded response
186 experiment, by virtue of not discarding right-censored observations. Trials with
187 longer RTs are not discarded, but instead contribute to the risk set in each time bin.
- 188 4. Hazard modeling allows one to incorporate time-varying explanatory covariates, such
189 as heart rate, electroencephalogram (EEG) signal amplitude, gaze location, etc.
190 (Allison, 2010). This is useful for linking physiological effects to behavioral effects
191 when performing cognitive psychophysiology (Meyer, Osman, Irwin, & Yantis, 1988).
- 192 5. EHA can help to solve the problem of model mimicry, i.e., the fact that different
193 computational models can often predict the same mean RTs as observed in the
194 empirical data, but not necessarily the detailed shapes of the empirical RT hazard
195 distributions. As such, EHA can be a tool to help distinguish between competing
196 theories of cognition and brain function.

197 2.3 Event history analysis in the context of experimental psychology

198 To make EHA more relevant to researchers studying cognitive psychology and
199 cognitive neuroscience, in this section we provide a relevant worked example and consider
200 implications that are relevant to that domain of research.

201 **2.3.1 A worked example.** In the context of experimental psychology, it is
202 common for participants to be presented with either a 1-button detection task or a
203 2-button discrimination task. For example, a task may involve choosing between two
204 response options with only one of them being correct. For such two-choice RT data, the
205 discrete-time EHA of the RT data (hazard and survivor functions) can be extended with a
206 discrete-time SAT analysis of the timed accuracy data. Specifically, the hazard function of
207 event occurrence can be extended with the discrete-time conditional accuracy function,
208 which gives you the probability that a response is correct given that it is emitted in time
209 bin t (Allison, 2010; Kantowitz & Pachella, 2021; Wickelgren, 1977). We refer to this
210 extended (hazard + conditional accuracy) analysis for choice RT data as EHA/SAT.

211 Integrating results between hazard and conditional accuracy functions for choice RT
212 data can be informative for understanding psychological processes. To illustrate, we
213 consider a hypothetical choice RT example that is inspired by real data (Panis & Schmidt,
214 2016), but simplified to make the main point clearer (Figure 3). In a standard priming
215 paradigm, there is a prime stimulus (e.g., an arrow pointing left or right) followed by a
216 target stimulus (another arrow pointing left or right). The prime can then be congruent or
217 incongruent with the target.

218 Figure 3 shows that the early upswing in hazard is equal for both priming conditions,
219 and that early responses are always correct in the congruent condition and always incorrect
220 in the incongruent condition. These results show that for short waiting times (< bin 6),
221 responses always follow the prime (and not the target, as instructed). Between 500 and 600
222 ms the target-triggered response channel is activated and causes response competition –

223 ca(6) = .5 – and a lower hazard probability in the incongruent condition. For waiting times
224 of 600 ms or more, the hazard of response occurrence is lower in incongruent compared to
225 congruent trials, and all responses emitted in these later bins are correct.

226 This joint pattern of results is interesting because it can provide meaningfully
227 different conclusions about psychological processes compared to conventional analyses, such
228 as computing mean-average RT across trials. Mean-average RT would only represent the
229 overall ability of cognition to overcome interference, on average, across trials. For instance,
230 if mean-average RT was higher in incongruent than congruent trials, one may conclude
231 that cognitive mechanisms that support interference control are working as expected across
232 trials, and are indexed by each recorded response. But such a conclusion is not supported
233 when the effects are explored over a timeline. Instead, the psychological conclusion is much
234 more nuanced and suggests that multiple states start, stop and possibly interact over a
235 particular temporal window.

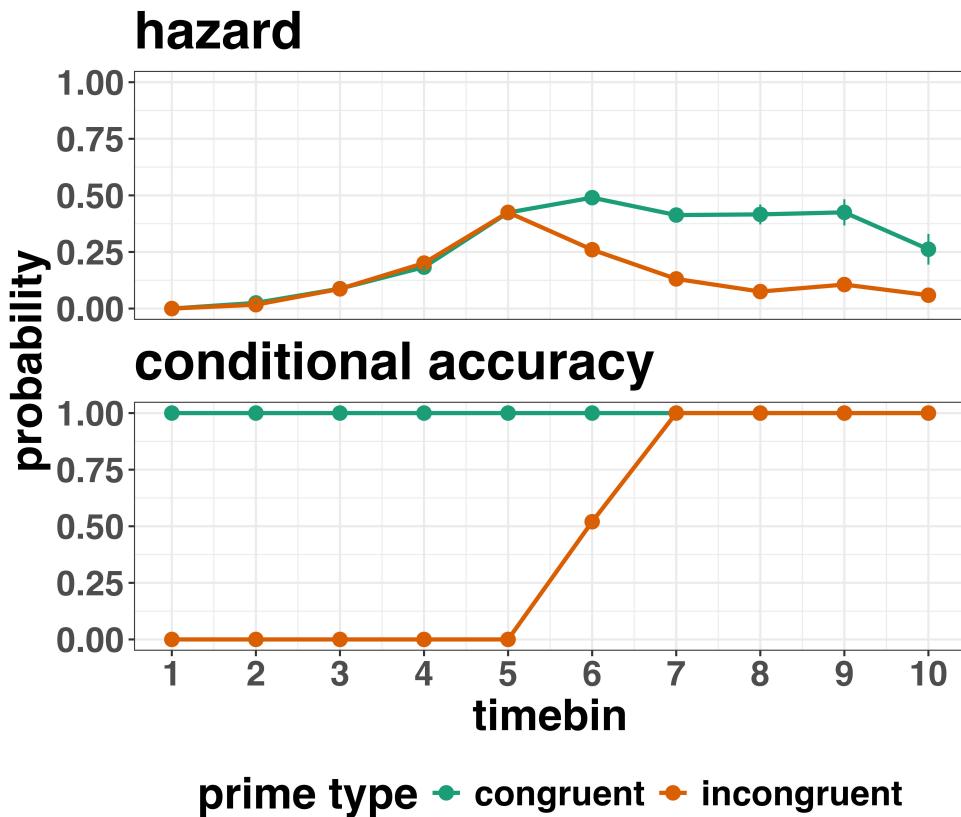


Figure 3. Discrete-time hazard and conditional accuracy functions. Discrete time-to-event and conditional accuracy data were simulated for 2000 trials for each of two priming conditions (congruent and incongruent prime stimuli).

236 Unlocking the temporal states of cognitive processes can be revealing for theory
 237 development and the understanding of basic psychological processes. Possibly more
 238 importantly, however, is that it simultaneously opens the door to address many new and
 239 previously unanswered questions. Do all participants show similar temporal states or are
 240 there individual differences? Do such individual differences extend to those individuals that
 241 have been diagnosed with some form of psychopathology? How do temporal states relate to
 242 brain-based mechanisms that might be studied using other methods from cognitive
 243 neuroscience? And how much of theory in cognitive psychology would be in need of
 244 revision if mean-average comparisons were supplemented with a temporal states approach?

245 **2.3.2 Implications for designing experiments.** Performing EHA in
246 experimental psychology has implications for how experiments are designed. Indeed, if
247 trials are categorised as a function of when responses occur, then each timebin will only
248 include a subset of the total number of trials. For example, let's consider an experiment
249 where each participant performs 2 conditions and there are 100 trial repetitions per
250 condition. Those 100 trials must be distributed in some manner across the chosen number
251 of bins.

252 In such experimental designs, since the number of trials per condition are spread
253 across bins, it is important to have a relatively large number of trial repetitions per
254 participant and per condition. Accordingly, experimental designs using this approach
255 typically focus on factorial, within-subject designs, in which a large number of observations
256 are made on a relatively small number of participants (so-called small- N designs). This
257 approach emphasizes the precision and reproducibility of data patterns at the individual
258 participant level to increase the inferential validity of the design (Baker et al., 2021; Smith
259 & Little, 2018).

260 In contrast to the large- N design that typically average across many participants
261 without being able to scrutinize individual data patterns, small- N designs retain crucial
262 information about the data patterns of individual observers. This can be advantageous
263 whenever participants differ systematically in their strategies or in the time courses of their
264 effects, so that averaging them would lead to misleading data patterns. Note that because
265 statistical power derives both from the number of participants and from the number of
266 repeated measures per participant and condition, small- N designs can still achieve what
267 are generally considered acceptable levels of statistical power, if they have a sufficient
268 amount of data overall (Baker et al., 2021; Smith & Little, 2018).

269 3. An overview of the general analytical workflow

270 Although the focus is on EHA/SAT, we also want to briefly comment on broader
271 aspects of our general analytical workflow, which relate more to data science and data
272 analysis workflows.

273 **3.1 Data science workflow and descriptive statistics**

274 Descriptive, data science workflow. We perform data wrangling following tidyverse
275 principles and a functional programming approach (Wickham, Çetinkaya-Rundel, &
276 Grolemund, 2023). Functional programming basically means you don't write your own
277 loops but instead use functions that have been built and tested by others. [[more here, as
278 necessary]].

279 **3.2 Inferential statistical approach**

280 Our lab adopts a estimation approach to multi-level regression (Kruschke & Liddell,
281 2018; Winter, 2019), which is heavily influenced by the Bayesian framework as suggested
282 by Richard McElreath (Kurz, 2023b; McElreath, 2018). We also use a “keep it maximal”
283 approach to specifying varying (or random) effects (Barr, Levy, Scheepers, & Tily, 2013).
284 This means that wherever possible we include varying intercepts and slopes per participant
285 To make inferences, we use two main approaches. We compare models of different
286 complexity, using information criteria (e.g., WAIC) and cross-validation (e.g., LOO), to
287 evaluate out-of-sample predictive accuracy (McElreath, 2018). We also take the most
288 complex model and evaluate key parameters of interest using point and interval estimates.

²⁸⁹ **3.3 Implementation**

²⁹⁰ We used R (Version 4.4.0; R Core Team, 2024)¹ for all reported analyses. The
²⁹¹ content of the tutorials, in terms of EHA and multi-level regression modelling, is mainly
²⁹² based on Allison (2010), Singer and Willett (2003), McElreath (2018), Kurz (2023a), and
²⁹³ Kurz (2023b).

²⁹⁴ **4. Tutorials**

²⁹⁵ Tutorials 1a and 1b show how to calculate and plot the descriptive statistics of
²⁹⁶ EHA/SAT when there are one and two independent variables, respectively. Tutorials 2a
²⁹⁷ and 2b illustrate how to use Bayesian multilevel modeling to fit hazard and conditional
²⁹⁸ accuracy models, respectively. Tutorials 3a and 3b show how to implement, respectively,
²⁹⁹ multilevel models for hazard and conditional accuracy in the frequentist framework.
³⁰⁰ Additionally, to further simplify the process for other users, the first two tutorials rely on a
³⁰¹ set of our own custom functions that make sub-processes easier to automate, such as data
³⁰² wrangling and plotting functions (see part B in the supplemental material for a list of the

¹ We, furthermore, used the R-packages *bayesplot* (Version 1.11.1; Gabry, Simpson, Vehtari, Betancourt, & Gelman, 2019), *brms* (Version 2.21.0; Bürkner, 2017, 2018, 2021), *citr* (Version 0.3.2; Aust, 2019), *cmdstanr* (Version 0.8.1.9000; Gabry, Češnovar, Johnson, & Broder, 2024), *dplyr* (Version 1.1.4; Wickham, François, Henry, Müller, & Vaughan, 2023), *forcats* (Version 1.0.0; Wickham, 2023a), *ggplot2* (Version 3.5.1; Wickham, 2016), *lme4* (Version 1.1.35.5; Bates, Mächler, Bolker, & Walker, 2015), *lubridate* (Version 1.9.3; Grolemund & Wickham, 2011), *Matrix* (Version 1.7.0; Bates, Maechler, & Jagan, 2024), *nlme* (Version 3.1.166; Pinheiro & Bates, 2000), *papaja* (Version 0.1.2.9000; Aust & Barth, 2023), *patchwork* (Version 1.2.0; Pedersen, 2024), *purrr* (Version 1.0.2; Wickham & Henry, 2023), *RColorBrewer* (Version 1.1.3; Neuwirth, 2022), *Rcpp* (Eddelbuettel & Balamuta, 2018; Version 1.0.12; Eddelbuettel & François, 2011), *readr* (Version 2.1.5; Wickham, Hester, & Bryan, 2024), *RJ-2021-048* (Bengtsson, 2021), *standist* (Version 0.0.0.9000; Girard, 2024), *stringr* (Version 1.5.1; Wickham, 2023b), *tibble* (Version 3.2.1; Müller & Wickham, 2023), *tidybayes* (Version 3.0.6; Kay, 2023), *tidyverse* (Version 2.0.0; Wickham et al., 2019), and *tinylabels* (Version 0.2.4; Barth, 2023).

303 custom functions).

304 Our list of tutorials is as follows:

- 305 • 1a. Wrangle raw data and calculate descriptive stats for one independent variable
- 306 • 1b. Wrangle raw data and calculate descriptive stats for two independent variables
- 307 • 2a. Bayesian multilevel modeling for $h(t)$
- 308 • 2b. Bayesian multilevel modeling for $ca(t)$
- 309 • 3a. Frequentist multilevel modeling for $h(t)$
- 310 • 3b. Frequentist multilevel modeling for $ca(t)$
- 311 • 4. Simulation and power analysis for planning experiments

312 **4.1 Tutorial 1a: Calculating descriptive statistics using a life table**

313 **4.1.1 Data wrangling aims.** Our data wrangling procedures serve two related
314 purposes. First, we want to summarise and visualise descriptive statistics that relate to our
315 main research questions about the time course of psychological processes. Second, we want
316 to produce two different data sets that can each be submitted to different types of
317 inferential modelling approaches. The two types of data structure we label as ‘person-trial’
318 data and ‘person-trial-bin’ data. The ‘person-trial’ data (Table 1) will be familiar to most
319 researchers who record behavioural responses from participants, as it represents the
320 measured RT and accuracy per trial within an experiment. This data set is used when
321 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.

