₁ **Supplemental Material**

₂ **A. RT tasks in the context of EHA: right-censoring versus random censoring**

₃ While people can die only once, in experimental RT tasks the events of interest are

₄ typically repeatable. In Supplementary Figure 1A, we present a hypothetical single-subject

₅ RT data set from a detection experiment with one response button and two experimental

₆ conditions: target present and target absent. For example, each trial in each condition

₇ might start with a fixation screen for 1 s, followed by a blank screen for 1 s, a target

₈ display screen for 100 ms in which the target stimulus is presented or not depending on the

₉ experimental condition the trials belongs to, and finally a blank screen until an overt

₁₀ response is detected or 500 ms have passed. Supplementary Figure 1A shows the typical

₁₁ pattern: the participants presses the button sometime before 500 ms in most trials in the

₁₂ target-present condition (black dots), while (s)he does not press the button in most trials

₁₃ of the target-absent condition (black crosses).

₁₄ In EHA parlance, the single event of interest is a button press response, *time zero* is defined

₁₅ as target display onset, the *observation or follow-up period* is 500 ms long in each trial or

₁₆ repeated measurement, in each trial the participant is *at risk* for response occurrence as

₁₇ long as the response has not occurred yet, the individual starts in an "idle" state in each

₁₈ trial and *transitions* to a "detected" state when a response occurs, all trials have a

₁₉ censoring time of 500 ms, and some trials result in observed event times (those with a RT

₂₀ below 500 ms), while other trials result in response times that are right-censored at 500 ms.

₂₁ Right-censoring occurs when we only know that the event time is larger than some value,

₂₂ i.e., the censoring time. The most common type of right-censoring is "singly Type I

₂₃ censoring" which applies when the experiment uses a fixed response deadline for all trials.

₂₄ "Type I" means that the censoring time is fixed and under the control of the experimenter;

₂₅ "Singly" refers to the fact that all observations have the same censoring time (Allison,

₂₆ 2010). Clearly, discarding right-censored observations before calculating the median or

27 mean RT will result in biased estimates, especially in the target-absent condition.
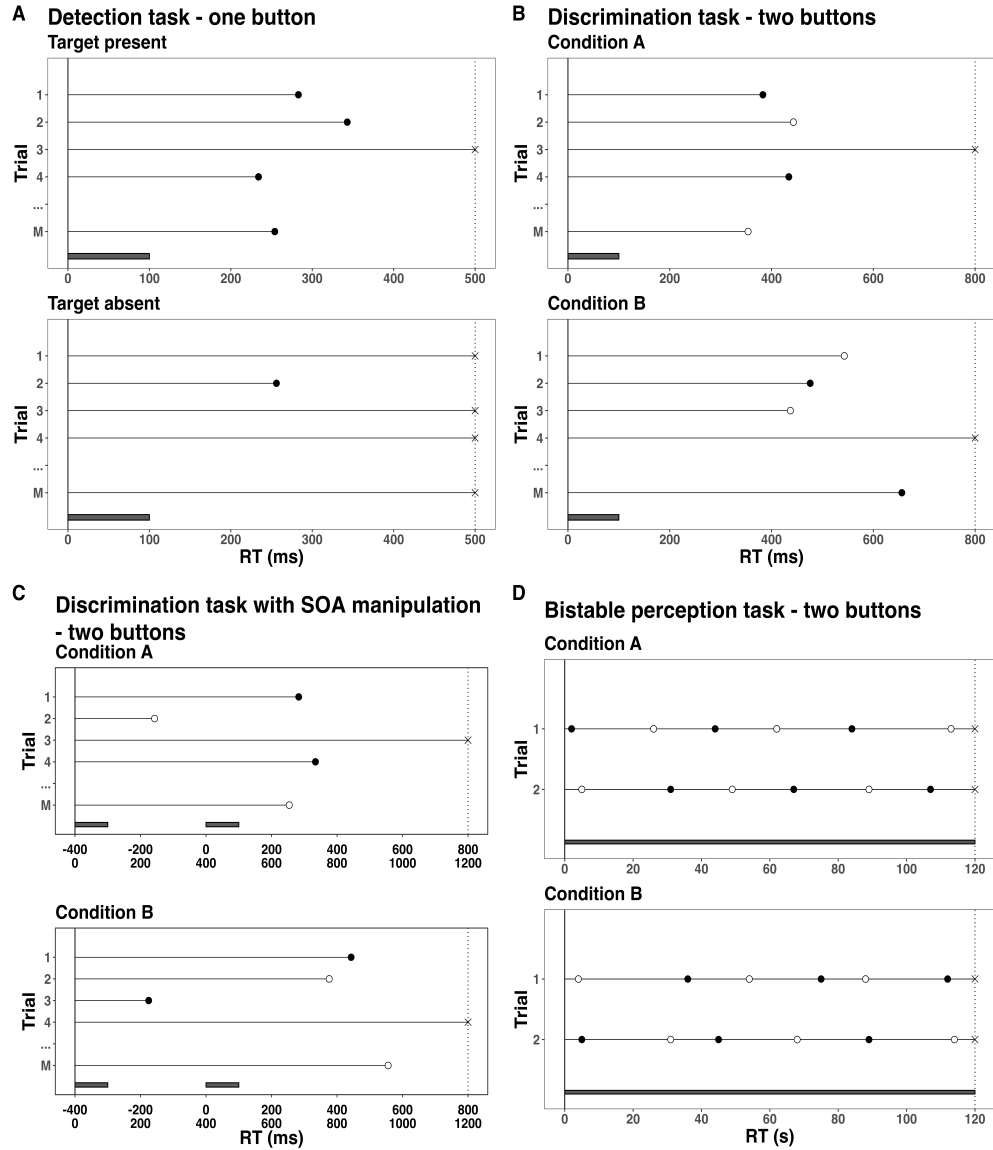
Supplementary Figure 1. RT tasks and EHA concepts. The timer starts at time 0 (black vertical line) and ends at the censoring time (dotted vertical line). In each trial of each condition of each task, an overt response time (RT) is indicated by a circle. (A) Single-button detection data requires a single-event 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. Black bars indicate stimulus presentation.

