



City Research Online

City St George's, University of London

Citation: Yarrow, K., Saurels, B., Arnold, D. H., Tal-Perry, N. & Solomon, J. A. (2026). How beliefs about the warning-target foreperiod inform temporal preparation to determine simple reaction times. Psychological review,

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/38356/>

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

How beliefs about the warning-target foreperiod inform temporal preparation to determine simple reaction times

Abbreviated Title: Modelling foreperiod effects on simple RT

Kielan Yarrow^{1*}, Blake Saurels², Derek H. Arnold², Noam Tal-Perry³, & Joshua A. Solomon⁴

¹ *Department of Psychology and Neuroscience, City St George's, University of London, London EC1V 0HB*

² *School of Psychology, The University of Queensland, Brisbane QLD 4072*

³ *Research Office and Bennett Institute for Public Policy, University of Cambridge, Cambridge, CB2 1TN*

⁴ *School of Health & Medical Sciences, City St George's, University of London, London EC1V 0HB*

* Author for correspondence:

Kielan Yarrow,
Rhind Building,
City St George's, University of London,
Northampton Square,
London EC1V 0HB

Tel: +44 (0)20 7040 8530

Email: kielan.yarrow.1@city.ac.uk

Author Note:

All raw data and analysis scripts are available at <https://doi.org/10.25383/city.30715439>.

Some elements of the research reported here were pre-registered

(<https://aspredicted.org/kt3qu.pdf>) as indicated more fully in the relevant methods sections.

Preliminary aspects of this research were presented at the ECVF 2024 conference, and the original manuscript (prior to peer review) was deposited at https://osf.io/preprints/psyarxiv/69xvk_v1.

31 Abstract

32 Events often unfold over a time course that is amenable to prediction, as when a sprinter learns the
33 constant foreperiod from set to go. This lets us temporally prepare. For many foreperiod
34 distributions, the conditional probability of an event (e.g. of a traffic light turning green) increases
35 over time, so preparation should similarly increase, improving reaction times (RTs). Modelling has
36 demonstrated such a quantitative match between *mean* RT across different foreperiods and the
37 (inverted) subjective hazard function (or some related preparation function). However, modelling
38 has provided few insights into how temporal preparation sculpts the tradeoff between RTs and
39 anticipatory errors. Furthermore, the standard simplifying assumption that the subjective foreperiod
40 distribution is stable offers no account of well-established history effects. To address these and
41 several further issues, here we present a modelling architecture that incorporates an initial Stimulus
42 Independent Process into a Non-Stationary Poisson account of simple Reaction Time (SIP-theN-
43 SPuRT). SIP-theN-SPuRT derives a temporal preparation signal from an estimated foreperiod
44 distribution that is updated from trial to trial. This signal, which may saturate at a physiological limit,
45 feeds into three processes, reflecting urgency, modulation of drive from change detectors, and
46 residual sensorimotor delays. We compared the predictions of nine SIP-theN-SPuRT model variants
47 with new and existing simple-RT datasets. Best fits were obtained by the model in which each trial's
48 putative preparation signal gradually asymptotes (or saturates), after initially following a subjective
49 hazard function. SIP-theN-SPuRT is, to our knowledge, unprecedented in explaining both foreperiod
50 and history effects across entire simple RT distributions.

51

52 Keywords

53 Temporal expectation, foreperiod effects, interval timing, simple RT, multilevel models

54

Events of interest often unfold over time in a predictable manner. Exploiting such temporal regularities lets us focus perceptual and motor resources at the right moment. This offers potential benefits for a range of behaviours, such as getting an opponent's timing down in competitive sport, predicting speech during conversation, or anticipating the cue from other musicians to start playing an instrument. Hence temporal preparation is a fundamental cognitive operation, studied by psychologists for over 100 years, yet the underlying mechanisms remain controversial (Nobre et al., 2007). In the remainder of the introduction, we briefly outline key empirical results obtained using classic approaches to the study of temporal preparation, alongside the formal models that have been proposed to explain them. We then introduce a new and more comprehensive modelling framework characterising temporal preparation when behaviour is measured using simple reaction times (RTs).

Temporal preparation: Theory and data

The classic approach to exploring temporal preparation manipulates the foreperiod between a warning and a go stimulus that must be judged and/or (typically) hastily reacted to (Niemi & Näätänen, 1981). For example, go stimuli might be presented either half a second, or a full second, after the warning stimulus. If a different foreperiod is used in each block of trials, the typical finding is that RTs slow with longer foreperiods. A common explanation relates this “fixed-foreperiod” effect to known limitations of human time perception, specifically to Weber’s law for time – we are less precise when timing longer intervals (Wearden & Lejeune, 2008) so we may be less able to achieve focussed preparation for more temporally distant events.

By contrast, when foreperiods vary randomly *within* a block of trials, longer foreperiods often yield faster RTs. This “variable-foreperiod” effect is traditionally explained by considering the conditional probability of an event occurring given that it has not already occurred (formally, the hazard function; Elithorn & Lawrence, 1955; Luce, 1986). For commonly used foreperiod

80 distributions (e.g. uniform), the hazard function increases across the whole stimulus range. This is
 81 easily appreciated by again considering a foreperiod distribution with just two possible foreperiods,
 82 selected by flipping a coin. At the start of the trial the outcome is unknown, so the chance of a
 83 stimulus occurring with the shorter foreperiod is 50%. However, if a trial progresses past this point
 84 and no stimulus has occurred, it becomes certain that the stimulus will occur at the longer
 85 foreperiod, so the hazard function will increase. An increasing hazard function implies that, for an
 86 optimized brain, preparation should similarly increase, and RTs should improve.

87 Marrying the explanations applied to fixed and variable-foreperiod effects suggests that the
 88 hazard function should be derived not from the *objective* probability density function (PDF) of
 89 foreperiods, as specified by the experimenter, but from the (noisy) *subjective* PDF as experienced by
 90 the participant (Janssen & Shadlen, 2005). To illustrate this idea, Figure 1A schematises a variable-
 91 foreperiod experiment with four equiprobable foreperiods. These discrete timings form a probability
 92 mass function (PMF, shown by black vertical lines). However, due to neural noise, over many
 93 repetitions they are likely experienced as a continuous probability density function, shown (at a
 94 different scale) in red. This, in turn, implies a growing subjective hazard function (shown in blue). In
 95 panel B, a prediction about mean RT is formed by assuming that RT equals a constant (when
 96 unprepared) minus a scaled version of the subjective hazard function, capturing the effect of
 97 preparation (Janssen & Shadlen, 2005; Tsunoda & Kakei, 2008).

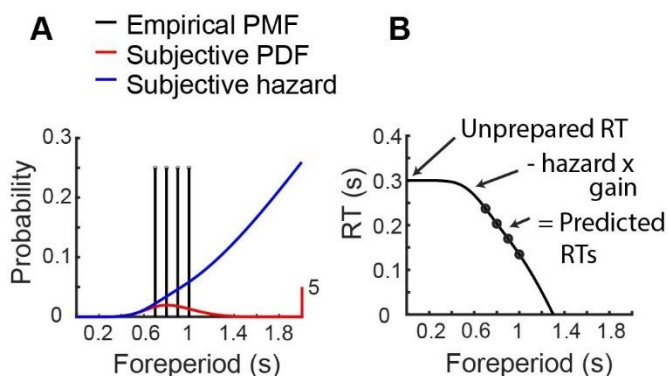


Figure 1. Schematic illustration of how temporal preparation might follow a function (such as a hazard) derived from the subjective PDF of foreperiods. A. Discrete foreperiods (black) are likely corrupted by internal noise before internal representation (red) from which subjective hazard can be derived (blue). B. Some previous studies have then modelled mean RT as a linear function of subjective hazard.

Other transforms of the PDF and/or hazard function have also been investigated, so as to better match particular data sets (Grabenhorst et al., 2019, 2021; Moon et al., 2019; Sharma et al., 2015). For example, Grabenhorst et al. (2019) suggest that changes in mean RT are better predicted by the reciprocal PDF than by the hazard function. Like them, we will consider several putative preparation functions derived from an internal representation of the foreperiod distribution.

However, from its inception, the foreperiod literature has also considered the strong influence of the foreperiod on trial $n - 1$ on the current trial (Woodrow, 1914). Some theorists consider such sequential effects as fundamental, positing that the balance of excitation and inhibition across a given trial reflects a recency-weighted sum of learnt instances (effectively a record of trial-by-trial Hebbian synaptic associations; Los & van den Heuvel, 2001; Salet et al., 2022). This instance-based perspective both explains sequential/learning effects and gives rise to an illusion of hazard-like preparation, with no need to cognitively maintain a subjective PDF of event times. By contrast, investigations incorporating detailed modelling of PDF-based or hazard-based preparation have thus far offered no account of sequential effects. Hence instance-based modelling can currently explain additional patterns seen in data relative to hazard-based accounts. Here, our first objective will be the development of a probabilistic-belief based account that incorporates belief updating, which is thus able to predict the same range of foreperiod-history effects as instance-based accounts.

This general framework, positing preparation based on the experienced distribution of foreperiods, is likely also relevant to a sister literature on temporal orienting (i.e. the probabilistic

cuing of one amongst several foreperiods; see Denison, 2024, for a recent review). The successful use of cues might be viewed as the ability to learn two or more subjective PDFs and switch between them. However, the exact relationship between the use of statistical expectations and of explicit temporal cues is currently a matter of debate (e.g. Duyar et al., 2024). Here, we limit ourselves to consideration of variable-foreperiod effects in temporally uncued contexts. We are agnostic about the extent to which modelled behavioural changes reflect implicit vs. explicit mental processes.

The existence of the variable-foreperiod effect raises a further question, however, for which our work here is relevant: If performance can be improved (to the level shown at longer foreperiods) why don't people just raise their game throughout the trial? The likely answer is that some trade-off exists such that what seems like better (e.g. faster) behaviour carries a cost that is unacceptably magnified by a continuous state of heightened preparation. For example, in RT tasks costs can manifest as additional errors: either selecting the wrong response (for choice RTs) or "jumping the gun" (for simple RTs).

This kind of trade-off (between speed and accuracy) is amenable to characterisation via formal models that predict RT distributions. However, for temporal preparation this approach has been applied only rarely, mainly in studies of cued or fixed rather than variable-foreperiods, and with mixed results (e.g. Grosjean et al., 2001; Jepma et al., 2012; van den Brink et al., 2021). Thus far, neither the hazard-based nor instance-based modelling approach has progressed much beyond fairly coarse predictions, most commonly at the level of mean RT (cf. Fig 1B). More generally, we note that distributional modelling of *choice* RTs is now commonplace, with model parameters from behavioural data fits aiding conceptual clarity when interpreting many kinds of experimental manipulation (and even being used to inform precise predictions about how relevant neural signals vary following stimulus onset, e.g. Kohl et al., 2019, 2020; Spieser et al., 2018). However, our impression is that models are less often applied to simple RTs. In this task, errors are instead often dealt with via trimming and are thus effectively ignored rather than treated as valuable information

in their own right (but see e.g. Carpenter & Noorani, 2023; Luce, 1986; Smith, 1995). Hence, our second objective is to embed competing accounts of the variable-foreperiod effect into a more complete (and multilevel) process model of simple RTs, predicting full RT distributions and thus better reflecting the current state of the art.

Looking beyond errors, sustaining maximal performance likely carries further costs that are either metabolic or reflect lost opportunities (Agrawal et al., 2022; Shenhav et al., 2017). We view the investigation of underlying cost-benefit trade-offs as critical to a fuller understanding of temporal preparation, not least because these trade-offs will determine which behavioural patterns we should consider rational/optimal. Hence, we will consider experimental manipulations that are relevant to opportunity costs alongside foreperiod effects.

A new architecture for models of how probabilistic beliefs inform temporal preparation to determine simple RTs

We now proceed to a formal presentation of the model architecture. As outlined above, in common with previous work we wanted to test how simple RTs are affected by the distribution of foreperiods, T . Whereas this distribution is established by the experimenter, participants may experience a somewhat different distribution, U . In the taxonomy of our model, T is an *objective* distribution and U is its *subjective* counterpart. Inferences regarding foreperiods are necessarily related to their objective distribution, but they are also affected by the variability inherent to perception and memory. Given the scalar property of interval time perception, the probability density function¹ for the subjective distribution (i.e., the subjective PDF) has previously been

¹ The probability density function (or probability mass function) for any random variable X is $P(X = x)$. The cumulative distribution function is $P(X \leq x)$.

170 modelled as the objective PDF, blurred with a Gaussian kernel that expands in width according to
 171 Weber's law (Janssen & Shadlen, 2005):

$$172 \quad P(U = u) = \int_{-\infty}^{\infty} P(T = t) \frac{1}{\varphi t \sqrt{2\pi}} \exp\left[\frac{-(u-t)^2}{2\varphi^2 t^2}\right] dt \quad (1)$$

173 where φ is the Weber fraction quantifying perceptual noise. We present Equation 1 here for
 174 completeness (and because it has often been presented with errors in previous work²) but will not
 175 utilise it. Instead, we want to note that the equation offers a sensible basis of prediction only if the
 176 objective PDF is believed to be stationary, and it has been fully learnt. However, it is well known that
 177 the foreperiod on trial $n - 1$ affects RT on trial n (Woodrow, 1914) and that foreperiod distributions
 178 in previous blocks affect RTs in the current block (Los et al., 2017). Previous accounts predicting
 179 foreperiod effects by utilising hazard functions based on the subjective PDF have not incorporated
 180 these findings (Grabenhorst et al., 2019; Janssen & Shadlen, 2005; Moon et al., 2019; Tsunoda &
 181 Kakei, 2008). Indeed, this has been noted as a critique of hazard-based accounts by researchers
 182 favouring alternative frameworks (Salet et al., 2022).

183

184 **Feature #1: The estimated PDF of foreperiods should change with experience**

185 In order to model $n - 1$ learning effects via a PDF or hazard-based account, the obvious
 186 solution is to specify how the subjective hazard is updated with each new experience. Although a
 187 Bayesian update (ideally incorporating beliefs about non-stationarity) might be optimal, this is
 188 challenging (both theoretically and computationally) when the objective PDF might take on almost

² Janssen and Shadlen (2005) used $f(t)$ to denote the foreperiods' PDF. For clarity, we'll define t as a sample of the random variable T . Subjective foreperiods are samples from a related but different random variable, which we call U . Janssen and Shadlen seem to have gotten confused because they used t to denote samples from U . (Their Eqn. 4 gives subjective foreperiods a PDF of t .) We won't make that mistake. We'll define u as a sample of the random variable U . The relationship between T and U can be written $U = T + V$, where V is a zero-mean normal random variable, with variance $\text{Var}(V) = (\varphi T)^2$. (NB: The PDF of U is thus a blurred version of the PDF of T . The blur kernel is Gaussian, with time constant φT .) The PDF of U can therefore be obtained by convolving the PDF of T with the PDF of V . That convolution is what has been notated on the right-hand side of our Eqn. 1.

any shape (but see Visalli et al., 2021, 2023). Here, we instead consider an alternative scheme (cf. Wang et al., 2020) in which a (non-parametric) estimate of the objective PDF evolves via a form of iterative kernel density estimation. This estimate of the objective PDF, which we term the *estimated* PDF, is distinct from the subjective PDF, which merely *summarises* the foreperiods experienced by the participant.

Let t_i represent the (objective, real-valued) foreperiod in any arbitrary trial i . Internal, Gaussian noise with scalar variability will produce a sample from the distribution of subjective foreperiods: $U_i \sim N(t_i, \varphi^2 t_i^2)$. The likelihood function is thus:

$$\mathcal{L}(t|u_i) = \frac{1}{\sqrt{2\pi\varphi^2 t^2}} \exp\left[\frac{-(u_i - t)^2}{2\varphi^2 t^2}\right] \quad (2)$$

(Jazayeri & Shadlen, 2010).

The estimated PDF for any trial $n > 1$ is formed from a weighted average between the previous trial's estimated PDF and its normalised³ likelihood:

$$P(\hat{T}_n = t) = (1 - \alpha)^{n-1} P(\hat{T}_1 = t) + \sum_{i=1}^{n-1} \alpha (1 - \alpha)^{n-1-i} \frac{\mathcal{L}(t|u_i)}{\int_0^\infty \mathcal{L}(v|u_i) dv}, \quad (3)$$

where α is a weight between zero and one, and $P(\hat{T}_1 = t)$ is the estimated PDF for trial $n = 1$ that may have evolved during practice. An example of the updating of the estimated PDF (and resulting hazard function) is shown in Figure 2.

³ This normalization step is tricky, because Equation 2 offers no closed-form integral. We used a polynomial approximation (based on t of 0.5 s numerically integrated out to one hour) suitable for low values of the Weber fraction. To make sure this approximation was valid when fitting models, we placed an arbitrary upper boundary of 0.3 on φ . For context, values inferred from Experiment 1's data had averages (for different variant models) ranging from 0.09 to 0.22.

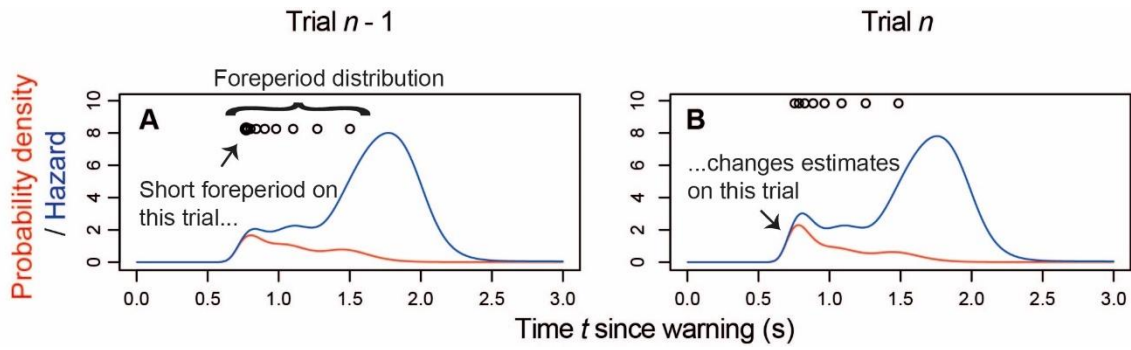


Figure 2. Evolution of the estimated PDF. **A & B.** The estimated PDF (red) and resulting hazard function (blue) on trials $n - 1$ (A) and n (B) in an experiment with exponentially distributed foreperiods. Because of the short foreperiod on trial $n - 1$, the estimated PDF and hazard function change on trial n .

Feature #2: Beliefs about whether the stimulus will *always* arrive matter

Based upon the estimated PDF of foreperiods, the participant can go on to determine a hazard function, describing the estimated likelihood of experiencing a go stimulus if it has not already occurred:

$$h_n(t) = \frac{(1 - \hat{b}_n) P(\hat{T}_n = t)}{1 - (1 - \hat{b}_n) P(\hat{T}_n \leq t)}. \quad (4)$$

Equation 4 expands the standard formula for the hazard function to explicitly incorporate the participant's estimated probability $\hat{b}_n$, that a stimulus will not be delivered at all on trial n (i.e. of an omission or "catch" trial). Note that the earlier Equation 3 describes only sequences of non-catch trials – a catch trial would imply no update to the estimated PDF.

