

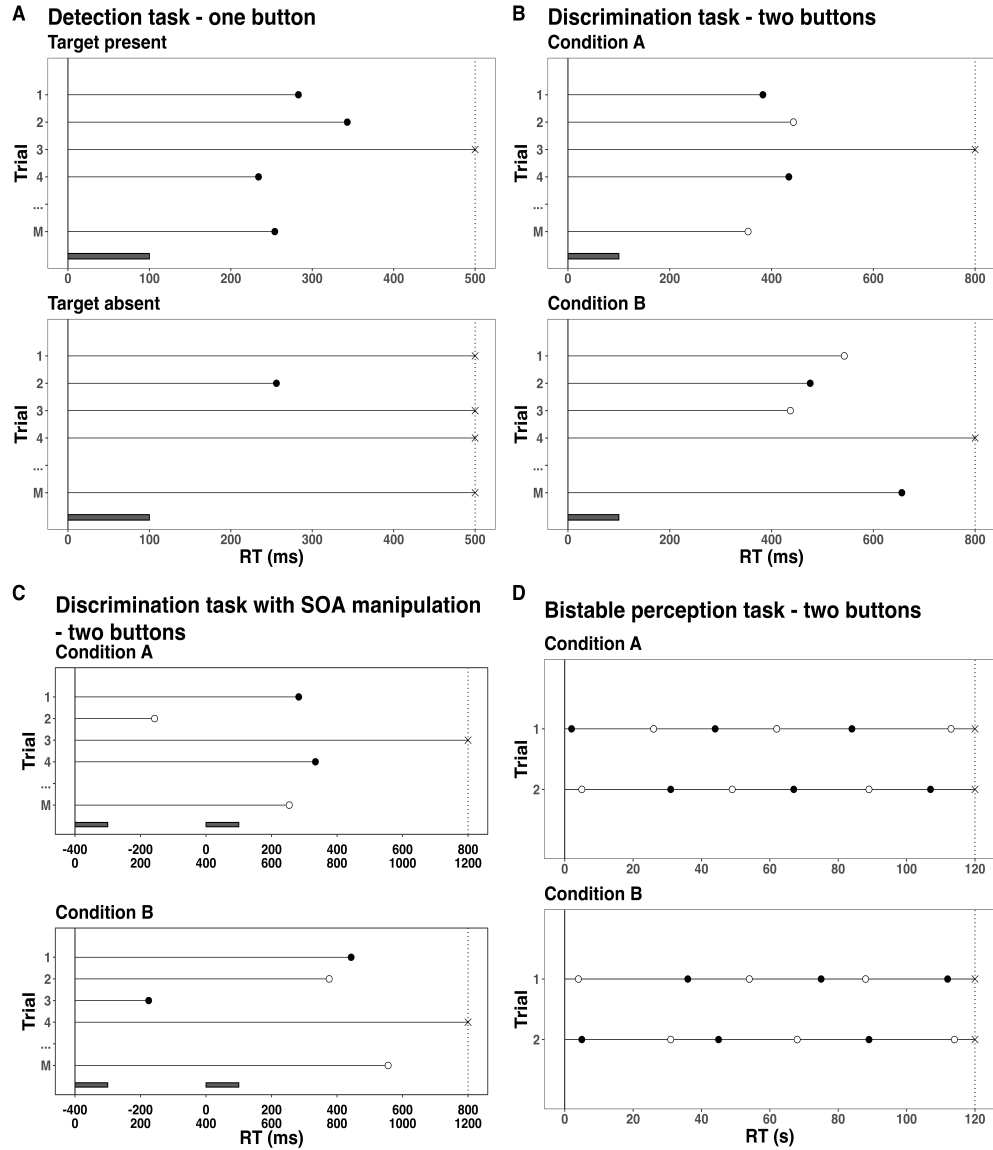
Supplemental Material

A. RT tasks in the context of EHA: right-censoring versus random censoring

While people can die only once, in experimental RT tasks the events of interest are typically repeatable. In Supplementary Figure 1A, we present a hypothetical single-subject RT data set from a detection experiment with one response button and two experimental conditions: target present and target absent. For example, each trial in each condition might start with a fixation screen for 1 s, followed by a blank screen for 1 s, a target display screen for 100 ms in which the target stimulus is presented or not depending on the experimental condition the trials belongs to, and finally a blank screen until an overt response is detected or 500 ms have passed. Supplementary Figure 1A shows the typical pattern: the participants presses the button sometime before 500 ms in most trials in the target-present condition (black dots), while (s)he does not press the button in most trials of the target-absent condition (black crosses).

In EHA parlance, the single event of interest is a button press response, *time zero* is defined as target display onset, the *observation period* is 500 ms long in each trial or repeated measurement, in each trial the participant is *at risk* for response occurrence as long as the response has not occurred yet, the individual starts in an “idle” state in each trial and *transitions* to a “detected” state when a response occurs, all trials have a censoring time of 500 ms, and some trials result in observed event times (those with a RT below 500 ms), while other trials result in response times that are right-censored at 500 ms. Right-censoring occurs when we only know that the event time is larger than some value, i.e., the censoring time. The most common type of right-censoring is “singly Type I censoring” which applies when the experiment uses a fixed response deadline for all trials. “Type I” means that the censoring time is fixed and under the control of the experimenter; “Singly” refers to the fact that all observations have the same censoring time (Allison, 2010). Clearly, discarding right-censored observations before calculating the median or

mean RT will result in biased estimates, especially in the target-absent condition.