322 In contrast, the ‘person-trial-bin’ data (Table 2) has a different, more extended
 323 structure, which indicates in which bin a response occurred, if at all, in each trial.
 324 Therefore, the ‘person-trial-bin’ data set generates a 0 in each bin until an event occurs
 325 and then it generates a 1 to signal an event has occurred in that bin. This data set is used
 326 when fitting hazard models (Tutorials 2a and 3a). It is worth pointing out that there is no
 327 requirement for an event to occur at all (in any bin), as maybe there was no response on
 328 that trial or the event occurred after the time window of interest. Likewise, when the event
 329 occurs in bin 1 there would only be one row of data for that trial in the person-trial-bin
 330 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.

331 **4.1.2 A real data wrangling example.** To illustrate how to quickly set up life
 332 tables for calculating the descriptive statistics (functions of discrete time), we use a
 333 published data set on masked response priming from Panis and Schmidt (2016). In their
 334 first experiment, Panis and Schmidt (2016) presented a double arrow for 94 ms that
 335 pointed left or right as the target stimulus with an onset at time point zero in each trial.
 336 Participants had to indicate the direction in which the double arrow pointed using their
 337 corresponding index finger, within 800 ms after target onset. Response time and accuracy
 338 were recorded on each trial. Prime type (blank, congruent, incongruent) and mask type
 339 were manipulated. Here we focus on the subset of trials in which no mask was presented.

340 The 13-ms prime stimulus was a double arrow presented 187 ms before target onset in the
 341 congruent (same direction as target) and incongruent (opposite direction as target) prime
 342 conditions.

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

346 The required column names are as follows:

- 347 • “pid”, indicating unique participant IDs;
- 348 • “trial”, indicating each unique trial per participant;
- 349 • “condition”, a factor indicating the levels of the independent variable (1, 2, ...) and
 350 the corresponding labels;
- 351 • “rt”, indicating the response times in ms;
- 352 • “acc”, indicating the accuracies (1/0).

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

```
data_wr <- read_csv("../Tutorial_1_descriptive_stats/data/DataExp1_6subjects_wrangled.csv")
colnames(data_wr) <- c("pid", "bl", "tr", "condition", "resp", "acc", "rt", "trial")
data_wr <- data_wr %>%
  mutate(condition = condition + 1, # original levels were 0, 1, 2.
        condition = factor(condition, levels=c(1,2,3), labels=c("blank", "congruent", "incongruent")))

```

354 Next, we can set up the life tables and plots of the discrete-time functions $h(t)$, $S(t)$,
 355 $ca(t)$, and $P(t)$ – see part A of the supplementary material for their definitions. To do so
 356 using a functional programming approach, one has to nest the data within participants
 357 using the group_nest() function, and supply a user-defined censoring time and bin width
 358 to our custom function “censor()”, as follows.

```

data_nested <- data_wr %>% group_nest(pid)

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

```

359 Note that the censoring time should be a multiple of the bin width (both in ms). The
 360 censoring time should be a time point after which no informative responses are expected
 361 anymore. In experiments that implement a response deadline in each trial the censoring
 362 time can equal that deadline time point. Trials with a RT larger than the censoring time,
 363 or trials in which no response is emitted during the data collection period, are treated as
 364 right-censored observations in EHA. In other words, these trials are not discarded, because
 365 they contain the information that the event did not occur before the censoring time.
 366 Removing such trials before calculating the mean event time will result in underestimation
 367 of the true mean.

368 The person-trial-bin oriented data set is created by our custom function `ptb()`, and it
 369 has one row for each time bin (of each trial) that is at risk for event occurrence. The
 370 variable “event” in the person-trial-bin oriented data set indicates whether a response
 371 occurs (1) or not (0) for each bin.

372 The next step is to set up the life table using our custom function `setup_lt()`,
 373 calculate the conditional accuracies using our custom function `calc_ca()`, add the $ca(t)$
 374 estimates to the life table using our custom function `join_lt_ca()`, and then plot the
 375 descriptive statistics using our custom function `plot_eha()`. When creating the plots, some
 376 warning messages will likely be generated, like these:

- 377 • Removed 2 rows containing missing values or values outside the scale range
 378 (`geom_line()`).

- 379 • Removed 2 rows containing missing values or values outside the scale range
380 (`geom_point()`).
- 381 • Removed 2 rows containing missing values or values outside the scale range
382 (`geom_segment()`).

383 The warning messages are generated because some bins have no hazard and $ca(t)$
384 estimates, and no error bars. They can thus safely be ignored. One can now inspect
385 different aspects, including the life table for a particular condition of a particular subject,
386 and a plot of the different functions for a particular participant.

387 In general, it is important to visually inspect the functions first for each participant,
388 in order to identify individuals that may be guessing (e.g., a flat conditional accuracy
389 function at .5 indicates that someone may be guessing), outlying individuals, and/or
390 different groups with qualitatively different behavior.

391 Table 3 shows the life table for condition “blank” (no prime stimulus presented) for
392 participant 6. A life table includes for each time bin, the risk set (i.e., the number of trials
393 that are event-free at the start of the bin), the number of observed events, and the
394 estimates of $h(t)$, $S(t)$, $P(t)$, possibly $ca(t)$, and their estimated standard errors (se). At
395 time point zero, no events can occur and therefore $h(t)$ and $ca(t)$ are undefined.

396 Figure 4 displays the discrete-time hazard, survivor, conditional accuracy, and
397 probability mass functions for each prime condition for participant 6. By using
398 discrete-time hazard functions of event occurrence – in combination with conditional
399 accuracy functions for two-choice tasks – one can provide an unbiased, time-varying, and
400 probabilistic description of the latency and accuracy of responses based on all trials of any
401 data set.

402 For example, for participant 6, the estimated hazard values in bin (240,280] are 0.03,
403 0.17, and 0.20 for the blank, congruent, and incongruent prime conditions, respectively. In
404 other words, when the waiting time has increased until *240 ms* after target onset, then the

405 conditional probability of response occurrence in the next 40 ms is more than five times
406 larger for both prime-present conditions, compared to the blank prime condition.

407 Furthermore, the estimated conditional accuracy values in bin (240,280] are 0.29, 1,
408 and 0 for the blank, congruent, and incongruent prime conditions, respectively. In other
409 words, if a response is emitted in bin (240,280], then the probability that it is correct is
410 estimated to be 0.29, 1, and 0 for the blank, congruent, and incongruent prime conditions,
411 respectively.

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. se = standard error. ca = conditional accuracy.

Descriptive stats for subject 6

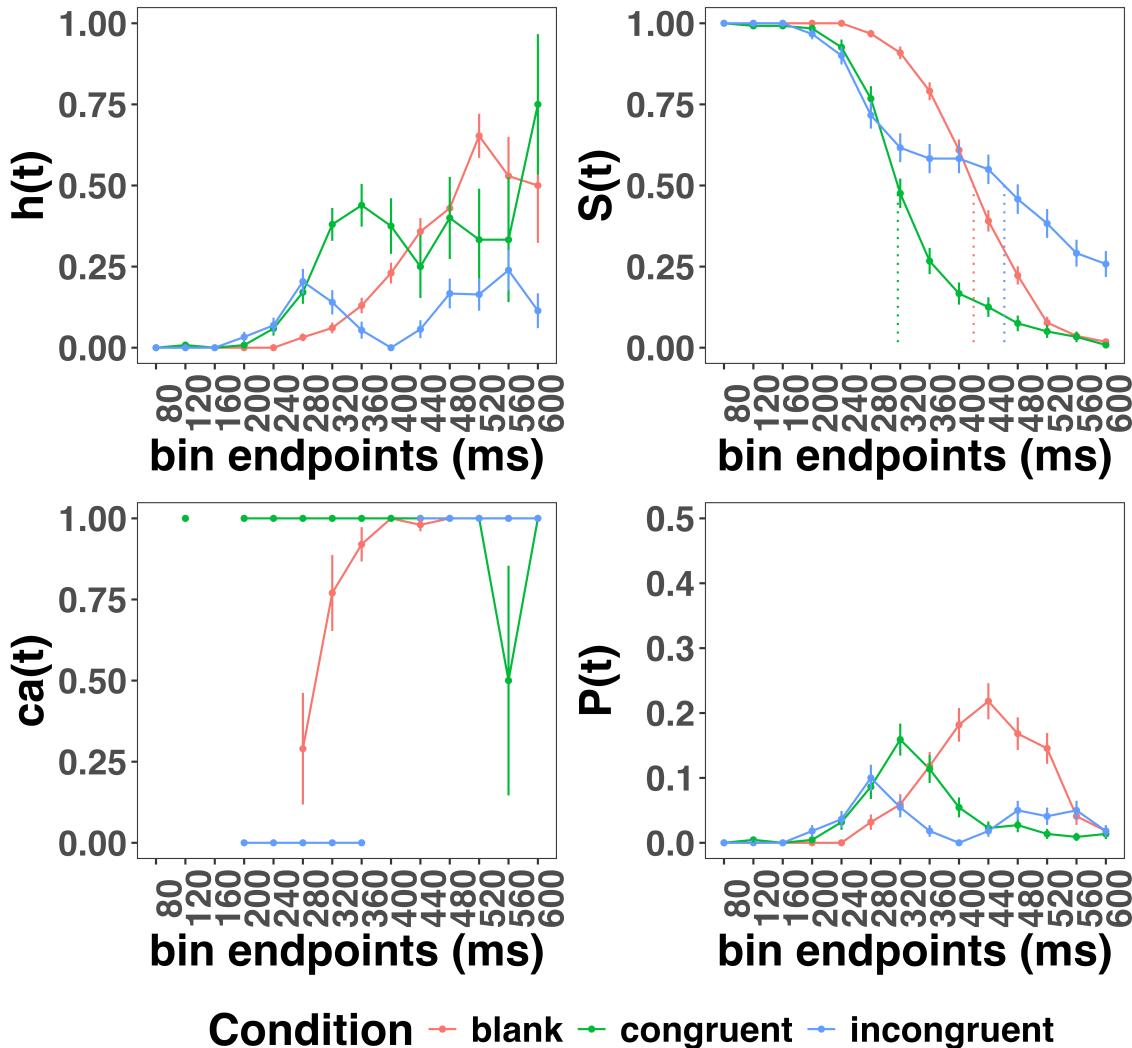


Figure 4. Estimated discrete-time hazard, survivor, probability mass, and conditional accuracy functions for participant 6, as a function of the passage of discrete waiting time.

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

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

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

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

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

417 conditions, respectively.

418 These distributional results suggest that the participant 6 is initially responding to
419 the prime even though (s)he was instructed to only respond to the target, that response
420 competition emerges in the incongruent prime condition around 300 ms, and that only
421 slower responses are fully controlled by the target stimulus. Qualitatively similar results
422 were obtained for the other five participants. When participants show qualitatively the
423 same distributional patterns, one might consider to aggregate their data and make one plot
424 (see Tutorial_1a.Rmd).

425 In general, these results go against the (often implicit) assumption in research on
426 priming that all observed responses are primed responses to the target stimulus. Instead,
427 the distributional data show that early responses are triggered exclusively by the prime
428 stimulus, while only later responses reflect primed responses to the target stimulus.

429 At this point, we have calculated, summarised and plotted descriptive statistics for
430 the key variables in EHA/SAT. As we will show in later Tutorials, statistical models for
431 $h(t)$ and $ca(t)$ can be implemented as generalized linear mixed regression models predicting
432 event occurrence (1/0) and conditional accuracy (1/0) in each bin of a selected time
433 window for analysis. But first we consider calculating the descriptive statistics for two
434 independent variables.

435 **4.2 Tutorial 1b: Generalising to a more complex design**

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

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

447 **4.3 Tutorial 2a: Fitting Bayesian hazard models to discrete time-to-event data**

448 In this third tutorial, we illustrate how to fit Bayesian multi-level regression models
449 to the RT data of the masked response priming data set used in Tutorial 1a. Fitting
450 (Bayesian or non-Bayesian) regression models to time-to-event data is important when you
451 want to study how the shape of the hazard function depends on various predictors (Singer
452 & Willett, 2003).

453 **4.3.1 Hazard model considerations.** There are several analytic decisions one
454 has to make when fitting a discrete-time hazard model. First, one has to select an analysis
455 time window, i.e., a contiguous set of bins for which there is enough data for each
456 participant. Second, given that the dependent variable (event occurrence) is binary, one
457 has to select a link function (see part C in the supplementary material). The cloglog link is
458 preferred over the logit link when events can occur in principle at any time point within a
459 bin, which is the case for RT data (Singer & Willett, 2003). Third, one has to choose
460 whether to treat discrete TIME (i.e., the time bin index t) as a discrete or continuous
461 predictor. And when you treat time as a discrete predictor, you can choose between
462 reference coding and index coding.

463 In the case of a large- N design without repeated measurements, the parameters of a
464 discrete-time hazard model can be estimated using standard logistic regression software
465 after expanding the typical person-trial data set into a person-trial-bin data set (Allison,
466 2010). When there is clustering in the data, as in the case of a small- N design with
467 repeated measurements, the parameters of a discrete-time hazard model can be estimated

468 using population-averaged methods (e.g., Generalized Estimating Equations), and Bayesian
 469 or frequentist generalized linear mixed models (Allison, 2010).