Objective stimulus omission rates are known to affect preparation functions (Grabenhorst et al., 2021). However, estimates of the omission rate have received little attention thus far. Indeed, where they have been considered at all, they have been assumed to exactly reflect the objectively correct value (e.g. Grabenhorst et al., 2019; Moon et al., 2019). Here, we adopt a normative

Bayesian learning model of how one should infer a probability across trials in a world that is either considered fixed or time-varying (Zhang et al., 2014). This is achieved based on updates to a belief that is formalised as a Beta distribution, with probability density function $P(\hat{B}_n = b)$. Its mean value $\hat{b}_n \equiv E(\hat{B}_n)$ is taken as the point estimate for trial n . Participants are assumed to believe that the world (here, the stimulus omission rate) may discretely change from time to time, with a probability $1 - \psi$. If and when this occurs, they will revert to their initially held prior regarding the stimulus-omission rate, $P(\hat{B}_0 = b)$. This implies that on each trial $n > 1$, a new belief is formulated from the combination of this long-term prior with the posterior probability of omission, given the preceding sequence of subjective foreperiods:

$$P(\hat{B}_n = b) = \psi P(\hat{B}_{n-1} = b | \mathbf{u}_{n-1}) + (1 - \psi) P(\hat{B}_0 = b), \quad (5)$$

where $\mathbf{u}_{n-1}$ denotes the preceding sequence of subjective foreperiods $(u_1, u_2, \dots, u_{n-1})$ and ψ is a weight between zero and one. Bayes' Rule is required to compute the posterior probability in Equation 5:

$$P(\hat{B}_{n-1} = b | \mathbf{u}_{n-1}) \propto \prod_i^{n-1} P(u_i | \hat{B}_i = b) P(\hat{B}_i = b). \quad (6)$$

Finally, note that trial i is either a catch trial (when $u_i = \infty$) or it is not (when $u_i < \infty$). Thus, the corresponding likelihood has only two values for $0 < b \leq 1$:

$$P(u_i | \hat{B}_i = b) = \begin{cases} 2(1-b) & u_i < \infty \\ 2b & u_i = \infty \end{cases}. \quad (7)$$

Examples of the resulting development of the estimated omission rate, across trials, and the effect this has on the hazard function, are presented in Figure 3. This figure illustrates how a participant's belief about the current omission rate would strongly affect their subjective hazard function, and thus, potentially their RTs at different foreperiods.

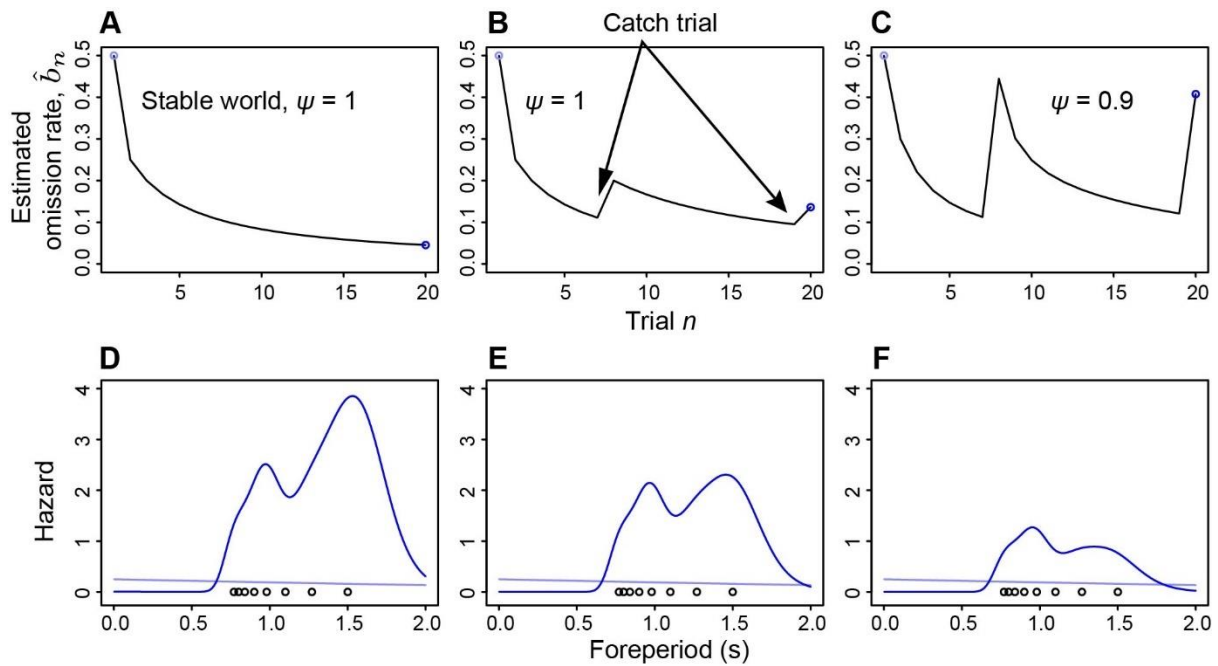


Figure 3. Predictions for the evolution and effect of estimated stimulus-omission rate across experimental trials. Prior to trial $n = 1$, the participant is assumed to have a flat prior belief regarding the omission rate, i.e. $P(\hat{B}_0 = b) = 1$, for all $0 \leq b < 1$. **A.** In an experiment with no catch trials (i.e. a stimulus on every single trial) where the participant assumes a stable rate of catch trials, the estimated omission rate rapidly converges towards zero. **B.** With occasional catch trials, the estimate converges towards the true omission rate. **C.** If participants additionally assume that the omission rate may change discretely (here on 10% of trials) they remain responsive to catch trials even in later trials, making them more accurate in a non-stationary environment (but less so in a stable one). **D-F.** Subjective hazard functions resulting from the estimated omission rates in the first (light blue) and twentieth (dark blue) trials of the experiments shown in parts A-C respectively (based on an experiment using exponentially distributed foreperiods; small circles). The estimated omission rate strongly affects both the magnitude and shape of the hazard function.

Feature #3: The whole RT distribution, not just its central tendency, is predicted, and at an individual level

261 We have thus far outlined two limitations of most previous computational accounts which
 262 assume that the foreperiod distribution is internalised and used to determine levels of preparation:
 263 failure to incorporate history dependence (but see Visalli et al., 2021, 2023) and failure to treat the
 264 rate of no-stimulus catch trials as a quantity that must be inferred. Two further limitations will be
 265 addressed here. One is a tendency to fit only (or primarily) at the group-mean level. This is
 266 problematic because when a model prediction is determined for each participant, and is then
 267 averaged across them, the average prediction can take forms unavailable to raw predictions. Hence,
 268 when a model of the individual is fit to the average of the group, a model that might actually be a
 269 good fit to data might be rejected. Here, we will fit data at the individual and group levels
 270 simultaneously via multilevel models, and we plot the average of individual data against the average
 271 of individual model predictions.

272 A further (final) limitation, shared also by computational accounts not premised on
 273 internalising the foreperiod distribution (Salet et al., 2022) is the tendency to characterise only
 274 changes in mean RT, rather than describing effects across the full RT distribution. In the case of
 275 simple RTs, responses likely reflect a trade-off between maximising speed and minimising the
 276 generation of erroneous anticipatory responses before the go stimulus appears. Erroneous
 277 anticipatory responses lie at the extreme left-hand tail of the RT distribution, and they are often
 278 censored out from simple-RT summary measures via trimming (or by a failure to instruct software to
 279 monitor for a response prior to the stimulus). However, their presence (and pattern) might be
 280 diagnostic from a mechanistic perspective. We therefore opted to implement models capable of
 281 characterising the full RT distribution with minimal data censoring.

282 How should these goals be accomplished? Although various distributions have been used to
 283 model simple RTs, several popular choices (such as the shifted log-normal model) struggle with RTs
 284 that are negative relative to the go signal, i.e. erroneous anticipations. This led us to take a bespoke

approach, wherein the simple RT on trial n is sampled from the distribution R_n of first-event times produced by an inhomogeneous (i.e. non-stationary) Poisson process. Its CDF is:

$$P(R_n \leq r) = \begin{cases} 1 - \exp\left[-\int_0^r \lambda_n(t) dt\right] & t \geq 0 \\ 0 & t < 0 \end{cases}, \quad (8)$$

where the Poisson rate $\lambda_n(t)$ varies as a function of time t since the warning signal. The timing of initial spikes from a leaky integrate-to-threshold-then-fire detection mechanism can form this type of distribution, provided it receives noisy input from sensory neurons (Stevens & Zador, 1996).

Recall that the go stimulus on trial n occurs at time t_n . Some time τ_n later, we assume a transient increase in Poisson rate, such that:

$$\lambda_n(t) = \begin{cases} f_n(t) + m_n \delta(t - w_n) \exp[1 - \delta(t - w_n)] & w_n \leq t \\ f_n(t) & 0 \leq t < w_n \\ 0 & t < 0 \end{cases}, \quad (9)$$

where the constants m_n and δ represent the maximum stimulus-dependent increase and its rate of decay, respectively; and $w_n = t_n + \tau_n$, where τ_n represents an inescapable residual delay⁴, somewhat analogous to the non-decision time parameter often incorporated in models of choice RT. The function $f_n(t)$ describes the background process operating throughout the foreperiod (somewhat like the “urgency” signals sometimes utilised in choice-RT modelling). We considered three variants of this model where urgency follows a profile that is based on either the subjective hazard function, the estimated foreperiod PDF, or the elapsed time, respectively:

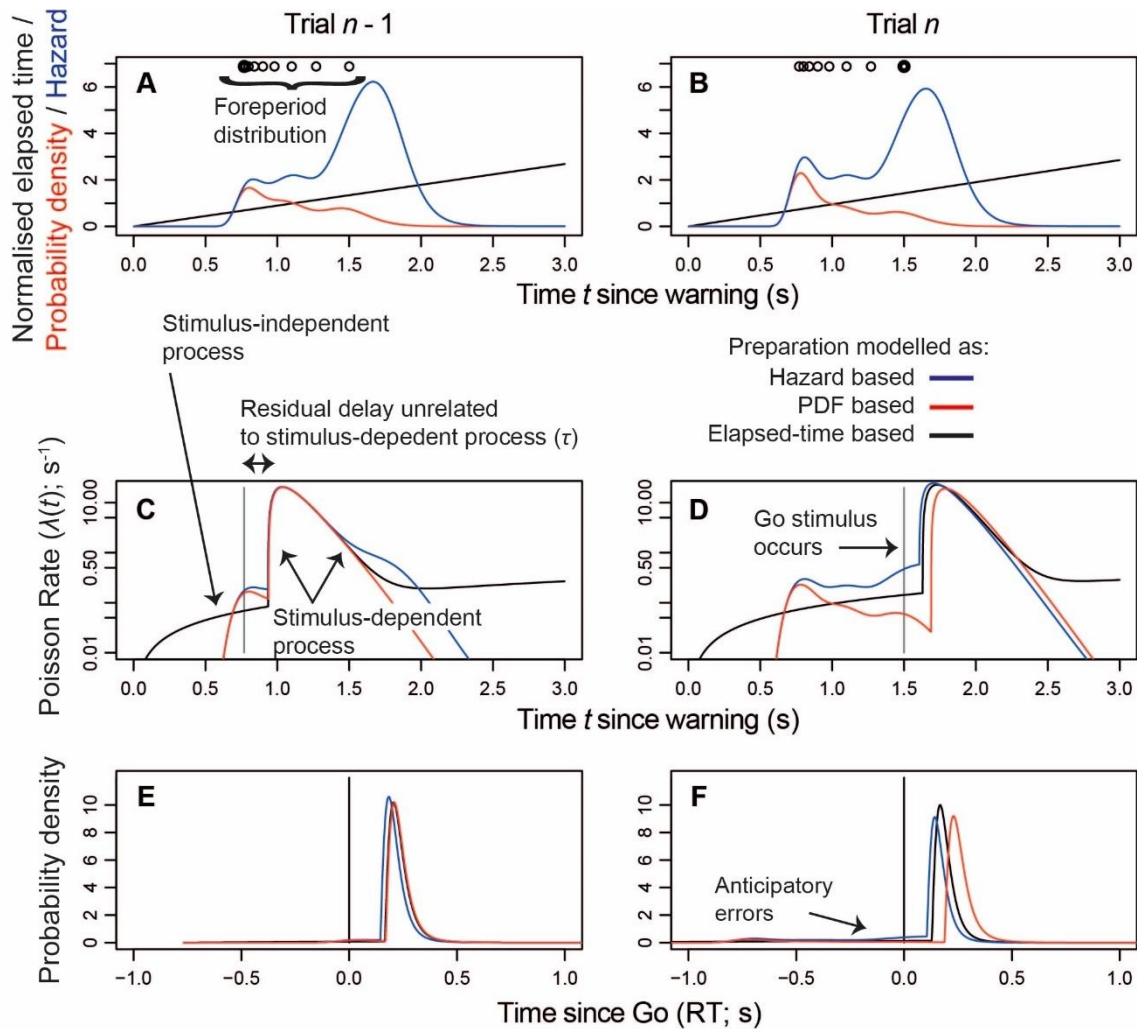
$$f_n(t) = b_f + a_f h_n(t) \quad (10)$$

$$f_n(t) = b_f + a_f P(\hat{T}_n = t) \quad (11)$$

$$f_n(t) = b_f + a_f t / E(\hat{T}_n) \quad (12)$$

⁴ Much of our code uses derivations from an earlier parameterization for the transient increase in Poisson rate, $\lambda_l(t - w_n) \exp[\delta(w_n - t)]$. The resulting parameters are translatable as $\lambda_l = m_n \delta e$.

304 where $E(\hat{T}_n)$ is the mean of the estimated PDF on trial n .⁵ The model is schematised in Figure 4.



305

306 Figure 4. Schematic of three models incorporating an initial Stimulus Independent Process into a Non-
 307 Stationary Poisson account of simple Reaction Time (SIP-theN-SPuRT). The figure shows predictions
 308 on two experimental trials, in which a go-stimulus occurs with either a short (A, C, E) or long (B, D, F)
 309 foreperiod. Model parameters vary either as a product of elapsed time since a warning stimulus
 310 (black), or in line with the estimated PDF (red) or the resulting hazard function (blue) arising from the
 311 foreperiods experienced so far. **A & B.** The estimated PDF (red) and resulting hazard functions (blue)
 312 on trials $n - 1$ and n , in an experiment with exponentially distributed foreperiods. Because of the

⁵ Division by the mean is introduced to allow our linear model variant to incorporate a (summary) belief about likely foreperiods based on recent experience, and by doing so to potentially capture learning effects.

short foreperiod on trial $n - 1$, the estimated PDF changes on trial n . This has consequences for the hazard function, as well as a simpler preparation function based on elapsed time but rescaled according to the estimated mean time of stimulus arrival (black). **C & D.** Changes in the Poisson rate of response triggering with time (note the log scale). Each SIP-theN-SPuRT model posits three cognitive components affecting how this rate evolves. The first, a slow rise in the tendency to respond that is independent of the occurrence of a go stimulus (cf. urgency), accounts primarily for very short RTs, including anticipatory errors. The second, a sudden increase (and subsequent slower decline) in the tendency to respond, loosely reflects the drive of a leaky integrate-to-threshold-then-fire detection mechanism. It determines the main portion of the RT distribution triggered by the go stimulus. The third process accounts for residual delays in triggered RT (i.e. all other sensorimotor delays). **E & F.** Predicted RT distributions. Model parameters capturing the translation of either elapsed time, estimated foreperiod PDF or hazard (onto each of the three processes illustrated in parts C & D) have been selected to roughly equate RT at the short foreperiod (E), whereas RT at the longer foreperiod (F) reflects the different levels that elapsed time, estimated PDF and hazard functions have reached by this later point of the trial.

Feature #4: Preparation affects both mode and spread of RT, and can saturate

We have thus far described how our modelling lets beliefs about the foreperiod distribution inspire three possible forms of stimulus-independent urgency (i.e., based on either the elapsed time, the estimated foreperiod PDF or a subjective hazard) that affect, primarily, the generation of anticipatory errors. However, Figure 4 also schematises how preparation (again in these three possible forms) can affect the bulk of RTs occurring in response to the go stimulus (see also Figure 5, later). Firstly, in analogy to previous work making quantitative predictions about mean simple RT, our models also allow for τ_n , the residual RT period (which essentially generates a horizontal translation in the stimulus-driven portion of the RT distribution), to vary according to the current

level of preparation. If preparation is hazard-based and thus adjusts τ_n in line with the estimated hazard function:

$$\tau_n = b_\tau - b_\tau a_\tau \tanh[g_\tau h_n(t_n)], \quad (13)$$

where b_τ represents τ_n at the beginning of the foreperiod (when a participant is relaxed), g_τ is a gain parameter modulating the (potentially saturating) effect of hazard on τ_n , and a_τ is a further gain parameter.⁶

Equation 13 may at first appear needlessly complex but note that by fixing g_τ to be very low we can essentially remove the effect of the saturating $\tanh$ function so that τ_n becomes linear with $h_n(t_n)$. This corresponds to most previous modelling in this area (where only mean RT has been predicted from, e.g., non-saturating subjective hazard; see Figures 1B and 5A). However, with higher values of g_τ , Equation 13 also offers the opportunity to model a hypothesis in which preparation follows the subjective hazard function before saturating at some limit. This would seem logically necessary, given that the residual RT τ_n is low-bounded, presumably at a value well above zero.

While Equation 13 describes a model in which τ_n is affected by hazard, we also considered models in which it was affected by either the estimated foreperiod PDF, or it was a simpler function of elapsed time. In these cases, the $h_n(t_n)$ function in Equation 13 is substituted by $P(\hat{T}_n = t_n)$ or by $t_n/E(\hat{T}_n)$, respectively.

We took a conceptually similar approach in allowing for a possible foreperiod effect on the maximum stimulus-dependent increment in drive (m_n from Equation 9). There is no analogy with previous work here, because previous work has considered an effect of temporal preparation only at

⁶ In our code, the parameter a_τ is slightly different to the one in Eqn. 13. In our code, the parameter a_τ satisfies the equation $\tau_n = b_\tau - b_\tau a_\tau \tanh[g_\tau h_n(t_n)] / \tanh[g_\tau h_{\max}]$, where $h_{\max} \equiv \max_n h_n(t_n)$ represents the hazard function's greatest value during the course of the experiment. This version of a_τ is more interpretable, representing the (approximately) greatest (proportional) reduction in τ_n that the model predicts. In our code, the analogous approach was taken when estimating a_λ in Equation 14, so that it represents the (approximately) greatest predicted increase in m_n .

the level of mean RT, whereas our generative model allows us to consider preparatory effects on all underlying parameters, including specifically the strength of input from change detectors, which can affect both the mean and spread of RTs. In the case of the hazard-based model, we implemented:

$$m_n = b_\lambda + a_\lambda \tanh[g_\lambda h_n(t_n)], \quad (14)$$

where b_λ is drive when in an unprepared state, and a_λ and g_λ are gain parameters describing how increasing subjective hazard raises this stimulus-dependent drive (until potential saturation), in line with our approach for τ_n . We again also tested models in which, instead of being hazard-dependent, drive was affected by the estimated foreperiod PDF or a simpler function of elapsed time. In these cases, the $h_n(t_n)$ function in Equation 14 is substituted by $P(\hat{T}_n = t_n)$ or by $t_n/E(\hat{T}_n)$, respectively.⁷

Summary of model architecture

Our model architecture incorporates an initial Stimulus Independent Process into a Non-Stationary Poisson account of simple Reaction Time (SIP-theN-SPuRT). Under this account, each simple RT is determined by three components: 1) an urgency process operating throughout the foreperiod that primarily generates anticipatory responses, 2) a stimulus-dependent drive, and 3) a residual delay (which, largely, just translates the RT distribution horizontally). Each of these components can vary, as a function of either how much time has elapsed since the warning signal, the participant's estimated PDF for foreperiods, or their resulting hazard. Each of these functions is specified with an intercept (the b parameters) and one or two gains (the a and g parameters). Further details are provided in the relevant methods sections. Table 1 summarises the model architecture's core parameters and their purposes.

⁷ Our code also allows tanh saturation to be incorporated in the urgency function $f_n(t)$, but exploratory fits found that this was not very useful. Consequently, we have omitted it from Equations 10-12.