Supplementary Figure 1. RT tasks and EHA concepts. The timer starts at time 0 (black vertical line) and ends at the censoring time (dotted vertical line). In each trial of each condition of each task, an overt response time (RT) is indicated by a circle. (A) Single-button detection data requires a single-event analysis. (B) Two-button discrimination data requires a competing risks analysis. (C) Competing risk data in combination with an stimulus-onset-asynchrony manipulation might require increasing the observation period in each trial. (D) Two-button bistable perception data requires a recurrent event / multistate analysis.

Importantly, when estimating the shape of the RT distribution in each experimental condition, EHA takes into account right-censored observations. As a result, EHA has a wider scope than current methods, as it can also be applied to RT data from experimental paradigms in which RTs are typically not measured, such as masking experiments and studies on the attentional blink. Also, EHA allows to increase experimental design efficiency, because we can introduce a response deadline instead of waiting until a response is emitted in each trial, so that we do not have to wait for very slow responses which would be trimmed anyway.

There are different types of censoring. Left censoring occurs when all that is known about an observation on a variable T is that it is less than some value. Interval censoring combines right and left censoring so that all you know about T is that $a < T < b$, for some values of a and b (Allison, 2010). Random censoring occurs when observations are terminated for reasons that are not under the control of the experimenter.

For example, in Supplementary Figure 1B, we present a hypothetical single-subject RT data set from a discrimination experiment with two response buttons and two experimental conditions: A and B. For example, each trial in each condition might start with a fixation screen for 1 s, followed by a blank screen for 1 s, a target display screen for 100 ms in which a female or male face is presented, and finally a blank screen until an overt response is detected or 800 ms have passed. Participants are required to press the left button when identifying a male, and the right button when identifying a female. In Supplementary Figure 1B, correct responses are indicated with black dots, error responses with white dots, and the few right-censored observations with crosses.

In EHA parlance this is known as a *competing risks* situation, because from target display screen onset (time zero) onwards, both correct and incorrect responses are competing with another in each trial until one occurs which determines the event type and latency that are recorded. In other words, in each trial the participant can transition from an idle state to

either a “correct response” state or an “incorrect response” state.