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

475 In tutorial_2a.Rmd we fit several Bayesian multilevel models (i.e., generalized linear
 476 mixed models) that differ in complexity to the person-trial-bin oriented data set that we
 477 created in Tutorial 1a. We decided to select the analysis time window (200,600] and the
 478 cloglog link. Below, we shortly discuss three of these models. The person-trial-bin data set
 479 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) %>%
    # continuous predictor for TIME named "period_9", centered on bin 9
  mutate(period_9 = period - 9,
    # categorical predictor for TIME named "timebin" with index coding
    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)))
```

480 **4.3.2 Prior distributions.** To get the posterior distribution of each model
 481 parameter given the data, we need to specify prior distributions for the model parameters
 482 which reflect our prior beliefs. In Tutorial_2a.Rmd we perform some prior predictive
 483 checks to make sure our selected prior distributions reflect our prior beliefs.

484 The middle column of Figure 12 in part E of the supplementary material shows seven
 485 examples of prior distributions for an intercept on the logit and/or cloglog scales. While a
 486 normal distribution with relatively large variance is often used as a weakly informative
 487 prior for continuous dependent variables, rows A and B in Figure 12 show that specifying
 488 such distributions on the logit and cloglog scales actually leads to rather informative
 489 distributions on the original probability scale, as most mass is pushed to probabilities of 0
 490 and 1.

491 **4.3.3 Model M0i: A general specification of TIME.** When you do not want
 492 to make assumptions about the shape of the hazard function, or its shape is not smooth
 493 but irregular, then you can use a general specification of TIME, i.e., fit one grand intercept
 494 per time bin. In this first model, we use a general specification of TIME using index
 495 coding, and call the model “M0i”.

496 Before we fit model M0i, we select the necessary columns from the data, and specify
 497 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")
```

498 After selecting the bernoulli family and the cloglog link, the model formula is
 499 specified. The specification “0 + ...” removes the default intercept in brm(). The fixed
 500 effects include an intercept for each level of timebin. Each of these intercepts is allowed to
 501 vary across individuals (variable pid). Estimating model M0i took about 30 minutes on a
 502 MacBook Pro (Sonoma 14.6.1 OS, 18GB Memory, M3 Pro Chip).

503 **4.3.4 Model M1i: Adding the effects of prime-target congruency.** Previous
 504 research has shown that psychological effects typically change over time (Panis, 2020;
 505 Panis, Moran, et al., 2020; Panis & Schmidt, 2022; Panis et al., 2017; Panis & Wagemans,
 506 2009). In the next model, we use index coding for both TIME (variable “timebin”) and the
 507 categorical predictor prime-target-congruency (variable “prime”), so that we get 30 grand
 508 intercepts, one for each combination of timebin level and prime level. Here is the model
 509 formula of this model that we call “M1i”.

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

510 Estimating model M1i took about 124 minutes.

511 **4.3.4 Model M1d: A more parsimonious model.** When the shape of the
 512 hazard function is rather smooth, as it is for behavioral RT data, one can fit a more
 513 parsimonious model by treating TIME as a continuous variable, and use a polynomial
 514 specification of the effect of TIME. Thus, if we want to make assumptions about (1) how
 515 hazard changes over TIME in the reference condition (blank prime), and (2) how the effect
 516 of congruent and incongruent primes change over TIME (relax the proportionality
 517 assumption), then we can switch to a dummy coding approach for prime-target congruency
 518 (variable “condition”) and treat TIME as a continuous variable (variable “period_9”).

519 For example, we may assume that hazard can change in a linear + quadratic fashion
 520 over time for a blank prime, and that the effects of congruent and incongruent primes
 521 relative to blank change in a linear + quadratic fashion, and fit the model called “M1d”.
 522 Here is its model formula.

```
event ~ 0 + Intercept + condition*period_9 + condition*period_9_sq +
      (1 + condition*period_9 + condition*period_9_sq | pid)
```

523 The specification “0 + Intercept + …” removes the default intercept in brm() and
 524 adds an explicit Intercept for which we can set the prior ourselves. Note that duplicate
 525 terms in the model formula (e.g., condition) are ignored. Because TIME is centered on bin
 526 9, the Intercept represents the estimated cloglog-hazard in bin 9 for the blank prime
 527 condition.

528 **4.3.5 Compare the models.** We can compare the three models using the Widely
 529 Applicable Information Criterion (WAIC) and Leave-One-Out (LOO) cross-validation, and
 530 look at model weights for both criteria (Kurz, 2023a; McElreath, 2018).

```
model_weights(model_M0i, model_M1i, model_M1d, weights = "loo")
```

```
531 ##      model_M0i      model_M1i      model_M1d
532 ## 2.462363e-183 1.000000e+00 5.739791e-17
```

```
model_weights(model_M0i, model_M1i, model_M1d, weights = "waic")
```

```
533 ##      model_M0i      model_M1i      model_M1d
534 ## 8.545948e-184 1.000000e+00 2.055767e-17
```

535 Clearly, both the loo and waic weighting schemes assign a weight of 1 to model M1i,
 536 and a weight of 0 to the other two simpler models.

537 **4.3.6 Interpreting model M1i.** To make inferences from the parameter estimates
 538 in model M1i, we first plot the densities of the draws from the posterior distributions of its
 539 population-level parameters in Figure 5, together with point (median) and interval
 540 estimates (80% and 95% credible intervals).

Posterior distributions for population-level effects in Model M1i

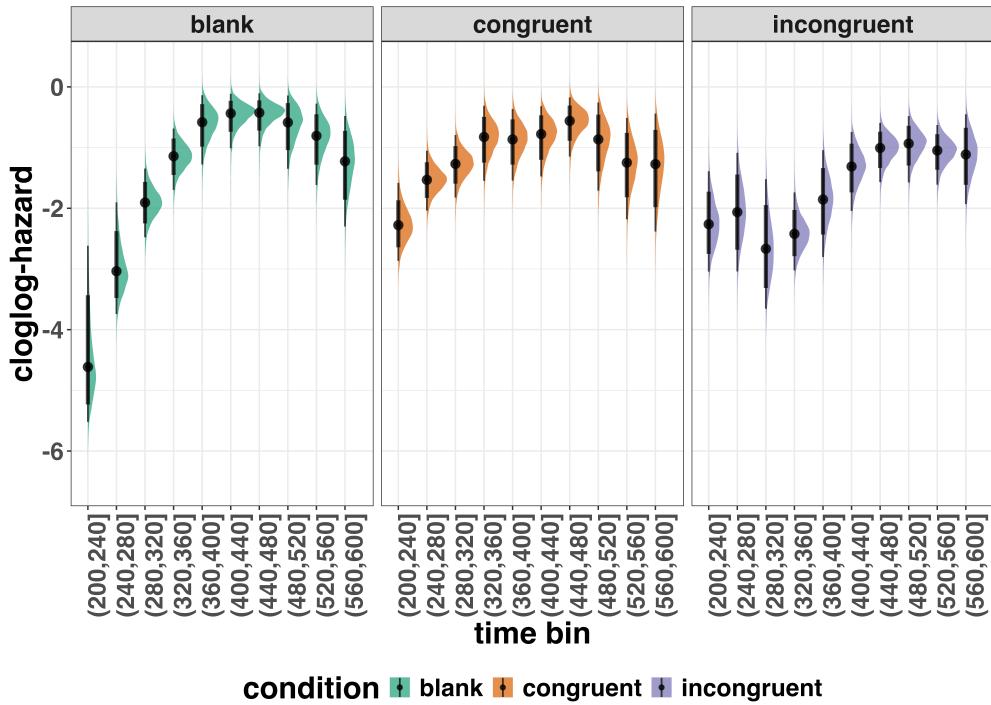


Figure 5. Medians and 80/95% credible intervals of the posterior distributions of the population-level parameters of model M1i.

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

interpretation by plotting the expected value of the posterior predictive distribution – the predicted hazard values – for the average participant (Figure 6), and for each participant in the data set (Figure 7).

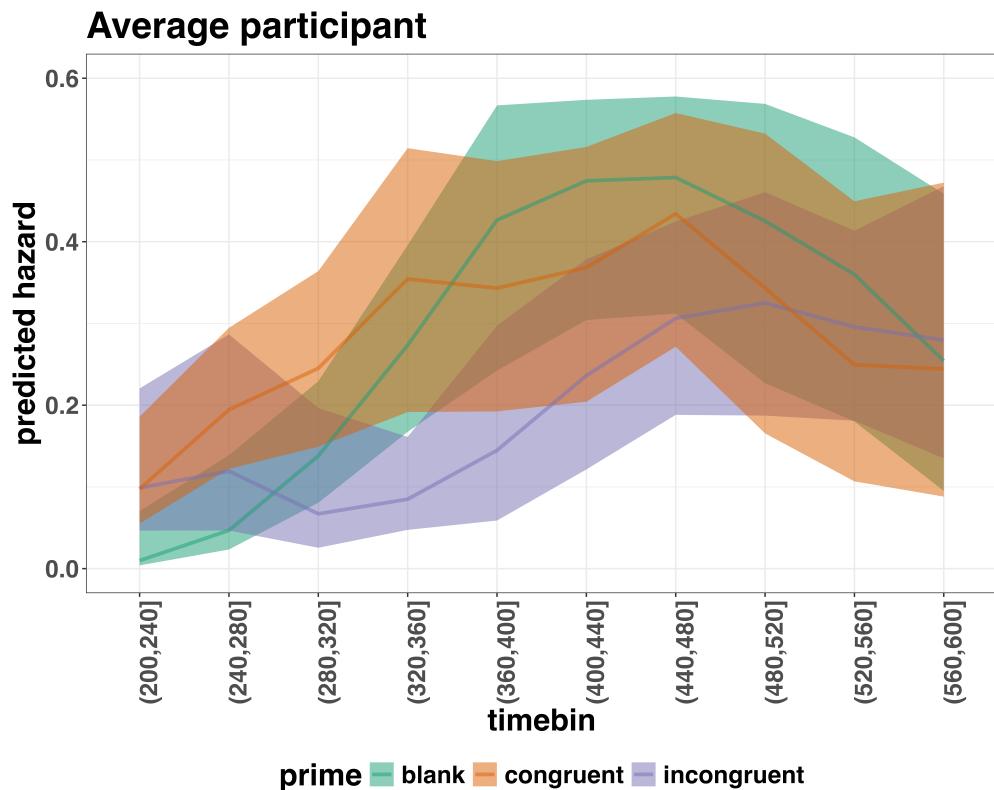


Figure 6. Point (median) and 80/95% credible interval summaries of the hazard estimates (expected values of the draws from the posterior predictive distributions) in each time bin for the average participant.

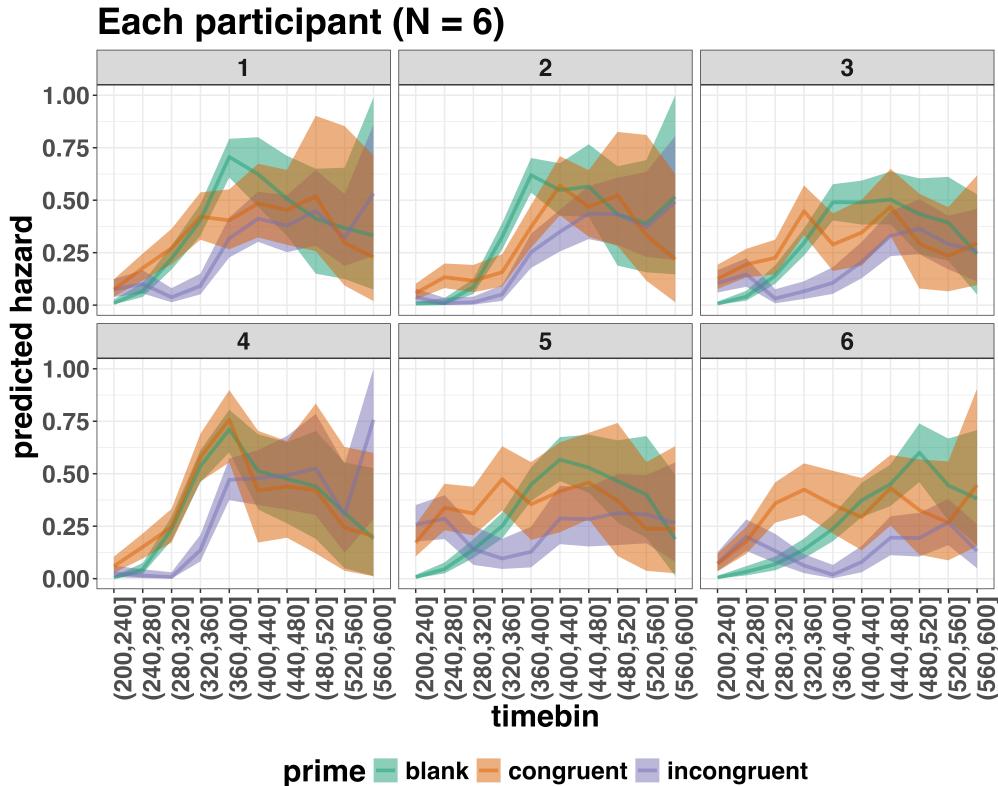


Figure 7. Point (median) and 80/95% credible interval summaries of the hazard estimates (expected values of the draws from the posterior predictive distributions) in each time bin for each participant.

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

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

547 incongruent-blank), and plot the posterior distributions of these contrast effects, both for

548 the average participant (Figure 8; grand average marginal effect) and for each participant

549 in the data set (Figure 9; subject-specific average marginal effect).