380

381 Table 1: *Core parameterisation of SIP-theN-SPuRT*

Model components	Parameters (single foreperiod design)	Further parameters (multiple foreperiod design)	Meaning	Bounds and typical value(s)*
Internal representation of foreperiod distribution	φ		Weber ratio (precision of interval timing; Eq. 2)	[0.05, 0.3], ~0.18
		α	Update rate for estimated foreperiod PDF (Eq. 3)	[0, 1], ~0.3
		ψ	Belief regarding probability that the rate of catch trials will remain the same (Eq. 5)	[0, 1], ~1 or 0.95 (for experiments without/with catch trials respectively)
Go-stimulus-independent urgency	b_f		Intercept and gain of $f_n(t)$, describing the background rate of action triggering (Eqs. 10-12)	[0, ∞], ~0.005
	a_f			[0, ∞], ~0.12
Go-stimulus-dependent drive	δ		Decay rate for drive generated by onset of go stimulus (Eq. 9).	[0, ∞], ~9
		b_λ	Intercept and gains determining m_n , the maximum drive generated by onset of go stimulus (Eq. 14)	[0, ∞], ~10
		a_λ		[0, ∞], ~23 (16)**
		g_λ		[0, ∞], ~0.5 ($m_n \approx 30$)
Residual reaction time	b_τ		Intercept and gains determining τ_n , the residual RT arising from all other sensorimotor delays (Eq. 13)	[0, ∞], ~0.25
		a_τ		[0, ∞], ~0.21 (0.18)**
		g_τ		[0, ∞], ~1 ($\tau_n \approx 0.22$)

382 *Note:* *Relates to saturating hazard-based model variants. ** Value in brackets shows the code's

383 variable as per Footnote 6.

384

385 **Fitting empirically derived reaction times with SIP-theN-SPuRT**

Evolution of the participant's estimate of the foreperiod PDF is one of the key features of SIP-theN-SPuRT. This estimate depends on the preceding sequence of foreperiods subjectively experienced by the participant. As experimenters, we cannot know the precise values of these subjective experiences, due to their stochastic nature. However, we can specify their distribution. Accordingly, we substitute our expectation

$$\langle \mathcal{L}(t|u_i) \rangle_{u_i} \propto \int_0^\infty \mathcal{L}(u_i|t_i) \mathcal{L}(t|u_i) du_i, \quad (15)$$

for each trial i 's foreperiod likelihood $\mathcal{L}(t|u_i)$, which serves as the basis of the participant's estimate in subsequent trials (Eq. 2).

Note that, although the evolving posterior probability of omission (Eq. 5) also depends on the preceding sequence of subjective foreperiods, all that matters is whether each of those foreperiods was experienced as a catch trial or not. For this, we simply assume that a catch trial was experienced if and only if there was, indeed, a catch trial.

Finally, note that some of the resulting RT CDFs (from substitution of Equation 10, 11 or 12 into Equations 9 then 8) contain integrals that will not simplify, so a numeric approach is required (in our case, trapezoid integration when evaluating the CDF, and two-point differencing of this CDF to evaluate the PDF).

Testing the nature of temporal preparation with SIP-theN-SpuRT

In the remainder of the paper, we use the SIP-theN-SPuRT model architecture to address two main questions. First, are the features that we have introduced (updating of the estimated foreperiod PDF, updating to beliefs about the rate of catch trials, the capacity to capture foreperiod effects on the full RT distribution rather than just mean RT, and the capacity to have the level of temporal preparation saturate) useful in providing a better fit to data? Second, which underlying

basis best reflects the effect of temporal preparation on simple RT: A subjective hazard function, the participant's estimated PDF of foreperiods, or simply the time elapsed since the warning signal?

To address these questions, we present fits to data from three experiments. The first is a new experiment manipulating both foreperiod and the opportunity cost of focussed preparation. After that, we revisit a published experiment with open-access data provided at the trial-by-trial level. In this experiment, foreperiods were manipulated alongside cues about the information value of the current trial's foreperiod (Visalli et al., 2021). Finally, we describe a further new experiment, manipulating foreperiod alongside several other factors, including the rate of stimulus omissions. To foreshadow our results – we find that SIP-theN-SPuRT usefully predicts many features of the experimental data, particularly when temporal preparation is modelled as a saturating hazard function.

Experiment 1

We have already noted our interest in considering how strategic influences might affect temporal preparation. For example, with sufficient motivation to react very quickly, one might imagine that simple RT would no longer vary with foreperiod but rather be pushed toward its minimum value from as near to trial onset as is physiologically possible. In the parlance of our modelling, we might expect a saturating form of temporal preparation to rapidly hit its limit via high values of parameters such as g_λ and g_τ (and perhaps concomitant changes in anticipatory errors via increases in b_f and/or a_f). Hence Experiment 1 is a new study testing how opportunity costs interact with the variable-foreperiod effect.

The literature appears to contain few studies along these lines, although Los and van den Heuvel (2001) have previously manipulated the presence vs. absence of a financial incentive for fast and accurate responding in a choice-RT task. In their Experiment 1, the “mixed” block combined with

neutral-cue conditions (similar to the classic, uncued, variable-foreperiod situation) suggests that payoffs may change the mean RT vs. foreperiod function, but not in an easily interpretable manner. The paper does not include formal modelling (but see Salet et al., 2022). Here, we utilise a simple-RT task appropriate to our modelling framework and introduce a different kind of strategic manipulation.

Methods

Given our modelling focus, full details of Experiment 1's methods (including transparency and openness information) are presented in Appendix A. Here, we present only a summary of the task design, but full details of the model-based data analysis.

Participants and task

We included data from 36 online participants making speeded responses to the appearance of a central visual stimulus, with small financial rewards and punishments for fast responses and anticipatory errors, respectively. The experimental task is schematised (later, alongside the results) in Figure 6A. Within each block of 144 trials, foreperiods varied randomly in eight steps between 0.77 and 1.5 seconds. The exception was a single 0.5 s "surprise" foreperiod at the very end of each block. Four blocks were completed, in one of four counterbalanced orders. In two of the blocks, the eight foreperiods clustered at shorter times to form an *exponential* foreperiod distribution. In the other two, they clustered at longer times to form a *flipped* distribution. In one block from each distribution, the always-present secondary task, in which participants had to monitor their visual periphery and count the appearance of targets there, while discounting non targets, was incentivised: participants were punished for failing at this secondary task, yielding an opportunity

cost to focussing central attention on the primary RT task. In alternative no-cost blocks, the periphery could safely be ignored.

Analysis

Data trimming

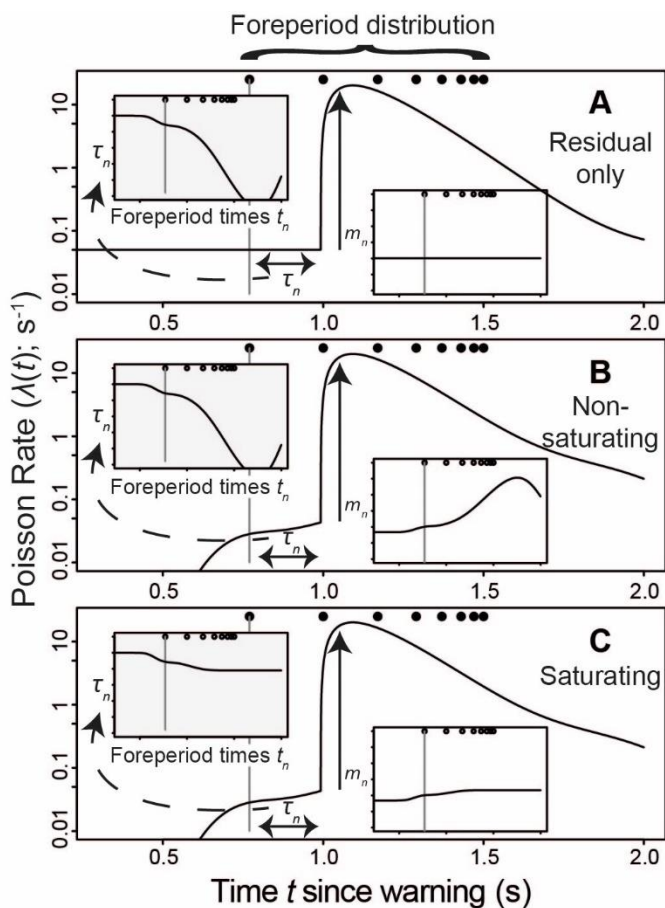
Following a pre-registered criterion, we excluded trials where a secondary (peripheral-track) task stimulus occurred during that trial prior to the primary task response. The foreperiods from all these trials were nonetheless utilised during modelling for updating the participant's estimated PDF.

Modelling

Further to the specification provided in the introduction, for models with hazard-like preparation functions, we fixed the belief that stimulus omission rate was unchanging, ψ , to equal 1, as initial fits with simplified models suggested that almost all participants behaved this way in this experiment without any catch trials. We also specified how the model's behaviour should be allowed to vary across experimental conditions. We did this by introducing effects-coded variation around the geometric mean value for a subset of the intercept and gain parameters, allowing them to vary according to the experimental manipulation of opportunity cost. For example, instead of having two separate intercept parameters for the residual RT (e.g. b_{tNC} for the no-cost conditions and b_{tHC} for the high-cost conditions) we used one parameter $\bar{b}_{\text{t}}$, equivalent to $\sqrt{b_{\text{tNC}}b_{\text{tHC}}}$, and another parameter β_{bt} , equivalent to $\ln(b_{\text{tHC}}/b_{\text{tNC}})$. This latter parameter quantifies the effect of cost. Analogously defined parameters $\bar{b}_{\text{f}}$, β_{bf} , $\bar{b}_{\lambda}$ and $\beta_{\text{b}\lambda}$ were used to specify intercepts for the background "urgency" signal and the maximum stimulus-dependent increase in Poisson rate (or "drive").

To assess the value of SIP-theN-SPuRT's innovations, we considered three variants of its hazard-based model, three variants of its PDF-based model, and three variants of the model based on elapsed time.

“Residual-only” variants permitted changes in RT with foreperiod in a manner as similar as possible to the RT modelling utilised in previous studies (Grabenhorst et al., 2019, 2021; Janssen & Shadlen, 2005; Salet et al., 2022; Tsunoda & Kakei, 2008) but within the expanded framework of SIP-theN-SPuRT. To this end, the gain parameters a_f and a_λ were fixed at 0. This prevented the urgency signal and the drive from varying between trials based on the foreperiod. Trial-by-trial variation in residual RT was constrained to be a linear function of the preparation basis (elapsed time, estimated PDF or subjective hazard) by fixing g_τ at an arbitrary but low value (that was nonetheless greater than zero). The residual-only variants' remaining parameters, $\bar{a}_\tau$ and β_{at} , correspond to the geometric mean gain for residual RT and the effects-coded variation around it, respectively. An illustration of a residual-only variant's prediction for the time-varying Poisson rate is provided in Figure 5A.



494

495 Figure 5. Schematic illustrating three variant models incorporating an initial Stimulus Independent
 496 Process into a Non-Stationary Poisson account of simple Reaction Time (SIP-theN-SPuRT). The main
 497 plot shows how the Poisson rate of response triggering varies with time, with an initial stimulus-
 498 independent (urgency) process that is then strongly augmented shortly after a go stimulus occurs
 499 (grey vertical line shows end of objective foreperiod on trial n ; see Figure 4 and main text for further
 500 details). **A.** Residual-only variants were designed to be as analogous as possible to previous work in
 501 which mean RT has been predicted as a linear function of preparation. Residual RT (τ_n) was allowed
 502 to vary with preparation (left-hand inset; preparation reflecting the estimated hazard rate
 503 illustrated), but other model components (urgency, indicated by the flat Poisson rate prior to
 504 registering the stimulus, and stimulus-dependent drive, m_n , shown in the right-hand inset) were
 505 modelled as constant across foreperiods. Note that the right-hand inset's horizontal axis label (not
 506 shown) is identical to that of the left-hand inset. **B.** Non-saturating variants more fully utilised all
 507 three of SIP-theN-SPuRT's core components by allowing urgency, residual RT, and stimulus-

508 *dependent drive to vary with preparation. C. Saturating variants additionally incorporated limits on*
 509 *the translation of preparatory intent into physiological effects. In this example, saturation becomes*
 510 *apparent a little over one second into the foreperiod as a plateau in the functions depicted in both*
 511 *insets (compared to those same insets in panel B).*

512 The utility of permitting preparation-dependent changes in urgency and drive was assessed
 513 by relaxing some of the constraints placed on parameter values in the residual-only variants. For the
 514 “non-saturating” variants, the parameters $\bar{a}_f$, β_{af} , $\bar{a}_\tau$, $\beta_{a\tau}$, $\bar{a}_\lambda$ and $\beta_{a\lambda}$ were free to vary, allowing
 515 linear gains for all three components of SIP-theN-SpuRT. However, as in the residual-only variants,
 516 trial-by-trial variations were constrained to be linear functions of the preparation basis by fixing g_f ,
 517 g_τ , and g_λ at arbitrary but low, non-zero values. Preliminary fits suggested that with evolving
 518 urgency ($\bar{a}_f > 0$) the rate of stimulus-independent response triggering began (i.e. had an intercept)
 519 very near zero, with little variation across participants or cost condition. Hence, when fitting the
 520 non-saturating models, different values of b_f were allowed only for the different preparation bases
 521 (elapsed time, estimated PDF, and subjective hazard). The other two intercepts were estimated
 522 individually and allowed to vary with cost condition and preparation basis. An illustration of a non-
 523 saturating variant’s prediction for the time-varying Poisson rate is provided in Fig. 5B.

524 Our final set of parameter constraints produced “saturating” variants. These constraints
 525 implemented the idea that the preparation affecting stimulus drive and residual RT might max out at
 526 some limit. Saturating variants again had the urgency intercept b_f estimated singly for the group and
 527 intercepts for the other two processes (stimulus drive and residual RT) estimated individually. We
 528 included free parameters for the geometric means of all five possible gains ($\bar{a}_f$, $\bar{a}_\lambda$, $\bar{a}_\tau$, $\bar{g}_\lambda$ and $\bar{g}_\tau$),
 529 but fixed $\beta_{a\lambda} = \beta_{a\tau} = 0$, because the saturating limit should be common to both no-cost and high-
 530 cost conditions. An illustration of a saturating variant’s prediction for the time-varying Poisson rate is
 531 provided in Fig. 5C.

Each model variant was instantiated using a multilevel framework. It was coded and fitted using the Stan programming language interfaced from R (R Core Team, 2021) via the RStan package (Stan Development Team 2022). Additional R packages including shinystan (Gabry & Veen, 2022) and LOO (Vehtari, Gabry, Magnusson, Yao, Bürkner, Paananen, & Gelman, 2022) were used to diagnose and evaluate goodness-of-fit. We fit a subset of model variants with a Bayesian approach, using four chains, each exploring the likelihood surface via the default Hamiltonian Monte-Carlo no U-turn sampling (HMC NUTS) algorithm, which retains samples in proportion to the height of the posterior distribution, and thus estimates it. All such reported fits used 1000 warmup iterations followed by 1000 post-warmup iterations per chain, with maximum tree depth set to 7.

However, because this approach was prohibitively time-consuming for general application, we mainly utilised maximum-likelihood point estimates (MLEs) for evaluating goodness-of-fit, searching the parameter space via Stan's L-BFGS quasi-Newton algorithm.⁸ The likelihood for most trials was determined by two-point differencing of the CDF, but trials with RTs > 1 second were all treated as members of a discrete category. Thus, for any combination of model variant and parameter values θ , the effective likelihood of RT r_n on trial n was defined as:

$$\tilde{l}_{\theta}(r_n) \equiv \begin{cases} P(R_n = r_n) & r_n \leq t_n + 1 \text{ s} \\ P(R_n > t_n + 1 \text{ s}) & r_n > t_n + 1 \text{ s} \end{cases} \quad (16)$$

⁸ In multilevel models, a given parameter's value for each participant is assumed to be drawn from a distribution, which must be specified so that its parameters can be estimated (to generate shrinkage and/or inform statistical inferences about the group). In general, we used gamma distributions for zero-bounded parameters (φ , δ , b_f , a_f , b_τ , a_τ , g_τ , b_λ , a_λ , g_λ), and logit-normal or beta distributions for 0-1 bounded parameters (α , ψ , respectively). The β parameters capturing changes across conditions for b , a and g parameters were drawn from normal distributions. Bayesian approaches encourage the use of hyperpriors on the parameters of all group-level distributions. In general, we specified diffuse hyperprior distributions. When fitting multilevel models via MLE, group-level distributions can generate problematic (potentially infinite) local maxima if variance is allowed to tend to zero, an issue we addressed by setting minimums on group-level parameters controlling variance and relying on mean-posterior Bayesian fits when making inferences about values of model parameters.

For model comparison in a context where models describe the behaviour of an individual but are fitted to multiple individuals, it has been argued that the simplest approach (summing the (penalized) log likelihood across participants and basing inferences on resulting differences) is only appropriate if we believe a single model should be best for *all* participants. If not (as seems plausible when comparing models describing cognitive processes) we should treat participants as a random sample from a potentially heterogeneous population in which some proportion of participants may be best described by each model (Rigoux et al., 2014; Stephan et al., 2009). We took the approach recommended in these two papers, utilising the spmBMS function from the SPM toolbox v25.01.02 for Matlab (applied to $AIC/-2$ and $BIC/-2$) for comparing variant models.

Results and discussion

Table 2 shows (relative) goodness-of-fit metrics for all nine SIP-theN-SPuRT model variants. Lower numbers denote better model fits. For each variant, we report “PEMP”, the posterior expectation of model probability, which is the mean posterior likelihood estimate of that variant’s multinomial probability parameter, given the data (note that the nine PEMP values must sum to 1).⁹ We also report “1-PEP” (for one minus the protected exceedance probability), which is like a traditional frequentist p-value. It allows us to assess whether we should reject the null hypothesis that a given variant will best fit the data from no greater a proportion of (hypothetical) participants than the other eight variants.

From Table 2, it is clear that the saturating hazard-based preparation model variant fits best. It has the lowest overall AIC and BIC, and it is estimated to be the best model for around 66% of participants (with the remaining 34% split among the other eight models). The associated protected exceedance probability supports the inference that this model is the most frequently correct of the

⁹ This quantity is denoted $\langle r_k \rangle$ for model k in Stephan et al. 2009, and `exp_r` (expectation for the posterior) in the spmBMS function’s help text.

models tested ($1-PEP < 0.001$). Looking beyond this best-fitting model, in general, saturating variants fit better than non-saturating variants, which outperform residual-only variants. Furthermore, hazard-based preparation variants fit better the PDF-based variants, which are poor even compared to models in which preparation depends simply on elapsed time.

