

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, Gloriustrasse 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. Moreover, 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 timeline 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 multi-level 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, top row). 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, bottom row). Therefore, a comparison of
62 means would lead to a similar conclusion in examples 1-3, as well as examples 4-6, whereas
63 a comparison of the distributions would lead to a different conclusion in every case.

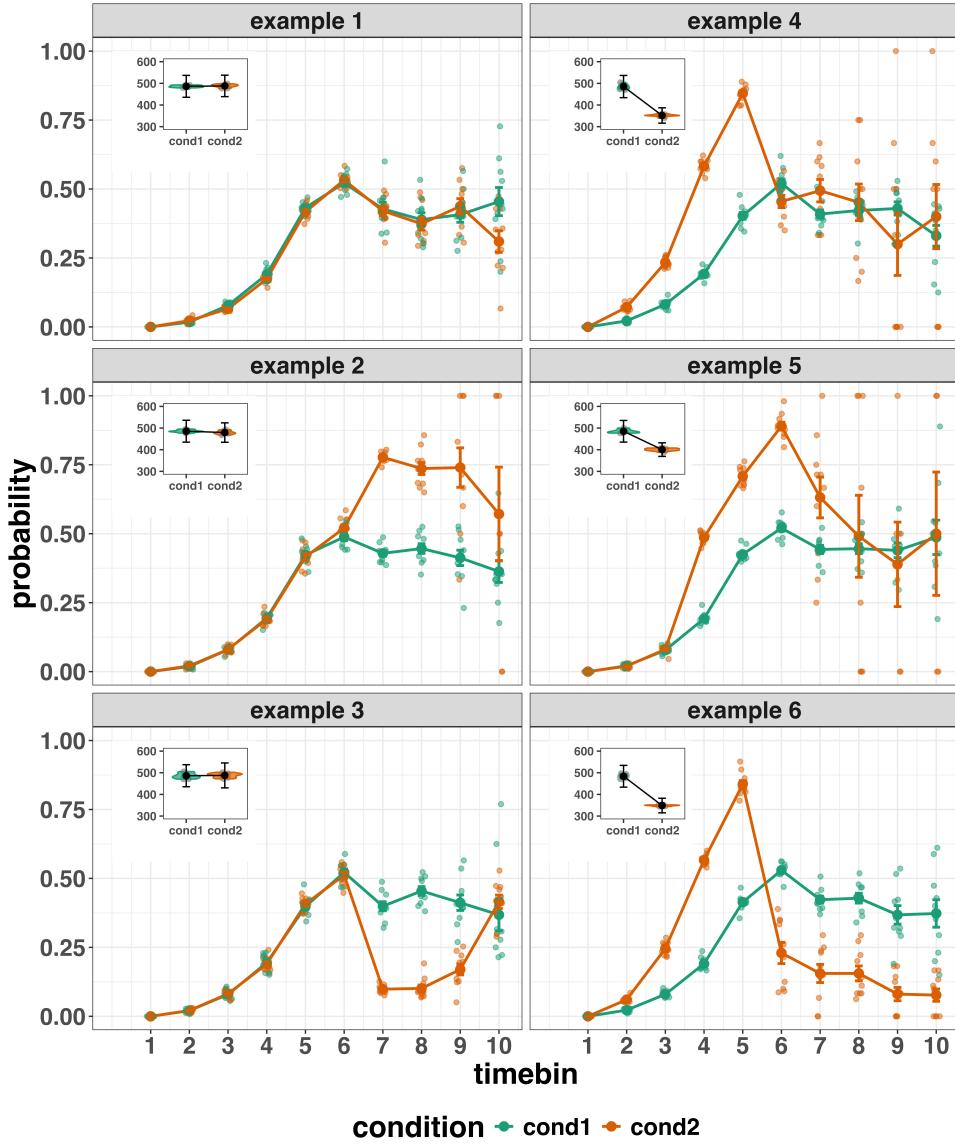


Figure 1. Means versus distributional shapes for six different simulated data set examples.

The first second after stimulus onset is divided in ten bins of 100 ms. 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.

64 Why does this matter for research in psychology? Compared to the aggregation of
 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 around 200 ms and is gone by
70 600 ms. In contrast, in Example 5, the effect starts around 400 ms and is gone by 800 ms.
71 And in the Example 6, the effect reverses between 500 and 600 ms. What kind of theory or
72 theories could account for such effects? Are there new auxiliary assumptions that theories
73 need to adopt? And are there new experiments that need to be performed to test the novel
74 predictions that follow from these analyses? As we show later using published examples,
75 for many psychological questions, such “temporal states” information can be theoretically
76 meaningful by leading to more fine-grained understanding of psychological processes, as
77 well as adding a relatively under-used dimension – the passage of time – to the theory
78 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

118 regression models to *timed* accuracy data to perform a micro-level speed-accuracy tradeoff
119 (SAT) analysis, which complements the EHA of RT data for choice RT data. In Tutorial
120 3a, we illustrate how to fit the same type of multilevel regression model for RT data in a
121 frequentist framework using the R package lme4. We then briefly compare and contrast
122 these inferential frameworks when applied to EHA. In Tutorial 3b, we illustrate how to
123 perform the SAT analysis in a frequentist framework.

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

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

139 2. A brief introduction to event history analysis

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

143 my comments in the notes file for relevant references here). Our focus here is not on
144 providing a detailed account of the underlying regression equations, since this topic has
145 been comprehensively covered many times before. Instead, we want to provide an intuition
146 regarding how EHA works in general, as well as in the context of experimental psychology.
147 As such, we only supply 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).

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 probability
¹⁷⁰ that the event occurs after bin t. In other words, the survivor function gives you for each
¹⁷¹ time bin the likelihood that the event occurs in the future, i.e., in one of the subsequent
¹⁷² timebins.

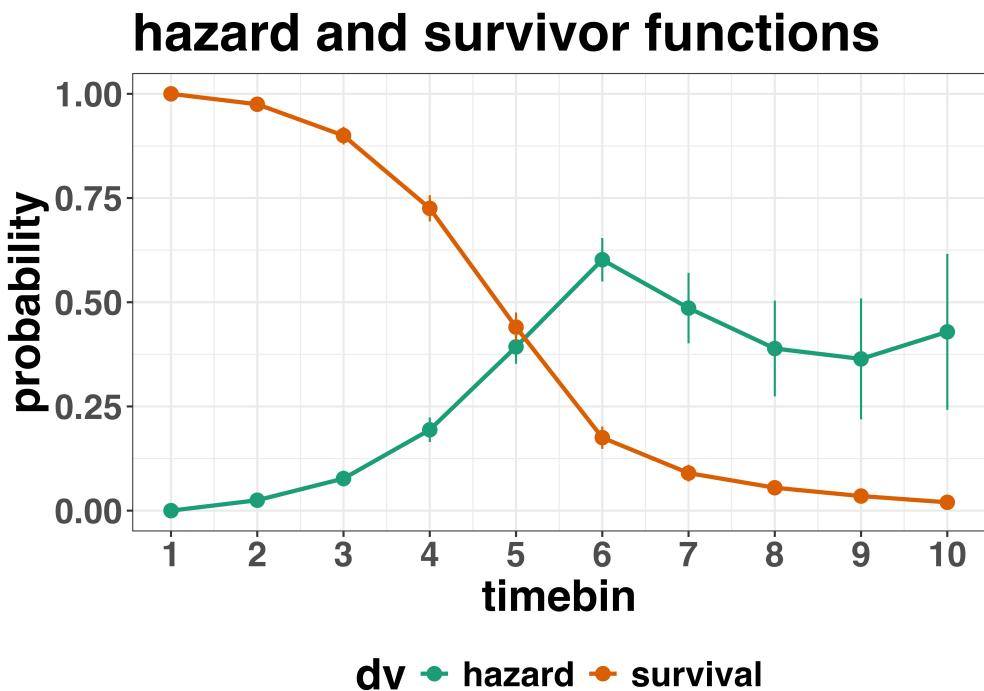


Figure 2. Discrete-time hazard and survivor functions. While the hazard function is the vehicle for inferring the time course of cognitive processes, the survivor function at $S(t-1)$ can help to qualify or provide context to the interpretation of the hazard function at $h(t)$. For example, the high hazard of $h(t=5) = .xx$ is experienced only by $S(t-1=4) = .yy$ percent of the trials. Because the survivor function is a decreasing function of time, the error bars in later parts of the hazard function will typically 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.

2.3 Event history analysis in the context of experimental psychology

To make EHA more relevant to researchers studying cognitive psychology and

cognitive neuroscience, in this section we provide a relevant worked example and consider

implications that are relevant to that domain of research.

2.3.1 A worked example. In the context of experimental psychology, it is

common for participants to be presented with either a 1-button detection task or a

2-button discrimination task, i.e., a task that has a right and a wrong answer. For

example, a task may involve choosing between two response options with only one of them

being correct. For such two-choice RT data, the discrete-time EHA of the RT data can be

extended with a discrete-time SAT analysis of the timed accuracy data. Specifically, the

hazard function of event occurrence can be extended with the discrete-time conditional

accuracy function, which gives you the probability that a response is correct given that it is

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

refer to this extended analysis for choice RT data as EHA/SAT.

Integrating results between hazard and conditional accuracy functions for choice RT

data can be informative for understanding psychological processes. To illustrate, we

consider a hypothetical choice RT example that is inspired by real data (Panis & Schmidt,

2016), but simplified to make the main point clearer (Figure 3). In a standard response

priming paradigm, there is a prime stimulus (e.g., an arrow pointing left or right) followed

by a target stimulus (another arrow pointing left or right). The prime can then be

congruent or incongruent with the target.

Figure 3 shows that the early upswing in hazard is equal for both prime conditions,

and that early responses are always correct in the congruent condition and always incorrect

in the incongruent condition. These results show that for short waiting times (< bin 6),

responses always follow the prime (and not the target, as instructed). And then for longer

waiting times, the response hazard is lower in incongruent compared to congruent trials,

²²³ and all responses emitted in these later bins are correct.