28 Importantly, when estimating the shape of the RT distribution in each experimental

29 condition, EHA takes into account right-censored observations. As a result, EHA has a

30 wider scope than current methods, as it can also be applied to RT data from experimental

31 paradigms in which RTs are typically not measured, such as masking experiments and

32 studies on the attentional blink. Also, EHA allows to increase experimental design

33 efficiency, because we can introduce a response deadline instead of waiting until a response

34 is emitted in each trial, so that we do not have to wait for very slow responses which would

35 be trimmed anyway.

36 There are different types of censoring. Left censoring occurs when all that is known about

37 an observation on a variable T is that it is less than some value. Interval censoring

38 combines right and left censoring so that all you know about T is that $a < T < b$, for some

39 values of a and b (Allison, 2010). Random censoring occurs when observations are

40 terminated for reasons that are not under the control of the experimenter.

41 In Supplementary Figure 1B, we present a hypothetical single-subject RT data set from a

42 discrimination experiment with two response buttons and two experimental conditions: A

43 and B. For example, each trial in each condition might start with a fixation screen for 1 s,

44 followed by a blank screen for 1 s, a target display screen for 100 ms in which a female or

45 male face is presented, and finally a blank screen until an overt response is detected or 800

46 ms have passed. Participants are required to press the left button when identifying a male,

47 and the right button when identifying a female. In Supplementary Figure 1B, correct

48 responses are indicated with black dots, error responses with white dots, and the few

49 right-censored observations with crosses.

50 In EHA parlance this is known as a *competing risks* situation, because from target display

51 screen onset (time zero) onwards, both correct and incorrect responses are competing with

52 another in each trial until one occurs which determines the event type and latency that are

53 recorded. In other words, in each trial the participant can transition from an idle state to

54 either a "correct response" state or an "incorrect response" state.

55 There are two ways to deal with a competing-risks situation. One approach is to model the

56 hazard function of correct response occurrence and treat incorrect responses as

57 right-censored observations, and vice versa. When estimating the hazard of correct

58 response occurrence, then the unpredictable error responses will introduce random

59 censoring (and vice versa). Importantly, all statistical methods for time-to-event data

60 require that random censoring be noninformative: a trial that is censored at time c should

61 be representative of all those trials with the same values of the explanatory variables that

62 survive to c (Allison, 2010). However, the random censoring by errors when estimating the

63 hazard function of correct responses (and vice versa) is very likely informative, because

64 response channels are known to compete with one another (Burle et al., 2004; Eriksen et

65 al., 1985; Praamstra & Seiss, 2005). Thus, another approach is to model the hazard of

66 (any) response occurrence and extend it with a speed-accuracy trade-off (SAT) analysis

67 (plotting accuracy as a function of latency). This is the approach taken here (see Figure 1),

68 and many experimental psychologists are familiar with SAT.