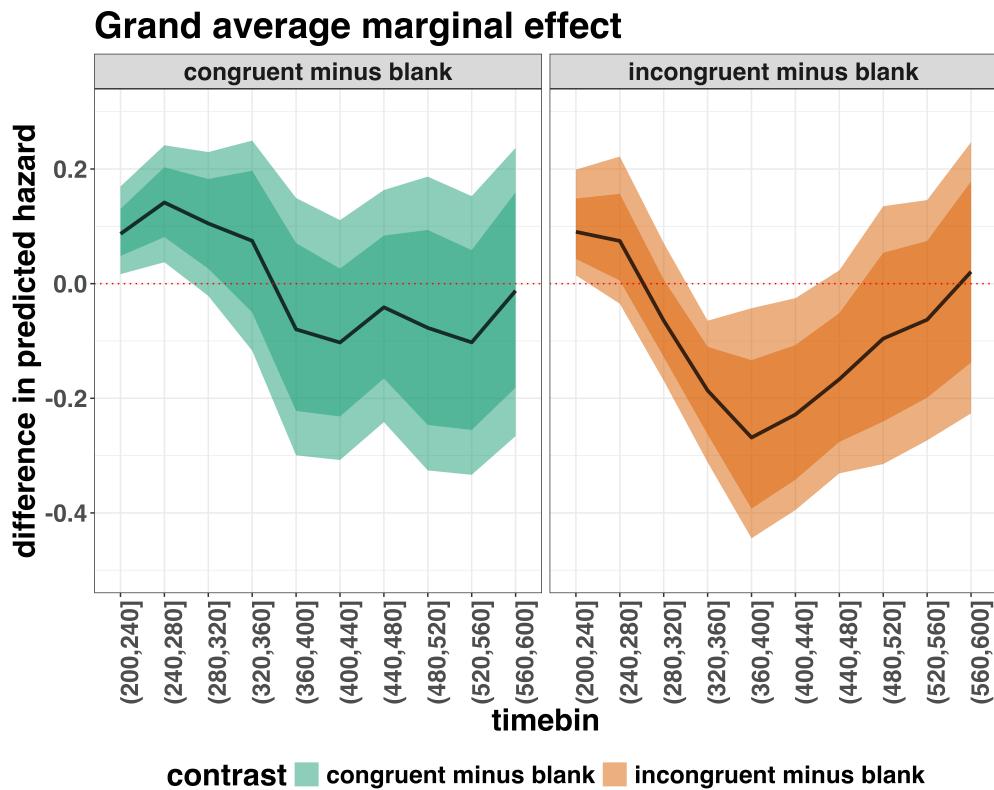


Figure 8. Point (mean) and 80/95% credible interval summaries of estimated differences in hazard in each time bin for the average participant.

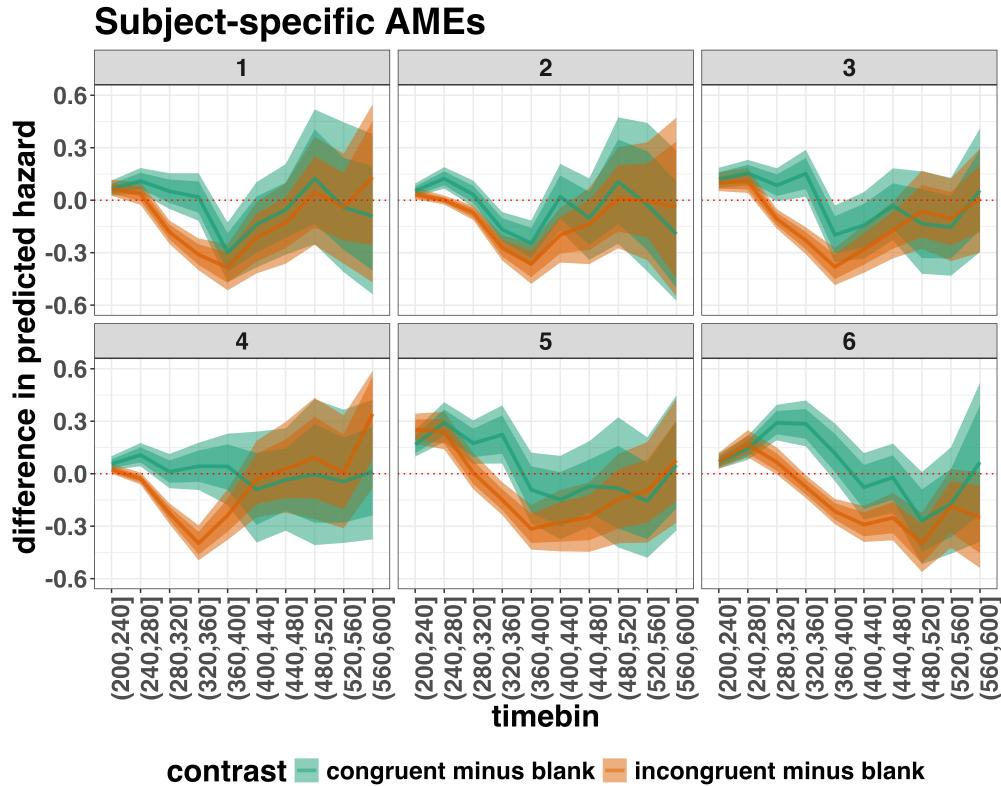


Figure 9. Point (mean) and 80/95% credible interval summaries of estimated differences in hazard in each time bin for each participant.

550 Table 4 shows the summaries of the estimated hazard differences for both contrasts in
 551 terms of a point estimate (the mean) and the upper and lower bounds of the 95% credible
 552 interval, for the average participant.

Table 4

*Point (mean) and 95% credible interval summary of
 estimated differences in hazard, for each time bin and
 contrast, in the average participant.*

contrast	timebin	diff	.lower	.upper
congruent minus blank	6	0.09	0.02	0.17

Table 4 continued

contrast	timebin	diff	.lower	.upper
congruent minus blank	7	0.14	0.04	0.25
congruent minus blank	8	0.11	-0.02	0.24
congruent minus blank	9	0.08	-0.12	0.27
congruent minus blank	10	-0.08	-0.30	0.15
congruent minus blank	11	-0.10	-0.31	0.11
congruent minus blank	12	-0.04	-0.24	0.17
congruent minus blank	13	-0.08	-0.33	0.20
congruent minus blank	14	-0.10	-0.33	0.15
congruent minus blank	15	-0.01	-0.27	0.27
incongruent minus blank	6	0.09	0.01	0.21
incongruent minus blank	7	0.08	-0.03	0.23
incongruent minus blank	8	-0.06	-0.17	0.07
incongruent minus blank	9	-0.19	-0.31	-0.06
incongruent minus blank	10	-0.27	-0.45	-0.04
incongruent minus blank	11	-0.23	-0.40	-0.03
incongruent minus blank	12	-0.17	-0.33	0.02
incongruent minus blank	13	-0.10	-0.31	0.14
incongruent minus blank	14	-0.06	-0.27	0.15
incongruent minus blank	15	0.03	-0.23	0.27

Note. diff = difference in predicted hazard.

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

555 our research question, i.e., the temporal dynamics of the effect of prime-target congruency
556 on RT? In other words, in which of the 40-ms time bins between 200 and 600 ms after
557 target onset does changing the prime from blank to congruent or incongruent affect the
558 hazard of response occurrence (for a prime-target SOA of 187 ms)?

559 If we want to study the average effect of prime type on hazard, uncontaminated by

560 inter-individual differences, we can base our conclusion on Figure 8 and Table 4. The
561 contrast “congruent minus blank” was estimated to be 0.09 hazard units in bin 6 (95% CrI
562 = [0.02, 0.17]), and 0.14 hazard units in bin 7 (95% CrI = [0.04, 0.25]). For the other bins,
563 the 95% credible interval contained zero. The contrast “incongruent minus blank” was
564 estimated to be 0.09 hazard units in bin 6 (95% CrI = [0.01, 0.21]), -0.19 hazard units in
565 bin 9 (95% CrI = [-0.31, -0.06]), -0.27 hazard units in bin 10 (95% CrI = [-0.45, -0.04]),
566 and -0.23 hazard units in bin 11 (95% CrI = [-0.40, -0.03]). For the other bins, the 95%
567 credible interval contained zero. Note that we could also have calculated hazard ratios
568 instead of hazard differences.

569 There are thus two phases of performance for the average person between 200 and

570 600 ms after target onset. In the first phase, the addition of a congruent or incongruent
571 prime stimulus increases the hazard of response occurrence compared to blank prime trials
572 in the time period (200, 240]. In the second phase, only the incongruent prime decreases
573 the hazard of response occurrence compared to blank primes, in the time period (320,440].
574 The sign of the effect of incongruent primes on the hazard of response occurrence thus
575 depends on how much waiting time has passed since target onset.

576 The posterior distribution of each contrast can also be summarized by considering its

577 proportion below or above some value, like zero. Table 5 shows the proportion of the
578 posterior distribution below or above zero, for each time bin and contrast.

Table 5

Summarizing the posterior distributions of each contrast by their proportions below and above zero.

timebin	contrast	prop_above	prop_below
6	congruent minus blank	0.99	0.01
7	congruent minus blank	0.99	0.01
8	congruent minus blank	0.95	0.05
9	congruent minus blank	0.79	0.21
10	congruent minus blank	0.24	0.76
11	congruent minus blank	0.15	0.85
12	congruent minus blank	0.33	0.67
13	congruent minus blank	0.28	0.72
14	congruent minus blank	0.19	0.81
15	congruent minus blank	0.47	0.53
6	incongruent minus blank	0.98	0.02
7	incongruent minus blank	0.92	0.08
8	incongruent minus blank	0.12	0.88
9	incongruent minus blank	0.00	1.00
10	incongruent minus blank	0.01	0.99
11	incongruent minus blank	0.02	0.98
12	incongruent minus blank	0.04	0.96
13	incongruent minus blank	0.20	0.80
14	incongruent minus blank	0.27	0.73
15	incongruent minus blank	0.58	0.42

Note. prop_below = proportion below zero: prop_above = proportion above zero.

579

580 Thus, the probability that the contrast “congruent minus blank” is larger than 0, is
581 larger than .9 in bins 6 to 8. And the probability that the contrast “incongruent minus
582 blank” is smaller than 0, is larger than .9 in bins 9 to 12.

583 If we want to focus more on inter-individual differences, we can study the
584 subject-specific differences in hazard in Figure 9. Note that three participants (1, 2, and 3)
585 show a negative difference for the contrast “congruent minus incongruent” in bin (360,400]
586 – subject 2 also in bin (320,360].

587 Future studies could (a) increase the number of participants to estimate the
588 proportion of “dippers” in the subject population, and/or (b) try to explain why this dip
589 occurs. For example, Panis and Schmidt (2016) concluded that active, top-down,
590 task-guided response inhibition effects emerge around 360 ms after the onset of the
591 stimulus following the prime (here: the target stimulus). Such a top-down inhibitory effect
592 might exist in our priming data set, because after some time participants will learn that the
593 first stimulus is not the one they have to respond to; To prevent a premature overt response
594 to the prime they thus might gradually increase a global response threshold during the
595 remainder of the experiment, which could result in a lower hazard in congruent trials
596 compared to blank trials, for bins after ~360 ms, and towards the end of the experiment.
597 This effect might be masked for incongruent primes by the response competition effect.

598 Interestingly, all subjects show a tendency in their mean difference (congruent minus
599 blank) to “dip” around that time (Figure 9). Therefore, future modeling efforts could
600 incorporate the trial number into the model formula, in order to also study how the effects
601 of prime type on hazard change on the long experiment-wide time scale, next to the short
602 trial-wide time scale. In Tutorial_2a.Rmd we provide a number of model formula that
603 should get you going.

604 **4.4 Tutorial 2b: Fitting Bayesian conditional accuracy models**

605 In this fourth tutorial, we illustrate how to fit a Bayesian multi-level regression model
 606 to the timed accuracy data from the masked response priming data set used in Tutorial 1a.
 607 The general process is similar to Tutorial 2a, except that (a) we use the person-trial data
 608 set, (b) we use the logit link function, and (c) we change the priors. For illustration
 609 purposes, we only fitted the effects of model M1i (see Tutorial 2a) in the conditional
 610 accuracy model called M1i_ca.

611 To make inferences from the parameter estimates in model M1i_ca, we first plot the
 612 densities of the draws from the posterior distributions of its population-level parameters in
 613 Figure 10, together with point (median) and interval estimates (80% and 95% credible
 614 intervals).

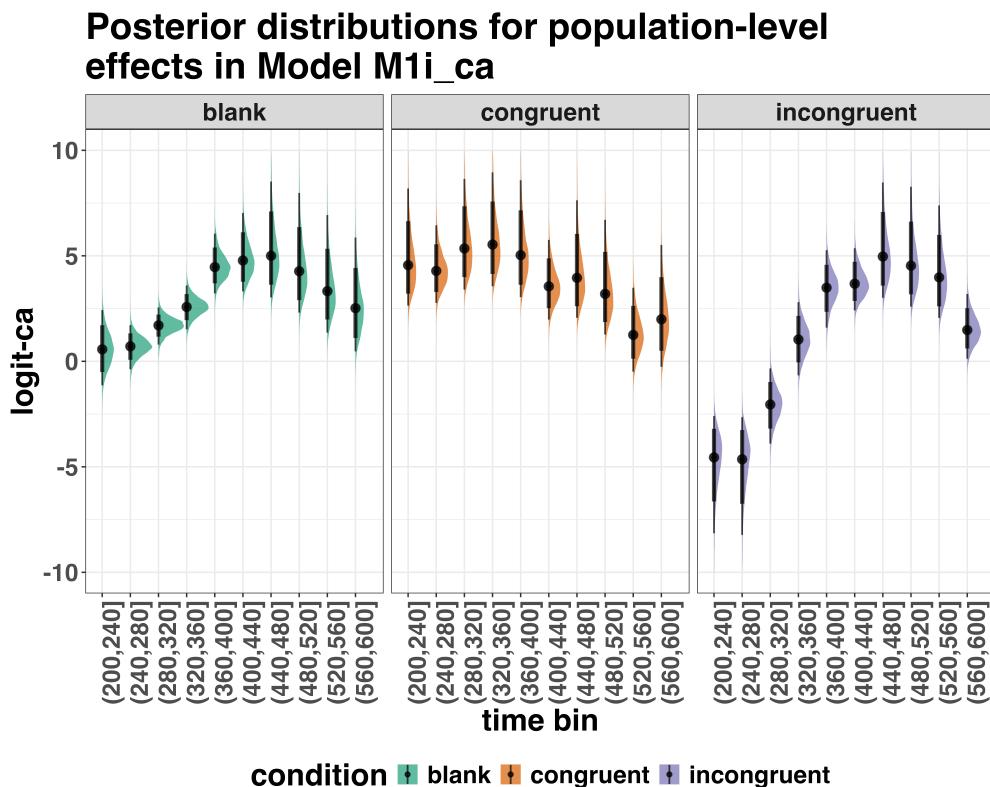


Figure 10. Medians and 80/95% credible intervals of the posterior distributions of the population-level parameters of model M1i_ca.

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 – for the average participant (Figure 11), and for each participant in the data set (Figure 12).