Table 2: *Fits of model variants in Experiment 1*

Model Variant		Parameters			ΔGoodness of fit*		
		Number		Individual parameters	MLE		Bayesian
Preparation function	Variant	Individual	Group-level			AIC (PEMP, 1–PEP)	BIC (PEMP, 1–PEP)
Hazard	Saturating	469	26	$\varphi, \alpha, \delta, \bar{b}_f^{**}, \bar{b}_\lambda, \bar{b}_\tau, \bar{a}_f, \beta_{af}, \bar{a}_\lambda, \bar{a}_\tau, \bar{g}_\lambda, \beta_{g\lambda}, \bar{g}_\tau, \beta_{g\tau}$	0 (0.6591, <0.001)	0 (0.6591, <0.001)	0
	Non-Sat	469	26	$\varphi, \alpha, \delta, \bar{b}_f^{**}, \bar{b}_\lambda, \beta_{b\lambda}, \bar{b}_\tau, \beta_{b\tau}, \bar{a}_f, \beta_{af}, \bar{a}_\lambda, \beta_{a\lambda}, \bar{a}_\tau, \beta_{a\tau}$	571.4 (0.1627, >0.999)	571.4 (0.1627, >0.999)	573.4 (se 72)
	Resid-O	396	22	$\varphi, \alpha, \delta, \bar{b}_f, \beta_{b\lambda}, \bar{b}_\lambda, \beta_{b\lambda}, \bar{b}_\tau, \beta_{b\tau}, \bar{a}_\tau, \beta_{a\tau}$	5120.1 (0.0222, >0.999)	4550.7 (0.0222, >0.999)	
PDF	Saturating	469	26	$\varphi, \alpha, \delta, \bar{b}_f^{**}, \bar{b}_\lambda, \bar{b}_\tau, \bar{a}_f, \beta_{af}, \bar{a}_\lambda, \bar{a}_\tau, \bar{g}_\lambda, \beta_{g\lambda}, \bar{g}_\tau, \beta_{g\tau}$	5442.9 (0.0222, >0.999)	5442.9 (0.0222, >0.999)	
	Non-Sat	469	26	$\varphi, \alpha, \delta, \bar{b}_f^{**}, \bar{b}_\lambda, \beta_{b\lambda}, \bar{b}_\tau, \beta_{b\tau}, \bar{a}_f, \beta_{af}, \bar{a}_\lambda, \beta_{a\lambda}, \bar{a}_\tau, \beta_{a\tau}$	5103.0 (0.0222, >0.999)	5103.0 (0.0222, >0.999)	4676.4 (se 143.4)
	Resid-O	396	22	$\varphi, \alpha, \delta, \bar{b}_f, \beta_{b\lambda}, \bar{b}_\lambda, \beta_{b\lambda}, \bar{b}_\tau, \beta_{b\tau}, \bar{a}_\tau, \beta_{a\tau}$	6905.2 (0.0222, >0.999)	6335.8 (0.0222, >0.999)	
Elapsed time	Saturating	469	26	$\varphi, \alpha, \delta, \bar{b}_f^{**}, \bar{b}_\lambda, \bar{b}_\tau, \bar{a}_f, \beta_{af}, \bar{a}_\lambda, \bar{a}_\tau, \bar{g}_\lambda, \beta_{g\lambda}, \bar{g}_\tau, \beta_{g\tau}$	1951.4 (0.0448, 1)	1951.4 (0.0448, 1)	1878.4 (se 113.2)
	Non-Sat	469	26	$\varphi, \alpha, \delta, \bar{b}_f^{**}, \bar{b}_\lambda, \beta_{b\lambda}, \bar{b}_\tau, \beta_{b\tau}, \bar{a}_f, \beta_{af}, \bar{a}_\lambda, \beta_{a\lambda}, \bar{a}_\tau, \beta_{a\tau}$	2758.9 (0.0222, >0.999)	2758.9 (0.0222, >0.999)	
	Resid-O	396	22	$\varphi, \alpha, \delta, \bar{b}_f, \beta_{b\lambda}, \bar{b}_\lambda, \beta_{b\lambda}, \bar{b}_\tau, \beta_{b\tau}, \bar{a}_\tau, \beta_{a\tau}$	5733.0 (0.0222, >0.999)	5163.6 (0.0222, >0.999)	

Note: Non-Sat = Non-saturating; Resid-O = Residual only; AIC = Akaike information criterion; BIC = Bayesian information criterion; PSIS-LOOIC = leave-one-out cross validation via Pareto smoothed

579 importance sampling information criterion; PEMP = posterior expectation of model probability; PEP
 580 = protected exceedance probability.

581 *Absolute difference from best-fitting model (best AIC = -51777.1 , BIC = -48118.7 , LOOIC = $-$
 582 52060.7). Measures exclude group-level parameter estimation.

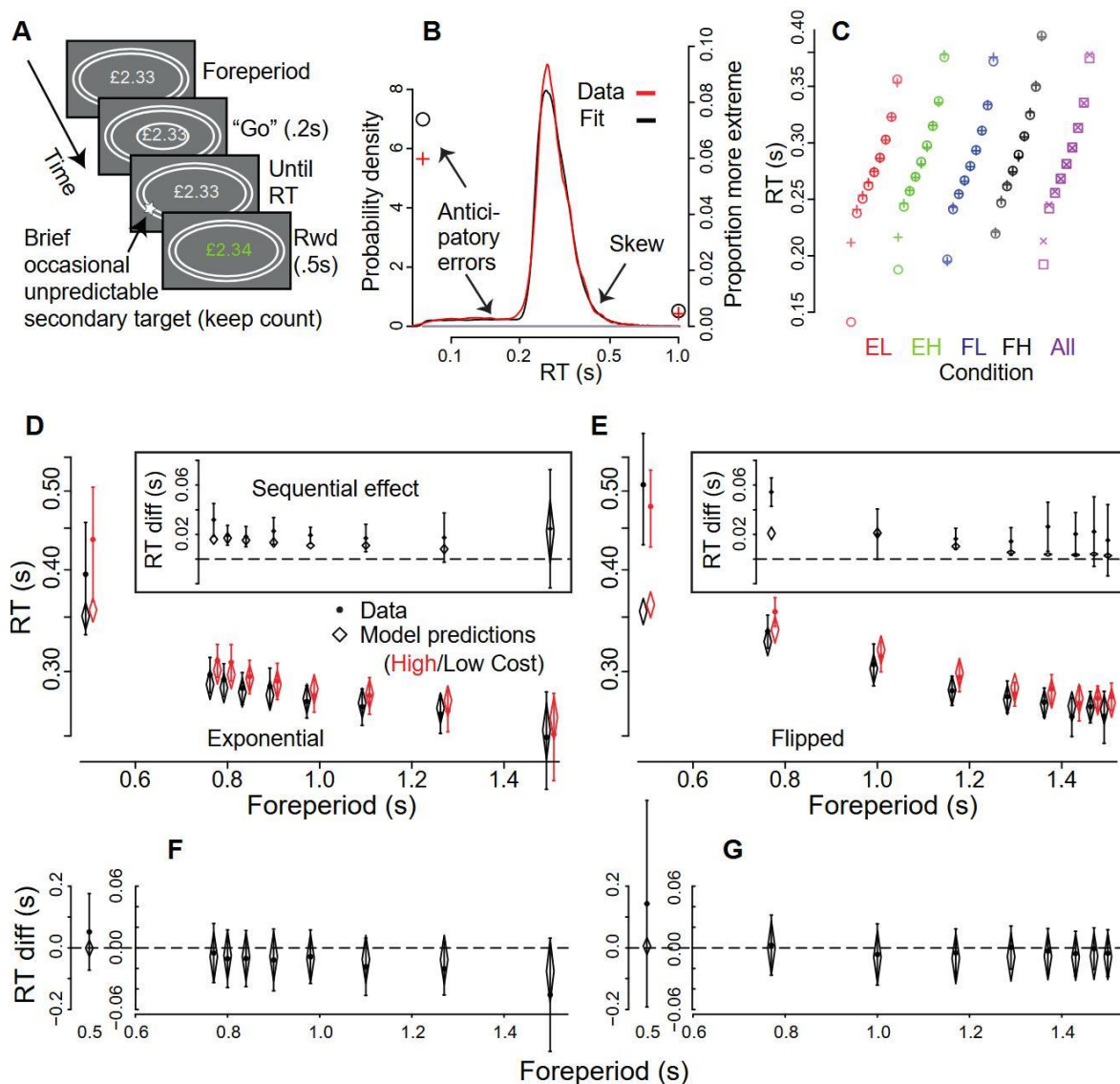
583 **Single parameter estimated for entire group.

584

585 The non-saturating hazard-based preparation model comes closest to matching
 586 performance of the saturating variant, but it is estimated to be best for only around 16% of
 587 participants. To compare the best two models, we supplemented the AIC and BIC estimates of leave-
 588 one-out goodness of fit with the PSIS-LOOIC estimate, which, although time-consuming to compute
 589 (for the current models), provides a method of model comparison that is both tailored to a
 590 multilevel context and considers the full posterior likelihood surface (Lambert, 2018; Vehtari et al.,
 591 2017). It too favoured the saturating model variant ($z = 7.96$, $p < 0.001$). This metric also rejected the
 592 most competitive models for other forms of preparation, linear preparation with saturation ($z =$
 593 16.59 , $p < 0.001$) and PDF preparation without saturation ($z = 32.60$, $p < 0.001$). We will consider
 594 why some of the other model variants failed later, after illustrating how the best model fared across
 595 our various experimental manipulations.

596 A group-level summary of several key results, focussing on foreperiod effects, trial-history
 597 effects, and collated RT distributions, is presented in Figure 6, alongside the MLE fit¹⁰ from the
 598 saturating, hazard-based variant.

¹⁰ The mean posterior fit looks virtually identical to the MLE fit in group-mean plots like these.



599

600 Figure 6. Experiment 1 procedure and results. Figure shows whole-group data along with equivalently

601 averaged or collated saturating hazard-based preparation SIP-thenN-SPuRT model predictions, but

602 fitting occurred via a multilevel model that provided separate fits for each individual. **A.** Schematic of603 procedure (single trial from high-cost condition, rewarded for a fast and accurate response). **B.**

604 Kernel-density estimate of the distribution of RTs (relative to go stimulus) collated across all levels of

605 experimental factors (red) along with equivalent model prediction (black). The left axis relates to

606 density for RTs both > 0.075 and < 1 second; right axis relates to discrete probability for RTs < 0.075 607 or > 1 second (shown as crosses and circles at $RT = 0.075$ and 1 , respectively). Note that the x axis608 uses a logarithmic scale that masks the heavy right tail characteristic of RT distributions. **C.** Quantiles

609 of distributions of RT (from 10th to 90th percentile, left to right offset within each condition) for trials
 610 collated across participants and foreperiods but separated by block (E = exponential distribution, F =
 611 flipped distribution, H denotes a high opportunity-cost block, L a low-cost condition). Open symbols
 612 are model predictions, crosses are data. **D-E.** Main plots show average effect of foreperiod on mean
 613 of median RTs in blocks utilising the exponential foreperiod distribution (D) or the flipped foreperiod
 614 distribution (E). Solid symbols are data, open diamonds are model predictions. The effect is shown
 615 separately for high (red) and low (black) cost blocks (x-axis offset added for clarity). Insets illustrate
 616 $n - 1$ sequential effects as the mean difference in median RTs (collated across high/low-cost blocks)
 617 between trials preceded by one of the two longest foreperiods and trials preceded by one of the two
 618 shortest foreperiods. Error bars (around data points) and top/bottom vertices (for model predictions)
 619 indicate 95% confidence intervals across the group. **F-G.** Prior-block effects. Mean difference in
 620 median RTs between two subgroups of participants who, as a result of counterbalancing,
 621 experienced the flipped vs. exponential blocks first. The surprise (0.5 s) foreperiod data point is shown
 622 at a different axis scale (far left) relative to more frequently experienced foreperiods. Error bars
 623 (around data points) and top/bottom vertices (for model predictions) indicate 95% confidence
 624 intervals around between-group differences.

625 It is apparent from Figure 6 panel B and C that the overall shape of the RT distribution is
 626 captured well by the saturating, hazard-based variant of SIP-theN-SPuRT. The quantile (decile) plots
 627 in panel C are more informative about the model's failings, however, which seem to relate mainly to
 628 simultaneously capturing the left-hand tail of the distribution (anticipatory errors) for the two
 629 contrasting kinds of foreperiod distribution (this model variant does well with the flipped
 630 distribution, but seems to overestimate the magnitude of anticipations in the exponential
 631 distribution, and thus also the data set as a whole).

632 The main (i.e. non-inset) parts of Figure 6 panels D and E focus on just the median RT (rather
 633 than the full distribution) but break the data down by the different levels of all three experimental

factors. In line with the classic variable-foreperiod effect, RTs decrease at greater foreperiods. The way in which RT varies as a function of foreperiod appears subtly different for the exponential vs. flipped foreperiod distribution blocks. RTs are also faster when participants do not experience any opportunity cost (via a requirement to monitor a peripheral track for target stimuli; see the procedural schematic in panel A). Aside from the surprise foreperiod at 0.5 s, the two most extreme foreperiods in this experiment were 0.77 s and 1.5 s. Because they uniquely appeared in blocks with both kinds of distribution (exponential and flipped), we entered them into a 2 (distribution) x 2 (cost) x 2 (foreperiod) repeated-measures ANOVA. This yielded a main effect of distribution ($F(1,35) = 25.33, p < 0.001, \omega^2 = 0.014$) along with main effects of cost ($F(1,35) = 6.32, p = 0.017, \omega^2 = 0.093$) and foreperiod ($F(1,35) = 95.13, p < 0.001, \omega^2 = 0.291$). There were no significant interactions. Hence the exact distribution of foreperiods indeed mattered.

Two further separate 2x8 ANOVAs¹¹ using all foreperiods except the 0.5-s surprise (for which the single relevant trial fell into the >1 s category for some participants/conditions so would create a missing data problem) confirmed that significant main effects of both cost (exponential: $F(1,35) = 5.65, p = 0.023, \omega^2 = 0.006$; flipped: $F(1,35) = 11.18, p = 0.002, \omega^2 = 0.019$) and foreperiod (exponential: $F(7,245) = 16.21, p < 0.001, \omega^2 = 0.098$; flipped: $F(7,245) = 94.70, p < 0.001, \omega^2 = 0.300$) were present in each distribution, with no significant interactions. Previous research using different kinds of dual-task manipulation has sometimes revealed an interaction with the foreperiod effect (e.g. Vallesi et al., 2014) and sometimes not (e.g. Van Lambalgen & Los, 2008). The exact timing(s) of the extra load(s) imposed by an additional task is an important consideration for interpreting findings of this kind, and we note that the adjustment to focussed attention employed here should, by design, be uniform across the foreperiod.

Although it struggles with the very rare surprise foreperiod of 0.5 seconds (experienced just once at the very end of each block), the best-fitting hazard-based preparation model (saturating)

¹¹ All reported ANOVA p-values employ the Greenhouse-Geisser correction for violations of sphericity.

658 clearly outperforms the best-fitting PDF-based preparation model (non-saturating). PDF-based
 659 preparation variants fail because they necessarily predict longer RTs for foreperiods whose
 660 estimated likelihood is low. Such is the case on the right tail of the exponential distribution, e.g. at a
 661 foreperiod of 1.5 s. However, as is apparent from Figure 6D, RTs are relatively short in this condition.
 662 The model of preparation based directly on elapsed time might, with saturation, approximate how
 663 RTs decrease towards a minimum. However, we find this model is less able to capture the exact
 664 pattern across all conditions.

665 From the insets to Figure 6 panels D and E, we can see the $n - 1$ sequential effects of
 666 preceding a trial with one of the two longest vs. the two shortest foreperiods (collated across those
 667 two foreperiods and both cost conditions). The difference is generally positive, indicating that RTs
 668 were speeded when the preceding foreperiod was short compared to long. This effect was
 669 confirmed via two separate 2 (preceding foreperiod) x 8 (foreperiod) repeated-measures ANOVAs,
 670 one per distribution, with main effects of the preceding foreperiod (exponential: $F(1,35) = 30.79$, $p <$
 671 0.001 , $\omega^2 = 0.060$; flipped: $F(1,35) = 35.60$, $p < 0.001$, $\omega^2 = 0.078$) and the current foreperiod
 672 (exponential: $F(7,245) = 18.40$, $p < 0.001$, $\omega^2 = 0.154$; flipped: $F(7,245) = 34.75$, $p < 0.001$, $\omega^2 = 0.226$)
 673 but no significant interactions. Because we included a process by which the estimated PDF of
 674 foreperiods is updated from trial to trial, the model captures this $n - 1$ pattern well, except on trials
 675 with a 0.77-s foreperiod in the flipped-distribution blocks.

676 Figure 6 panels F and G show history effects resulting from past blocks of trials with a
 677 different foreperiod distribution. Our design was not optimised to elicit such effects, which are
 678 subtle. In panel F, negative values indicate that prior experience with the flipped distribution
 679 (compared to no such experience), which should shift preparation relatively later, made participants
 680 trend slower. In panel G, negative values indicate that prior experience with the exponential
 681 distribution, which should shift preparation earlier, made participants trend faster. However, two
 682 separate 2 (prior block) x 8 (foreperiod) mixed between/within-participant ANOVAs yielded only

683 significant effects of foreperiod (exponential: $F(7,238) = 14.19$, $p < 0.001$, $\omega^2 = 0.084$; flipped:
 684 $F(7,238) = 131.33$, $p < 0.001$, $\omega^2 = 0.320$). The model nonetheless provides predictions that are
 685 broadly consistent with the non-significant trend towards effects of a prior block.

686 We considered three model variants, depicted in Figure 5, which quantify the value of our
 687 modelling innovations relative to previous work (which focussed largely on changes in mean RT). The
 688 residual-only variants are closest to the spirit of previous work, as they permit only variations in
 689 residual RT with foreperiod. However, these variants are not competitive compared to the winning
 690 model variant, even when utilizing hazard-based preparation, despite their reduced AIC/BIC
 691 correction for model complexity. Figure 7 elucidates one factor in this failure. It plots the mean
 692 distance between upper/lower quartiles of the RT distribution and the median, as foreperiod varies,
 693 characterising changes in the spread of the RT distribution. We assessed these changes via four 2
 694 (cost) x 8 (foreperiod) repeated-measures ANOVAs (one for each kind of foreperiod distribution and
 695 quartile). For the lower-quartile-to-median distance, there were main effects of foreperiod in all
 696 blocks (exponential: $F(7,245) = 20.23$, $p < 0.001$, $\omega^2 = 0.305$; flipped: $F(7,245) = 4.65$, $p = 0.002$, $\omega^2 =$
 697 0.066). For the upper-quartile-to-median distance, there was a main effect foreperiod, but only in
 698 the flipped distribution blocks ($F(7,245) = 7.86$, $p = 0.001$, $\omega^2 = 0.096$). Hence the shape of the RT
 699 distribution is changing significantly with foreperiod. There were no other significant effects. The
 700 saturating model variant captures these changes well (Figure 7, open diamonds). The residual-only
 701 model variant (Figure 7, grey regions) has very little capacity to model these effects.