69 Studying the effect of stimulus-onset-asynchrony (SOA) on mean RT is still a popular

70 approach to study the time course of cognitive processes. For example, a target stimulus

71 can be presented with a certain SOA in relation to another stimulus (e.g., prime, mask,

72 distractor, cue). However, mean RT conceals the possibe presence of responses triggered by

73 the non-target stimulus (Panis, 2020; Panis & Schmidt, 2016). For example, in

74 Supplementary Figure 1C a prime is presented before a target with a SOA of -400 ms

75 (upper x-axis labels). To capture responses triggered by the prime (and emitted before

76 target onset) we have to start the timer at trial onset. Also, when RTs are originally

77 recorded relative to target onset, we have to add a constant time to each (here: + 400 ms)

78 in order to include these prime-triggered "negative" RTs in the analysis (lower x-axis

79 labels). In general, all RTs must be positive for EHA to work.

80   Finally, in a bistable perception task (Supplementary Figure 1D), the participant is looking

81   at an ambiguous stimulus (e.g., the duck-rabbit illusion, the Necker cube) for two minutes

82   in each trial, for example, and asked to press a button each time when her/his perception

83   switches from one possibe interpretation to the other possible interpretation. In this task,

84   there are two events (percept A switches to percept B, percept B switches to percept A)

85   that can recur within the same observation period of two minutes, so that the individual

86   transitions back and forth between two states. Note that we do not analyse recurrent

87   events in this tutorial. More information about recurrent events analysis can be found in

88   Stoolmiller and Snyder (2006), Mills (2011), and Lougheed, Benson, Cole, and Ram (2019).

89   **B. Definitions of discrete-time hazard, surivor, probability mass, and**

90   **conditional accuracy functions**

91   The shape of a distribution of waiting times can be described in multiple ways (Luce,

92   1991). After dividing time in discrete, contiguous time bins indexed by t, let RT be a

93   discrete random variable denoting the rank of the time bin in which a particular person's

94   response occurs in a particular trial. Because waiting times can only increase, discrete-time

95   EHA focuses on the discrete-time hazard function

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

97   and the discrete-time survivor function

98   $S(t) = P(RT > t) = [1\text{-}h(t)].[1\text{-}h(t\text{-}1)].[1\text{-}h(t\text{-}2)]\dots[1\text{-}h(1)]$ (2)

99   and not on the probability mass function

100 $P(t) = P(RT = t) = h(t).S(t\text{-}1)$ (3)

101 nor the cumulative distribution function

102 $F(t) = P(RT \leq t) = 1\text{-}S(t)$ (4)

103 The discrete-time hazard function of event occurrence gives you for each bin the

104 probability that the event occurs (sometime) in that bin, given that the event has not

105 occurred yet in previous bins. This conditionality in the definition of hazard is what makes

106 the hazard function so diagnostic for studying event occurrence, as an event can physically

107 not occur when it has already occurred before.

108 While the discrete-time hazard function assesses the unique risk of event occurrence

109 associated with each time bin, the discrete-time survivor function cumulates the bin-by-bin

110 risks of event *non*occurrence to obtain the probability that the event does not occur before

111 the endpoint of bin t.

112 The probability mass function cumulates the risk of event occurrence in bin t with the risks

113 of event nonoccurrence in bins 1 to t-1. From equation 3 we find that hazard in bin t is

114 equal to P(t)/S(t-1).

115 The survivor function provides a context for the hazard function, as $S(t-1) = P(RT > t-1)$

116 $= P(RT \geq t)$ tells you on how many percent of the trials the estimate $h(t) = P(RT = t|$

117 $RT \geq t)$ is based. The probability mass function provides a context for the conditional

118 accuracy function, as $P(t) = P(RT = t)$ tells you on how many percent of the trials the

119 estimate $ca(t) = P(correct \mid RT = t)$ is based.