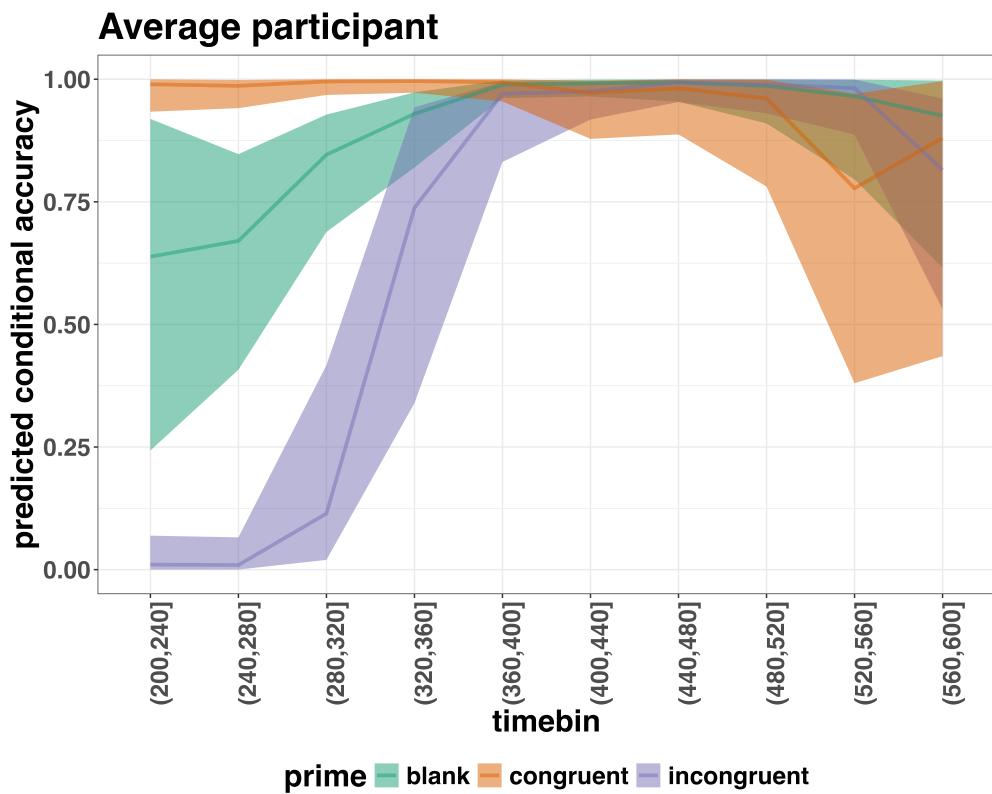


Figure 11. Point (median) and 80/95% credible interval summaries of the conditional accuracy estimates (expected values of the draws from the posterior predictive distributions) in each time bin for the average participant.

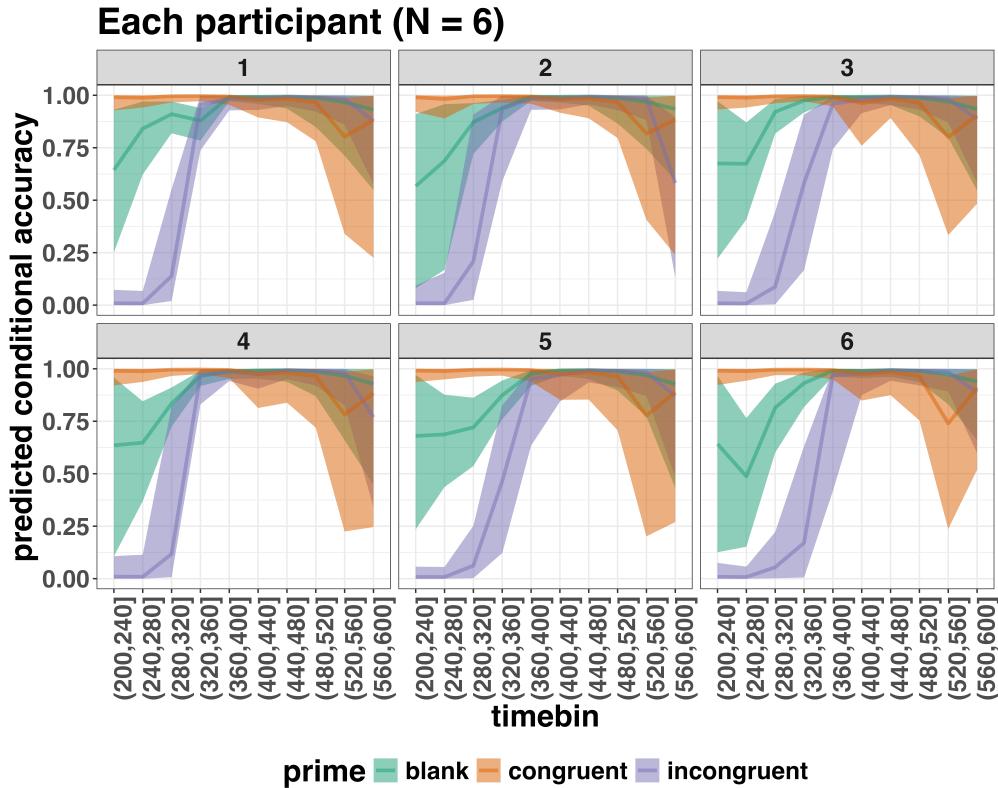


Figure 12. Point (median) and 80/95% credible interval summaries of the conditional accuracy estimates (expected values of the draws from the posterior predictive distributions) in each time bin for each participant.

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 for the average participant (Figure 13; grand average marginal effect).

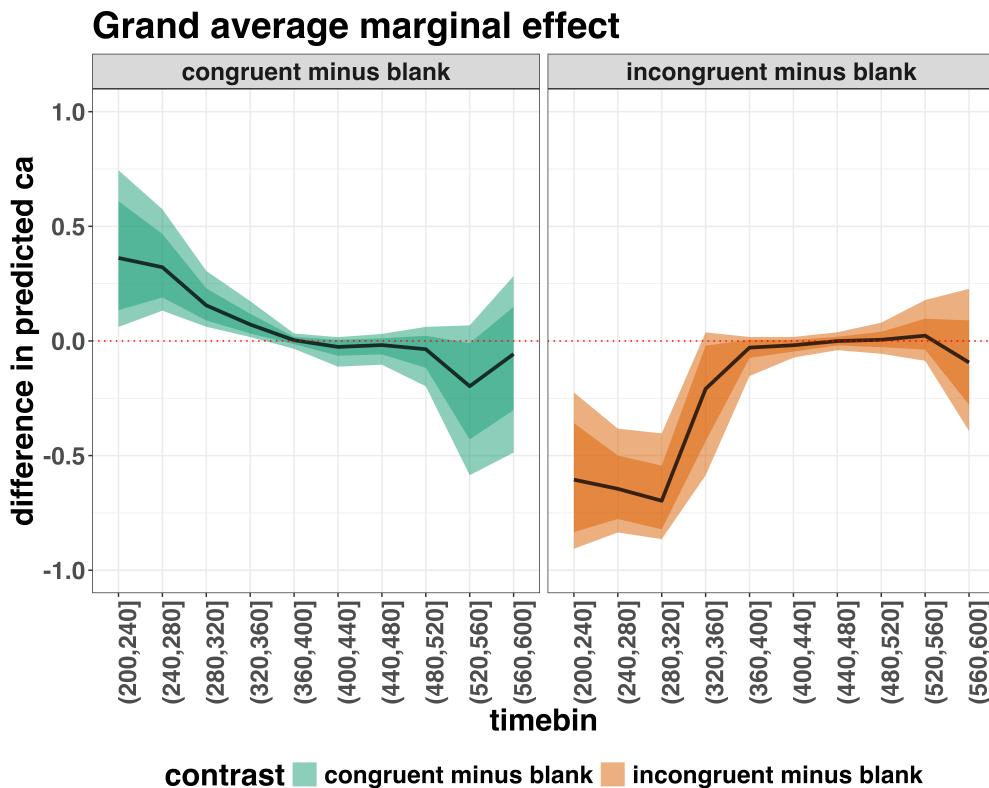


Figure 13. Point (mean) and 80/95% credible interval summaries of estimated differences in conditional accuracy in each time bin for the average participant.

623 Table 6 shows the summaries of the estimated differences in conditional accuracy for
 624 both contrasts in terms of a point estimate (the mean) and the upper and lower bounds of
 625 the 95% credible interval, for the average participant.

Table 6

Point (mean) and 95% credible interval summary of estimated differences in conditional accuracy, for each time bin and contrast, in the average participant.

contrast	timebin	diff_ca	.lower	.upper
congruent minus blank	6	0.36	0.06	0.74

Table 6 continued

contrast	timebin	diff_ca	.lower	.upper
congruent minus blank	7	0.32	0.13	0.57
congruent minus blank	8	0.16	0.06	0.31
congruent minus blank	9	0.07	0.02	0.17
congruent minus blank	10	0.00	-0.03	0.03
congruent minus blank	11	-0.03	-0.11	0.02
congruent minus blank	12	-0.02	-0.10	0.03
congruent minus blank	13	-0.04	-0.20	0.06
congruent minus blank	14	-0.20	-0.59	0.07
congruent minus blank	15	-0.06	-0.49	0.28
incongruent minus blank	6	-0.61	-0.91	-0.22
incongruent minus blank	7	-0.64	-0.84	-0.38
incongruent minus blank	8	-0.70	-0.86	-0.40
incongruent minus blank	9	-0.21	-0.59	0.04
incongruent minus blank	10	-0.03	-0.15	0.02
incongruent minus blank	11	-0.02	-0.07	0.02
incongruent minus blank	12	0.00	-0.04	0.04
incongruent minus blank	13	0.01	-0.06	0.08
incongruent minus blank	14	0.02	-0.09	0.18
incongruent minus blank	15	-0.09	-0.39	0.23

Note. diff = difference in predicted conditional accuracy.

628 on the conditional accuracy of emitted responses in time bins (200,240], (240,280], and
629 (280,320], relative to the estimates in the baseline condition (blank prime; red dashed lines
630 in Figure 14). Incongruent primes have a negative effect on the conditional accuracy of
631 emitted responses in those time bins, relative to the estimates in the baseline condition.

632 **4.5 Tutorial 3a: Fitting Frequentist hazard models**

633 In this fifth tutorial we illustrate how to fit a multilevel regression model to RT data
634 in the frequentist framework, for the data set used in Tutorial 1a. The general process is
635 similar to that in Tutorial 2a, except that there are no priors to set.

636 For illustration purposes, we only fitted the effects from model M1i (see Tutorial 2a)
637 using the function `glmer()` from the R package `lme4`. Alternatively, one could also use the
638 function `glmmPQL()` from the R package `MASS` (Ripley et al., 2024). The resulting hazard
639 model is called `M1i_f`.

640 In Figure 14 we compare the parameter estimates of model M1i from `brm()` with
641 those of model `M1i_f` from `glmer()`.

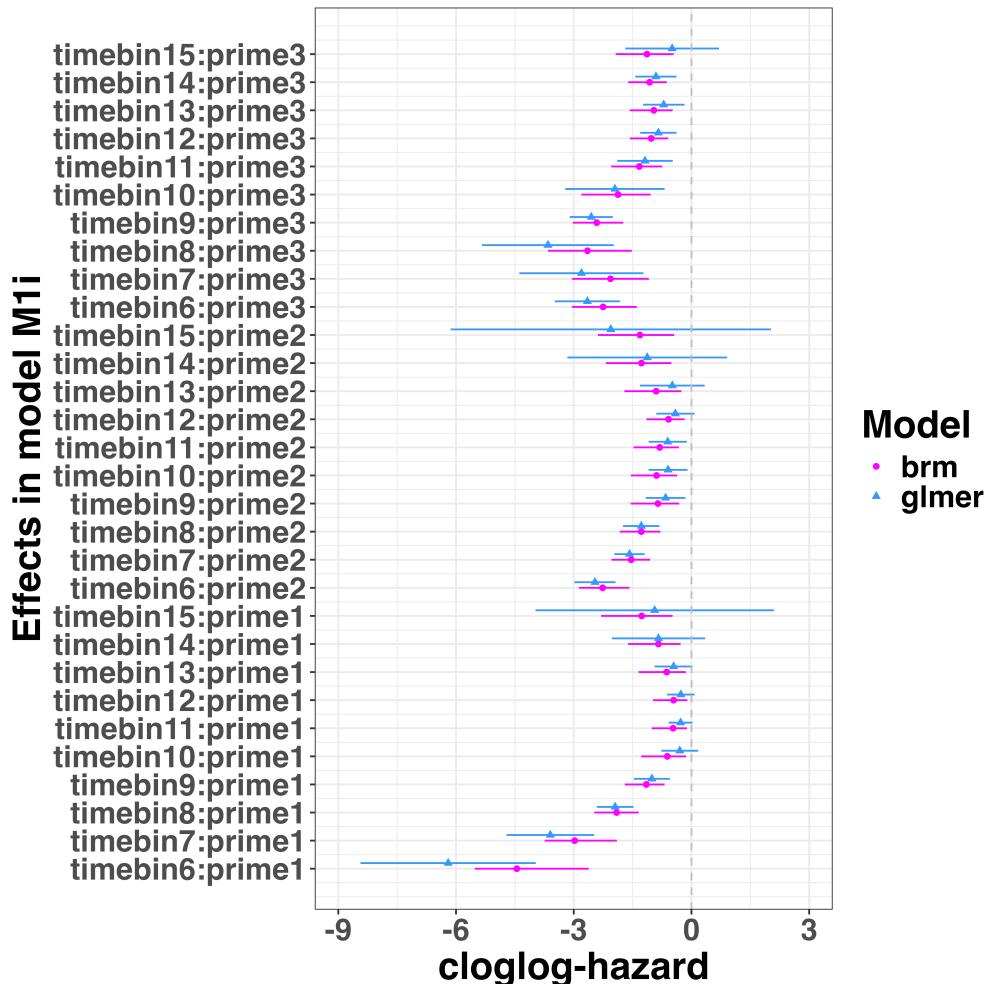


Figure 14. Parameter estimates for model M1i from `brm()` – means and 95% credible intervals – and model M1i_f from `glmer()` – maximum likelihood estimates and 95% confidence intervals.