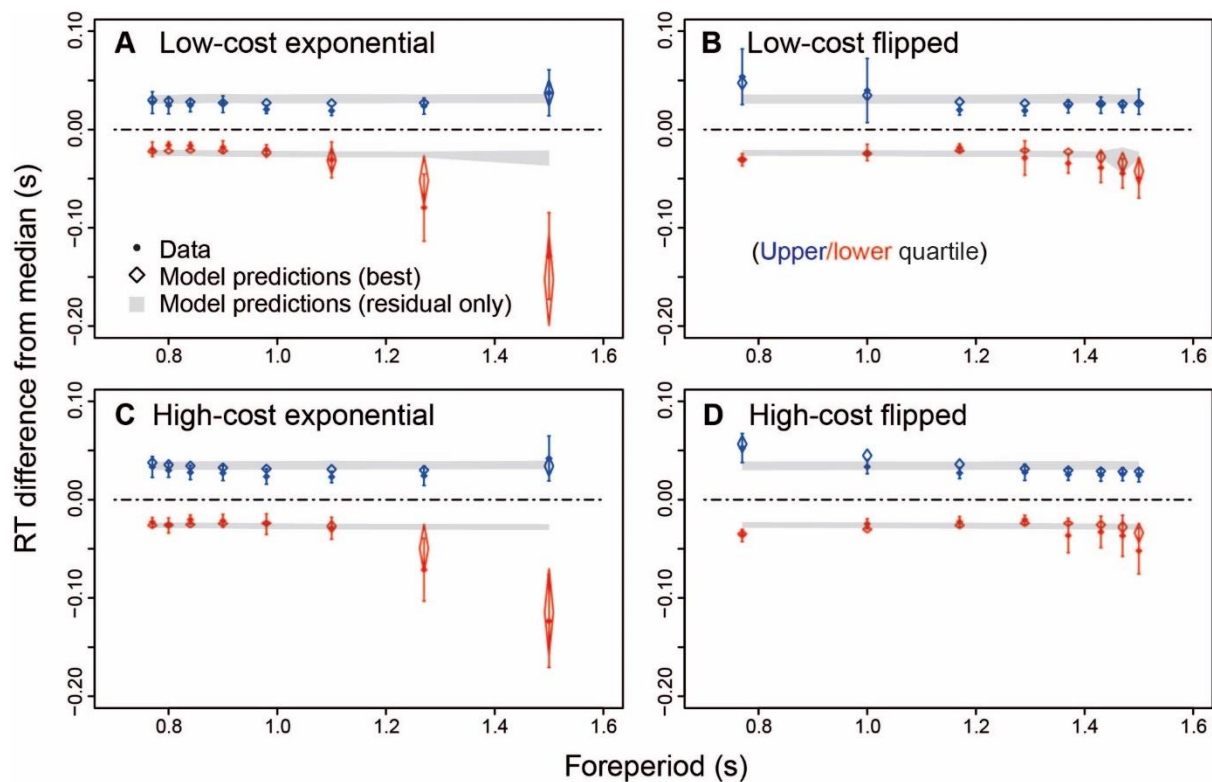
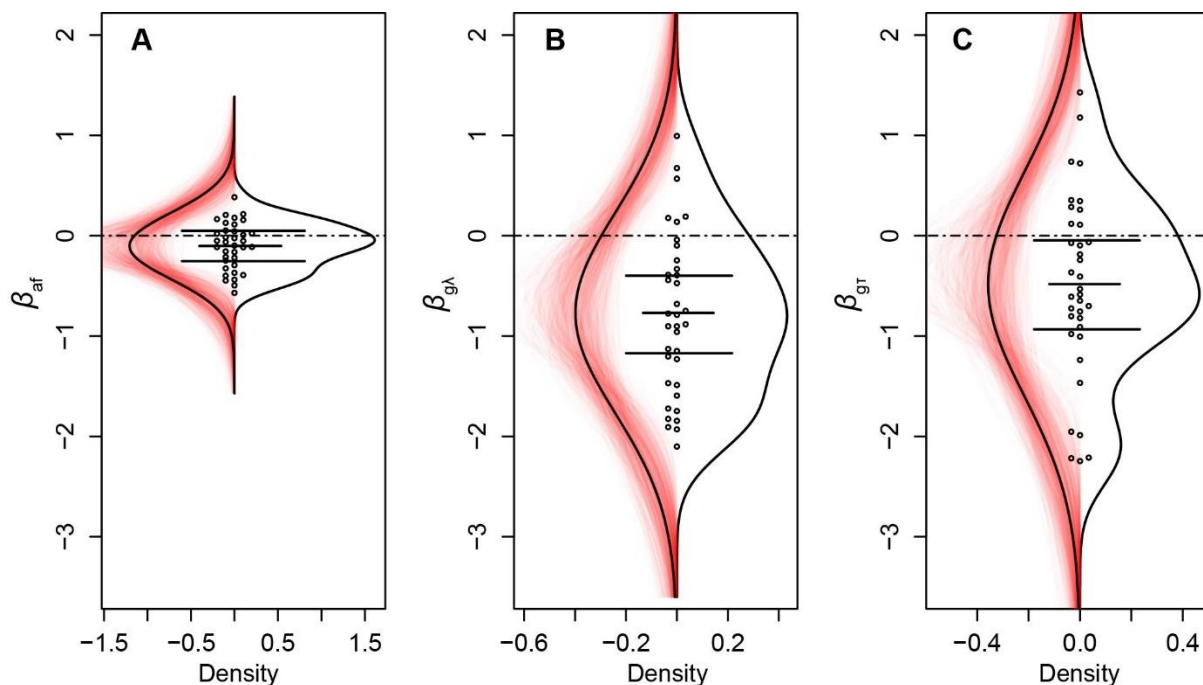


Figure 7. Variable-foreperiod effects on the shape of the RT distribution. Blue symbols show data (solid points) and saturating hazard-based model predictions (open symbols) for the 50th to 75th percentile of RTs. Red symbols denote the 50th to 25th percentile. Error bars show the 95% confidence interval on the data, while upper and lower diamond vertices show the equivalent confidence interval around model predictions. Grey shading denotes the (interpolated) 95% confidence region for the residual-only hazard-based model prediction. **A.** Exponential distribution low-cost block. **B.** Flipped distribution low-cost block. **C.** Exponential distribution high-cost block. **D.** Flipped distribution high-cost block.

How is the saturating model variant able to explain the changes in behaviour observed when opportunity costs are varied by having participants pay attention to a peripheral track in order to avoid financial penalties? As noted earlier, the cost-dependent elevation in median RT was fairly constant across foreperiods (see Figure 6D,E and associated ANOVA results). However, the winning variant generates this change via condition-dependent adjustments to the gain with which subjective hazard is translated into three processes: Urgency, stimulus-dependent drive, and

717 residual RT (formally, non-zero values for parameters β_{af} , $\beta_{g\lambda}$, $\beta_{g\tau}$ respectively). Hence, an outcome
 718 that might at first glance appear independent of and thus additive with the process of temporal
 719 preparation can in fact be modelled as a consequence of adjustments to temporal preparation. As
 720 illustrated in Figure 8, the presence of an opportunity cost generally results in a reduced gain of
 721 translation for all three model processes, i.e. less aggressive preparation. Credible intervals around
 722 the mean posterior group-average parameter values imply a non-significant trend towards a
 723 reduction to urgency (of 9.6%) and significant reductions of 53.7% to the gain of stimulus-dependent
 724 drive and 38.3% to the gain of residual RT.



726 *Figure 8. Saturating hazard-dependent SIP-theN-SPuRT model parameters quantifying changes*
 727 *across opportunity-cost conditions in Experiment 1. Circles indicate parameter estimates for*
 728 *individual participants. Shorter central solid horizontal lines indicate the group mean. Longer central*
 729 *solid horizontal lines indicate the 95% Bayesian credible interval around the group mean. The left-*
 730 *hand butterfly wings describe the model's estimate of the group-level distribution, with the black line*
 731 *illustrating the mean posterior estimate while red shading is derived from the entire posterior to*
 732 *illustrate uncertainty. Right-hand butterfly wings denote kernel-density estimates of individual*

parameter values. **A.** β_{af} , the (logarithmic) change in the rise of go-stimulus-independent urgency. **B.** $\beta_{g\lambda}$, the (logarithmic) change in the rate with which preparation affects go-stimulus-dependent drive (prior to saturation). **C.** $\beta_{g\tau}$, the (logarithmic) change in the rate with which preparation affects residual RT (prior to saturation).

The SIP-theN-SPuRT modelling architecture assumes that the speed-accuracy tradeoff will be captured by changes in urgency, perhaps in some relationship with changes in stimulus-dependent drive or residual RT, but it does not enforce this mechanistically. However, we noted an interesting positive correlation ($r(34) = 0.46$, $p = 0.005$) between the parameter quantifying cost-related changes in urgency (β_{af}) and that quantifying changes in residual RT ($\beta_{g\tau}$). Perhaps, then, more aggressive preparation-dependent reductions in residual RT in low-cost conditions (e.g. putting response-relevant muscles on more of a hair trigger) carries an unavoidable cost in terms of greater increases in the go-stimulus-independent rate of response triggering.

For our best-fitting saturating hazard-based model variant, data (from 36 experimental conditions) are explained using 13 parameters for each participant, plus a further parameter fixed for the whole group. Lest this should seem to offer excessive parametric flexibility, we fitted a further model based on our winning model, but making a larger part of the fit occur at the group, rather than at the individual, level. This analysis is presented in Appendix B. Substantially reducing model parameters (to 10 per participant, with four parameters fixed for the whole group) did not affect any of our conclusions about the model's capacity to explain key trends illustrated in Figures 6 and 7.

753

754 **Experiment 2**

Experiment 1 provided initial evidence supporting the inclusion of several features of our new modelling approach. Benefits were shown to accrue from 1) multilevel modelling at the whole-

RT-distribution level to explain the variable-foreperiod effect on multiple distribution quantiles; 2) trial-by-trial updating of the estimated PDF to explain foreperiod history effects; and 3) saturating hazard-based preparation to maximise overall explanatory power. However, it is unwise to attempt to draw generally applicable conclusions from a single experimental scenario. Hence, we now turn to a published data set from Visalli et al. (2021) and consider a situation where participants are provided with information on some trials regarding enhanced or diminished relevance of the current foreperiod for those that follow.

Methods

A full description of methods is provided by Visalli et al. (2021). Briefly – 27 participants completed a simple RT experiment that might be characterised as a blend of classic fixed-foreperiod and variable-foreperiod designs. On each trial a warning stimulus (a cross) was followed by a go stimulus (a filled circle; see Figure 9A). A continuous sequence of 790 trials was created by appending 72 miniblocks, each of 8-13 trials, one after another. Most of the foreperiods in each miniblock were drawn from a Gaussian distribution, but with a mean and standard deviation that changed from miniblock to miniblock (9 means from 500-2100 ms, combined with 8 standard deviations, from 25 to 200 ms). Go-stimuli in each miniblock had the same colour, which changed from the previous miniblock, alerting participants to the change in mean foreperiod. Additionally, 20% of trials were drawn from a wide and uniformly distributed range of foreperiods (250-2500 ms) and used a white circle as a go stimulus, alerting participants that they were not relevant to estimating the mean foreperiod in the current miniblock. All participants experienced the same sequence of trials.

Data were modelled using an approach similar to Experiment 1 but matched to this particular experimental design. For all three model variants (residual-only, non-saturating and saturating) we permitted changes to the trial-by-trial rate at which the estimated PDF of foreperiods

was updated (α) via two change parameters, $\beta_{\alpha FT}$ and $\beta_{\alpha WG}$, applied via dummy coding. The former coefficient allows the update rate to change on “update” trials, i.e. the first trial of each new miniblock (indicated by a change in the colour of the go stimulus). The latter coefficient allows the update rate to change on “uniform” trials (indicated by a white go stimulus, and not relevant to estimating the current miniblock’s statistics). Both coefficients adjusted a logit version of α utilised for parameter estimation prior to transformation onto a 0-1 range.

Results and Discussion

Table 3 shows goodness of fit metrics for all nine model variants. Although the saturating version of the simple, elapsed-time-based model variant fits best (and is estimated to be the best model for around 42% of participants based on AIC and 30% based on BIC) it does not best suit a significantly greater proportion of participants than all other models (1-PEP > 0.05). This is likely because the saturating hazard-based model (which won in Experiment 1) is a very close second, estimated to fit best for around 31% and 21% of participants based on AIC and BIC respectively. These two models do not differ significantly based on PSIS-LOOIC ($z = 1.53$, $p = 0.126$). Other model variants, especially PDF-based ones, performed less well.

Table 3: *Fits of model variants to Visalli et al. (2021).*

Model Variant		Parameters			Δ Goodness of fit*		
		Number		Individual parameters	MLE		Bayesian
Preparation function	Variant	Individual	Group-level		AIC (PEMP, 1-PEP)	BIC (PEMP, 1-PEP)	PSIS-LOOIC
	Saturating	325	24	$\varphi, \alpha, \delta, b_f^{**}, a_f, b_\lambda, a_\lambda, g_\lambda, b_\tau, a_\tau, g_\tau, \beta_{\alpha FT}, \beta_{\alpha WG}$	24.2 (0.3088, 0.796)	24.2 (0.2100, 0.824)	130.2 (se 85.0)
Hazard	Non-Sat	271	20	$\varphi, \alpha, \delta, b_f^{**}, a_f, b_\lambda, a_\lambda, b_\tau, a_\tau, \beta_{\alpha FT}, \beta_{\alpha WG}$	539.5 (0.0724, >0.999)	109.3 (0.1824, 0.898)	

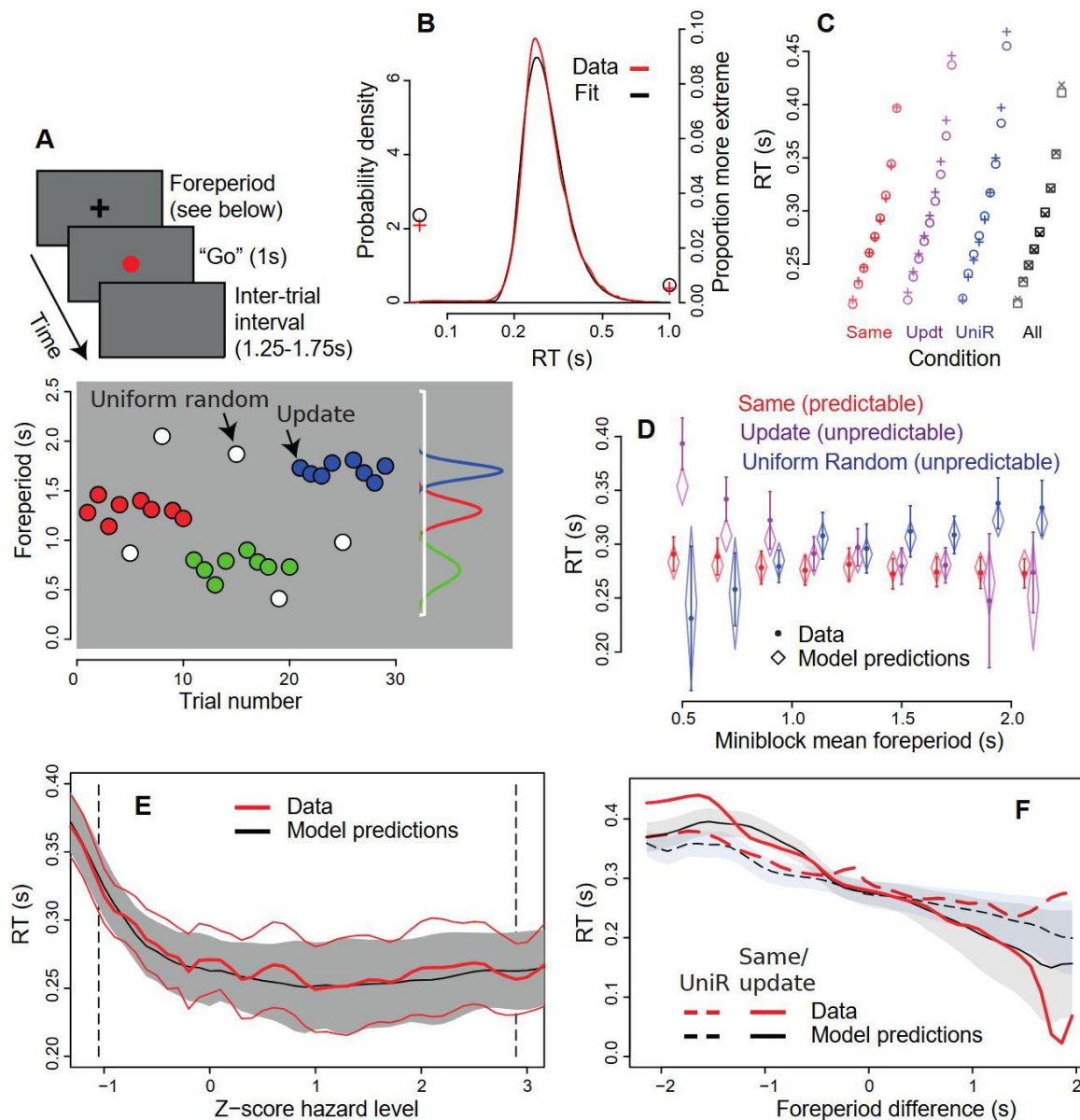
PDF	Resid-O	243	18	$\varphi, \alpha, \delta, b_f, b_\lambda, a_\tau, b_\tau, \beta_{\alpha FT}, \beta_{\alpha WG}$	1986 (0.0278, >0.999)	1332.6 (0.0292, >0.999)
	Saturating	325	24	$\varphi, \alpha, \delta, b_f^{**}, a_f, b_\lambda, a_\lambda, g_\lambda, b_\tau, a_\tau, g_\tau, \beta_{\alpha FT}, \beta_{\alpha WG}$	2194.9 (0.0278, >0.999)	2194.9 (0.0278, >0.999)
	Non-Sat	271	20	$\varphi, \alpha, \delta, b_f^{**}, a_f, b_\lambda, a_\lambda, b_\tau, a_\tau, \beta_{\alpha FT}, \beta_{\alpha WG}$	2587.1 (0.0278, >0.999)	2156.9 (0.0278, >0.999)
	Resid-O	243	18	$\varphi, \alpha, \delta, b_f, b_\lambda, a_\tau, b_\tau, \beta_{\alpha FT}, \beta_{\alpha WG}$	3216.6 (0.0278, >0.999)	2563.3 (0.0278, >0.999)
Elapsed time	Saturating	325	24	$\varphi, \alpha, \delta, b_f^{**}, a_f, b_\lambda, a_\lambda, g_\lambda, b_\tau, a_\tau, g_\tau, \beta_{\alpha FT}, \beta_{\alpha WG}$	0 (0.4246, 0.204)	0 (0.3014, 0.313)
	Non-Sat	271	20	$\varphi, \alpha, \delta, b_f^{**}, a_f, b_\lambda, a_\lambda, b_\tau, a_\tau, \beta_{\alpha FT}, \beta_{\alpha WG}$	1033.3 (0.0552, >0.999)	603.0 (0.1386, 0.967)
	Resid-O	243	18	$\varphi, \alpha, \delta, b_f, b_\lambda, a_\tau, b_\tau, \beta_{\alpha FT}, \beta_{\alpha WG}$	2456.0 (0.0278, >0.999)	1802.7 (0.0551, >0.999)

Note: Non-Sat = Non-saturating; Resid-O = Residual only; AIC = Akaike information criterion; BIC = Bayesian information criterion; PSIS-LOOIC = leave-one-out cross validation via Pareto smoothed importance sampling information criterion; PEMP = posterior expectation of model probability; PEP = protected exceedance probability.

*Absolute difference from best-fitting model (best AIC = -54584.2, BIC = -51994.7, LOOIC = -55012.4). Measures exclude group-level parameter estimation.

**Single parameter estimated for entire group.

A summary of group-level distributions and the effects of different trial types is presented in Figure 9. Although two models were viable, we again use MLE fits from the saturating hazard-based model variant in illustrations, for reasons that will become clear after we present Experiment 3.



811

812 *Figure 9. Experiment 2 procedure and results. A. Schematic of procedure and examples of different*
 813 *trial types across three mini-blocks of the experiment. B. Kernel-density estimate of the distribution*
 814 *of RTs collated across all participants and trial types (red) along with equivalent distribution*
 815 *predicted by model (black). The left axis relates to density for RTs both > 0.075 and < 1 second; right*
 816 *axis relates to discrete probability for RTs < 0.075 or > 1 second (shown as crosses and circles at RT =*
 817 *0.075 and 1, respectively). Note that the x axis uses a logarithmic scale that masks the heavy right*
 818 *tail characteristic of RT distributions. C. Quantiles of distributions of RT (from 10th to 90th percentile,*
 819 *left to right offset within each condition) for trials collated across participants and foreperiods but*

820 *separated by trial type (UniR = uniform random, Updt = update). Open symbols are model*
 821 *predictions, crosses are data. D. Average effect of the mean of the current miniblock on mean of*
 822 *median RTs. Note that the mean of the current miniblock only predicts the actual trial foreperiod for*
 823 *same and update trials, not for uniform-random trials. Solid symbols are data, open diamonds are*
 824 *model predictions. The effect is shown separately for same (red), update (purple) and uniform*
 825 *random (blue) trials (x-axis offset added for clarity). Error bars (around data points) and top/bottom*
 826 *vertices (for model predictions) indicate 95% confidence intervals across the group. E. Gaussian-*
 827 *smoothed moving average RT derived by sorting trials according to their z-scored level of (model-*
 828 *predicted) hazard. Grey region (for model prediction) and thinner red lines (for data) indicate 95%*
 829 *confidence intervals. F. Gaussian-smoothed moving average RT derived by sorting trials according to*
 830 *the difference between the current and previous foreperiod. Data are shown separately for uniform*
 831 *random trials compared to all other trials. Grey regions indicate 95% confidence intervals on model*
 832 *predictions.*

833

834 Figure 9 panels B and C indicate that the overall shape of the RT distribution is once again
 835 captured well by the saturating variant of the hazard-based preparation SIP-theN-SPuRT model.
 836 Panel D shows how that model also captures differences between trial types across miniblocks. For
 837 the majority of trials (red data) the current foreperiod is predictable to within a tight range due to
 838 the miniblock structure (similar to classic blocked designs yielding fixed-foreperiod effects). Here,
 839 this yields little change in RT with mean miniblock foreperiod. The first “update” trial of a new
 840 miniblock is relatively unexpected, yielding slow reactions if it has a short foreperiod (where it will
 841 typically occur at an earlier than expected time and thus before preparation has built up) and
 842 (somewhat) faster reactions if it has a long foreperiod (where the opposite is true). Uniform-random
 843 trials are also relatively unexpected, but unlike update trials they have a foreperiod that bears no
 844 relation to the mean of the current miniblock. Hence when the current miniblock has a short mean

foreperiod, uniform-random trials tend to have longer than expected foreperiods (yielding fast RTs) and vice versa when the current miniblock has a long foreperiod. (Note that, if plotted against foreperiod, rather than the mean miniblock foreperiod, the function would be downward sloping, as is typically observed for uniform foreperiod distributions). The model captures the trends in both update and uniform-random trials well. It also nicely predicts the reduction in RT for trials with foreperiods implying a greater subjective hazard level (panel E) and the effect of the previous trial's foreperiod relative to the current trial's foreperiod (panel F) – a shortening of foreperiod across trials yields a higher RT compared to a lengthening of foreperiod.