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

121 discrete-time conditional accuracy function

122 $$ca(t) = P(correct \mid RT = t) \tag{5}$$

123 which gives you for each bin the probability that a response is correct given that it is

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

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

126 While psychological RT data is typically measured in small, continuous units (e.g.,

127 milliseconds), discrete-time EHA treats the RT data as interval-censored data, because it

128 only uses the information that the response occurred sometime in a particular bin of time

129 (x,y]: $x < RT \leq y$. If we want to use the exact event times, then we treat time as a

130 continuous variable, and let RT be a continous random variable denoting a particular

131 person's response time in a particular trial. Continuous-time EHA does not focus on the

132 cumulative distribution function $F(t) = P(RT \leq t)$ and its derivative, the probability

133 density function $f(t) = F(t)'$, but on the survivor function $S(t) = P(RT > t)$ and the

134 hazard rate function $\lambda(t) = f(t)/S(t)$. The hazard rate function gives you the instantaneous

135 *rate* of event occurrence at time point t, given that the event has not occurred yet. Models

136 for continous-time event data come in three forms: parametric models, Cox regression, and

137 piece-wise exponential models (Singer & Willett, 2003).

C. Custom functions for descriptive discrete-time hazard analysis

139 We defined 12 custom functions that we list here.

140 • censor(df,timeout,bin_width) : divide the time segment (0,timeout] in bins, identify

141   any right-censored observations, and determine the discrete RT (time bin rank)

142 • ptb(df) : transform the person-trial data set to the person-trial-bin data set

143 • setup_lt(ptb) : set up a life table for each level of 1 independent variable

144 • setup_lt_2IV(ptb) : set up a life table for each combination of levels of 2

145   independent variables

146 • calc_ca(df) : estimate the conditinal accuracies when there is 1 independent variable

147 • calc_ca_2IV(df) : estimate the conditional accuraies when there are 2 independent

148   variables

149 • join_lt_ca(df1,df2) : add the ca(t) estimates to the life tables (1 independent

150   variable)

151 • join_lt_ca_2IV(df1, df2) : add the ca(t) estimates to the life tables (2 independent

152   variables)

153 • extract_median(df) : estimate quantiles $S(t)_{.50}$ (1 independent variable)

154 • extract_median_2IV(df) : estimate quantiles $S(t)_{.50}$ (2 independent variables)

155   • plot_eha(df, subj, haz_yaxis=1, first_bin_shown=1, aggregated_data=F, Nsubj=6)

156   : create plots of the discrete-time functions (1 independent variable), and specify the

157   upper limit of the y-axis in the hazard plot, with which bin to start plotting, whether

158   the data is aggregated across participants, and across how many participants

159   • plot_eha_2IV(df, subj, haz_yaxis=1, first_bin_shown=1, aggregated_data=F,

160   Nsubj=6) : create plots of the discrete-time functions (2 independent variables), and

161   specify the upper limit of the y-axis in the hazard plot, with which bin to start

162   plotting, whether the data is aggregated across participants, and across how many

163   participants

164   When you want to analyse simple RT data from a detection experiment with one

165   independent variable, the functions calc_ca() and join_lt_ca() should not be used, and

166   the code to plot the conditional accuracy functions should be removed from the function

167   plot_eha(). When you want to analyse simple RT data from a detection experiment with

168   two independent variables, the functions calc_ca_2IV() and join_lt_ca_2IV() should not

169   be used, and the code to plot the conditional accuracy functions should be removed from

170   the function plot_eha_2IV().