642 Figure 14 confirms that the parameter estimates from both Bayesian and frequentist

643 models are pretty similar, which makes sense given the close similarity in model structure.

644 However, model M1i_f did not converge and resulted in a singular fit. This is of course one

645 of the reasons why Bayesian modeling has become so popular in recent years. But the price

646 you pay for being able to fit more complex random effects models in a Bayesian framework

647 is computation time. In other words, as we have noted throughout, some of the Bayesian

648 models in Tutorials 2a took several hours to build.

649 4.6 Tutorial 3b: Fitting Frequentist conditional accuracy models

650 In this sixth tutorial we illustrate how to fit a multilevel regression model to the
651 timed accuracy data in the frequentist framework, for the data set used in Tutorial 1a. For
652 illustration purposes, we only fitted the effects from model M1i_ca (see Tutorial 2b) using
653 the function glmer() from the R package lme4. Alternatively, one could also use the
654 function glmmPQL() from the R package MASS (Ripley et al., 2024). Again, the resulting
655 conditional accuracy model M1i_ca_f did not converge and resulted in a singular fit.

656 4.4 Tutorial 4: Planning

657 This needs adding... RR to do this.

658 5. Discussion

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

672 interpretability, individual differences, as well as limitations of the approach and future
673 extensions.

674 **5.1 What are the main use-cases of EHA for understanding cognition and brain
675 function?**

676 For those researchers, like ourselves, who are primarily interested in understanding
677 human cognitive and brain systems, we consider two broadly-defined, main use-cases of
678 EHA. First, as we hope to have made clear by this point, EHA is one way to investigating
679 a “temporal states” approach to cognitive processes. EHA provides one way to uncover
680 when cognitive states may start and stop, as well as what they may be tied to or interact
681 with. Therefore, if your research questions concern **when** and **for how long** psychological
682 states occur, our EHA tutorials could be useful tools for you to use.

683 Second, even if you are not primarily interested in studying the temporal states of
684 cognition, EHA could still be a useful tool to consider using, in order to qualify inferences
685 that are being made based on mean-average comparisons. Given that distinctly different
686 inferences can be made from the same data based on whether one computes a
687 mean-average across trials or a RT distribution of events (Figure 1), it may be important
688 for researchers to supplement mean-average comparisons with EHA. One could envisage
689 scenarios where the implicit assumption of an effect manifesting across all of the time bins
690 measured would not be supported by EHA. Therefore, the conclusion of interest would not
691 apply to all responses, but instead it would be restricted to certain aspects of time.

692 **5.2 Model complexity versus interpretability**

693 EHA can quickly become very complex when adding more than 1 time scale, due to
694 the many possible higher-order interactions. For example, model M4 contains two time
695 scales as covariates: the passage of time on the within-trial time scale, and the passage of

696 time on the across-trial (or within-experiment) time scale. However, when trials are
697 presented in blocks, and blocks of trials within sessions, and when the experiment
698 comprises three sessions, then four time scales can be defined (within-trial, within-block,
699 within-session, and within-experiment). From a theoretical perspective, adding more than
700 1 time scale – and their interactions – can be important to capture plasticity and other
701 learning effects that may play out on such longer time scales, and that are probably present
702 in each experiment in general. From a practical perspective, therefore, some choices need
703 to be made to balance the amount of data that is being collected per participant, condition
704 and across the varying timescales. As one example, if there are several timescales of
705 relevance, then it might be prudent for interpretational purposes to limit the number of
706 experimental predictor variables (conditions). This is of course where planning and data
707 simulation efforts would be important to provide a guide to experimental design choices.

708 5.3 Individual differences

709 One important issue is that of possible individual differences in the overall location of
710 the distribution, and the time course of psychological effects. For example, when you wait
711 for a response of the participant on each trial, you allow the participant to have control
712 over the trial duration, and some participants might respond only when they are confident
713 that their emitted response will be correct. These issues can be avoided by introducing a
714 (relatively short) response deadline in each trial, e.g., 600 ms for simple detection tasks,
715 1000 ms for more difficult discrimination tasks, or 2 s for tasks requiring extended high-level
716 processing. Because EHA can deal in a straightforward fashion with right-censored
717 observations (i.e., trials without an observed response), introducing a response deadline is
718 recommended when designing RT experiments. Furthermore, introducing a response
719 deadline and asking participants to respond before the deadline as much as possible, will
720 also lead to individual distributions that overlap in time, which is important when selecting
721 a common analysis time window when fitting hazard and conditional accuracy models.

722 But even when using a response deadline, participants can differ qualitatively in the
723 effects they display (see Panis, 2020). One way to deal with this is to describe and
724 interpret the different patterns. Another way is to run a clustering algorithm on the
725 individual hazard estimates across all conditions. The obtained dendrogram can then be
726 used to identify a (hopefully big) cluster of participants that behave similarly, and to
727 identify a (hopefully small) cluster of participants with different behavioral patterns. One
728 might then exclude the smaller sub-group of participants before fitting a hazard model or
729 consider the possibility that different cognitive processes may be at play during task
730 performance across the different sub-groups.

731 Another approach to deal with individual differences is Bayesian prevalence (Ince,
732 Paton, Kay, & Schyns, 2021), which is a from of Small-N approach (Smith & Little, 2018).
733 This method looks at effects within each individual in the study and asks how likely it
734 would be to see the same result if the experiment was repeated with a new person chosen
735 from the wider population at random. This approach allows one to quantify how typical or
736 uncommon an observed effect is in the population, and the uncertainty around this
737 estimate.

738 5.4 Limitations

739 Compared to the orthodox method – comparing mean-averages between conditions –
740 the most important limitation of multi-level hazard and conditional accuracy modeling is
741 that it might take a long time to estimate the parameters using Bayesian methods or the
742 model might have to be simplified significantly to use frequentist methods.

743 Another issue is that you need a relatively large number of trials per condition to
744 estimate the hazard function with high temporal resolution, which is required when testing
745 predictions of process models of cognition. Indeed, in general, there is a trade-off between
746 the number of trials per condition and the temporal resolution (i.e., bin width) of the

747 hazard function. Therefore, we recommend researchers to collect as many trials as possible
748 per experimental condition, given the available resources and considering the participant
749 experience (e.g., fatigue and boredom). For instance, if the maximum session length
750 deemed reasonable is between 1 and 2 hours, what is the maximum number of trials per
751 condition that you could reasonably collect? After consideration, it might be worth
752 conducting multiple testing sessions per participant and/or reducing the number of
753 experimental conditions. Finally, there is a user-friendly online tool for calculating
754 statistical power as a function of the number of trials as well as the number of participants,
755 and this might be worth consulting to guide the research design process (Baker et al., 2021).

756 We did not discuss continuous-time EHA, nor continuous-time SAT analysis. As
757 indicated by Allison (2010), learning discrete-time EHA methods first will help in learning
758 continuous-time methods. Given that RT is typically treated as a continuous variable, it is
759 possible that continuous-time methods will ultimately prevail. However, they require much
760 more data to estimate the continuous-time hazard (rate) function well. Thus, by trading a
761 bit of temporal resolution for a lower number of trials, discrete-time methods seem ideal for
762 dealing with typical psychological time-to-event data sets for which there are less than
763 ~200 trials per condition per experiment.

764 **5.5 Extensions**

765 The hazard models in this tutorial assume that there is one event of interest. For RT
766 data, this event constitutes a single transition between an “idle” state and a “responded”
767 state. However, in certain situations, more than one event of interest might exist. For
768 example, in a medical or health-related context, an individual might transition back and
769 forth between a “healthy” state and a “depressed” state, before being absorbed into a final
770 “death” state. When you have data on the timing of these transitions, one can apply
771 multi-state hazard models, which generalize EHA to transitions between three or more
772 states (Steele, Goldstein, & Browne, 2004). Also, the predictor variables in this tutorial are

773 time-invariant, i.e., their value did not change over the course of a trial. Thus, another
774 extension is to include time-varying predictors, i.e., predictors whose value can change
775 across the time bins within a trial (Allison, 2010). For example, when gaze position is
776 tracked during a visual search trial, the gaze-target distance will vary during a trial when
777 the eyes move around before a manual response is given; shorter gaze-target distances
778 should be associated with a higher hazard of response occurrence. Note that the effect of a
779 time-varying predictor (e.g., an occipital EEG signal) can itself vary over time.

780 **6. Conclusions**

781 Estimating the temporal distributions of RT and accuracy provide a rich source of
782 information on the time course of cognitive processing, which have been largely
783 undervalued in the history of experimental psychology and cognitive neuroscience.
784 Statistically controlling for the passage of time during data analysis is equally important as
785 experimental control during the design of an experiment, to better understand human
786 behavior in experimental paradigms. We hope that by providing a set of hands-on,
787 step-by-step tutorials, which come with custom-built and freely available code, researchers
788 will feel more comfortable embracing EHA and investigating the temporal profile of
789 cognitive states. On a broader level, we think that wider adoption of such approaches will
790 have a meaningful impact on the inferences drawn from data, as well as the development of
791 theories regarding the structure of cognition.

792

References

- 793 Allison, P. D. (1982). Discrete-Time Methods for the Analysis of Event Histories.
- 794 *Sociological Methodology*, 13, 61. <https://doi.org/10.2307/270718>
- 795 Allison, P. D. (2010). *Survival analysis using SAS: A practical guide* (2. ed). Cary, NC:
- 796 SAS Press.
- 797 Aust, F. (2019). *Citr: 'RStudio' add-in to insert markdown citations*. Retrieved from
- 798 <https://github.com/crsh/citr>
- 799 Aust, F., & Barth, M. (2023). *papaja: Prepare reproducible APA journal articles with R*
- 800 *Markdown*. Retrieved from <https://github.com/crsh/papaja>
- 801 Aust, F., & Barth, M. (2024). *papaja: Prepare reproducible APA journal articles with R*
- 802 *Markdown*. <https://doi.org/10.32614/CRAN.package.papaja>
- 803 Baker, D. H., Vilidaite, G., Lygo, F. A., Smith, A. K., Flack, T. R., Gouws, A. D., &
- 804 Andrews, T. J. (2021). Power contours: Optimising sample size and precision in
- 805 experimental psychology and human neuroscience. *Psychological Methods*, 26(3),
- 806 295–314. <https://doi.org/10.1037/met0000337>
- 807 Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for
- 808 confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*,
- 809 68(3), 10.1016/j.jml.2012.11.001. <https://doi.org/10.1016/j.jml.2012.11.001>
- 810 Barth, M. (2023). *tinylabes: Lightweight variable labels*. Retrieved from
- 811 <https://cran.r-project.org/package=tinylabes>
- 812 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects
- 813 models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- 814 <https://doi.org/10.18637/jss.v067.i01>
- 815 Bates, D., Maechler, M., & Jagan, M. (2024). *Matrix: Sparse and dense matrix classes and*
- 816 *methods*. Retrieved from <https://CRAN.R-project.org/package=Matrix>
- 817 Bengtsson, H. (2021). A unifying framework for parallel and distributed processing in r
- 818 using futures. *The R Journal*, 13(2), 208–227. <https://doi.org/10.32614/RJ-2021-048>

- 819 Blossfeld, H.-P., & Rohwer, G. (2002). *Techniques of event history modeling: New
820 approaches to causal analysis, 2nd ed* (pp. x, 310). Mahwah, NJ, US: Lawrence
821 Erlbaum Associates Publishers.
- 822 Box-Steffensmeier, J. M. (2004). Event history modeling: A guide for social scientists.
823 Cambridge: University Press.
- 824 Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan.
825 *Journal of Statistical Software*, 80(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>
- 826 Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms.
827 *The R Journal*, 10(1), 395–411. <https://doi.org/10.32614/RJ-2018-017>
- 828 Bürkner, P.-C. (2021). Bayesian item response modeling in R with brms and Stan. *Journal
829 of Statistical Software*, 100(5), 1–54. <https://doi.org/10.18637/jss.v100.i05>
- 830 Eddelbuettel, D., & Balamuta, J. J. (2018). Extending R with C++: A Brief Introduction
831 to Rcpp. *The American Statistician*, 72(1), 28–36.
832 <https://doi.org/10.1080/00031305.2017.1375990>
- 833 Eddelbuettel, D., & François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal
834 of Statistical Software*, 40(8), 1–18. <https://doi.org/10.18637/jss.v040.i08>
- 835 Gabry, J., Češnovar, R., Johnson, A., & Brodner, S. (2024). *Cmdstanr: R interface to
836 'CmdStan'*. Retrieved from <https://github.com/stan-dev/cmdstanr>
- 837 Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization
838 in bayesian workflow. *J. R. Stat. Soc. A*, 182, 389–402.
839 <https://doi.org/10.1111/rssa.12378>
- 840 Gelman, A., Hill, J., & Vehtari, A. (2020). Regression and Other Stories.
841 <https://www.cambridge.org/highereducation/books/regression-and-other-stories/DD20DD6C9057118581076E54E40C372C>; Cambridge University Press.
842 <https://doi.org/10.1017/9781139161879>
- 843 Girard, J. (2024). *Standist: What the package does (one line, title case)*. Retrieved from
844 <https://github.com/jmgirard/standist>