There are two ways to deal with a competing-risks situation. One approach is to model the hazard function of correct response occurrence and treat incorrect responses as right-censored observations, and vice versa. When estimating the hazard of correct response occurrence, then the unpredictable error responses will introduce random censoring (and vice versa). Importantly, all statistical methods for time-to-event data require that random censoring be noninformative: a trial that is censored at time c should be representative of all those trials with the same values of the explanatory variables that survive to c (Allison, 2010). However, the random censoring by errors when estimating the hazard function of correct responses (and vice versa) is very likely informative, because response channels are known to compete with one another (Burle et al., 2004; Eriksen et al., 1985; Praamstra & Seiss, 2005). Thus, another approach is to model the hazard of (any) response occurrence and extend it with a speed-accuracy trade-off (SAT) analysis (plotting accuracy as a function of latency). This is the approach taken here (see Figure 1), and many experimental psychologists are familiar with SAT.

Studying the effect of stimulus-onset-asynchrony (SOA) on mean RT is still a popular approach to study the time course of cognitive processes (ref). For example, a target stimulus can be presented with a certain SOA in relation to another stimulus (e.g., prime, mask, distractor, cue). However, mean RT conceals the possible presence of responses triggered by the non-target stimulus (Panis, 2020; Panis & Schmidt, 2016). For example, in Supplementary Figure 1C a prime is presented before a target with a SOA of -400 ms (upper x-axis labels). To capture responses triggered by the prime (and emitted before target onset) we have to start the timer at trial onset. Also, when RTs are originally recorded relative to target onset, we have to add a constant time to each (here: + 400 ms) in order to include these “negative” RTs in the analysis (lower x-axis labels). In general, all RTs must be positive for EHA to work.

Finally, in a bistable perception task (Supplementary Figure 1D), the participant is looking at an ambiguous stimulus (e.g., the duck-rabbit illusion, the Necker cube) for two minutes in each trial, for example, and asked to press a button each time when her/his perception switches from one possible interpretation to the other possible interpretation. In this task, there are two events (percept A switches to percept B, percept B switches to percept A) that can recur within the same observation period of two minutes, so that the individual transitions back and forth between two states. Note that we do not analyse recurrent events in this tutorial. More information about recurrent events analysis can be found in REF and REF...

B. Definitions of discrete-time hazard, survivor, probability mass, and 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

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

and the discrete-time survivor function

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

and not on the probability mass function

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

nor the cumulative distribution function

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

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

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

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

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

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

While psychological RT data is typically measured in small, continuous units (e.g., milliseconds), discrete-time EHA treats the RT data as interval-censored data, because it only uses the information that the response occurred sometime in a particular bin of time $(x,y]: x < RT \leq y$. If we want to use the exact event times, then we treat time as a continuous variable, and let RT be a continuous random variable denoting a particular

person's response time in a particular experimental condition. Continuous-time EHA does not focus on the cumulative distribution function $F(t) = P(RT \leq t)$ and its derivative, the probability density function $f(t) = F(t)'$, but on the survivor function $S(t) = P(RT > t)$ and the hazard rate function $\lambda(t) = f(t)/S(t)$. The hazard rate function gives you the instantaneous *rate* of event occurrence at time point t , given that the event has not occurred yet.

C. Custom functions for descriptive discrete-time hazard analysis

We defined 12 custom functions that we list here.

- `censor(df, timeout, bin_width)` : divide the time segment $(0, \text{timeout}]$ in bins, identify any right-censored observations, and determine the discrete RT (time bin rank)
- `ptb(df)` : transform the person-trial data set to the person-trial-bin data set
- `setup_lt(ptb)` : set up a life table for each level of 1 independent variable
- `setup_lt_2IV(ptb)` : set up a life table for each combination of levels of 2 independent variables
- `calc_ca(df)` : estimate the conditinal accuracies when there is 1 independent variable
- `calc_ca_2IV(df)` : estimate the conditional accuracies when there are 2 independent variables
- `join_lt_ca(df1, df2)` : add the $ca(t)$ estimates to the life tables (1 independent variable)
- `join_lt_ca_2IV(df1, df2)` : add the $ca(t)$ estimates to the life tables (2 independent variables)
- `extract_median(df)` : estimate quantiles $S(t)_{.50}$ (1 independent variable)
- `extract_median_2IV(df)` : estimate quantiles $S(t)_{.50}$ (2 independent variables)
- `plot_eha(df, subj, haz_yaxis=1, first_bin_shown=1, aggregated_data=F, Nsubj=6)` : create plots of the discrete-time functions (1 independent variable), and specify the