²²⁴ This joint pattern of results is interesting because it can provide meaningfully
²²⁵ different conclusions about psychological processes compared to conventional analyses, such
²²⁶ as computing mean-average RT across trials. Mean-average RT would only represent the
²²⁷ overall ability of cognition to overcome interference, on average, across trials. For instance,
²²⁸ if mean-average RT was higher in incongruent than congruent trials, one may conclude
²²⁹ that cognitive mechanisms that support interference control are working as expected across
²³⁰ trials. But such a conclusion is not supported when the effects are explored over a timeline.
²³¹ Instead, the psychological conclusion is much more nuanced and suggests that multiple
²³² states start, stop and possibly interact over a particular temporal window.

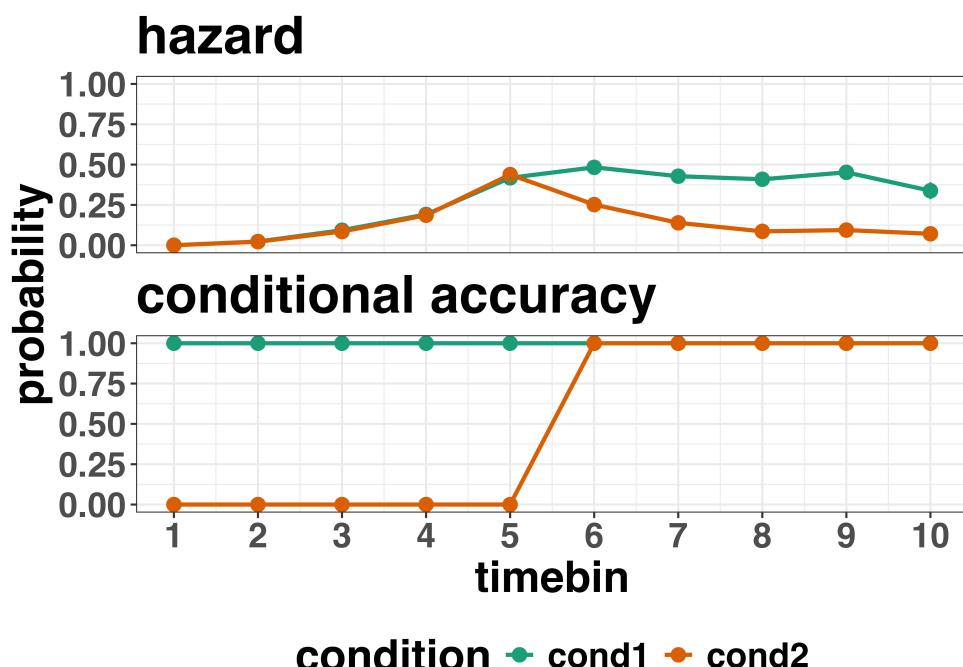


Figure 3. Discrete-time hazard and conditional accuracy functions.

²³³ Unlocking the temporal states of cognitive processes can be revealing for theory
²³⁴ development and the understanding of basic psychological processes. Possibly more
²³⁵ importantly, however, is that it simultaneously opens the door to address many new and

236 previously unanswered questions. Do all participants show similar temporal states or are
237 there individual differences? Do such individual differences extend to those individuals that
238 have been diagnosed with some form of psychopathology? How do temporal states relate to
239 brain-based mechanisms that might be studied using other methods from cognitive
240 neuroscience? And how much of theory in cognitive psychology would be in need of
241 revision if mean-average comparisons were supplemented with a temporal states approach?

242 **2.3.2 Implications for designing experiments.** Performing EHA in
243 experimental psychology has implications for how experiments are designed. Indeed, if
244 trials are categorised as a function of when responses occur, then each timebin will only
245 include a subset of the total number of trials. For example, let's consider an experiment
246 where each participant performs 2 conditions and there are 100 trial repetitions per
247 condition. Those 100 trials must be distributed in some manner across the chosen number
248 of bins.

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

257 In contrast to the large- N design that typically average across many participants
258 without being able to scrutinize individual data patterns, small- N designs retain crucial
259 information about the data patterns of individual observers. This can be advantageous
260 whenever participants differ systematically in their strategies or in the time courses of their
261 effects, so that averaging them would lead to misleading data patterns. Note that because
262 statistical power derives both from the number of participants and from the number of

263 repeated measures per participant and condition, small- N designs can still achieve what
264 are generally considered acceptable levels of statistical power, if they have have a sufficient
265 amount of data overall (Baker et al., 2021; Smith & Little, 2018).

266 **3. An overview of the general analytical workflow**

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

270 **3.1 Data science workflow and descriptive statistics**

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

276 **3.2 Inferential statistical approach**

277 Our lab adopts a estimation approach to multi-level regression (Kruschke & Liddell,
278 2018) ; Winter, 2019), which is heavily influenced by the Bayesian framework as suggested
279 by Richard McElreath (Kurz, 2023b; McElreath, 2018). We also use a “keep it maximal”
280 approach to specifying varying (or random) effects (Barr, Levy, Scheepers, & Tily, 2013).
281 This means that wherever possible we include varying intercepts and slopes per participant
282 To make inferences, we use two main approaches. We compare models of different
283 complexity, using information criteria (e.g., WAIC) and cross-validation (e.g., LOO), to
284 evaluate out-of-sample predictive accuracy (McElreath, 2018). We also take the most
285 complex model and evaluate key parameters of interest using point and interval estimates.

286 3.3 Implementation

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

291 4. Tutorials

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

300 custom functions).

301 Our list of tutorials is as follows:

- 302 • 1a. Wrangle raw data and calculate descriptive stats for one independent variable.
- 303 • 1b. Wrangle raw data and calculate descriptive stats for two independent variables.
- 304 • 2a. Bayesian multilevel modeling for $h(t)$
- 305 • 2b. Bayesian multilevel modeling for $ca(t)$
- 306 • 3a. Frequentist multilevel modeling for $h(t)$
- 307 • 3b. Frequentist multilevel modeling for $ca(t)$

308 Planning (T4) - if we get a simulation and power analysis script working, which we
309 are happy with then we could include it here.

310 **4.1 Tutorial 1a: Calculating descriptive statistics using a life table**

311 **4.1.1 Data wrangling aims.** Our data wrangling procedures serve two related
312 purposes. First, we want to summarise and visualise descriptive statistics that relate to our
313 main research questions about the time course of psychological processes. Second, we want
314 to produce two different data sets that can each be submitted to different types of
315 inferential modelling approaches. The two types of data structure we label as ‘person-trial’
316 data and ‘person-trial-bin’ data. The ‘person-trial’ data (Table 1) will be familiar to most
317 researchers who record behavioural responses from participants, as it represents the
318 measured RT and accuracy per trial within an experiment. This data set is used when
319 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.

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

329 **4.1.2 A real data wrangling example.** To illustrate how to quickly set up life
 330 tables for calculating the descriptive statistics (functions of discrete time), we use a
 331 published data set on masked response priming from Panis and Schmidt (2016). In their
 332 first experiment, Panis and Schmidt (2016) presented a double arrow for 94 ms that
 333 pointed left or right as the target stimulus with an onset at time point zero in each trial.
 334 Participants had to indicate the direction in which the double arrow pointed using their
 335 corresponding index finger, within 800 ms after target onset. Response time and accuracy
 336 were recorded on each trial. Prime type (blank, congruent, incongruent) and mask type
 337 were manipulated. Here we focus on the subset of trials in which no mask was presented.

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

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

344 The required column names are as follows:

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

351 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")))

```

352 Next, we can set up the life tables and plots of the discrete-time functions $h(t)$, $S(t)$,
 353 $ca(t)$, and $P(t)$ – see part A of the supplementary material for their definitions. To do so
 354 using a functional programming approach, one has to nest the data within participants
 355 using the group_nest() function, and supply a user-defined censoring time and bin width
 356 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

```

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

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

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

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

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

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

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

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

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

400 For example, for participant 6, the estimated hazard values in bin (240,280] are 0.03,
401 0.17, and 0.20 for the blank, congruent, and incongruent prime conditions, respectively. In
402 other words, when the waiting time has increased until *240 ms* after target onset, then the

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

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

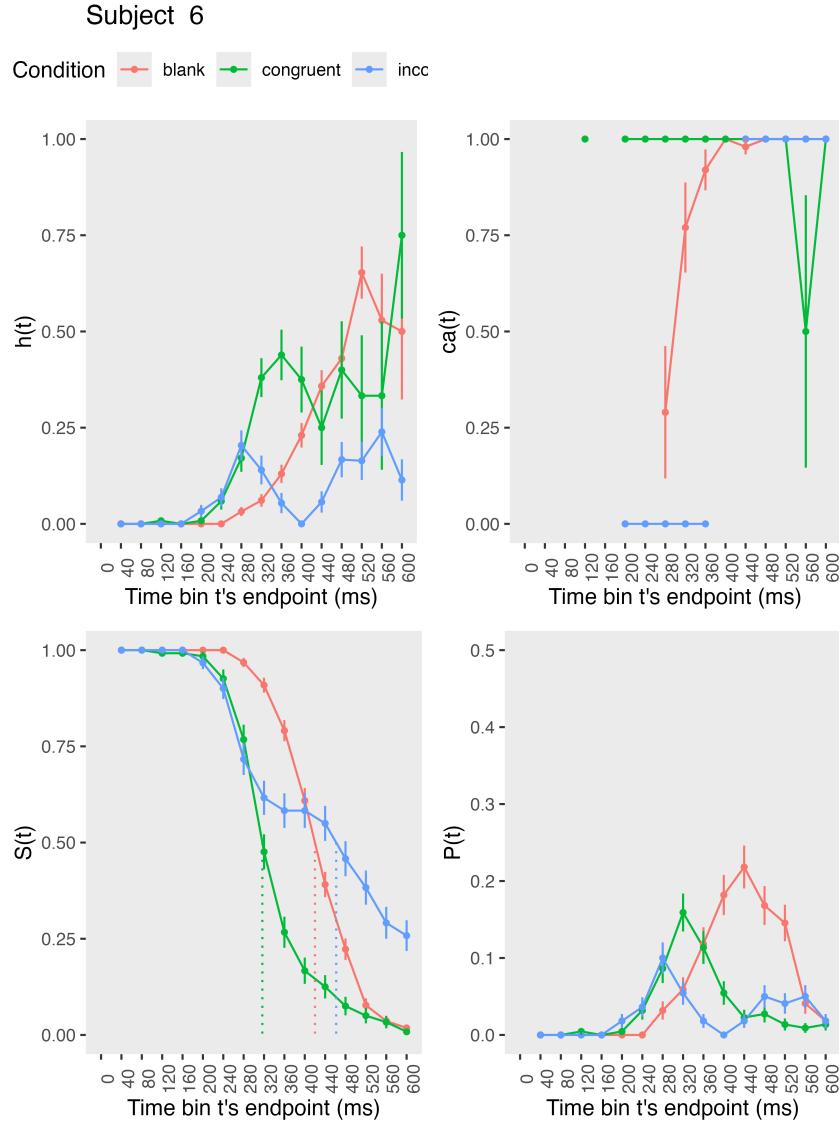


Figure 4. Estimated discrete-time hazard, survivor, probability mass, and conditional accuracy functions for participant 6, as a function of the passage of discrete waiting time.