- 846 Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate.
- 847 *Journal of Statistical Software*, 40(3), 1–25. Retrieved from
848 <https://www.jstatsoft.org/v40/i03/>
- 849 Holden, J. G., Van Orden, G. C., & Turvey, M. T. (2009). Dispersion of response times
850 reveals cognitive dynamics. *Psychological Review*, 116(2), 318–342.
851 <https://doi.org/10.1037/a0014849>
- 852 Hosmer, D. W., Lemeshow, S., & May, S. (2011). *Applied Survival Analysis: Regression
853 Modeling of Time to Event Data* (2nd ed). Hoboken: John Wiley & Sons.
- 854 Ince, R. A., Paton, A. T., Kay, J. W., & Schyns, P. G. (2021). Bayesian inference of
855 population prevalence. *eLife*, 10, e62461. <https://doi.org/10.7554/eLife.62461>
- 856 Kantowitz, B. H., & Pachella, R. G. (2021). The Interpretation of Reaction Time in
857 Information-Processing Research 1. *Human Information Processing*, 41–82.
858 <https://doi.org/10.4324/9781003176688-2>
- 859 Kay, M. (2023). *tidybayes: Tidy data and geoms for Bayesian models*.
860 <https://doi.org/10.5281/zenodo.1308151>
- 861 Kelso, J. A. S., Dumas, G., & Tognoli, E. (2013). Outline of a general theory of behavior
862 and brain coordination. *Neural Networks: The Official Journal of the International
863 Neural Network Society*, 37, 120–131. <https://doi.org/10.1016/j.neunet.2012.09.003>
- 864 Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing,
865 estimation, meta-analysis, and power analysis from a Bayesian perspective.
866 *Psychonomic Bulletin & Review*, 25(1), 178–206.
867 <https://doi.org/10.3758/s13423-016-1221-4>
- 868 Kurz, A. S. (2023a). *Applied longitudinal data analysis in brms and the tidyverse* (version
869 0.0.3). Retrieved from <https://bookdown.org/content/4253/>
- 870 Kurz, A. S. (2023b). *Statistical rethinking with brms, ggplot2, and the tidyverse: Second
871 edition* (version 0.4.0). Retrieved from <https://bookdown.org/content/4857/>
- 872 Landes, J., Engelhardt, S. C., & Pelletier, F. (2020). An introduction to event history

- 873 analyses for ecologists. *Ecosphere*, 11(10), e03238. <https://doi.org/10.1002/ecs2.3238>
- 874 Luce, R. D. (1991). *Response times: Their role in inferring elementary mental organization*
875 (1. issued as paperback). Oxford: Univ. Press.
- 876 Makeham, William M. (1860). *On the Law of Mortality and the Construction of Annuity*
877 *Tables*. The Assurance Magazine, and Journal of the Institute of Actuaries.
- 878 McElreath, R. (2018). *Statistical Rethinking: A Bayesian Course with Examples in R and*
879 *Stan* (1st ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781315372495>
- 880 Meyer, D. E., Osman, A. M., Irwin, D. E., & Yantis, S. (1988). Modern mental
881 chronometry. *Biological Psychology*, 26(1-3), 3–67.
882 [https://doi.org/10.1016/0301-0511\(88\)90013-0](https://doi.org/10.1016/0301-0511(88)90013-0)
- 883 Müller, K., & Wickham, H. (2023). *Tibble: Simple data frames*. Retrieved from
884 <https://CRAN.R-project.org/package=tibble>
- 885 Neuwirth, E. (2022). *RColorBrewer: ColorBrewer palettes*. Retrieved from
886 <https://CRAN.R-project.org/package=RColorBrewer>
- 887 Panis, S. (2020). How can we learn what attention is? Response gating via multiple direct
888 routes kept in check by inhibitory control processes. *Open Psychology*, 2(1), 238–279.
889 <https://doi.org/10.1515/psych-2020-0107>
- 890 Panis, S., Moran, R., Wolkersdorfer, M. P., & Schmidt, T. (2020). Studying the dynamics
891 of visual search behavior using RT hazard and micro-level speed–accuracy tradeoff
892 functions: A role for recurrent object recognition and cognitive control processes.
893 *Attention, Perception, & Psychophysics*, 82(2), 689–714.
894 <https://doi.org/10.3758/s13414-019-01897-z>
- 895 Panis, S., Schmidt, F., Wolkersdorfer, M. P., & Schmidt, T. (2020). Analyzing Response
896 Times and Other Types of Time-to-Event Data Using Event History Analysis: A Tool
897 for Mental Chronometry and Cognitive Psychophysiology. *I-Perception*, 11(6),
898 2041669520978673. <https://doi.org/10.1177/2041669520978673>
- 899 Panis, S., & Schmidt, T. (2016). What Is Shaping RT and Accuracy Distributions? Active

- 900 and Selective Response Inhibition Causes the Negative Compatibility Effect. *Journal of*
901 *Cognitive Neuroscience*, 28(11), 1651–1671. https://doi.org/10.1162/jocn_a_00998
- 902 Panis, S., & Schmidt, T. (2022). When does “inhibition of return” occur in spatial cueing
903 tasks? Temporally disentangling multiple cue-triggered effects using response history
904 and conditional accuracy analyses. *Open Psychology*, 4(1), 84–114.
905 <https://doi.org/10.1515/psych-2022-0005>
- 906 Panis, S., Torfs, K., Gillebert, C. R., Wagemans, J., & Humphreys, G. W. (2017).
907 Neuropsychological evidence for the temporal dynamics of category-specific naming.
908 *Visual Cognition*, 25(1-3), 79–99. <https://doi.org/10.1080/13506285.2017.1330790>
- 909 Panis, S., & Wagemans, J. (2009). Time-course contingencies in perceptual organization
910 and identification of fragmented object outlines. *Journal of Experimental Psychology:
911 Human Perception and Performance*, 35(3), 661–687.
912 <https://doi.org/10.1037/a0013547>
- 913 Pedersen, T. L. (2024). *Patchwork: The composer of plots*. Retrieved from
914 <https://CRAN.R-project.org/package=patchwork>
- 915 Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in s and s-PLUS*. New York:
916 Springer. <https://doi.org/10.1007/b98882>
- 917 R Core Team. (2024). *R: A language and environment for statistical computing*. Vienna,
918 Austria: R Foundation for Statistical Computing. Retrieved from
919 <https://www.R-project.org/>
- 920 Ripley, B., Venables, B., Bates, D. M., ca 1998), K. H. (partial. port, ca 1998), A. G.
921 (partial. port, & polr), D. F. (support. functions for. (2024). *MASS: Support Functions
922 and Datasets for Venables and Ripley’s MASS*.
- 923 Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling
924 Change and Event Occurrence*. Oxford, New York: Oxford University Press.
- 925 Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-N design.
926 *Psychonomic Bulletin & Review*, 25(6), 2083–2101.

- 927 <https://doi.org/10.3758/s13423-018-1451-8>
- 928 Steele, F., Goldstein, H., & Browne, W. (2004). A general multilevel multistate competing
929 risks model for event history data, with an application to a study of contraceptive use
930 dynamics. *Statistical Modelling*, 4(2), 145–159.
931 <https://doi.org/10.1191/1471082X04st069oa>
- 932 Teachman, J. D. (1983). Analyzing social processes: Life tables and proportional hazards
933 models. *Social Science Research*, 12(3), 263–301.
934 [https://doi.org/10.1016/0049-089X\(83\)90015-7](https://doi.org/10.1016/0049-089X(83)90015-7)
- 935 Townsend, J. T. (1990). Truth and consequences of ordinal differences in statistical
936 distributions: Toward a theory of hierarchical inference. *Psychological Bulletin*, 108(3),
937 551–567. <https://doi.org/10.1037/0033-2909.108.3.551>
- 938 Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics.
939 *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)
- 940 Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New
941 York. Retrieved from <https://ggplot2.tidyverse.org>
- 942 Wickham, H. (2023a). *Forcats: Tools for working with categorical variables (factors)*.
943 Retrieved from <https://forcats.tidyverse.org/>
- 944 Wickham, H. (2023b). *Stringr: Simple, consistent wrappers for common string operations*.
945 Retrieved from <https://stringr.tidyverse.org>
- 946 Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., ...
947 Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43),
948 1686. <https://doi.org/10.21105/joss.01686>
- 949 Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). *R for data science: Import,
950 tidy, transform, visualize, and model data* (2nd edition). Beijing Boston Farnham
951 Sebastopol Tokyo: O'Reilly.
- 952 Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *Dplyr: A
953 grammar of data manipulation*. Retrieved from <https://dplyr.tidyverse.org>

- 954 Wickham, H., & Henry, L. (2023). *Purrr: Functional programming tools*. Retrieved from
955 <https://purrr.tidyverse.org/>
- 956 Wickham, H., Hester, J., & Bryan, J. (2024). *Readr: Read rectangular text data*. Retrieved
957 from <https://readr.tidyverse.org>
- 958 Wickham, H., Vaughan, D., & Girlich, M. (2024). *Tidyr: Tidy messy data*. Retrieved from
959 <https://tidyr.tidyverse.org>
- 960 Winter, B. (2019). *Statistics for Linguists: An Introduction Using R*. New York:
961 Routledge. <https://doi.org/10.4324/9781315165547>
- 962 Wolkersdorfer, M. P., Panis, S., & Schmidt, T. (2020). Temporal dynamics of sequential
963 motor activation in a dual-prime paradigm: Insights from conditional accuracy and
964 hazard functions. *Attention, Perception, & Psychophysics*, 82(5), 2581–2602.
965 <https://doi.org/10.3758/s13414-020-02010-5>

Supplementary material

967 A. Definitions of discrete-time hazard, survivor, probability mass, and
968 conditional accuracy functions

The shape of a distribution of waiting times can be described in multiple ways (Luce, 1991). After dividing time in discrete, contiguous time bins indexed by t , let RT be a discrete random variable denoting the rank of the time bin in which a particular person's response occurs in a particular experimental condition. Because waiting times can only increase, discrete-time EHA focuses on the discrete-time hazard function

$$974 \quad h(t) = P(RT = t | RT \geq t) \quad (1)$$

and the discrete-time survivor function

$$S(t) = P(RT > t) = [1-h(t)][1-h(t-1)][1-h(t-2)] \dots [1-h(1)] \quad (2)$$

⁹⁷⁷ and not on the probability mass function

$$P(t) = P(RT = t) = h(t).S(t-1) \quad (3)$$

979 nor the cumulative distribution function

$$F(t) = P(RT < t) = 1 - S(t) \quad (4)$$

The discrete-time hazard function of event occurrence gives you for each bin the probability that the event occurs (sometime) in that bin, given that the event has not occurred yet in previous bins. This conditionality in the definition of hazard is what makes the hazard function so diagnostic for studying event occurrence, as an event can physically not occur when it has already occurred before. While the discrete-time hazard function assesses the unique risk of event occurrence associated with each time bin, the discrete-time survivor function cumulates the bin-by-bin risks of event *non*occurrence to obtain the probability that the event occurs after bin t. The probability mass function cumulates the risk of event occurrence in bin t with the risks of event nonoccurrence in

990 bins 1 to $t-1$. From equation 3 we find that hazard in bin t is equal to $P(t)/S(t-1)$.

991 For two-choice RT data, the discrete-time hazard function can be extended with the
 992 discrete-time conditional accuracy function

993 $ca(t) = P(\text{correct} \mid RT = t)$ (5)

994 which gives you for each bin the probability that a response is correct given that it is
 995 emitted in time bin t (Allison, 2010; Kantowitz & Pachella, 2021; Wickelgren, 1977). This
 996 latter function is also known as the micro-level speed-accuracy tradeoff (SAT) function.

997 The survivor function provides a context for the hazard function, as $S(t-1) = P(RT >$
 998 $t-1) = P(RT \geq t)$ tells you on how many percent of the trials the estimate $h(t) = P(RT =$
 999 $t \mid RT \geq t)$ is based. The probability mass function provides a context for the conditional
 1000 accuracy function, as $P(t) = P(RT = t)$ tells you on how many percent of the trials the
 1001 estimate $ca(t) = P(\text{correct} \mid RT = t)$ is based.

1002 While psychological RT data is typically measured in small, continuous units (e.g.,
 1003 milliseconds), discrete-time EHA treats the RT data as interval-censored data, because it
 1004 only uses the information that the response occurred sometime in a particular bin of time
 1005 $(x,y]: x < RT \leq y$. If we want to use the exact event times, then we treat time as a
 1006 continuous variable, and let RT be a continuous random variable denoting a particular
 1007 person's response time in a particular experimental condition. Continuous-time EHA does
 1008 not focus on the cumulative distribution function $F(t) = P(RT \leq t)$ and its derivative, the
 1009 probability density function $f(t) = F(t)'$, but on the survivor function $S(t) = P(RT > t)$
 1010 and the hazard rate function $\lambda(t) = f(t)/S(t)$. The hazard rate function gives you the
 1011 instantaneous *rate* of event occurrence at time point t , given that the event has not
 1012 occurred yet.

1013 **B. Custom functions for descriptive discrete-time hazard analysis**

1014 We defined 13 custom functions that we list here.

- censor(df,timeout,bin_width) : divide the time segment $(0, \text{timeout}]$ in bins, identify any right-censored observations, and determine the discrete RT (time bin rank)
- ptb(df) : transform the person-trial data set to the person-trial-bin data set
- setup_lt(ptb) : set up a life table for each level of 1 independent variable
- setup_lt_2IV(ptb) : set up a life table for each combination of levels of 2 independent variables
- calc_ca(df) : estimate the conditional accuracies when there is 1 independent variable
- calc_ca_2IV(df) : estimate the conditional accuracies when there are 2 independent variables
- join_lt_ca(df1,df2) : add the $\text{ca}(t)$ estimates to the life tables (1 independent variable)
- join_lt_ca_2IV(df1,df2) : add the $\text{ca}(t)$ estimates to the life tables (2 independent variables)
- extract_median(df) : estimate quantiles $S(t)._{50}$ (1 independent variable)
- extract_median_2IV(df) : estimate quantiles $S(t)._{50}$ (2 independent variables)
- plot_oha(df,subj,haz_yaxis) : create plots of the discrete-time functions (1 independent variable)
- plot_oha_2IV(df,subj,haz_yaxis) : create plots of the discrete-time functions (2 independent variables)
- plot_oha_agg(df,subj,haz_yaxis) : create 1 plot for data aggregated across participants (1 independent variable)

When you want to analyse simple RT data from a detection experiment with one independent variable, the functions `calc_ca()` and `join_lt_ca()` should not be used, and the code to plot the conditional accuracy functions should be removed from the function `plot_oha()`. When you want to analyse simple RT data from a detection experiment with two independent variables, the functions `calc_ca_2IV()` and `join_lt_ca_2IV()` should not be used, and the code to plot the conditional accuracy functions should be removed from

₁₀₄₂ the function `plot_eha_2IV()`.