171   **D. Link functions**

172   Popular link functions include the logit link and the complementary log-log link, as shown

173   in Supplementary Figure 2.
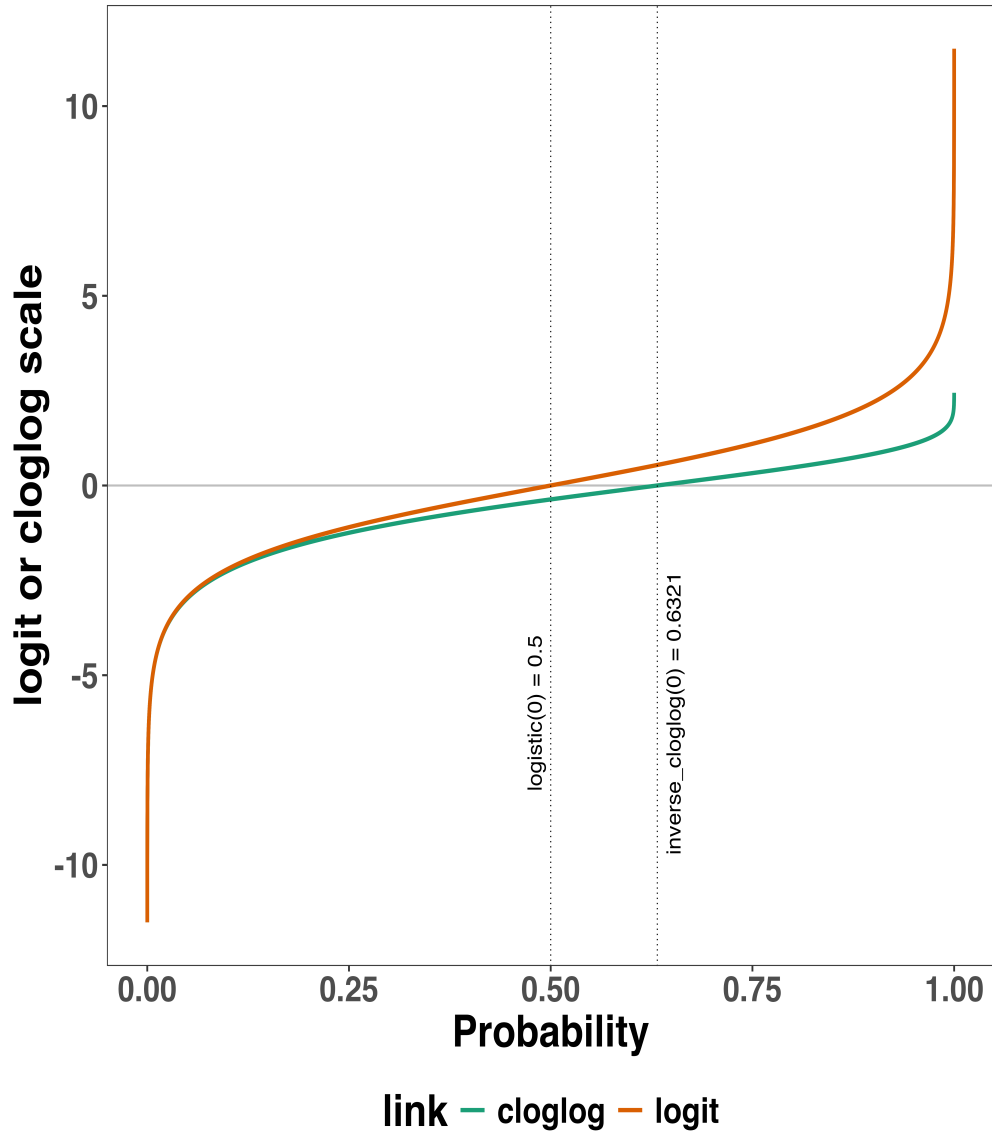
Supplementary Figure 2. The logit and complementary log-log (cloglog) link functions.

E. Regression equations

An example (single-level) discrete-time hazard model for individual i in bin t of trial j 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:

$$\text{cloglog}[h_{ij}(t)] = \ln(-\ln[1-h_{ij}(t)]) = [\beta_0 \text{ONE} + \beta_1(\text{TIME-9}) + \beta_2(\text{TIME-9})^2] + [\beta_3 X_{1ij} + \beta_4 X_{2ij} + \beta_5 X_{2ij}(\text{TIME-9})] \tag{6}$$

180   The main predictor variable TIME is the time bin index t that is centered on value 9 in

181   this example. The first set of terms within brackets, the parameters β_0 to β_2 multiplied by

182   their polynomial specifications of (centered) time, represents the shape of the baseline

183   cloglog-hazard function (i.e., when all predictors X_{ij} take on a value of zero). The second

184   set of terms (the beta parameters β_3 to β_5) represents the vertical shift in the baseline

185   cloglog-hazard for a 1 unit increase in the respective predictor variable. Predictors can be

186   discrete, continuous, and time-varying or time-invariant. For example, the effect of a 1 unit

187   increase in X_{1ij} is to vertically shift the whole baseline cloglog-hazard function by β_3

188   cloglog-hazard units. However, if the predictor interacts linearly with TIME (see X_{2ij} in the

189   example), then the effect of a 1 unit increase in X_{2ij} is to vertically shift the predicted

190   cloglog-hazard in bin 9 by β_4 cloglog-hazard units (when TIME-9 = 0), in bin 10 by $\beta_4 +$

191   β_5 cloglog-hazard units (when TIME-9 = 1), and so forth. To interpret the effects of a

192   predictor, its β parameter is exponentiated, resulting in a hazard ratio (due to the use of

193   the cloglog link). When using the logit link, exponentiating a β parameter results in an

194   odds ratio.