410 However, when the waiting time has increased until 400 ms after target onset, then
 411 the conditional probability of response occurrence in the next 40 ms is estimated to be
 412 0.36, 0.25, and 0.06 for the blank, congruent, and incongruent prime conditions,
 413 respectively. And when a response does occur in bin (400,440], then the probability that it
 414 is correct is estimated to be 0.98, 1, and 1 for the blank, congruent, and incongruent prime

415 conditions, respectively.

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

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

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

433 4.2 Tutorial 1b: Generalising to a more complex design

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

440 To this end, Tutorial 1b illustrates how to calculate and plot the descriptive statistics

441 for the full data set of Experiment 1 of Panis and Schmidt (2016), which includes two

442 independent variables: mask type and prime type. As we use the same functional

443 programming approach as in Tutorial 1a, we simply present the sample-based functions for

444 participant 6 as part of Tutorial 1b for those that are interested.

445 **4.3 Tutorial 2a: Fitting Bayesian hazard models to time-to-event data**

446 In this third tutorial, we illustrate how to fit Bayesian multi-level regression models

447 to the RT data of the masked response priming data set used in Tutorial 1a. Fitting

448 (Bayesian or non-Bayesian) regression models to time-to-event data is important when you

449 want to study how the shape of the hazard function depends on various predictors (Singer

450 & Willett, 2003).

451 **4.3.1 Hazard model considerations.** There are several analytic decisions one

452 has to make when fitting a hazard model. First, one has to select an analysis time window,

453 i.e., a contiguous set of bins for which there is enough data for each participant. Second,

454 given that the dependent variable (event occurrence) is binary, one has to select a link

455 function (see part C in the supplementary material). The cloglog link is preferred over the

456 logit link when events can occur in principle at any time point within a bin, which is the

457 case for RT data (Singer & Willett, 2003). Third, one has to choose a specification of the

458 effect of discrete TIME (i.e., the time bin index t) in a selected baseline condition. One can

459 choose a general specification (one intercept per bin) or a functional specification, such as a

460 polynomial one (compare model 1 with models 2, 3, and 4 below; see also part D of the

461 supplementary material).

462 In the case of a large- N design without repeated measurements, the parameters of a

463 discrete-time hazard model can be estimated using standard logistic regression software

464 after expanding the typical person-trial data set into a person-trial-bin data set (Allison,

465 2010). When there is clustering in the data, as in the case of a small- N design with

⁴⁶⁶ repeated measurements, the parameters of a discrete-time hazard model can be estimated
⁴⁶⁷ using population-averaged methods (e.g., Generalized Estimating Equations), and Bayesian
⁴⁶⁸ or frequentist generalized linear mixed models (Allison, 2010).

⁴⁶⁹ In general, there are three assumptions one can make or relax when adding
⁴⁷⁰ experimental predictor variables and other covariates: The linearity assumption for
⁴⁷¹ continuous predictors (the effect of a 1 unit change is the same anywhere on the scale), the
⁴⁷² additivity assumption (predictors do not interact), and the proportionality assumption
⁴⁷³ (predictors do not interact with TIME).

⁴⁷⁴ In this tutorial we will fit four Bayesian multilevel models (i.e., generalized linear
⁴⁷⁵ mixed models) that differ in complexity to the person-trial-bin oriented data set that we
⁴⁷⁶ created in Tutorial 1a. We decided to select the analysis time window (200,600] and the
⁴⁷⁷ cloglog link. The data is prepared as follows.

```
# load person-trial-bin oriented data set
ptb_data <- read_csv("../Tutorial_1_descriptive_stats/data/inputfile_hazard_modeling.csv")

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

# create factor condition, with "blank" as the reference level
ptb_data <- ptb_data %>% mutate(condition = factor(condition, labels = c("blank", "congruent", "incongruent")))

# center TIME (variable period) on bin 9, and variable trial on number 1000 and rescale; Add dummy variables for each bin.
ptb_data <- ptb_data %>%
  mutate(period_9 = period - 9,
         trial_c = (trial - 1000)/1000,
         d6 = if_else(period == 6, 1, 0),
         d7 = if_else(period == 7, 1, 0),
         d8 = if_else(period == 8, 1, 0),
         d9 = if_else(period == 9, 1, 0),
         d10 = if_else(period == 10, 1, 0),
         d11 = if_else(period == 11, 1, 0),
         d12 = if_else(period == 12, 1, 0),
         d13 = if_else(period == 13, 1, 0),
         d14 = if_else(period == 14, 1, 0),
         d15 = if_else(period == 15, 1, 0))
```

478 **4.3.2 Prior distributions.** To get the posterior distribution of each model

479 parameter given the data, we need to specify a prior distribution for each parameter. The
 480 middle column of Figure 12 in part E of the supplementary material shows seven examples
 481 of prior distributions on the logit and/or cloglog scales.

482 While a normal distribution with relatively large variance is often used as a weakly
 483 informative prior for continuous dependent variables, rows A and B in Figure 12 show that
 484 specifying such distributions on the logit and cloglog scales leads to rather informative
 485 distributions on the original probability scale, as most mass is pushed to probabilities of 0
 486 and 1. The other rows in Figure 12 show prior distributions on the logit and cloglog scale
 487 that we use instead.

488 **4.3.3 Model 1: A general specification of TIME, and main effects of**

489 **congruency and trial number.** When you do not want to make assumptions about the
 490 shape of the hazard function in the selected baseline condition, or its shape is not smooth
 491 but irregular, then you can use a general specification of TIME, i.e., fit one intercept per
 492 time bin. In this first model, we use a general specification of TIME for the selected
 493 baseline condition (blank prime), and assume that the effects of prime-target congruency
 494 and trial number are proportional and additive, and that the effect of trial number is
 495 linear. Before we fit model 1, we remove unnecessary columns from the data, and specify
 496 our priors. In the code of Tutorial 2a, model M1 is specified as follows.

```
model_M1 <-
  brm(data = M1_data,
       family = binomial(link="cloglog"),
       event | trials(1) ~ 0 + d6 + d7 + d8 + Intercept + d10 + d11 + d12 + d13 + d14 + d15 +
              condition + trial_c +
              (d6 + d7 + d8 + 1 + d10 + d11 + d12 + d13 + d14 + d15 +
               condition + trial_c | pid),
       prior = priors_M1,
       chains = 4, cores = 4, iter = 3000, warmup = 1000,
       control = list(adapt_delta = 0.999, step_size = 0.04, max_treedepth = 12),
       seed = 12, init = "O",
```

```
file = "Tutorial_2_Bayesian/models/model_M1")
```

497 After selecting the binomial family and the cloglog link, the model formula is
 498 specified. The fixed effects include 9 dummy variables, the explicit Intercept variable
 499 (which represents bin 9 in this example), and the main effects of prime-target congruency
 500 (variable condition) and centered trial number (variable trial_c). Each of these effects is
 501 allowed to vary across individuals (variable pid). Estimating model M1 took about 70
 502 minutes on a MacBook Pro (Sonoma 14.6.1 OS, 18GB Memory, M3 Pro Chip).

503 **4.3.4 Model 2: A polynomial specification of TIME, and main effects of**
 504 **congruency and trial number.** When the shape of the hazard function is rather
 505 smooth, as it is for behavioral RT data, one can fit a more parsimonious model by using a
 506 polynomial specification of TIME. For our second example model, we thus use a
 507 third-order polynomial specification of TIME for the selected baseline condition (blank
 508 prime), and again assume that the effects of prime-target congruency and centered trial
 509 number are proportional and additive, and that the effect of trial number is linear. The
 510 model formula for model M2 looks as follows.

```
event | trials(1) ~ 0 + Intercept + period_9 + I(period_9^2) + I(period_9^3) +  

       condition + trial_c +  

       (1 + period_9 + I(period_9^2) + I(period_9^3) +  

       condition + trial_c | pid),
```

511 Because TIME is centered on bin 9, and trial number on trial 1000, the Intercept
 512 represents the cloglog-hazard in bin 9 for the blank prime condition in trial 1000.
 513 Estimating model M2 took about 2.5 hours.