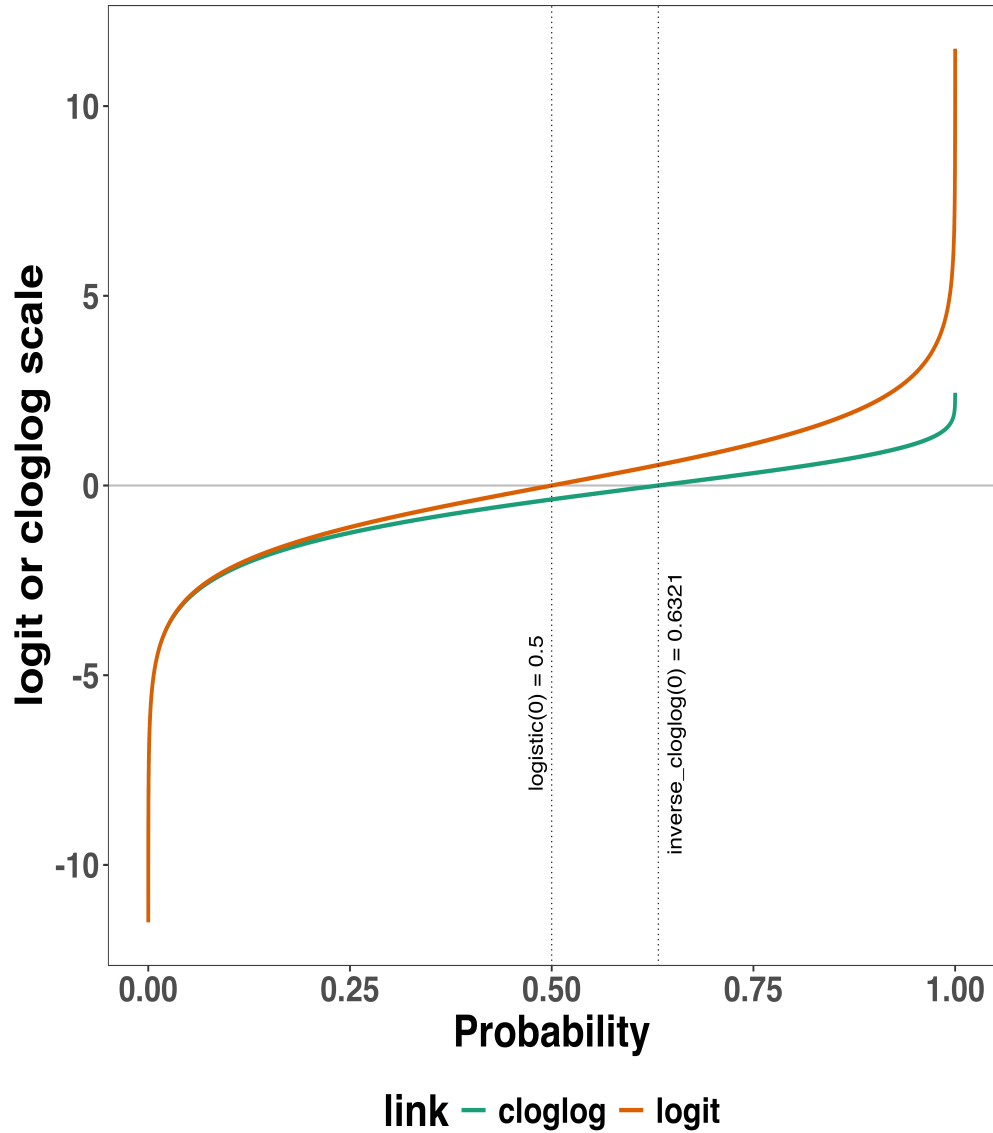
upper limit of the y-axis in the hazard plot, with which bin to start plotting, whether the data is aggregated across participants, and across how many participants