195   An example (single-level) discrete-time hazard model for individual i in bin t of trial j with

196   a general specification for TIME (separate intercepts for each of six bins, where D1 to D6

197   are binary indicator variables identifying each bin) and a single predictor (X_{1ij}) can be

198   written as follows:

199    $$\text{cloglog}[h_{ij}(t)] = [\beta_0 D1 + \beta_1 D2 + \beta_2 D3 + \beta_3 D4 + \beta_4 D5 + \beta_5 D6] + [\beta_6 X_{1ij}] \tag{7}$$

200   **F. Prior distributions**

201   To gain a sense of what prior *logit* values would approximate a uniform distribution on the

202   probability scale, Kurz (2023) simulated a large number of draws from the Uniform(0,1)

203   distribution, converted those draws to the log-odds metric, and fitted a Student's t

204   distribution. Row C in Supplementary Figure 3 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 a uniform distribution on the probability scale. According to Kurz (2023), such a prior might be a good prior for the intercept(s) in a logit-hazard model, while the N(0,1) prior in row D might be a good prior for the non-intercept parameters in a logit-hazard model, as it gently regularizes p towards .5 (i.e., a zero effect on the logit scale).
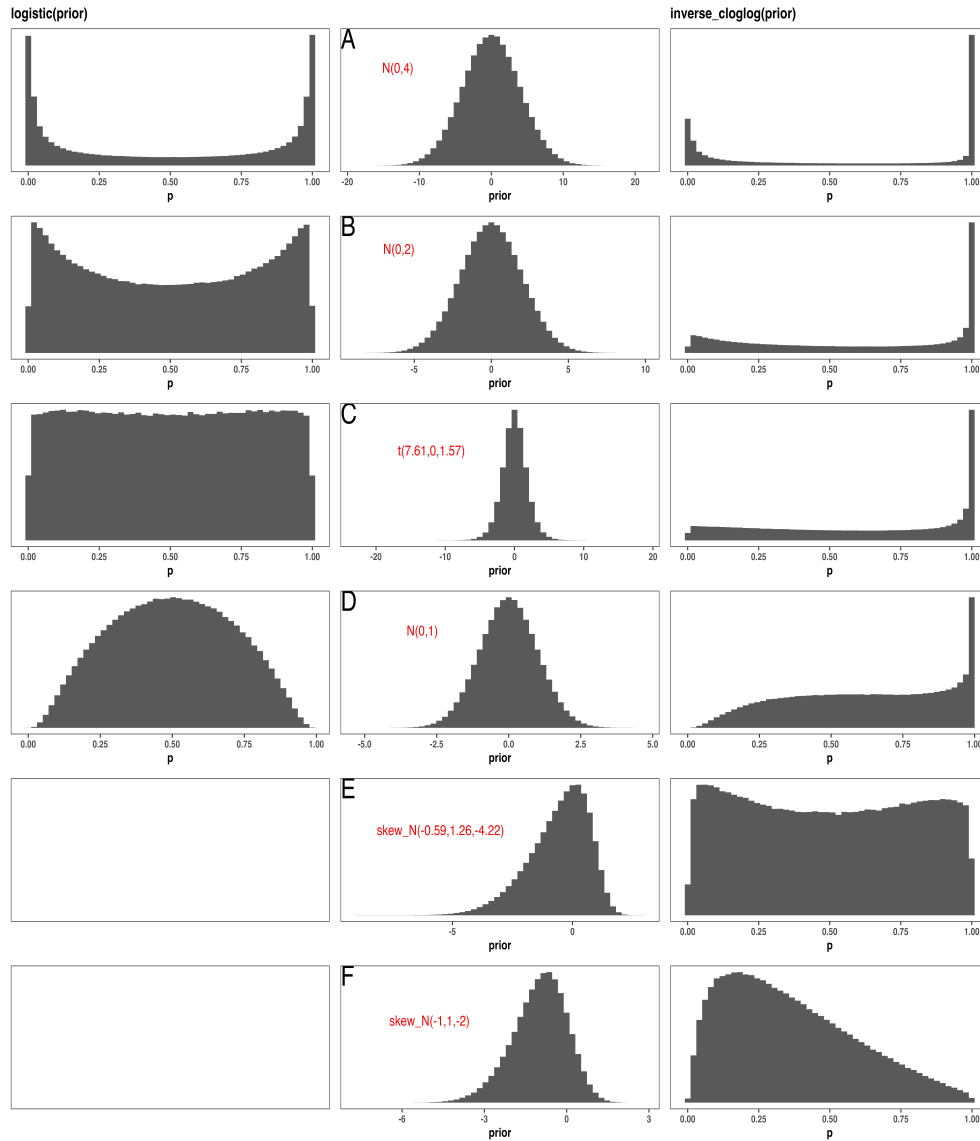
Supplementary Figure 3. Prior distributions for the Intercept on the logit and/or cloglog scales (middle column), and their implications on the probability scale after applying the inverse-logit (or logistic) transformation (left column), and the inverse-cloglog transformation (right column).