Interestingly, the influence of the previous trial's foreperiod seems similar regardless of what kind of trial it was (illustrated for uniform random vs. all other trials by the dashed vs. solid lines in Figure 9F). The model had a capacity to capture differences via our inclusion of two parameters that could modify α , the weight given to each new foreperiod when it is combined into a participant's estimated PDF. These parameters allow different weightings to be applied following update and uniform-random trials (compared to same trials). Values recovered for these coefficients are presented in Figure 10.

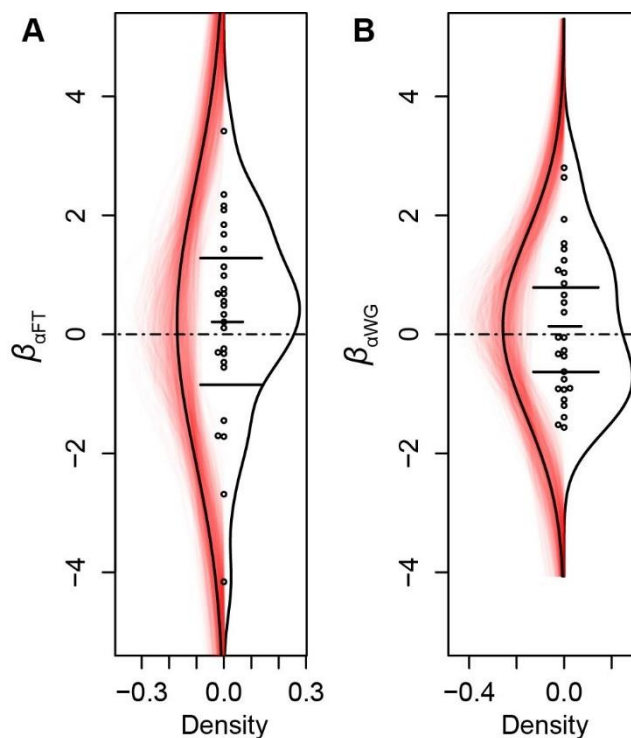


Figure 10. Saturating hazard-dependent SIP-theN-SPuRT model parameters quantifying changes across trial types in Experiment 2. Circles indicate parameter estimates for individual participants. Shorter central solid horizontal lines indicate the group mean. Longer central solid horizontal lines indicate the 95% Bayesian credible interval around the group mean. The left-hand butterfly wings describe the model's estimate of the group-level distribution, with the black line illustrating the mean posterior estimate while red shading is derived from the entire posterior to illustrate uncertainty. Right-hand butterfly wings denote kernel-density estimates of individual parameter values. These parameters quantify additive adjustments to the log odds in baseline "same" trials, i.e. those within a miniblock and adhering to the miniblock's dominant statistics, prior to a conversion into a proportional trial-by-trial update to the estimated PDF. **A.** $\beta_{\alpha_{FT}}$, the change for update trials initiating a new miniblock. **B.** $\beta_{\alpha_{WG}}$, the change for uniform-random trials occurring sporadically throughout the experiment.

From Figure 10, credible intervals around the group-average change scores include zero, indicating that the participant's estimated PDF of foreperiods is, on average, modified similarly following update trials compared to same trials, and also following uniform random trials compared to same trials. These modifications equate to a mean update rate (α) of 0.581 in same trials, 0.566 in update trials, and 0.592 in uniform-random trials. Hence the model's α parameter does not appear to have changed reliably based on the instructions and cues provided to participants, suggesting a largely implicit updating process (or else a lack of engagement, or a recognition that the wider distribution of foreperiods generating uniform-random trials is actually just as relevant to optimising preparation).

Experiment 3

Experiments 1 and 2 provide converging evidence relevant to our two key questions. We asked firstly whether the features that we introduced helped provide a better fit to data. We now have evidence to support the usefulness of most of these model features (e.g. updating of the estimated PDF of foreperiods; the capacity to capture foreperiod effects on the full RT distribution rather than just the mean RT; the capacity to have the level of temporal preparation saturate) but we have not yet considered updates to beliefs about the rate of catch trials. This is because, thus far, we have only considered experiments that did not include any such catch trials. We also asked what function derived from or influenced by the estimated PDF of foreperiods (elapsed time, PDF, or hazard) best reflected the effect of temporal preparation on simple RT. The experiments we have reviewed thus far consistently rejected PDF-based accounts, but both saturating-hazard and saturating-elapsed-time accounts remain somewhat viable.

For our final experiment we present new data collected to address these outstanding issues. The key change is the introduction of catch trials on which no stimulus was delivered. As illustrated in Figure 3, under a hazard-based preparation account, one's belief about the rate of such trials necessarily affects preparation and thus RTs. In the immediate aftermath of a catch trial, the estimated rate of catch trials should increase (particularly if the world is considered non-stationary in this respect) leading to reduced and temporally reshaped hazard-based preparation. By contrast, as currently specified, our PDF and elapsed-time models have no mechanism by which beliefs about the rate of catch trials can affect behaviour. Hence incorporating catch trials and then partitioning data based on time since the last catch trial should allow us to illustrate any history effect and determine if our hazard-based model can approximate it. This should also help differentiate hazard-based from elapsed-time accounts, as the latter accounts will fail to predict catch-trial effects.

Like Experiment 1, all participants in Experiment 3 experienced two different foreperiod distributions, to challenge the models. However, unlike Experiment 1, these two distributions had the same mean, which minimises predicted differences in the case of the elapsed-time model (for

which preparation is only influenced by this summary statistic, albeit on a trial-by-trial basis). We also included two further positive-control manipulations (of stimulus intensity and the effector of response) to assess the extent to which allowing appropriate model parameters from SIP-theN-SPuRT to vary could capture basic manipulations of this kind.

Methods

Briefly – Eight participants experienced a sequence of warning thuds which, on most trials, preceded imperative beeps instructing a response (Figure 11a). In each of 16 blocks of at least 192 trials, foreperiod was randomised between 0.793 and 1.607 s in 24 steps. The intensity of the imperative beep was randomly determined to be strong or weak. In separate blocks, we factorially (2x2x2) varied whether 1) steps could cluster centrally to form a *Gaussian* foreperiod distribution, or not cluster at all to form a *uniform* distribution; 2) catch trials, in which the warning thud simply recurred after 2 s of silence, could either be absent, or could occur on average every one in nine trials; and 3) responses could be made either with a pinchable button or via a foot pedal. Further details are provided in Appendix A.

Analysis

Data trimming

On the trials immediately following a catch trial there was a tendency for responses to be triggered by the warning stimulus. There is no mechanism in our models to generate these responses (which would contaminate results by masquerading as anticipatory errors) so RTs from these trials were not included in the analysis.

933 Modelling

934 Like Experiments 1-2, RT likelihoods from non-catch trials were determined as
 935 described in Equation 16. For catch trials, anticipatory-error likelihoods were calculated similarly,
 936 while correctly withheld responses were treated as members of a discrete category. Thus, for any
 937 combination of model variant and parameter values θ , the effective likelihood of RT r_j on catch-trial
 938 j was defined as:

$$939 \quad \tilde{l}_{\theta}(r_j) \equiv \begin{cases} P(R_j = r_j) & r_j \leq t_j + 2 s \\ P(R_j > t_j + 2 s) & r_j > t_j + 2 s \end{cases}. \quad (17)$$

940 Given the strong evidence from Experiments 1 and 2 that preparation affects all model
 941 processes and was best modelled (with the exception of PDF-based variants) using a saturating
 942 function, we dropped the systematic comparison of saturating, non-saturating and residual-only
 943 model variants in Experiment 3. That is, we only tested saturating variants of elapsed-time and
 944 hazard-based models, along with saturating and non-saturating variants of the PDF-based model.
 945 However, with the introduction of catch trials we now tested two variants of saturating hazard-like
 946 preparation. The first fixed the belief that stimulus omission rate was unchanging, ψ , to equal 1, as
 947 in previous experiments. The second freely estimated ψ , to model the inferred probability (per trial)
 948 that the catch-rate remains the same (in an assumed non-stationary world where the rate of catch
 949 trials can suddenly change).

950 All model variants had the urgency intercept b_f estimated singly for the group and intercepts
 951 for the other two processes (stimulus drive and residual RT) estimated individually. We included all
 952 five possible gain parameters (a_f , a_{λ} , a_{τ} , g_{λ} and g_{τ}), except for the non-saturating PDF-based
 953 variant where pre-saturating gains were dropped. To be consistent with the logical ways in which
 954 different model components might be affected by Experiment 3's two positive-control
 955 manipulations, we allowed only the residual RT to vary with effector and thus differ between pinch
 956 and pedal responses (hence $\beta_{Ea\tau}$, $\beta_{Eb\tau}$ and $\beta_{Eg\tau}$, where E denotes effector, for saturating variants).

For the effect of intensity, we permitted variation in both sensory drive and residual RT, as the residual RT can incorporate both sensory and motor delays (hence $\beta_{Ia\lambda}$, $\beta_{Ib\lambda}$, $\beta_{Ig\lambda}$, $\beta_{Ia\tau}$, $\beta_{Ib\tau}$, and $\beta_{Ig\tau}$ for saturating variants).

Results and discussion

Table 4 shows goodness of fit metrics for the five model variants. Consistent with Experiment 1, the saturating hazard-based model variant fitted best, in particular the version with an additional parameter modelling beliefs that the world was non-stationary regarding the rate of catch trials. This variant was estimated to be best for around 69% of participants, and to be best for a higher proportion of participants than any other model (1-PEP = 0.008). It significantly outperformed the stable-world variant (PSIS-LOOIC $z = 11.97$, $p < 0.001$), which was the next best model. The saturating hazard-based non-stationary variant also clearly outperformed the saturating elapsed-time model, and it was similarly dominant against the best of the two PDF-based models.

Table 4: *Fits of model variants in Experiment 3*

Model Variant		Parameters			Goodness of fit*		
		Number		Individual parameters	MLE		Bayesian
Preparation function	Variant	Individual	Group-level		AIC (PEMP, 1-PEP)	BIC (PEMP, 1-PEP)	PSIS-LOOIC
Hazard	Sat non-stable	161	40	$\varphi, \alpha, \psi, \delta, \bar{b}_f^{**}, \bar{a}_f, \bar{b}_\lambda, \bar{a}_\lambda, \bar{g}_\lambda, \bar{b}_\tau, \bar{a}_\tau, \bar{g}_\tau, \beta_{Ib\lambda}, \beta_{Ia\lambda}, \beta_{Ig\lambda}, \beta_{Eb\tau}, \beta_{Ea\tau}, \beta_{Eg\tau}, \beta_{Ib\tau}, \beta_{Ia\tau}, \beta_{Ig\tau}$	0 (0.6923, 0.008)	0 (0.6923, 0.008)	0
	Saturating stable	153	38	$\varphi, \alpha, \delta, \bar{b}_f^{**}, \bar{a}_f, \bar{b}_\lambda, \bar{a}_\lambda, \bar{g}_\lambda, \bar{b}_\tau, \bar{a}_\tau, g_\tau, \beta_{Ib\lambda}, \beta_{Ia\lambda}, \beta_{Ig\lambda}, \beta_{Eb\tau}, \beta_{Ea\tau}, \beta_{Eg\tau}, \beta_{Ib\tau}, \beta_{Ia\tau}, \beta_{Ig\tau}$	795.3 (0.0769, 0.998)	729.4 (0.0769, 0.998)	819.0 (se 68.4)
PDF	Saturating	153	38	$\varphi, \alpha, \delta, \bar{b}_f^{**}, \bar{a}_f, \bar{b}_\lambda, \bar{a}_\lambda, \bar{g}_\lambda, \bar{b}_\tau, \bar{a}_\tau, g_\tau, \beta_{Ib\lambda}, \beta_{Ia\lambda}, \beta_{Ig\lambda}, \beta_{Eb\tau}, \beta_{Ea\tau}, \beta_{Eg\tau}, \beta_{Ib\tau}, \beta_{Ia\tau}, \beta_{Ig\tau}$	2989.9 (0.0769, 0.998)	2924.1 (0.0769, 0.998)	

	Non sat	113	28	$\varphi, \alpha, \delta, \bar{b}_f^{**}, \bar{a}_f, \bar{b}_\lambda, \bar{a}_\lambda,$ $\bar{b}_\tau, \bar{a}_\tau, \beta_{Ib\lambda}, \beta_{Ia\lambda}, \beta_{Eb\tau},$ $\beta_{Ea\tau}, \beta_{Ib\tau}, \beta_{Ia\tau}$	3050.7 (0.0769, 0.998)	2655.9 (0.0769, 0.998)
Elapsed time	Saturat- ing	153	38	$\varphi, \alpha, \delta, \bar{b}_f^{**}, \bar{a}_f, \bar{b}_\lambda, \bar{a}_\lambda,$ $\bar{g}_\lambda, \bar{b}_\tau, \bar{a}_\tau, g_\tau, \beta_{Ib\lambda},$ $\beta_{Ia\lambda}, \beta_{Ig\lambda}, \beta_{Eb\tau}, \beta_{Ea\tau},$ $\beta_{Eg\tau}, \beta_{Ib\tau}, \beta_{Ia\tau}, \beta_{Ig\tau}$	2448.3 (0.0769, 0.998)	2382.5 (0.0769, 0.998)

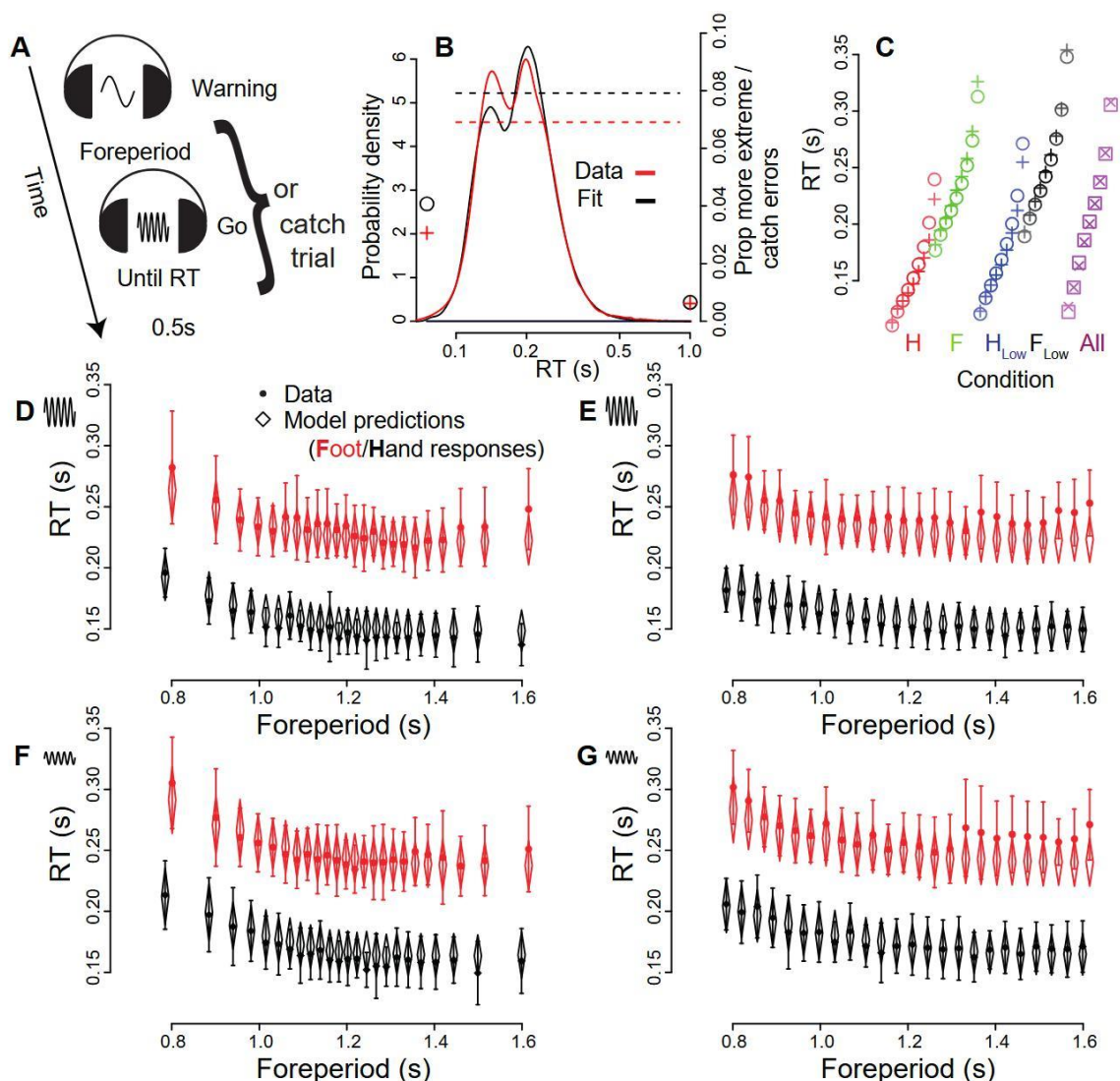
972 *Note:* Sat non-stable = Saturating non-stable; Non sat = Non saturating; AIC = Akaike information
973 criterion; BIC = Bayesian information criterion; PSIS-LOOIC = leave-one-out cross validation via
974 Pareto smoothed importance sampling information criterion; PEMP = posterior expectation of
975 model probability; PEP = protected exceedance probability.

976 * Absolute difference from best-fitting model (best AIC = -79956.7, BIC = -78632.2, LOOIC = -
977 79821.6). Measures exclude group-level parameter estimation

978 **Single parameter estimated for entire group

979

980 In Experiment 3, our key new manipulation was the occurrence of catch trials. We will turn
981 to this soon. However, we first illustrate the effect of the various positive-control manipulations that
982 were also included. A summary of group-level RT distributions and the effects of these positive-
983 controls is presented in Figure 11, alongside the MLE fit from the saturating hazard-based non-stable
984 world SIP-theN-SPuRT model variant.