- `plot_aha_2IV(df, subj, haz_yaxis=1, first_bin_shown=1, aggregated_data=F, Nsubj=6)` : create plots of the discrete-time functions (2 independent variables), and specify the upper limit of the y-axis in the hazard plot, with which bin to start plotting, whether the data is aggregated across participants, and across how many participants

When you want to analyse simple RT data from a detection experiment with one independent variable, the functions `calc_ca()` and `join_lt_ca()` should not be used, and the code to plot the conditional accuracy functions should be removed from the function `plot_aha()`. When you want to analyse simple RT data from a detection experiment with two independent variables, the functions `calc_ca_2IV()` and `join_lt_ca_2IV()` should not be used, and the code to plot the conditional accuracy functions should be removed from the function `plot_aha_2IV()`.

D. Link functions

Popular link functions include the logit link and the complementary log-log link, as shown in Supplementary Figure 1.



Supplementary Figure 2. The logit and cloglog link functions.

E. Regression equations

An example (single-level) discrete-time hazard model with three predictors (TIME, X_1 , X_2), the cloglog link function, and a second-order polynomial specification for TIME can be written as follows:

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

The main predictor variable TIME is the time bin index t that is centered on value 9 in this example. The first set of terms within brackets, the parameters β_0 to β_2 multiplied by their polynomial specifications of (centered) time, represents the shape of the baseline cloglog-hazard function (i.e., when all predictors X_i take on a value of zero). The second set of terms (the beta parameters β_3 to β_5) represents the vertical shift in the baseline cloglog-hazard for a 1 unit increase in the respective predictor variable. Predictors can be discrete, continuous, and time-varying or time-invariant. For example, the effect of a 1 unit increase in X_1 is to vertically shift the whole baseline cloglog-hazard function by β_3 cloglog-hazard units. However, if the predictor interacts linearly with TIME (see X_2 in the example), then the effect of a 1 unit increase in X_2 is to vertically shift the predicted cloglog-hazard in bin 9 by β_4 cloglog-hazard units (when $\text{TIME}-9 = 0$), in bin 10 by $\beta_4 + \beta_5$ cloglog-hazard units (when $\text{TIME}-9 = 1$), and so forth. To interpret the effects of a predictor, its β parameter is exponentiated, resulting in a hazard ratio (due to the use of the cloglog link). When using the logit link, exponentiating a β parameter results in an odds ratio.