210 To gain a sense of what prior *cloglog* values would approximate a uniform distribution on
211 the hazard probability scale, we followed Kurz's approach and simulated a large number of
212 draws from the Uniform(0,1) distribution, converted them to the cloglog metric, and fitted

213 a skew-normal model (due to the asymmetry of the cloglog link function). Row E shows

214 that using a skew-normal distribution with a mean of -0.59, a standard deviation of 1.26,

215 and a skewness of -4.22 as a prior on the cloglog scale, approximates a uniform distribution

216 on the probability scale. However, because hazard values below .5 are more likely in RT

217 studies, using a skew-normal distribution with a mean of -1, a standard deviation of 1, and

218 a skewness of -2 as a prior on the cloglog scale (row F), might be a good weakly

219 informative prior for the intercept(s) in a cloglog-hazard model.

G. Advantages of EHA

221 Statisticians and mathematical psychologists recommend focusing on the hazard function

222 when analyzing time-to-event data for various reasons. First, as discussed by Holden, Van

223 Orden, and Turvey (2009), "probability density [and mass] functions can appear nearly

224 identical, both statistically and to the naked eye, and yet are clearly different on the basis

225 of their hazard functions (but not vice versa). Hazard functions are thus more diagnostic

226 than density functions" (p. 331) when one is interested in studying the detailed shape of a

227 RT distribution (see also Figure 1 in Panis, Schmidt, Wolkersdorfer, & Schmidt, 2020).

228 Therefore, when the goal is to study how psychological effects change over time, hazard and

229 conditional accuracy functions are the preferred ways to describe the RT + accuracy data.

230 Second, because RT distributions may differ from one another in multiple ways, Townsend

231 (1990) developed a dominance hierarchy of statistical differences between two arbitrary

232 distributions A and B. For example, if $h_A(t) > h_B(t)$ for all t, then both hazard functions

233 are said to show a complete ordering. Townsend (1990) concluded that stronger

234 conclusions can be drawn from data when comparing the hazard functions using EHA. For

235 example, when mean A < mean B, the hazard functions might show a complete ordering

236 (i.e., for all t), a partial ordering (e.g., only for t > 300 ms, or only for t < 500 ms), or they

237 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 observation 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. Ignoring trials without a response leads to underestimation of the true mean. By introducing a fixed censoring time for all trials, 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 respone 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/

Lougheed, J. P., Benson, L., Cole, P. M., & Ram, N. (2019). Multilevel survival analysis: Studying the timing of children's recurring behaviors. *Developmental Psychology, 55*(1), 53–65. https://doi.org/10.1037/dev0000619

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

Mills, M. (2011). *Introducing Survival and Event History Analysis.* 1 Oliver's Yard, 55 City Road, London EC1Y 1SP United Kingdom: SAGE Publications Ltd. https://doi.org/10.4135/9781446268360

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

290 functions: A role for recurrent object recognition and cognitive control processes.

291 *Attention, Perception, & Psychophysics*, *82*(2), 689–714.

292 https://doi.org/10.3758/s13414-019-01897-z

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

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

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

296 2041669520978673. https://doi.org/10.1177/2041669520978673

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

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

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

300 Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling*

301 *Change and Event Occurrence.* Oxford, New York: Oxford University Press.

302 Stoolmiller, M., & Snyder, J. (2006). Modeling heterogeneity in social interaction processes

303 using multilevel survival analysis. *Psychological Methods*, *11*(2), 164–177.

304 https://doi.org/10.1037/1082-989X.11.2.164

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

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

307 551–567. https://doi.org/10.1037/0033-2909.108.3.551

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

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