514 **4.3.5 Model 3: A polynomial specification of TIME, and relaxing the**
 515 **proportionality assumption.** So far, we assumed that the effect of our predictors
 516 prime-target congruency and centered trial number are the same in each time bin. However,
 517 the descriptive plots (e.g., Figure 4) suggest that the effect of prime-target congruency

518 varies across time bins. Previous research has shown that psychological effects typically
 519 change over time (Panis, 2020; Panis, Moran, et al., 2020; Panis & Schmidt, 2022; Panis et
 520 al., 2017; Panis & Wagemans, 2009). For the third model, we thus use a third-order
 521 polynomial specification of TIME for the baseline condition (blank prime), and relax the
 522 proportionality assumption for the predictor variables prime-target congruency (variable
 523 condition) and centered trial number (variable trial_c).

```
event | trials(1) ~ 0 + Intercept +
      condition*period_9 +
      condition*I(period_9^2) +
      condition*I(period_9^3) +
      trial_c*period_9 +
      trial_c*I(period_9^2) +
      trial_c*I(period_9^3) +
      (1 + condition*period_9 +
      condition*I(period_9^2) +
      condition*I(period_9^3) +
      trial_c*period_9 +
      trial_c*I(period_9^2) +
      trial_c*I(period_9^3) | pid),
```

524 Note that duplicate terms in the model formula are ignored. Estimating model M3
 525 took about 4.5 hours.

526 **4.3.6 Model 4: A polynomial specification of TIME, and relaxing all three
 527 assumptions.** Based on previous work (e.g., Panis, 2020) we expect nonlinear effects of
 528 trial number on the hazard of response occurrence. We thus relax all three assumptions in
 529 model 4. We add a squared term for the continuous predictor centered trial number –
 530 $I(trial_c^2)$ – and include interaction terms. For example, how the effect of congruent
 531 primes changes across time bins within a trial might change across the trials within an
 532 experiment.

```
event | trials(1) ~ 0 + Intercept +
      condition*period_9*trial_c +
      condition*period_9*I(trial_c^2) +
```

```

condition*I(period_9^2)*trial_c +
condition*I(period_9^2)*I(trial_c^2) +
condition*I(period_9^3) +
trial_c*I(period_9^3) +
(1 + condition*period_9*trial_c +
condition*period_9*I(trial_c^2) +
condition*I(period_9^2)*trial_c +
condition*I(period_9^2)*I(trial_c^2)
condition*I(period_9^3) +
trial_c*I(period_9^3) | pid)

```

533 Again, duplicate terms in the model formula are ignored. Estimating model M4 took
 534 about 8 hours.

535 **4.3.7 Compare the models.** We can compare the four models using the Widely
 536 Applicable Information Criterion (WAIC) and Leave-One-Out (LOO) cross-validation, and
 537 look at model weights for both criteria (Kurz, 2023a; McElreath, 2018).

```

model_weights(model_M1, model_M2, model_M3, model_M4, weights = "loo") %>% round(digits = 3)

538 ## model_M1 model_M2 model_M3 model_M4
539 ##      0      0      0      1

model_weights(model_M1, model_M2, model_M3, model_M4, weights = "waic") %>% round(digits = 3)

540 ## model_M1 model_M2 model_M3 model_M4
541 ##      0      0      0      1

```

542 Clearly, both the loo and waic weighting schemes assign a weight of 1 to model M4,
 543 and a weight of 0 to the other three simpler models.

544 **4.3.8 Evaluate parameter estimates.** To make inferences from the parameter
 545 estimates in model M4, we summarize the draws from the posterior distributions of the
 546 effects of congruent and incongruent primes relative to the blank prime condition, in each
 547 time bin for trial numbers 500, 1000, and 1500, in terms of point and interval estimates.

548 Figure 6 shows one point (mean) and three highest posterior density interval

549 (50/80/95%) estimates for the effects of congruent and incongruent primes relative to

550 neutral primes, for each time bin in trial numbers 500, 1000, and 1500.

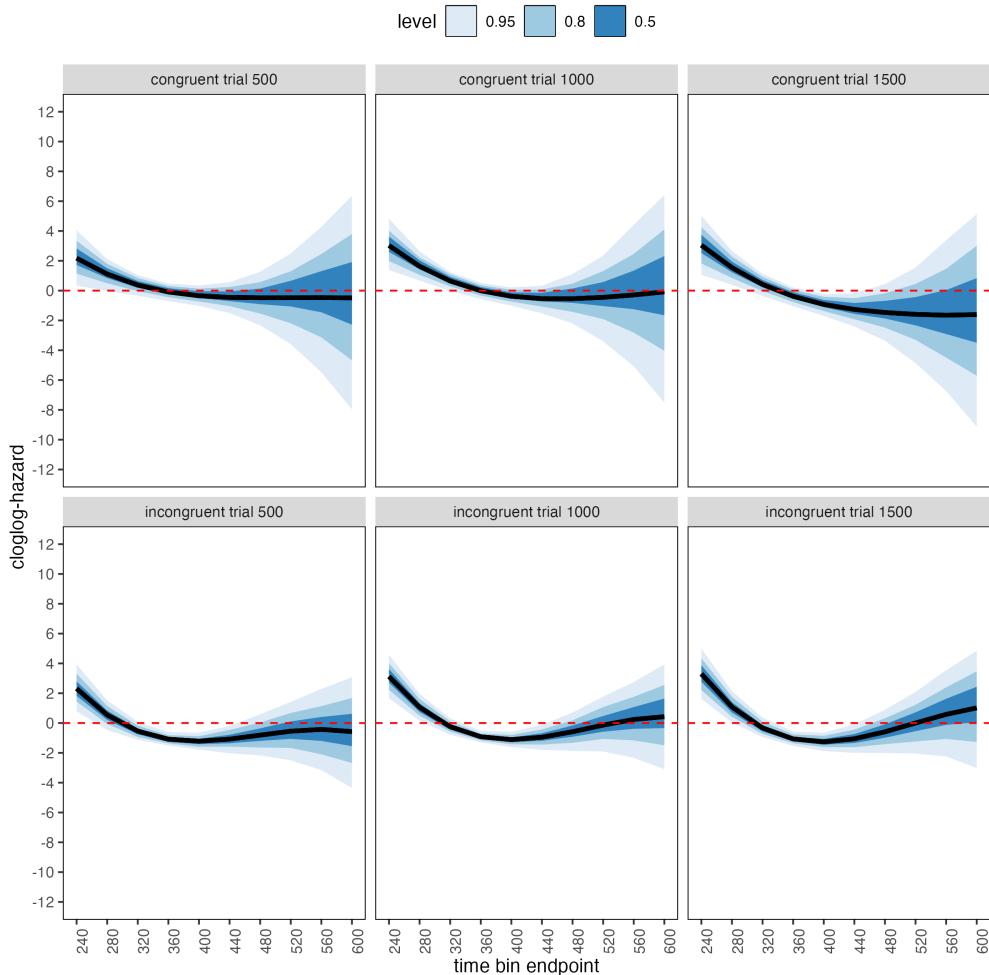


Figure 5. Means and 50/80/95% highest posterior density intervals of the draws from the posterior distributions representing the effect of congruent and incongruent primes relative to neutral primes in each bin for trial numbers 500, 1000, and 1500.

551 Table 4 shows the summaries of the draws from the posterior distributions of the

552 effects of congruent and incongruent primes relative to the blank prime condition in trials

553 500, 1000, and 1500, in terms of a point estimate (the mean) and the upper and lower

554 bounds of the 95% highest posterior density interval. Also, by exponentiating the mean we

555 obtain an effect size in terms of a hazard ratio. For simplicity and ease of presentation, we
556 only tabulate data for a subset of the design in the main text (trial 500 for congruent and
557 incongruent conditions). For the full table, see Supplementary materials or Tutorial X.

558 Based on Figure 6 and Supplementary Table XX, we see that at the beginning of the
559 experiment (trial 500), congruent and incongruent primes have a positive effect in time bin
560 (200,240] on cloglog-hazard, relative to the cloglog-hazard estimate in the baseline
561 condition (no prime; red striped lines in Figure 6). For example, the hazard ratio shows
562 that the hazard of response occurrence for congruent primes is estimated to be 8.82 times
563 higher than that for no-prime trials in bin (200,240] of trial 500. Incongruent primes also
564 have a negative effect on cloglog-hazard in bins (320,360], (360,400], and (400,440]. For
565 example, in bin (320,360], the hazard ratio shows that the hazard of response occurrence
566 for incongruent prime is estimated to be .34 times smaller than that for no-prime trials.
567 While the early positive effects reflect responses to the prime stimulus, the later negative
568 effect for incongruent primes likely reflects response competition between the
569 prime-triggered response (e.g., left) and the target-triggered response (e.g., right)

570 In the middle of the experiment (trial 1000), both congruent and incongruent primes
571 have positive effects in bins (200,240] and (240,280], while incongruent primes again have
572 negative effects in bins (320,360], (360,400], and (400,440]. Probably due to practicing
573 stimulus-response associations, the primes generate a higher hazard of response occurrence
574 for 80 ms early in a trial (compared to 40 ms at the beginning of the experiment)
575 compared to the blank prime condition.

576 Towards the end of the experiment (trial 1500), both congruent and incongruent
577 primes have positive and negative effects. Positive effects are present in bins (200,240] and
578 (240,280]. Incongruent primes again have negative effects in bins (320,360], (360,400], and
579 (400,440], and congruent primes now also have negative effects in bins (360,400] and
580 (400,440].

These results show that the effect of prime-target congruency changes not only on the across-bin/within-trial time scale (variable period_9), but also on the across-trial/within-experiment time scale (variable trial_c). The fact that congruent primes generate negative effects for 80 ms (compared to no-prime trials) towards the end of the experiment, while incongruent primes generate negative effects for 120 ms throughout the experiment, suggests the involvement of separate cognitive processes.

Panis and Schmidt (2016) distinguished between automatic response competition (bottom-up lateral inhibition between response channels), active and global inhibition (top-down nonselective response inhibition), and active and selective inhibition (top-down selective response inhibition). While automatic response competition can be expected to be present in the incongruent trials throughout the experiment, active and global response inhibition effects might be present in both congruent and incongruent (unmasked) prime trials. In other words, people learn that the prime-triggered response is premature and that they have to temporarily slow down (increase the global response threshold) in order to allow gating of the correct response to the target stimulus. Thus, it seems that this global inhibitory effect becomes visible in the congruent (compared to no-prime) trials towards the end of the experiment, while it might be masked by the automatic inhibitory effect of response competition in the incongruent trials. Interestingly, while Panis and Schmidt (2016) did not test interactions between prime-target congruency and trial number, they concluded that active (i.e., top-down) response inhibition starts around 360 ms after the onset of the second stimulus (the target stimulus in no-mask trials), which nicely coincides with the onset of the negative effect of congruent primes observed here in trial 1500.