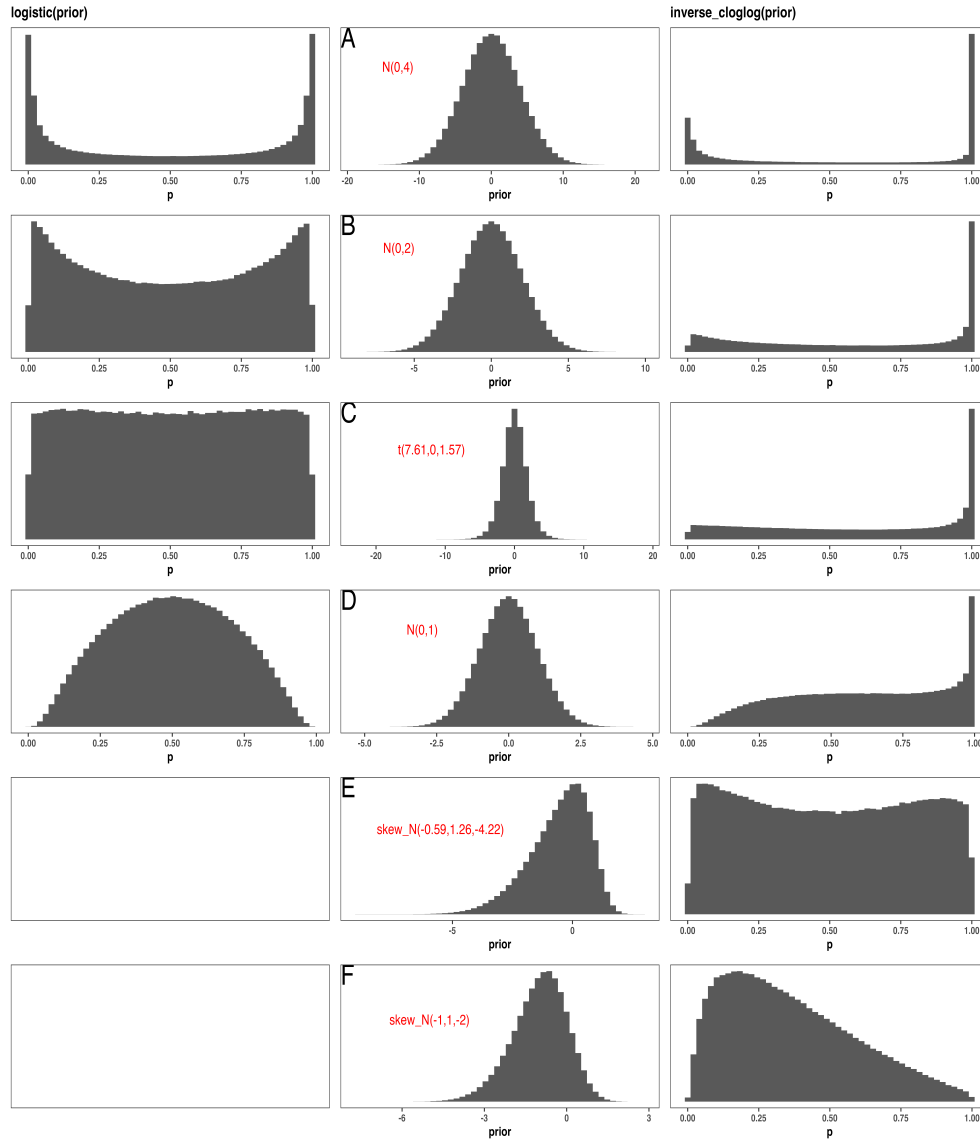
An example (single-level) discrete-time hazard model with a general specification for TIME (separate intercepts for each of six bins, where D1 to D6 are binary indicator variables identifying each bin) and a single predictor (X_1) can be written as follows:

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

F. Prior distributions

To gain a sense of what prior *logit* values would approximate a uniform distribution on the probability scale, Kurz (2023) simulated a large number of draws from the Uniform(0,1) distribution, converted those draws to the log-odds metric, and fitted a Student's t distribution. Row C in Supplementary Figure 2 shows that using a t -distribution with 7.61 degrees of freedom and a scale parameter of 1.57 as a prior on the logit scale, approximates

203 a uniform distribution on the probability scale. According to Kurz (2023), such a prior
 204 might be a good prior for the intercept(s) in a logit-hazard model, while the $N(0,1)$ prior in
 205 row D might be a good prior for the non-intercept parameters in a logit-hazard model, as it
 206 gently regularizes p towards .5 (i.e., a zero effect on the logit scale).



Supplementary Figure 3. Prior distributions for the Intercept 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).

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.

G. Advantages of hazard analysis

Statisticians and mathematical psychologists recommend focusing on the hazard function when analyzing time-to-event data for various reasons. First, as discussed by Holden, Van Orden, and Turvey (2009), “probability density [and mass] functions can appear nearly identical, both statistically and to the naked eye, and yet are clearly different on the basis of their hazard functions (but not vice versa). Hazard functions are thus more diagnostic than density functions” (p. 331) when one is interested in studying the detailed shape of a RT distribution (see also Figure 1 in Panis, Schmidt, Wolkersdorfer, & Schmidt, 2020). Therefore, when the goal is to study how psychological effects change over time, hazard and conditional accuracy functions are the preferred ways to describe the RT + accuracy data. Second, because RT distributions may differ from one another in multiple ways, Townsend (1990) developed a dominance hierarchy of statistical differences between two arbitrary distributions A and B. For example, if $h_A(t) > h_B(t)$ for all t , then both hazard 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 one to incorporate time-varying explanatory covariates, such as heart rate, electroencephalogram (EEG) signal amplitude, gaze location, pupil dilation, etc. (Allison, 2010). This is useful for linking physiological effects to behavioral effects when performing cognitive psychophysiology (Meyer, Osman, Irwin, & Yantis, 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 or computational models to data without studying the shape of the empirical discrete-time $h(t)$ and $ca(t)$ functions can miss important features in the data, such as short-lived “dips”