985

986 Figure 11. Experiment 3 procedure and results. **A.** Schematic of procedure. **B.** Kernel-density estimate
 987 of the distribution of RTs collated across all participants and trial types (red) along with equivalent
 988 distribution predicted by model (black). The left axis relates to density for RTs both > 0.075 and < 1
 989 second; right axis relates to discrete probability for RTs < 0.075 or > 1 second (shown as crosses and
 990 circles at RT = 0.075 and 1, respectively) and for catch trials yielding a commission error (shown as
 991 horizontal dashed lines). Note that the x axis uses a logarithmic scale that masks the heavy right tail
 992 characteristic of RT distributions. **C.** Quantiles of distributions of RT (from 10th to 90th percentile, left
 993 to right offset within each condition) for trials collated across participants and foreperiods but
 994 separated by effector (H = hand, F = foot) and stimulus intensity (high unless denoted low). Open
 995 symbols are model predictions, crosses are data. **D-G.** Average effect of foreperiod on mean of

median RTs in blocks utilising the Gaussian foreperiod distribution (D, F) or the uniform foreperiod distribution (E, G) and for high (D-E) and low (F-G) stimulus intensity trials. Solid symbols are data, open diamonds are model predictions. The effect is shown separately for foot (red) and hand (black) response blocks (x-axis offset added for clarity). Error bars (around data points) and top/bottom vertices (for model predictions) indicate 95% confidence intervals across the group.

Figure 11 panels B and C indicate that the overall shape of the RT distribution is once again captured well by the saturating hazard-based preparation SIP-theN-SPuRT model. The distribution is bimodal, resulting from the marked difference in RT between hand and foot responses. However, while fitting well for the dataset as a whole, the model slightly underestimates dispersion in foot conditions and overestimates it in hand conditions (panel C) because the parameters that are allowed to vary between these conditions affect only the non-decision time. This highlights one way in which SIP-theN-SpuRT's assumptions are too simplistic – empirically, the effector of response significantly influenced the distance between 75th percentile and median of the RT distribution in the uniform distribution blocks ($F(1,7) = 5.98, p = 0.044, \omega^2 = 0.096$) but the model cannot capture this effect.¹² Panels D-G show how RTs are reduced at longer foreperiods, higher stimulus intensities, and, most dramatically, for pinch vs. pedal responses. The model captures all these trends well. Inferential statistics assessing the reliability of these effects, along with an examination of how the model captured them via changes in relevant parameters, are presented in Appendix C.

We now turn to the key question for this experiment: How catch trials influenced behaviour. Figure 12 illustrates these effects averaged across types of response and stimulus intensities. Data are separated by foreperiod distribution (Gaussian and uniform to the left and right respectively). In

¹² There were similar but non-significant trends for the 25th percentile to median distance, and in the Gaussian distribution blocks. We collapsed over the two intensities for these inferential tests. As in Experiment 1 there were also trends towards effects of foreperiod on distribution shape, but with sphericity corrections these did not reach significance. We have not illustrated these effects for concision.

the upper panels, they are further separated by catch-trial recency, illustrating the clear trend for longer RTs after a catch trial occurs. RTs are most elevated at the earliest interval for which data were not contaminated by spuriously triggered responses and thus included in the analysis (i.e. two trials after a catch trial) and seem to level off by the time a catch trial hasn't occurred for around 30 trials. Note that the apparently linear decline in RTs with log trials would appear more exponential, had it been plotted against a linear function of trial number. The best-fitting model (saturating hazard with a parameter, ψ , to capture a non-stationary rate of catch trials) fits the trend in the data fairly well, but it slightly underestimates its magnitude (diamonds). When the catch trial rate is assumed to be constant, the otherwise-identical model (grey region) predicts an even shallower decline in RT, illustrating how including the ψ parameter improved fit.

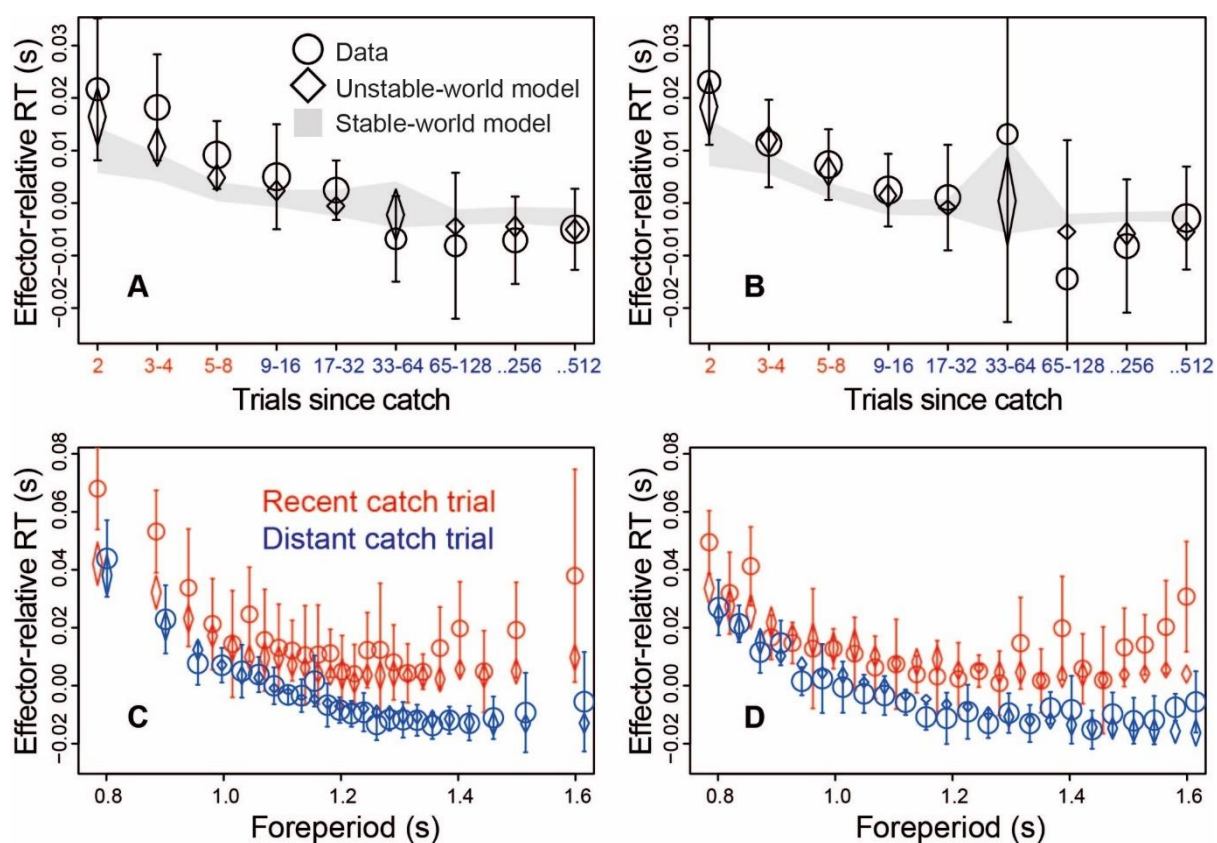


Figure 12. Average effect of catch-trial recency on mean of median RT (expressed as differences relative to the current effector) in blocks utilising the Gaussian foreperiod distribution (A,C) or the uniform foreperiod distribution (B,D) from Experiment 3. Circles are data, with circle size proportional

to average number of contributing trials. Open diamonds are model predictions for the best-fitting (saturating hazard-based) model, which incorporated an estimate of how likely the rate of catch trials is to suddenly change. Grey solid regions are the (interpolated) predictions of the same model but with catch-trial rate assumed to be stationary. Error bars (around data points) and top/bottom vertices (for model predictions) indicate 95% confidence intervals across the group. In parts C and D, data are collapsed into two recency categories (red for 2-8 trials since a catch trial, blue for 9+ trials, x-axis offset added for clarity) but additionally separated by foreperiod.

In the lower two panels of Figure 12, data are collated less finely for catch-trial recency, into just two (recent and non-recent) bins, but are further partitioned by foreperiod. Here, only the best-fitting model fit is shown, capturing data fairly well and predicting a trend toward larger catch-trial recency effects at long foreperiods. SIP-theN-SpuRT makes this prediction because the estimated catch-trial rate, $\hat{b}_n$, which is higher following a catch trial, forms a lower bound on the denominator for the estimated hazard function. It thus typically has a greater impact at higher foreperiods where the function might otherwise tend towards infinity (see Equations 4 and 5, and Figure 3).

For inferential tests, relating to the effect of catch-trial recency, we restricted our analyses to ensure sufficient data from all participants in all ANOVA cells.¹³ To test the trends shown in Figure 12 panels A and B, we tested differences across only the first five levels of catch-trial recency (so for catch trials occurring up to 32 trials ago) in a 2 (distribution) x 5 (catch-trial recency) repeated-measures ANOVA. There was a main effect of catch-trial recency ($F(4,28) = 14.65$, $p < 0.001$, $\omega^2 = 0.413$) that did not interact with distribution. Holm-corrected contrasts showed significant differences ($p_{adj} < 0.05$) for levels 1 vs. 3, 4 and 5, and level 2 vs. 4 and 5. The data shown in Figure 12 panels C and D were subjected to two separate 2 (catch-trial recency) x 24 (foreperiod) ANOVAs,

¹³ We first collated data across both intensity levels, so do not include this factor. We also expressed RTs relative to the average for a given effector (hand or foot), effectively removing this factor.

which showed, for both distributions, a main effect of catch-trial recency (Gaussian: $F(1,7) = 24.70$, $p = 0.002$, $\omega^2 = 0.650$; uniform: $F(1,7) = 15.85$, $p = 0.005$, $\omega^2 = 0.602$) but the interactions with foreperiod failed to reach significance.¹⁴

To achieve the exponential decline in RT with increasing time since the last catch trial, our winning model was free to estimate a belief that the rate of catch trials would remain stable from trial to trial (parameter ψ). Values for this belief ranged from 0.924 to 0.978 across participants (mean 0.959, SD 0.027). Interestingly, to best capture data these estimates fell well below the correct value for our experiment (which discretely changed catch-trial rate only once every 768 trials, implying $\psi = 0.9987$). However, our experiment sat within a wider real-world context that might be less stable. In future work, it would be interesting to compare a condition that varies catch-trial rates more frequently and see if parameter ψ varies accordingly.

Overall, Experiment 3 supported the idea that preparation is best modelled via a saturating form of subjective hazard, with part of the success of this account coming down to its capacity to model the effect of catch-trial recency. Hazard-based accounts naturally predict such recency effects, whereas our PDF-based and elapsed-time-based models do not. It would of course be possible to add some tweak to the otherwise competitive saturating elapsed-time model, rather as we did in order to let it deal with effects of past foreperiods, in order to model an effect of catch-trial recency. However, other issues would likely ultimately arise, for example in dealing with fixed, rather than variable, foreperiod effects (which hazard-based accounts can in principle model, but the elapsed-time account is probably less well equipped for).

General Discussion

¹⁴ Both ANOVAs also showed significant main effects of foreperiod ($F(23,161) = 13.54$, $p < 0.001$; $F(23,161) = 16.45$, $p < 0.001$).

When imperative stimuli might appear at one of several foreperiods after a warning, behaviour varies with the exact foreperiod on both current and recent trials. However, the basis of this behavioural variation remains controversial. In this paper we have introduced the SIP-theN-SPuRT model architecture and have applied it to predict simple RT data from three diverse experiments sharing the feature that they randomised foreperiods in-between warnings and go stimuli. In SIP-theN-SPuRT, reacting is modelled as an inhomogeneous Poisson process, which combines slowly accumulating urgency with a transient input triggered by the stimulus. The three components of the model (urgency, drive and the leftover residual RT) are each potentially influenced by temporal preparation. We aimed to assess the usefulness of several of these modelling innovations, and to evaluate which of three functions derived from or influenced by the estimated PDF of foreperiods (elapsed time, PDF or hazard) best described temporal preparation under our expanded modelling framework. We summarise what we have learnt in relation to these questions next, before moving on to discuss more general issues and limitations.

Saturating preparation

A key difference between our approach, attempting to formally specify the temporal preparation function, and contributions of previous authors is the incorporation of (potential) saturation at some clearly defined limit. We achieved this via a hyperbolic tangent function, although this choice was quite arbitrary. Incorporating saturation dramatically improved the fit of our models, particularly for hazard-based and elapsed-time-based preparations. Without it, the hazard function can explode (although this is already mitigated by basing the hazard function on the participant's estimated PDF of foreperiods and, in particular, by assuming that participants have a non-zero belief about the rate of catch trials). Any simple elapsed-time-based preparation would also trend unrealistically to infinity, albeit at a rate that would not be problematic in the context of a

1101 typical experiment. However, even PDF-based preparation, which is less inherently problematic from
 1102 this perspective, benefitted from saturation about as often as not across the three experiments.

1103 We acknowledge that the idea that there is a further linking function between the inferred
 1104 likelihood of stimulus occurrence and RTs is not a natural starting point for formal investigations, but
 1105 it seems to us physiologically and conceptually more plausible than a direct translation of probability
 1106 or time onto RTs. We find it physiologically plausible because RTs are necessarily low-bounded. We
 1107 find it conceptually plausible because if performance could reach a certain level late in a trial, then
 1108 with sufficient incentive or a disregard of costs, it seems likely that one should be able to attain that
 1109 level earlier in the trial and then maintain it. Of course, this is not certain – there may be some
 1110 structural limitation to maintaining the highest levels of temporal preparation (Denison, 2024).
 1111 However, the notion of a saturating limit to the effectiveness of increased preparation has been
 1112 touched upon already in purely descriptive models of the variable-foreperiod effect (e.g.
 1113 Houshmand Chatroudi et al., 2024) although this has not been combined with the principled notion
 1114 of hazard-based preparation. We do, of course, recognise that permitting linking functions of this
 1115 kind opens up an almost limitless space of alternatives that might ultimately make it impossible to
 1116 distinguish between different accounts of the underlying computation. Note, however, that we are
 1117 certainly not the first to try out tweaks to the estimated PDF or hazard function, or indeed to
 1118 consider transforms that compress it at the high end, a point we return to later when considering
 1119 how our conclusions compare to those of other authors.

1120

1121 **Modelling the full distribution of RTs via preparatory influences on several processes**

1122 The SIP-theN-SPuRT model architecture allowed us to predict the entire distribution of
 1123 simple RTs, something that has received much less attention than trimmed central tendency
 1124 estimates in the variable-foreperiod literature to date (but see e.g. Tsunoda & Kakei, 2008). In so
 1125 doing, we have described a previously underappreciated variable-foreperiod effect, specifically on

1126 the *dispersion* of RTs. Experiment 1 illustrated how the shape of the RT distribution (quantified as
 1127 distances from the median to the 25th or 75th percentiles) changed in a way that is complex and
 1128 nonlinear with foreperiod, shrinking and then growing again. SIP-theN-SPuRT was able to capture
 1129 these trends by permitting an effect of preparation on parameters reflecting sensory drive and
 1130 stimulus-independent urgency. As far as we are aware, no other formal modelling work has
 1131 attempted to quantitatively relate preparation functions such as subjective hazard to this aspect of
 1132 data, so this represents an important step forward.

1133 More generally, modelling frameworks like SIP-theN-SPuRT allow us to relate experimental
 1134 manipulations to a set of putative underlying cognitive operations. In Experiment 1, an additional
 1135 opportunity cost to maintaining focussed attention (via a requirement to attend peripherally for
 1136 high-value targets) yielded changes in RT. These could be explained via reduced values for
 1137 parameters capturing the aggressiveness with which preparation ramps up early in a trial.
 1138 Interestingly, the strategic adjustment of perceptual and motor readiness was accompanied by
 1139 adjustments to urgency, and hence anticipatory errors. The possible yoking of residual RT (with a
 1140 likely motor component) and urgency bears further investigation. It sits comfortably with recent
 1141 intracranial EEG data suggesting widely dispersed but seemingly motoric brain responses correlated
 1142 with an increase in anticipatory errors at a longer foreperiod, but not other aspects of temporal
 1143 preparation (Ramayya et al., 2025).

1144 As in Experiment 1, in Experiment 3 changes to particular model parameters, some purely
 1145 sensory (determining the level of sensory drive) and some either motoric or sensory (determining
 1146 the residual RT), allowed SIP-theN-SPuRT to capture changes across experimental conditions. Results
 1147 like these complement previous literature attempting to determine the locus of foreperiod effects
 1148 (see Rolke and Ulrich, 2010, for a concise review). Techniques have included the incorporation of
 1149 additional factors with known or presumed loci (to permit application of additive-factors logic), the
 1150 use of unspeeded perceptual discrimination tasks with no motor element, or the measurement of

EEG event-related potentials (with inferred links to particular processes). These studies have, like ours, suggested foreperiod effects on both perception and action (and indeed the translation of one into the other), although these studies have often been informed by fixed rather than by variable-foreperiod paradigms (Bausenhart et al., 2006, 2007; Burle et al., 2010; Rolke & Hofmann, 2007).

However, our “proof of concept” demonstrations here were conservative, in that we only permitted change in a subset of parameters consistent with the experimental situation and model mechanics. The capacity of SIP-theN-SPuRT to allocate cognitive meaning by showing that an effect is captured best by a given parameter when many parameters might vary is currently less explored, and we would be quite cautious about this approach in general.¹⁵ For example, if a manipulation affects RT dispersion it is likely to be captured, at least in part, by stimulus-dependent drive or urgency parameters (if they are allowed to vary). However, this would arise from the particular simplifying assumptions we have made (i.e. no variability in residual RTs) to produce a model that is tractable, with parameters that are identifiable, rather than representing ground truth. Similar issues have been highlighted for interpretations arising from widely used models of choice RTs, for example the inferred impact of endogenous factors on non-decision times (Bompas et al., 2024). Such approaches applied to foreperiod effects have yielded a mixed picture (Grosjean et al., 2001; Jepma et al., 2012; van den Brink et al., 2021; Xiong et al., 2025).

Trial-by-trial learning

One of our key goals here was to formulate a model in which preparation reflects beliefs derived from the distribution of experienced foreperiods, but which still accounts for the strong influence of prior foreperiods on the current response. Previously, such effects have been explained in alternative accounts where the foreperiod PDF is not represented cognitively (Los & van den

¹⁵ Experiment 3 did give broadly sensible results in this regard; see Appendix B

Heuvel, 2001; Salet et al., 2022), or it has been treated as a distinct process riding atop hazard-based preparation (Vallesi & Shallice, 2007). SIP-theN-SPuRT allows the participant's estimated foreperiod PDF to vary from trial to trial as a recency-weighted average of the likelihood functions derived from individual stimuli. In common with the non-cognitive, single-process perspective (e.g. Salet et al. 2022), this approach requires some further squaring with neuropsychological and behavioural dissociations between foreperiod and prior-foreperiod effects (Vallesi & Shallice, 2007; Vallesi et al., 2007a, 2007b, 2013). However, our approach did generate robust prior-foreperiod effects with correct qualitative trends without needing to propose a further (or alternative) mechanism aside from the representation of the foreperiod PDF. Our modelling also assumed that a participant's estimated foreperiod distribution was retained across breaks between blocks or sessions, as there is evidence that foreperiod-learning effects can last a very considerable time (Mattiesing et al., 2017). In this way model predictions matched the behavioural trend of Experiment 1 toward an effect of the previous block. However, there is also evidence for a reduction of n-1 foreperiod effects when very long inter-trial intervals are used, which our current modelling would not capture (Vallesi et al., 2013).

We have also considered, to our knowledge for the first time, another form of learning – relating to the updating of beliefs about the rate at which catch trials occur. Recent catch trials affect behaviour, as shown in Experiment 3. If preparation is based on a hazard function, a basic account of such changes comes for free via the insertion of a plausible Bayesian update to the belief about the rate of catch trials. However, our results suggest that behaviour changes markedly in the immediate aftermath of a catch trial and then rebounds quickly. This trend was captured best by incorporating an additional parameter modelling a further belief that the underlying rate of catch trials could change (i.e. that the world was non-stationary in this respect, see Zhang et al., 2014). Such a framework makes hazard-based preparation more plausible with a range of catch-trial rates, a point we return to shortly in the section on comparisons with other modelling work.