To conclude this Tutorial 2a, Figure 7 shows the model-based hazard functions for each prime type for participant 6, in trials 500, 1000, and 1500.

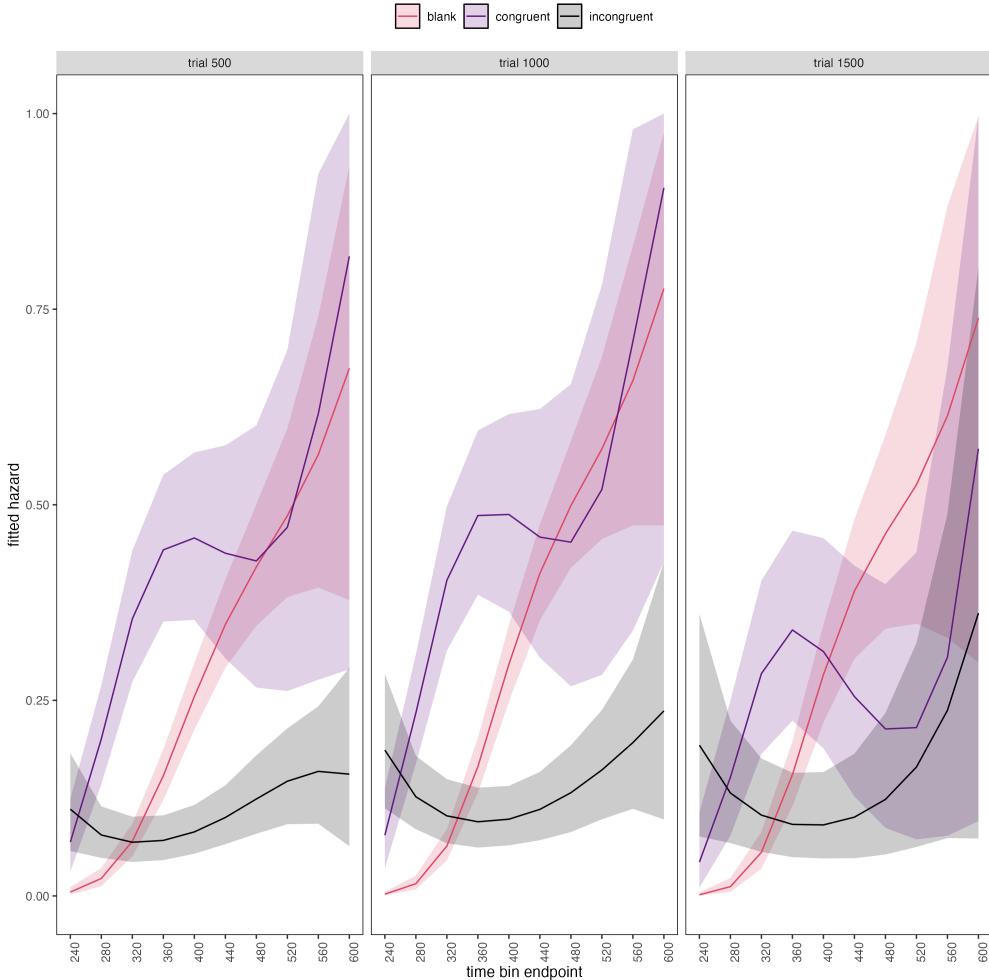


Figure 6. Model-based hazard functions for each prime type for participant 6 in trials 500, 1000, and 1500.

605 4.4 Tutorial 2b: Fitting Bayesian conditional accuracy models

606 In this fourth tutorial, we illustrate how to fit a Bayesian multi-level regression model
 607 to the timed accuracy data from the masked response priming data set used in Tutorial 1a.
 608 The general process is similar to Tutorial 2a, except that (a) we use the person-trial data
 609 set, (b) we use the logit link function, and (c) we change the priors. For illustration
 610 purposes, we only fitted the effects of model M4 (see Tutorial 2a) in the conditional
 611 accuracy model called M4_ca.

612 To make inferences from the parameter estimates in model M4_ca, we summarize the
 613 draws from the posterior distributions of the effects of congruent and incongruent primes
 614 on logit-ca relative to the blank prime condition, in each time bin for trial numbers 500,
 615 1000, and 1500, in terms of point and interval estimates.

616 Figure 8 shows one point (mean) and three highest posterior density interval
 617 (50/80/95%) estimates for the effects of congruent and incongruent primes relative to
 618 neutral primes on logit-ca, for each time bin in trial numbers 500, 1000, and 1500.

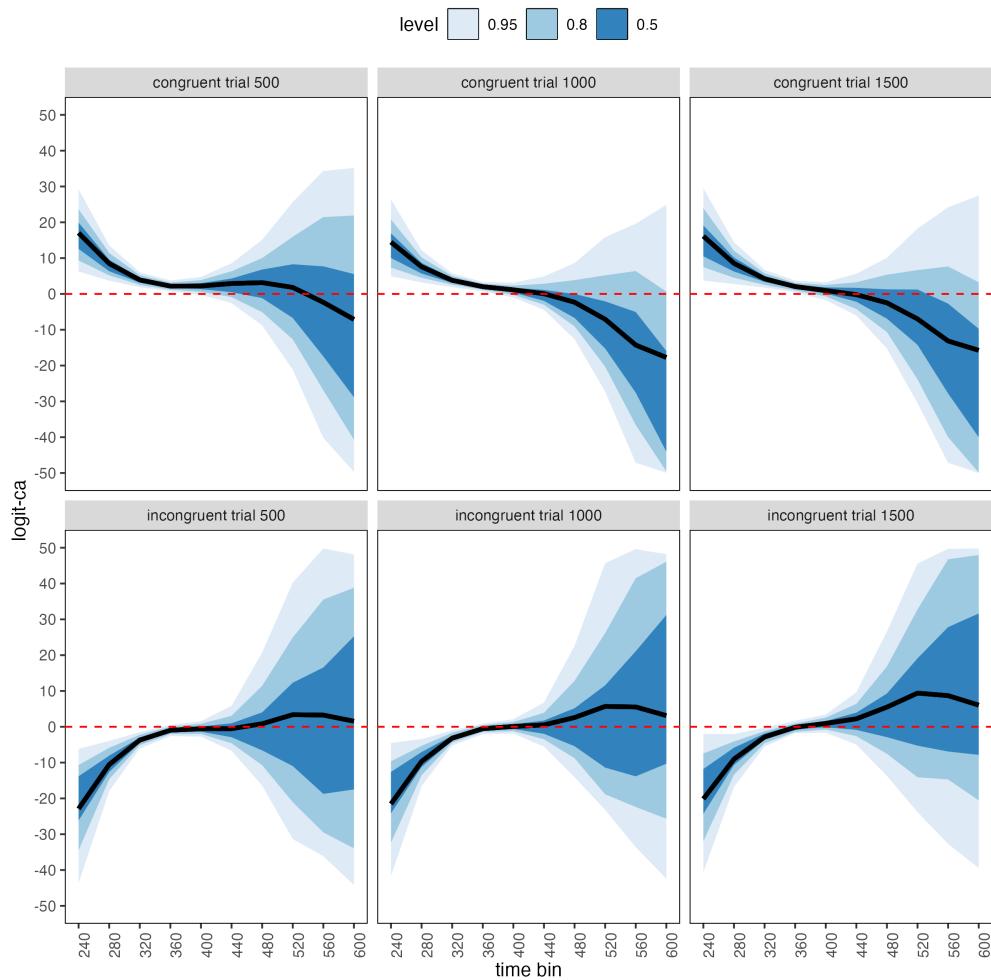


Figure 7. Means and 50/80/95% highest posterior density intervals of the draws from the posterior distributions representing the effect of congruent and incongruent primes relative to neutral primes in each bin for trial numbers 500, 1000, and 1500.

619 Supplementary Table XX shows the summaries of the draws from the posterior
620 distributions of the effects of congruent and incongruent primes relative to the blank prime
621 condition in trials 500, 1000, and 1500, in terms of a point estimate (the mean) and the
622 upper and lower bounds of the 95% highest posterior density interval. Also, by
623 exponentiating the mean we obtain an effect size in terms of an odds ratio.

624 Based on Figure 8 and Supplementary Table XX, we see that throughout the
625 experiment (trials 500, 1000, and 1500), congruent primes have a positive effect on
626 logit-ca(t) in time bins (200,240], (240,280], (280,320], and (320,360], relative to the
627 logit-ca(t) estimates in the baseline condition (blank prime; red dashed lines in Figure 8).
628 For example, the odds ratio for congruent primes in bin (320,360] in trial 500 shows that
629 the odds of a correct response are estimated to be 8.89 times higher than the odds of a
630 correct response when there is no prime. Incongruent primes have a negative effect on
631 logit-ca(t) in time bins (200,240], (240,280], and (280,320] throughout the experiment,
632 relative to the logit-ca(t) estimates in the baseline condition (no prime; red striped lines).

633 To conclude this Tutorial 2b, Figure 9 shows the model-based ca(t) functions for each
634 prime type for participant 6, in trials 500, 1000, and 1500.