that could signal the presence of active top-down response inhibition processes (Panis, Moran, Wolkersdorfer, & Schmidt, 2020; Panis & Schmidt, 2016). As such, EHA can be a tool to help distinguish between competing theories of cognition and brain function.

References

- Allison, P. D. (2010). *Survival analysis using SAS: A practical guide* (2. ed). Cary, NC: SAS Press.
- Holden, J. G., Van Orden, G. C., & Turvey, M. T. (2009). Dispersion of response times reveals cognitive dynamics. *Psychological Review*, 116(2), 318–342.
<https://doi.org/10.1037/a0014849>
- Kantowitz, B. H., & Pachella, R. G. (2021). The Interpretation of Reaction Time in Information-Processing Research 1. *Human Information Processing*, 41–82.
<https://doi.org/10.4324/9781003176688-2>
- Kelso, J. A. S., Dumas, G., & Tognoli, E. (2013). Outline of a general theory of behavior and brain coordination. *Neural Networks: The Official Journal of the International Neural Network Society*, 37, 120–131. <https://doi.org/10.1016/j.neunet.2012.09.003>
- Kurz, A. S. (2023). *Applied longitudinal data analysis in brms and the tidyverse* (version 0.0.3). Retrieved from <https://bookdown.org/content/4253/>
- Luce, R. D. (1991). *Response times: Their role in inferring elementary mental organization* (1. issued as paperback). Oxford: Univ. Press.
- Meyer, D. E., Osman, A. M., Irwin, D. E., & Yantis, S. (1988). Modern mental chronometry. *Biological Psychology*, 26(1-3), 3–67.
[https://doi.org/10.1016/0301-0511\(88\)90013-0](https://doi.org/10.1016/0301-0511(88)90013-0)
- Panis, S., Moran, R., Wolkersdorfer, M. P., & Schmidt, T. (2020). Studying the dynamics of visual search behavior using RT hazard and micro-level speed–accuracy tradeoff functions: A role for recurrent object recognition and cognitive control processes. *Attention, Perception, & Psychophysics*, 82(2), 689–714.

285 <https://doi.org/10.3758/s13414-019-01897-z>

286 Panis, S., Schmidt, F., Wolkersdorfer, M. P., & Schmidt, T. (2020). Analyzing Response

287 Times and Other Types of Time-to-Event Data Using Event History Analysis: A Tool

288 for Mental Chronometry and Cognitive Psychophysiology. *I-Perception*, 11(6),

289 2041669520978673. <https://doi.org/10.1177/2041669520978673>

290 Panis, S., & Schmidt, T. (2016). What Is Shaping RT and Accuracy Distributions? Active

291 and Selective Response Inhibition Causes the Negative Compatibility Effect. *Journal of*

292 *Cognitive Neuroscience*, 28(11), 1651–1671. https://doi.org/10.1162/jocn_a_00998

293 Townsend, J. T. (1990). Truth and consequences of ordinal differences in statistical

294 distributions: Toward a theory of hierarchical inference. *Psychological Bulletin*, 108(3),

295 551–567. <https://doi.org/10.1037/0033-2909.108.3.551>

296 Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics.

297 *Acta Psychologica*, 41(1), 67–85. [https://doi.org/10.1016/0001-6918\(77\)90012-9](https://doi.org/10.1016/0001-6918(77)90012-9)