₁₀₄₃ C. Link functions

₁₀₄₄ Popular link functions include the logit link and the complementary log-log link, as
₁₀₄₅ shown in Figure 11.

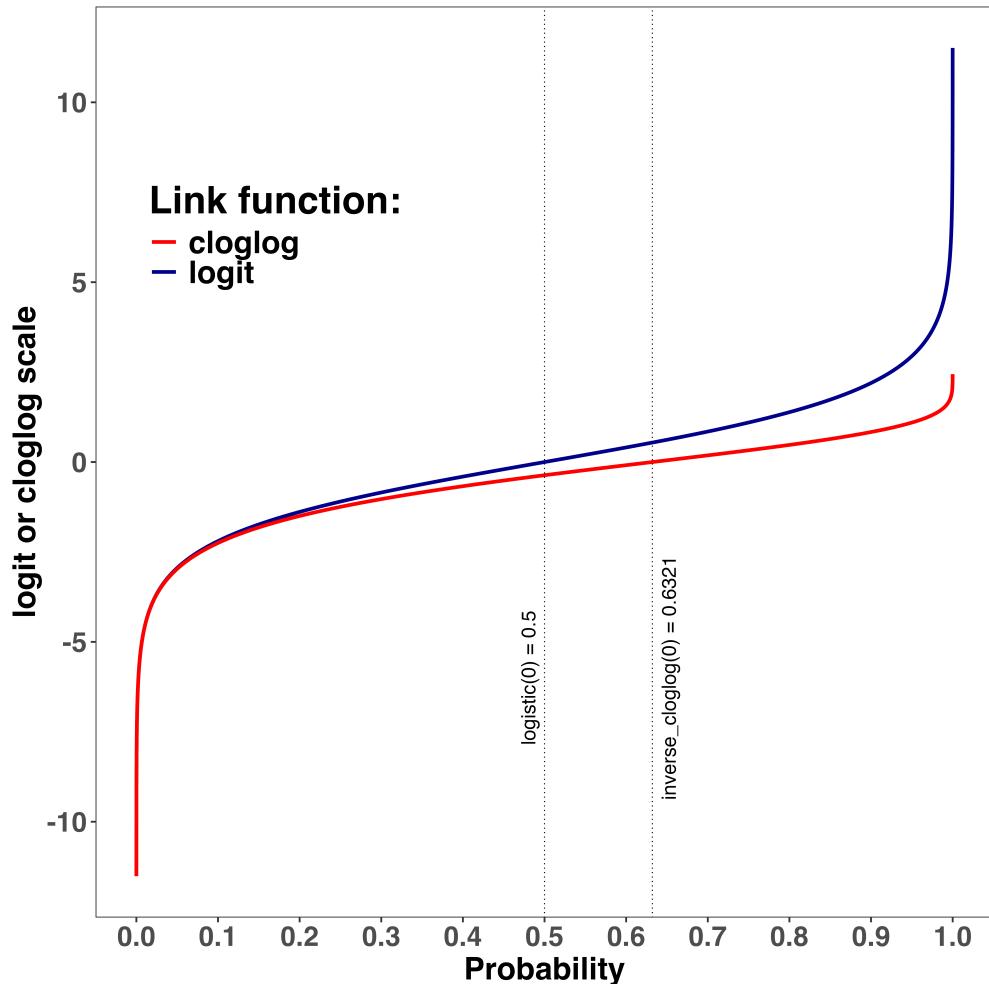


Figure 15. The logit and cloglog link functions.

₁₀₄₆ D. Regression equations

₁₀₄₇ An example (single-level) discrete-time hazard model with three predictors (TIME,
₁₀₄₈ X₁, X₂), the cloglog link function, and a third-order polynomial specification for TIME can

1049 be written as follows:

$$\begin{aligned} \text{cloglog}[h(t)] &= \ln(-\ln[1-h(t)]) = [\beta_0 \text{ONE} + \beta_1(\text{TIME}-9) + \beta_2(\text{TIME}-9)^2] + [\beta_3 X_1 + \beta_4 X_2 \\ &\quad + \beta_5 X_2(\text{TIME}-9)] \end{aligned} \quad (6)$$

1052 The main predictor variable TIME is the time bin index t that is centered on value 9
1053 in this example. The first set of terms within brackets, the parameters β_0 to β_2 multiplied
1054 by their polynomial specifications of (centered) time, represents the shape of the baseline
1055 cloglog-hazard function (i.e., when all predictors X_i take on a value of zero). The second
1056 set of terms (the beta parameters β_3 to β_5) represents the vertical shift in the baseline
1057 cloglog-hazard for a 1 unit increase in the respective predictor variable. Predictors can be
1058 discrete, continuous, and time-varying or time-invariant. For example, the effect of a 1 unit
1059 increase in X_1 is to vertically shift the whole baseline cloglog-hazard function by β_1
1060 cloglog-hazard units. However, if the predictor interacts linearly with TIME (see X_2 in the
1061 example), then the effect of a 1 unit increase in X_2 is to vertically shift the predicted
1062 cloglog-hazard in bin 9 by β_2 cloglog-hazard units (when $\text{TIME}-9 = 0$), in bin 10 by $\beta_2 + \beta_3$ cloglog-hazard units (when $\text{TIME}-9 = 1$), and so forth. To interpret the effects of a
1064 predictor, its β parameter is exponentiated, resulting in a hazard ratio (due to the use of
1065 the cloglog link). When using the logit link, exponentiating a β parameter results in an
1066 odds ratio.

1067 An example (single-level) discrete-time hazard model with a general specification for
1068 TIME (separate intercepts for each of six bins, where D1 to D6 are binary variables
1069 identifying each bin) and a single predictor (X_1) can be written as follows:

$$\text{cloglog}[h(t)] = [\beta_0 D1 + \beta_1 D2 + \beta_2 D3 + \beta_3 D4 + \beta_4 D5 + \beta_5 D6] + [\beta_6 X_1] \quad (7)$$

1071 E. Prior distributions

1072 To gain a sense of what prior *logit* values would approximate a uniform distribution
1073 on the probability scale, Kurz (2023a) simulated a large number of draws from the

1074 Uniform(0,1) distribution, converted those draws to the log-odds metric, and fitted a
 1075 Student's t distribution. Row C in Figure 12 shows that using a t-distribution with 7.61
 1076 degrees of freedom and a scale parameter of 1.57 as a prior on the logit scale, approximates
 1077 a uniform distribution on the probability scale. According to Kurz (2023a), such a prior
 1078 might be a good prior for the intercept(s) in a logit-hazard model, while the N(0,1) prior in
 1079 row D might be a good prior for the non-intercept parameters in a logit-hazard model, as it
 1080 gently regularizes p towards .5 (i.e., a zero effect on the logit scale).

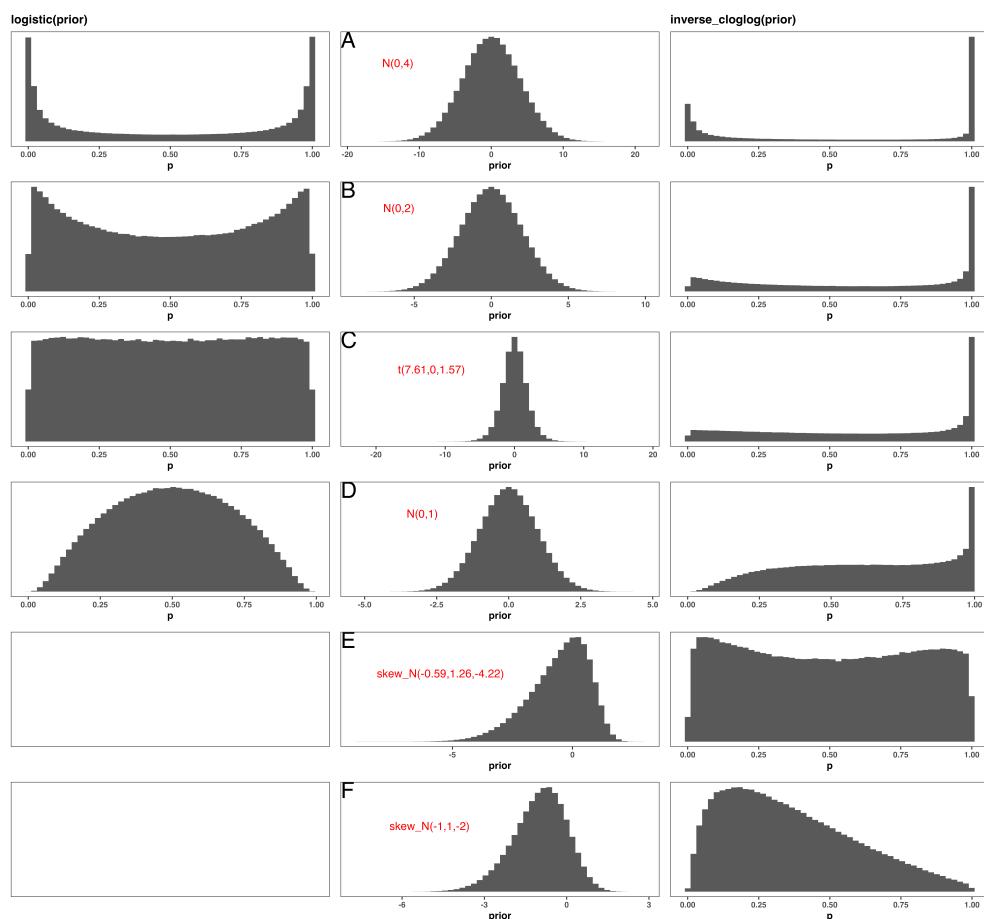


Figure 16. Prior distributions on the logit and/or cloglog scales (middle column), and their implications on the probability scale after applying the inverse-logit (or logistic) transformation (left column), and the inverse-cloglog transformation (right column).

1081 To gain a sense of what prior *cloglog* values would approximate a uniform distribution

on the hazard probability scale, we followed Kurz's approach and simulated a large number of draws from the Uniform(0,1) distribution, converted them to the cloglog metric, and fitted a skew-normal model (due to the asymmetry of the cloglog link function). Row E shows that using a skew-normal distribution with a mean of -0.59, a standard deviation of 1.26, and a skewness of -4.22 as a prior on the cloglog scale, approximates a uniform distribution on the probability scale. However, because hazard values below .5 are more likely in RT studies, using a skew-normal distribution with a mean of -1, a standard deviation of 1, and a skewness of -2 as a prior on the cloglog scale (row F), might be a good weakly informative prior for the intercept(s) in a cloglog-hazard model. A skew-normal distribution with a mean of -0.2, a standard deviation of 0.71, and a skewness of -2.2 might be a good weakly informative prior for the non-intercept parameters in a cloglog-hazard model as it gently regularizes p towards .6321 (i.e., a zero effect on the cloglog scale).

1094 F. Advantages of hazard analysis

1095 Statisticians and mathematical psychologists recommend focusing on the hazard
1096 function when analyzing time-to-event data for various reasons. First, as discussed by
1097 Holden, Van Orden, and Turvey (2009), "probability density [and mass] functions can
1098 appear nearly identical, both statistically and to the naked eye, and yet are clearly different
1099 on the basis of their hazard functions (but not vice versa). Hazard functions are thus more
1100 diagnostic than density functions" (p. 331) when one is interested in studying the detailed
1101 shape of a RT distribution (see also Figure 1 in Panis, Schmidt, et al., 2020). Therefore,
1102 when the goal is to study how psychological effects change over time, hazard and
1103 conditional accuracy functions are preferred.

1104 Second, because RT distributions may differ from one another in multiple ways,
1105 Townsend (1990) developed a dominance hierarchy of statistical differences between two
1106 arbitrary distributions A and B. For example, if $h_A(t) > h_B(t)$ for all t, then both hazard
1107 functions are said to show a complete ordering. Townsend (1990) concluded that stronger

conclusions can be drawn from data when comparing the hazard functions using EHA. For example, when mean A < mean B, the hazard functions might show a complete ordering (i.e., for all t), a partial ordering (e.g., only for $t > 300$ ms, or only for $t < 500$ ms), or they may cross each other one or more times.

Third, EHA does not discard right-censored observations when estimating hazard functions, that is, trials for which we do not observe a response during the data collection period in a trial so that we only know that the RT must be larger than some value (e.g., the response deadline). This is important because although a few right-censored observations are inevitable in most RT tasks, a lot of right-censored observations are expected in experiments on masking, the attentional blink, and so forth. In other words, by using EHA you can analyze RT data from experiments that typically do not measure response times. As a result, EHA can also deal with long RTs in experiments without a response deadline, which are typically treated as outliers and are discarded before calculating a mean. This orthodox procedure leads to underestimation of the true mean. By introducing a fixed censoring time for all trials at the end of the analysis time window, trials with long RTs are not discarded but contribute to the risk set of each bin.

Fourth, hazard modeling allows incorporating time-varying explanatory covariates such as heart rate, electroencephalogram (EEG) signal amplitude, gaze location, etc. (Allison, 2010). This is useful for linking physiological effects to behavioral effects when performing cognitive psychophysiology (Meyer et al., 1988).

Finally, as explained by Kelso, Dumas, and Tognoli (2013), it is crucial to first have a precise description of the macroscopic behavior of a system (here: $h(t)$ and possibly $ca(t)$ functions) in order to know what to derive on the microscopic level. EHA can thus solve the problem of model mimicry, i.e., the fact that different computational models can often predict the same mean RTs as observed in the empirical data, but not necessarily the detailed shapes of the empirical RT hazard distributions. Also, fitting parametric functions

₁₁₃₄ or computational models to data without studying the shape of the empirical discrete-time
₁₁₃₅ $h(t)$ and $ca(t)$ functions can miss important features in the data (Panis, Moran, et al.,
₁₁₃₆ 2020; Panis & Schmidt, 2016).