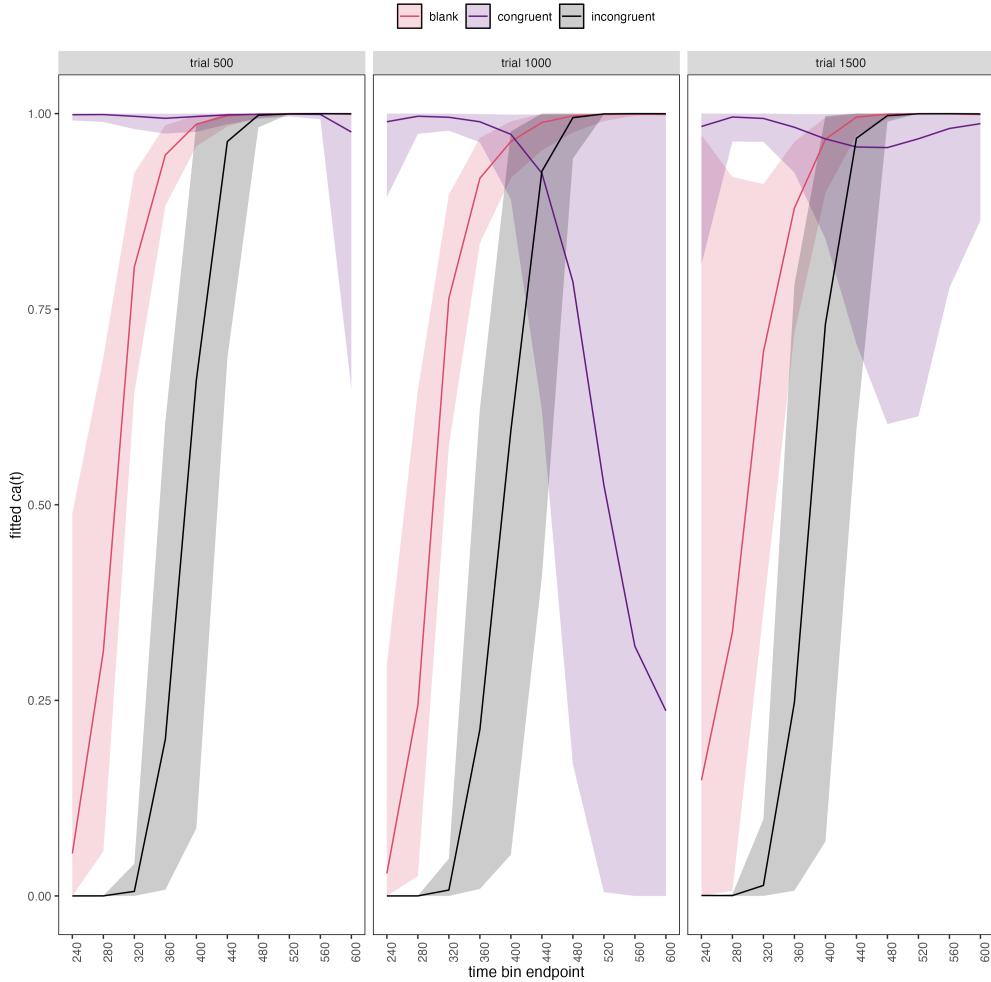


Figure 8. Model-based $ca(t)$ functions for each prime type for participant 6 in trials 500, 1000, and 1500.

635 4.5 Tutorial 3a: Fitting Frequentist hazard models

636 In this fifth tutorial we illustrate how to fit a multilevel regression model to RT data
 637 in the frequentist framework, for the data set used in Tutorial 1a. The general process is
 638 similar to that in Tutorial 2a, except that there are no priors to set. For illustration
 639 purposes, we only fitted the effects from model M3 (see Tutorial 2a) using the function
 640 `glmer()` from the R package `lme4`. Alternatively, one could also use the function
 641 `glmmPQL()` from the R package `MASS` (Ripley et al., 2024). The resulting hazard model

642 is called M3_f.

643 In Figure 10 we compare the parameter estimates of model M3 from brm() with those
 644 of model M3_f from glmer().

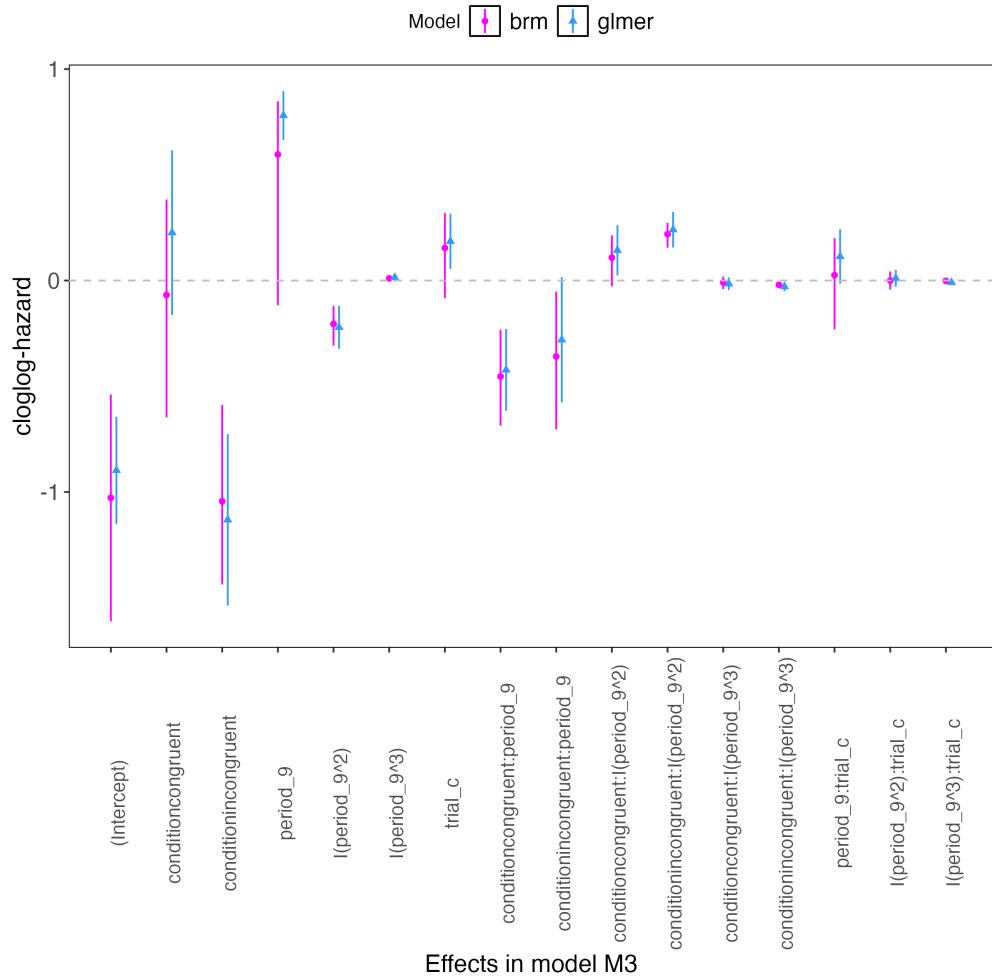


Figure 9. Parameter estimates for model M3 from brm() – means and 95% quantile intervals – and model M3_f from glmer() – maximum likelihood estimates and 95% confidence intervals.

645 Figure 10 confirms that the parameter estimates from both Bayesian and frequentist
 646 models are pretty similar,, which makes sense given the close similarity in model structure.
 647 However, the random effects structure of model M3_f was already too complex for the
 648 frequentist model as it did not converge and resulted in a singular fit. This is of course one

649 of the reasons why Bayesian modeling has become so popular in recent years. But the price
650 you pay for being able to fit more complex random effects models in a Bayesian framework
651 is computation time. In other words, as we have noted throughout, some of the Bayesian
652 models in Tutorials 2a and 2b took several hours to build.

653 **4.6 Tutorial 3b: Fitting Frequentist conditional accuracy models**

654 In this sixth tutorial we illustrate how to fit a multilevel regression model to the
655 timed accuracy data in the frequentist framework, for the data set used in Tutorial 1a. For
656 illustration purposes, we only fitted the effects from model M3 (see Tutorial 2a) using the
657 function glmer() from the R package lme4. Alternatively, one could also use the function
658 glmmPQL() from the R package MASS (Ripley et al., 2024). Again, the resulting
659 conditional accuracy model M3_ca_f did not converge and resulted in a singular fit.

660 **4.4 Tutorial 4: Planning**

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

662 **5. Discussion**

663 This main motivation for writing this paper is the observation that EHA and SAT
664 analysis remain under-used in psychological research. As a consequence, the field of
665 psychological research is not taking full advantage of the many benefits EHA/SAT provides
666 compared to more conventional analyses. By providing a freely available set of tutorials,
667 which provide step-by-step guidelines and ready-to-use R code, we hope that researchers
668 will feel more comfortable using EHA/SAT in the future. Indeed, we hope that our
669 tutorials may help to overcome a barrier to entry with EHA/SAT, which is that such
670 approaches require more analytical complexity compared to mean-average comparisons.
671 While we have focused here on within-subject, factorial, small- N designs, it is important to

realize that EHA/SAT can be applied to other designs as well (large- N designs with only one measurement per subject, between-subject designs, etc.). As such, the general workflow and associated code can be modified and applied more broadly to other contexts and research questions. In the following, we discuss issues relating to model complexity and interpretability, individual differences, as well as limitations of the approach and future extensions.

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

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

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

696 5.2 Model complexity versus interpretability

697 EHA can quickly become very complex when adding more than 1 time scale, due to
698 the many possible higher-order interactions. For example, model M4 contains two time
699 scales as covariates: the passage of time on the within-trial time scale, and the passage of
700 time on the across-trial (or within-experiment) time scale. However, when trials are
701 presented in blocks, and blocks of trials within sessions, and when the experiment
702 comprises three sessions, then four time scales can be defined (within-trial, within-block,
703 within-session, and within-experiment). From a theoretical perspective, adding more than
704 1 time scale – and their interactions – can be important to capture plasticity and other
705 learning effects that may play out on such longer time scales, and that are probably present
706 in each experiment in general. From a practical perspective, therefore, some choices need
707 to be made to balance the amount of data that is being collected per participant, condition
708 and across the varying timescales. As one example, if there are several timescales of
709 relevance, then it might be prudent for interpretational purposes to limit the number of
710 experimental predictor variables (conditions). This is of course where planning and data
711 simulation efforts would be important to provide a guide to experimental design choices.

712 5.3 Individual differences

713 One important issue is that of possible individual differences in the overall location of
714 the distribution, and the time course of psychological effects. For example, when you wait
715 for a response of the participant on each trial, you allow the participant to have control
716 over the trial duration, and some participants might respond only when they are confident
717 that their emitted response will be correct. These issues can be avoided by introducing a
718 (relatively short) response deadline in each trial, e.g., 600 ms for simple detection tasks,
719 1000 ms for more difficult discrimination tasks, or 2 s for tasks requiring extended high-level
720 processing. Because EHA can deal in a straightforward fashion with right-censored

721 observations (i.e., trials without an observed response), introducing a response deadline is
722 recommended when designing RT experiments. Furthermore, introducing a response
723 deadline and asking participants to respond before the deadline as much as possible, will
724 also lead to individual distributions that overlap in time, which is important when selecting
725 a common analysis time window when fitting hazard and conditional accuracy models.

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

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

742 **5.4 Limitations**

743 Compared to the orthodox method – comparing mean-averages between conditions –
744 the most important limitation of multi-level hazard and conditional accuracy modeling is
745 that it might take a long time to estimate the parameters using Bayesian methods or the

746 model might have to be simplified significantly to use frequentist methods.

747 Another issue is that you need a relatively large number of trials per condition to
748 estimate the hazard function with high temporal resolution, which is required when testing
749 predictions of process models of cognition. Indeed, in general, there is a trade-off between
750 the number of trials per condition and the temporal resolution (i.e., bin width) of the
751 hazard function. Therefore, we recommend researchers to collect as many trials as possible
752 per experimental condition, given the available resources and considering the participant
753 experience (e.g., fatigue and boredom). For instance, if the maximum session length
754 deemed reasonable is between 1 and 2 hours, what is the maximum number of trials per
755 condition that you could reasonably collect? After consideration, it might be worth
756 conducting multiple testing sessions per participant and/or reducing the number of
757 experimental conditions. Finally, there is a user-friendly online tool for calculating
758 statistical power as a function of the number of trials as well as the number of participants,
759 and this might be worth consulting to guide the research design process (Baker et al., 2021).

760 We did not discuss continuous-time EHA, nor continuous-time SAT analysis. As
761 indicated by Allison (2010), learning discrete-time EHA methods first will help in learning
762 continuous-time methods. Given that RT is typically treated as a continuous variable, it is
763 possible that continuous-time methods will ultimately prevail. However, they require much
764 more data to estimate the continuous-time hazard (rate) function well. Thus, by trading a
765 bit of temporal resolution for a lower number of trials, discrete-time methods seem ideal for
766 dealing with typical psychological time-to-event data sets for which there are less than
767 ~200 trials per condition per experiment.

768 5.5 Extensions

769 The hazard models in this tutorial assume that there is one event of interest. For RT
770 data, this event constitutes a single transition between an “idle” state and a “responded”

771 state. However, in certain situations, more than one event of interest might exist. For
772 example, in a medical or health-related context, an individual might transition back and
773 forth between a “healthy” state and a “depressed” state, before being absorbed into a final
774 “death” state. When you have data on the timing of these transitions, one can apply
775 multi-state hazard models, which generalize EHA to transitions between three or more
776 states (Steele, Goldstein, & Browne, 2004). Also, the predictor variables in this tutorial are
777 time-invariant, i.e., their value did not change over the course of a trial. Thus, another
778 extension is to include time-varying predictors, i.e., predictors whose value can change
779 across the time bins within a trial (Allison, 2010). For example, when gaze position is
780 tracked during a visual search trial, the gaze-target distance will vary during a trial when
781 the eyes move around before a manual response is given; shorter gaze-target distances
782 should be associated with a higher hazard of response occurrence. Note that the effect of a
783 time-varying predictor (e.g., an occipital EEG signal) can itself vary over time.

784

6. Conclusions

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

796

References

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

- 823 Blossfeld, H.-P., & Rohwer, G. (2002). *Techniques of event history modeling: New*
824 *approaches to causal analysis, 2nd ed* (pp. x, 310). Mahwah, NJ, US: Lawrence
825 Erlbaum Associates Publishers.
- 826 Box-Steffensmeier, J. M. (2004). Event history modeling: A guide for social scientists.
827 Cambridge: University Press.
- 828 Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan.
829 *Journal of Statistical Software*, 80(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>
- 830 Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms.
831 *The R Journal*, 10(1), 395–411. <https://doi.org/10.32614/RJ-2018-017>
- 832 Bürkner, P.-C. (2021). Bayesian item response modeling in R with brms and Stan. *Journal*
833 *of Statistical Software*, 100(5), 1–54. <https://doi.org/10.18637/jss.v100.i05>
- 834 Eddelbuettel, D., & Balamuta, J. J. (2018). Extending R with C++: A Brief Introduction
835 to Rcpp. *The American Statistician*, 72(1), 28–36.
836 <https://doi.org/10.1080/00031305.2017.1375990>
- 837 Eddelbuettel, D., & François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal*
838 *of Statistical Software*, 40(8), 1–18. <https://doi.org/10.18637/jss.v040.i08>
- 839 Gabry, J., Češnovar, R., Johnson, A., & Brodner, S. (2024). *Cmdstanr: R interface to*
840 *'CmdStan'*. Retrieved from <https://github.com/stan-dev/cmdstanr>
- 841 Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization
842 in bayesian workflow. *J. R. Stat. Soc. A*, 182, 389–402.
843 <https://doi.org/10.1111/rssa.12378>
- 844 Girard, J. (2024). *Standist: What the package does (one line, title case)*. Retrieved from
845 <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 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
- 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
- 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 Wolkersdorfer, M. P., Panis, S., & Schmidt, T. (2020). Temporal dynamics of sequential
961 motor activation in a dual-prime paradigm: Insights from conditional accuracy and
962 hazard functions. *Attention, Perception, & Psychophysics*, 82(5), 2581–2602.
963 <https://doi.org/10.3758/s13414-020-02010-5>

Supplementary material

965 A. Definitions of discrete-time hazard, survivor, probability mass, and
966 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

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

⁹⁷³ and the discrete-time survivor function

$$974 \quad S(t) = P(RT > t) = [1-h(t)].[1-h(t-1)].[1-h(t-2)] \dots [1-h(1)] \quad (2)$$

⁹⁷⁵ and not on the probability mass function

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

977 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

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

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

991 $ca(t) = P(\text{correct} \mid RT = t)$ (5)

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

995 The survivor function provides a context for the hazard function, as $S(t-1) = P(RT >$
 996 $t-1) = P(RT \geq t)$ tells you on how many percent of the trials the estimate $h(t) = P(RT =$
 997 $t \mid RT \geq t)$ is based. The probability mass function provides a context for the conditional
 998 accuracy function, as $P(t) = P(RT = t)$ tells you on how many percent of the trials the
 999 estimate $ca(t) = P(\text{correct} \mid RT = t)$ is based.

1000 While psychological RT data is typically measured in small, continuous units (e.g.,
 1001 milliseconds), discrete-time EHA treats the RT data as interval-censored data, because it
 1002 only uses the information that the response occurred sometime in a particular bin of time
 1003 $(x,y]: x < RT \leq y$. If we want to use the exact event times, then we treat time as a
 1004 continuous variable, and let RT be a continuous random variable denoting a particular
 1005 person's response time in a particular experimental condition. Continuous-time EHA does
 1006 not focus on the cumulative distribution function $F(t) = P(RT \leq t)$ and its derivative, the
 1007 probability density function $f(t) = F(t)'$, but on the survivor function $S(t) = P(RT > t)$
 1008 and the hazard rate function $\lambda(t) = f(t)/S(t)$. The hazard rate function gives you the
 1009 instantaneous *rate* of event occurrence at time point t, given that the event has not
 1010 occurred yet.

1011 **B. Custom functions for descriptive discrete-time hazard analysis**

1012 We defined 13 custom functions that we list here.

- 1013 • censor(df,timeout,bin_width) : divide the time segment $(0, \text{timeout}]$ in bins, identify
1014 any right-censored observations, and determine the discrete RT (time bin rank)
- 1015 • ptb(df) : transform the person-trial data set to the person-trial-bin data set
- 1016 • setup_lt(ptb) : set up a life table for each level of 1 independent variable
- 1017 • setup_lt_2IV(ptb) : set up a life table for each combination of levels of 2
1018 independent variables
- 1019 • calc_ca(df) : estimate the conditinal accuracies when there is 1 independent variable
- 1020 • calc_ca_2IV(df) : estimate the conditional accuraiies when there are 2 independent
1021 variables
- 1022 • join_lt_ca(df1,df2) : add the ca(t) estimates to the life tables (1 independent
1023 variable)
- 1024 • join_lt_ca_2IV(df1,df2) : add the ca(t) estimates to the life tables (2 independent
1025 variables)
- 1026 • extract_median(df) : estimate quantiles $S(t)._{50}$ (1 independent variable)
- 1027 • extract_median_2IV(df) : estimate quantiles $S(t)._{50}$ (2 independent variables)
- 1028 • plot_oha(df,subj,haz_yaxis) : create plots of the discrete-time functions (1
1029 independent variable)
- 1030 • plot_oha_2IV(df,subj,haz_yaxis) : create plots of the discrete-time functions (2
1031 independent variables)
- 1032 • plot_oha_agg(df,subj,haz_yaxis) : create 1 plot for data aggregated across
1033 participants (1 independent variable)

1034 When you want to analyse simple RT data from a detection experiment with one
1035 independent variable, the functions calc_ca() and join_lt_ca() should not be used, and
1036 the code to plot the conditional accuracy functions should be removed from the function
1037 plot_oha(). When you want to analyse simple RT data from a detection experiment with
1038 two independent variables, the functions calc_ca_2IV() and join_lt_ca_2IV() should not
1039 be used, and the code to plot the conditional accuracy functions should be removed from

₁₀₄₀ the function `plot_eha_2IV()`.

₁₀₄₁ C. Link functions

₁₀₄₂ Popular link functions include the logit link and the complementary log-log link, as
₁₀₄₃ shown in Figure 11.

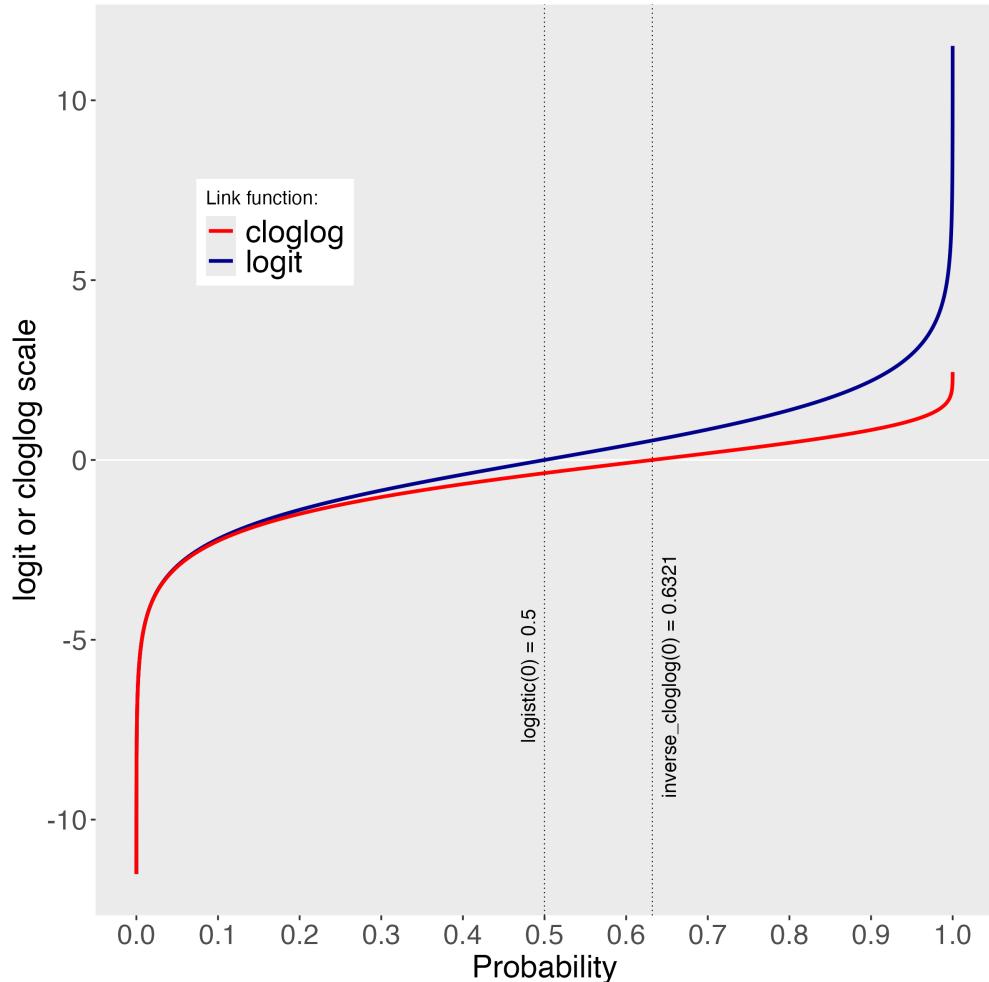


Figure 10. The logit and cloglog link functions.

₁₀₄₄ D. Regression equations

₁₀₄₅ An example (single-level) discrete-time hazard model with three predictors (TIME,
₁₀₄₆ X₁, X₂), the cloglog link function, and a third-order polynomial specification for TIME can

1047 be written as follows:

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

1050 The main predictor variable TIME is the time bin index t that is centered on value 9
1051 in this example. The first set of terms within brackets, the parameters β_0 to β_2 multiplied
1052 by their polynomial specifications of (centered) time, represents the shape of the baseline
1053 cloglog-hazard function (i.e., when all predictors X_i take on a value of zero). The second
1054 set of terms (the beta parameters β_3 to β_5) represents the vertical shift in the baseline
1055 cloglog-hazard for a 1 unit increase in the respective predictor variable. Predictors can be
1056 discrete, continuous, and time-varying or time-invariant. For example, the effect of a 1 unit
1057 increase in X_1 is to vertically shift the whole baseline cloglog-hazard function by β_1
1058 cloglog-hazard units. However, if the predictor interacts linearly with TIME (see X_2 in the
1059 example), then the effect of a 1 unit increase in X_2 is to vertically shift the predicted
1060 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
1062 predictor, its β parameter is exponentiated, resulting in a hazard ratio (due to the use of
1063 the cloglog link). When using the logit link, exponentiating a β parameter results in an
1064 odds ratio.

1065 An example (single-level) discrete-time hazard model with a general specification for
1066 TIME (separate intercepts for each of six bins, where D1 to D6 are binary variables
1067 identifying each bin) and a single predictor (X_1) can be written as follows:

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

1069 E. Prior distributions

1070 To gain a sense of what prior *logit* values would approximate a uniform distribution
1071 on the probability scale, Kurz (2023a) simulated a large number of draws from the

1072 Uniform(0,1) distribution, converted those draws to the log-odds metric, and fitted a
 1073 Student's t distribution. Row C in Figure 12 shows that using a t-distribution with 7.61
 1074 degrees of freedom and a scale parameter of 1.57 as a prior on the logit scale, approximates
 1075 a uniform distribution on the probability scale. According to Kurz (2023a), such a prior
 1076 might be a good prior for the intercept(s) in a logit-hazard model, while the N(0,1) prior in
 1077 row D might be a good prior for the non-intercept parameters in a logit-hazard model, as it
 1078 gently regularizes p towards .5 (i.e., a zero effect on the logit scale).

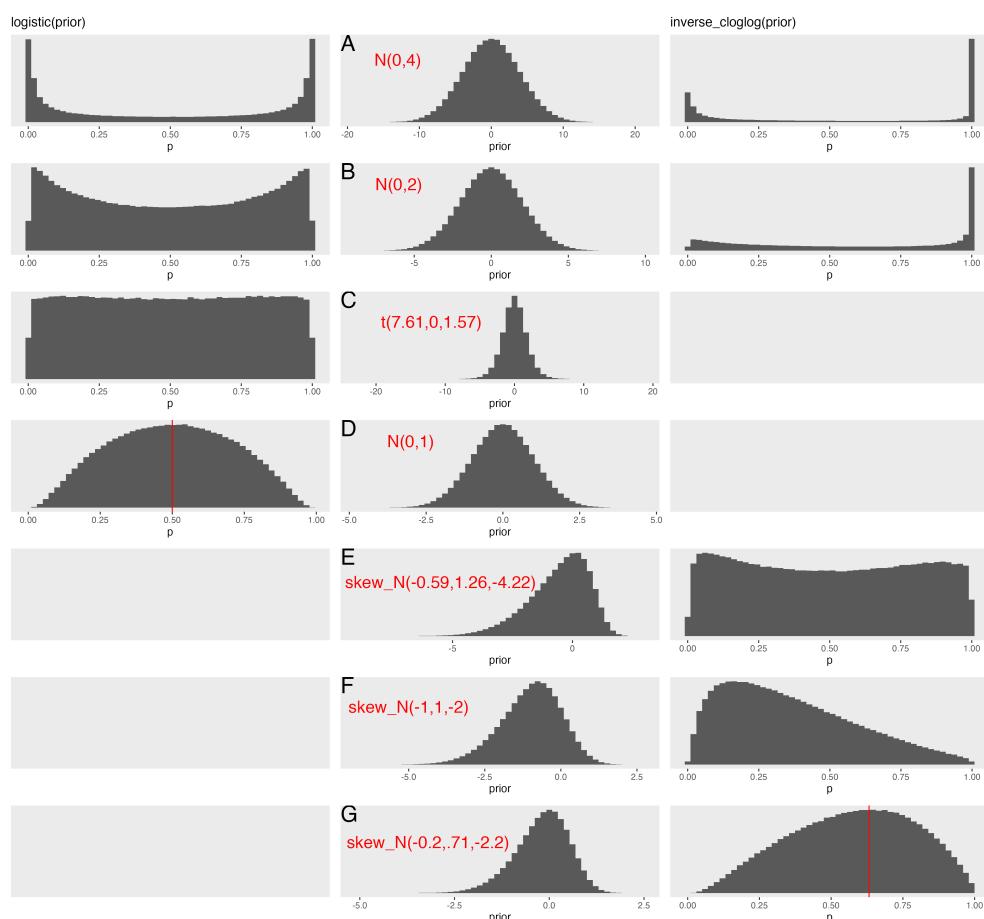


Figure 11. Prior distributions on the logit and/or cloglog scales (middle column), and their implications on the probability scale after applying the inverse-logit (or logistic) transformation (left column), and the inverse-cloglog transformation (right column).

1079 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).

1092 F. Advantages of hazard analysis

1093 Statisticians and mathematical psychologists recommend focusing on the hazard
1094 function when analyzing time-to-event data for various reasons. First, as discussed by
1095 Holden, Van Orden, and Turvey (2009), "probability density [and mass] functions can
1096 appear nearly identical, both statistically and to the naked eye, and yet are clearly different
1097 on the basis of their hazard functions (but not vice versa). Hazard functions are thus more
1098 diagnostic than density functions" (p. 331) when one is interested in studying the detailed
1099 shape of a RT distribution (see also Figure 1 in Panis, Schmidt, et al., 2020). Therefore,
1100 when the goal is to study how psychological effects change over time, hazard and
1101 conditional accuracy functions are preferred.

1102 Second, because RT distributions may differ from one another in multiple ways,
1103 Townsend (1990) developed a dominance hierarchy of statistical differences between two
1104 arbitrary distributions A and B. For example, if $h_A(t) > h_B(t)$ for all t, then both hazard
1105 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

₁₁₃₂ or computational models to data without studying the shape of the empirical discrete-time
₁₁₃₃ $h(t)$ and $ca(t)$ functions can miss important features in the data (Panis, Moran, et al.,
₁₁₃₄ 2020; Panis & Schmidt, 2016).