For recency effects of both kinds (foreperiod and catch trials) our best-fitting model tended not to predict effects of sufficient magnitude when examined separately across foreperiods. In the case of prior-foreperiod effects, it appears that increasing the update-rate of learning offers the capacity to generate larger effects, but that this may have deleterious effects on other aspects of the model fit. It therefore remains to be seen whether a more accurate specification of the preparation function would allow the model to do better at predicting foreperiod effects. For the effect of catch trials, it is less clear that decreasing the belief that catch-trial rate will remain stable could capture the finer details regarding how RTs vary as a function of both foreperiod and catch-trial recency. This issue would benefit from further empirical work and modelling.

Comparison to other models

Currently, there are two classes of model providing formal predictions for the effect of foreperiods on behaviour. The first class, to which our work belongs, considers preparation to build based on information derived from a cognitive representation of the PDF of foreperiods. The second class does not require any such representation. In the past, models from both classes have mainly considered fits of a given function to the whole-group central tendency of RT behaviour. This may underestimate how well a given function might perform if fitted to individual data, as implemented here, where we have presented the average of individual model predictions against equivalently averaged data.

Previous modelling of the first class, building from the PDF of foreperiods to predict RTs, began by considering fits of the subjective hazard function (see Equation 1) to simple RT datasets. Data from small samples of monkeys completing many trials have fit subjective hazard functions either very well (Janssen & Shadlen, 2005), somewhat well (Tsunoda & Kakei, 2008) or very well with a switch to a subjective *log* hazard (Sharma et al., 2015). This basic idea has been further developed in a number of ways. Grabenhorst et al. (2019) performed experiments in human participants,

1224 focussing on exponential and flipped exponential foreperiod distributions (see also Grabenhorst et
 1225 al., 2025). They considered the subjective hazard rate, but also the subjective PDF, as the basis of
 1226 preparation. Moreover, they considered several variants: A subjective PDF where noise is
 1227 proportional to (inverse) frequency of exposure rather than following Weber's law, as well as inverse
 1228 likelihood in place of the raw PDF or hazard function. They concluded in favour of the inverse PDF
 1229 with probabilistic (rather than Weber-like) noise as the basis of preparation.

1230 This finding is at odds with ours here, based on very similar kinds of data from the same
 1231 species. Our results clearly favoured hazard-based preparation. However, we think there are several
 1232 reasons to question Grabenhorst et al.'s (2019) conclusions. First, the rejection of Weber's law flies
 1233 in the face of expectations derived from over 100 years of psychophysics, and extraordinary claims
 1234 should generally be held to high evidential standards. Their position is also harder to square with
 1235 widely reported fixed-foreperiod effects, suggesting less focussed preparation at longer foreperiods
 1236 despite experiencing a similar number of trials. Second, recent results with an experiment, models
 1237 and fitting methods generally very similar to those of Grabenhorst et al. (2019) (but using choice RTs
 1238 and a near-threshold discrimination) have reached a different conclusion, favouring log hazard as
 1239 the basis of preparation (Vishne et al., 2025).¹⁶ Third, Grabenhorst et al.'s (2019) methods differed
 1240 from those employed here in some ways that are clearly up for debate (e.g. our introduction of
 1241 saturation) but some of Grabenhorst et al.'s (2019) approaches seem less complete (e.g. fitting a
 1242 prediction that should operate at the individual level to group-average data; fitting effects on central
 1243 tendency but not dispersion; assuming perfect knowledge of the catch-trial rate).

1244 Fourth, and finally, their result seems dependent on the exact catch-trial rate they initially
 1245 investigated (1/11). In subsequent work, the same group found mismatches between the reciprocal
 1246 PDF and RTs at other catch-trial rates (although these were less dramatic than those seen here and

¹⁶ These authors additionally introduced a second level of "decoding" noise in the blurring of preparation functions.

in the wider literature) leading these authors to introduce an additional arbitrary modifying function arising from a hypothesised second process (Grabenhorst et al., 2021). Notably, as we illustrated in Figure 3, hazard naturally varies with the rate of catch trials, in a way that is qualitatively consistent with the effects Grabenhorst et al. (2021) report and that we have observed in Experiment 3 – i.e. particularly at longer foreperiods. This kind of interaction, between catch-trial rate (or recency) and the foreperiod effect, has been observed in several different kinds of task (Correa et al., 2004; Los, 2004).¹⁷

For all these reasons, we are sceptical about rejecting both hazard (as the basis of temporal prediction) and Weber-based noise (to model subjective experiences) in favour of probabilistically blurred PDF-based predictions. However, the idea of a transform of the raw PDF or hazard rate certainly seems worthy of further investigation. We have noted several failures of our models to capture exactly how foreperiod influences RTs, particularly a consistent tendency of our hazard-based models to underestimate RTs at short foreperiods. This might in part reflect the need for modality specific elements specifically affecting short-foreperiod preparations (Grabenhorst et al., 2019). However, the log transform to hazard also seems worth looking at. This would complicate interpretation of intercept parameters in our modelling (and the setting of appropriate parameter ranges and hyperpriors) but the interpretation of many SIP-theN-SPuRT parameters is already non-intuitive, so this is a minor concern. Somewhat similar ideas to additionally smooth the hazard function have been proposed by Wang et al. (2020), who smoothed by discounting the future (to predict choice RTs) and by Moon et al. (2019) and by Vangkilde et al. (2013), both of whom additionally smoothed the hazard function via an exponential function (to predict stimulus occurrence either now *or soon* for tasks focussing on perceptual discrimination sensitivity and

¹⁷ We note, however, that the impact of catch trials in simple RT tasks may be more profound than we have suggested here (via a mere reshaping of hazard) and might even transform the task into something closer to choice RT as the foreperiod elapses. Further theoretical and empirical work on this issue seems warranted.

1269 encoding rate respectively). However, log hazard has some nice theoretical properties compared to
 1270 these alternatives, being related to Shannon information.

1271 Before concluding this section, and lest we appear unduly critical of the work of Grabenhorst
 1272 et al. (2019), it is worth reiterating an important logical concern they raise regarding the plausibility
 1273 that the brain explicitly derives a hazard function. This is a non-trivial computation, requiring an
 1274 additional step over and above representing the foreperiod PDF. Our model makes this calculation,
 1275 and we are not aware of a mechanism by which a simple neural network might do so. Nonetheless,
 1276 human behaviour seems more consistent with hazard-function computations than with the other
 1277 types of computation we have simulated. Even if the brain can merely approximate a hazard
 1278 function, a model based on actual hazard functions may maximise both explanatory power and
 1279 simplicity.

1280 We now turn to the second class of models (which do not require cognitive representations
 1281 of the foreperiod PDF), currently represented by the formalised multiple trace theory of temporal
 1282 preparation (fMTP; Salet et al., 2022). Based on the principal of associative learning between a
 1283 population representation of time and inhibitory or excitatory motor units, it predicts a preparation
 1284 function sharing many of the features of a (saturating) hazard function. Now that we have described
 1285 a version of PDF-based preparation that predicts foreperiod history effects, the two classes of model
 1286 are on a roughly equal explanatory footing. Hypotheses regarding the basis of temporal preparation
 1287 could be quantitatively compared in future work, for example using simple RT data and the SIP-theN-
 1288 SPuRT framework, thus permitting the influence of each putative preparation function (e.g. fMTP vs.
 1289 saturating hazard) on one or more parameters describing urgency, sensory drive and residual RTs.¹⁸
 1290 We note that the two classes of model posit quite different mechanistic effects from receiving no
 1291 stimulus on a given trial (either an update to a beta distribution representing beliefs about the

¹⁸ fMTP focuses on the motor level so the language of sensory drive may not be appropriate, but fMTP would need to be given some mechanism for predicting the influence of foreperiod on RT-distribution shape, in a fair model comparison.

proportion of catch trials, or the reinforcement of synapses between temporally-tuned clock units and inhibitory motor units). The experimental manipulation of catch trials might therefore be important to discriminate between different accounts. Another useful manipulation, highlighted by Salet et al. (2022) and tested by Los (2013), is the introduction of no-go trials, which seem to yield a different $n-1$ effect (that is more uniform across foreperiods) compared to go-trials. This is not predicted by the most straightforward extensions of the types of account we have tested here (in which no-go trials either would or would not update the estimated PDF, thus yielding either a standard $n-1$ effect or none at all), but the effect could be handled by fMTP.

Relation to the wider literature on temporal preparation

Here, we have focussed closely on one behavioural manifestation of the variable-foreperiod effect, specifically effects on simple RTs. However, foreperiod effects can be observed for many other dependent variables, including choice RT, perceptual discrimination, spiking activity in single neurones, the fMRI BOLD signal, and pre-stimulus oculomotor inhibition. Moreover, temporal preparation is studied in ways other than (or complementary to) varying the foreperiod, such as explicit temporal cuing.

Broadly speaking, our conclusions here are quite consistent with this wider literature. For example, we have argued for a saturating transform of subjective hazard as the basis of temporal preparation. The link function from hazard to a given manifestation of preparation is unlikely to be exactly the same for different kinds of dependent variable, but the underlying principle should be in common. Indeed, in monkeys the spiking rates of neurones in intraparietal sulcus track subjective hazard (Janssen & Shadlen, 2005), while neurones in monkey V1 are modulated at a preferred spatial location in line with log subjective hazard (Sharma et al., 2015). In humans, levels of fMRI BOLD signal evoked by a go stimulus have been found to vary in a way well predicted by cumulative hazard, in both supplementary motor area and the superior temporal gyrus (Cui et al., 2009).

1317 Furthermore, MEPs evoked by single-pulse TMS to primary motor cortex show a fairly close
 1318 correspondence with a hazard function. (It is closer than classic EEG-based measures like the CNV,
 1319 including dips between potential foreperiods when ambiguity about time is absent; Trillenberg et al.,
 1320 2000; van Elswijk et al., 2007.)

1321 However, in support of their alternative position, Grabenhorst et al. (2025) have recently
 1322 reported correlations between several features of MEG data and RTs, in a context where RTs are
 1323 correlated with the subjective PDF, and thus they have argued that these effects are neural
 1324 correlates of PDF-based temporal preparation. We look forward to modelling efforts exploring
 1325 further correlates of temporal preparation, for example the rate of microsaccade generation. This
 1326 decreases with preparation even when targets are non-visual (see Abeles et al., 2020; Amit et al.,
 1327 2020; Badde et al., 2020; Tal-Perry & Yuval-Greenberg, 2020), but, to our knowledge, has not yet
 1328 been quantitatively related to either the subjective PDF or hazard rate.

1329 In the introduction, we excused ourselves from describing the ever-growing literature on
 1330 temporal preparation following from explicit cues or rhythmic expectancies, which is largely beyond
 1331 our scope (and well described in recent reviews, e.g. Denison, 2024; Nobre & Van Ede, 2018). Hence,
 1332 on this front, we limit ourselves to the comment that temporal preparation of the type revealed by
 1333 variable-foreperiod paradigms is, presumably, the background process against which further
 1334 preparatory manipulations must be assessed. Resolving the exact form of temporal preparation
 1335 effects might be important for the interpretation of a range of related findings.

1336

1337 **Wider uses for SIP-theN-SpuRT**

1338 Our focus here has been on understanding processes underlying temporal preparation, and
 1339 we have developed the SIP-theN-SpuRT modelling framework to assist in this endeavour. However,
 1340 SIP-theN-SpuRT will have applications in any context where simple RTs are used as a dependent

variable, and it offers an alternative to data trimming prior to calculating central tendency as a means of data interpretation. SIP-theN-SpuRT offers several benefits, most particularly the capacity to assess behaviour at any level of the speed-accuracy tradeoff (or across multiple such levels), thereby providing a fuller account of what is changing across experimental conditions. As we noted in the introduction, consideration of the speed-accuracy tradeoff (Garrett, 1922; Hick, 1952) has been central to the increasing prominence of modelling in the interpretation of choice RT data (e.g. Smith & Ratcliff, 2004; Spieser et al., 2018). Errors are also an important feature of simple RT data, albeit in the form of anticipations rather than incorrect responses. Our modelling offers a principled approach to explaining errors in simple RT data. However, it is worth noting that the mechanics of our model do not generate a tradeoff in the way that popular accumulation-to-bound models of choice RTs do, because urgency parameters are independent from those affecting sensory gain and residual RTs. Exploring the way in which tradeoffs occur would represent an interesting avenue for model development.

In terms of practical advice for the application of SIP-theN-SpuRT, the exact selection of model parameters should depend on the finer details of any given experiment (see Table 1). For example, in an experiment without foreperiod variation, a maximum of six parameters per condition could be used to characterise the full RT distribution, including anticipatory errors, but we would suggest fixing α_f , the urgency intercept, at zero, or estimating just one value for the whole group. This approach could also be applied for the Weber ratio. The resulting four-parameter fit would be comparable to existing modelling schemes such as the LATER model (Carpenter & Noorani, 2023) or to the use of one or more distributions, such as the exGaussian, which would require at least 3-4 parameters to deal with the main RT distribution plus anticipatory errors (cf. Ramayya et al., 2025). With a small number of foreperiods, it would likely be best to incorporate some, but not all, of the additional parameters involved in estimating the preparation function and its translation onto model processes, to avoid issues with parameter degeneracy. Alternatively, one might just fit the core

1366 model at each foreperiod, via duplicates on parameters – such as intercepts for sensory gain and
 1367 residual RT.

1368 **Limitations and constraints on generality**

1369 Many of our results depend on model comparisons, and this invites controversy. First, there
 1370 is the method of evaluating models. Ideally, we would compare models based on their leave-one-out
 1371 log-likelihood, to control for differences in model complexity. However, computational demands and
 1372 practical constraints mean that this was not viable here. By reporting both the AIC and BIC
 1373 approximations from MLE fits, supplemented by Bayesian PSIS-LOOIC from the full posterior where
 1374 possible, we believe we have still offered solid evidence for our conclusions.

1375 Another issue is the almost infinite space of models and model parameterisations that are
 1376 available. For example, we have tested models where a given underlying computation (hazard-
 1377 based, PDF-based or elapsed-time-based) is applied across all three model components (urgency,
 1378 sensory drive and residual RT) but it is possible that better fits could be achieved by applying
 1379 different computations to each. In any modelling effort a limited number of models can be tested,
 1380 and not all readers will agree with our particular selections. However, our approach represents a
 1381 step forward (in terms of modelling the full RT distribution and working from an individual up to a
 1382 group level when characterising model predictions) and it could be applied in future to test a
 1383 different range of models and perhaps challenge the conclusions reached here.

1384 Although not strictly a limitation, it seems worth flagging an unusual aspect of our
 1385 manipulation of the foreperiod distribution in Experiments 1 and 3. To our knowledge, most past
 1386 papers have drawn foreperiods from the same set of possible values in each condition (but with
 1387 different frequencies of occurrence at each possible foreperiod). Some have employed a discrete
 1388 (and typically small) set of possible foreperiods, and others have drawn from a continuous
 1389 distribution. By contrast, we used a discrete (but quite large) set of possible foreperiods, with (often)
 1390 non-uniform spacing but a uniform number of trials at each one. In some cases, these sorts of

1391 distinction will have implications under current theories (e.g. the estimated foreperiod PDF in SIP-
 1392 theN-SpuRT will not be exactly the same when foreperiods are drawn from a continuous distribution
 1393 vs. a discrete approximation with large steps – the latter would generate a scalloped shape to a
 1394 degree dependent on the participant's Weber ratio). It remains a question for future research
 1395 whether different sampling methods for stimulus selection have the effects predicted by particular
 1396 models, or even generate/minimise particular established effects from the foreperiod literature.

1397 Finally, while we have increased generalisability by utilising data from different
 1398 experiments/labs, each with different balances of gender/sex, the samples were all essentially
 1399 WEIRD (western, educated, industrialised, rich and democratic), and we have not looked to identify,
 1400 for example, sex differences. Hence a full generalisation of results across diverse groups cannot be
 1401 assumed.

1402

1403 **Conclusions**

1404 An organism's state of preparation evolves over time in a way that is influenced by the
 1405 history of foreperiods that they have experienced. Here, we have used the most complete model-
 1406 fitting methodology to date to compare predictions from several putative preparation functions to
 1407 describe a full distribution of simple RTs. We have found variable-foreperiod effects on both the
 1408 dispersion and central tendency of RTs and have tested models which incorporated updates to the
 1409 estimated PDF of foreperiods and to beliefs about the rate of catch trials. Of the model variants we
 1410 tested, a saturating transform of the subjective hazard function best predicted simple RTs, by
 1411 modulating urgency, sensory gain, and residual RT. However, fits were not perfect, suggesting
 1412 further work needs to be done in terms of fine-tuning or supplementing the transformation of
 1413 hazard into preparatory intent. In future, it will also be important to test this general account against
 1414 one in which associative processes mimic a cognitive representation of the timing of imperative
 1415 events.

1416 **References**

- 1417 Agrawal, M., Mattar, M. G., Cohen, J. D., & Daw, N. D. (2022). The temporal dynamics of opportunity
 1418 costs: A normative account of cognitive fatigue and boredom. *Psychological Review*, 129(3),
 1419 564-585.
- 1420 Amit, R., Abeles, D., Carrasco, M., & Yuval-Greenberg, S. (2019). Oculomotor inhibition reflects
 1421 temporal expectations. *Neuroimage*, 184, 279-292.
- 1422 Bausenhardt, K. M., Rolke, B., Hackley, S. A., & Ulrich, R. (2006). The locus of temporal preparation
 1423 effects: Evidence from the psychological refractory period paradigm. *Psychonomic Bulletin &*
 1424 *Review*, 13, 536–542.
- 1425 Bausenhardt, K. M., Rolke, B., & Ulrich, R. (2007). Short article: Knowing when to hear aids what to
 1426 hear. *Quarterly Journal of Experimental Psychology*, 60(12), 1610–1615.
- 1427 Bompas, A., Sumner, P., & Hedge, C. (2024). Non-decision time: The Higgs Boson of decision.
 1428 *Psychological Review*, 132(2), 330–363.
- 1429 Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433–436.
- 1430 Burle, B., Tandonnet, C., & Hasbroucq, T. (2010). Excitatory and inhibitory motor mechanisms of
 1431 temporal preparation. In Anna C. Nobre, & Jennifer T. Coull (Eds.), *Attention and time* (pp. 244–
 1432 255). Oxford University Press.
- 1433 Carpenter, R. H., & Noorani, I. (2023). *LATER: The Neurophysiology of Decision-making*. Cambridge
 1434 University Press.
- 1435 Correa, A., Lupiáñez, J., Milliken, B., & Tudela, P. (2004). Endogenous temporal orienting of attention
 1436 in detection and discrimination tasks. *Perception & Psychophysics*, 66, 264-278.

- 1437 Cui, X., Stetson, C., Montague, P. R., & Eagleman, D. M. (2009). Ready... go: amplitude of the fMRI
1438 signal encodes expectation of cue arrival time. *PLoS Biology*, 7(8), e1000167.
- 1439 Denison, R. N. (2024). Visual temporal attention from perception to computation. *Nature Reviews*
1440 *Psychology*, 3(4), 261–274. <https://10.1038/s44159-024-00294-0>
- 1441 Duyar, A., Ren, S., & Carrasco, M. (2024). When temporal attention interacts with expectation.
1442 *Scientific Reports*, 14(1), 4624. <https://10.1038/s41598-024-55399-6>
- 1443 Elithorn, A., & Lawrence, C. (1955). Central inhibition-some refractory observations. *Quarterly*
1444 *Journal of Experimental Psychology*, 7(3), 116–127.
- 1445 Gabry J, & Veen D (2022). shinystan: Interactive Visual and Numerical Diagnostics and Posterior
1446 Analysis for Bayesian Models. R package version 2.6.0, <https://github.com/stan-dev/shinystan>.
- 1447 Garrett, H. E. (1922). A study of the relation of accuracy to speed. *Archives of Psychology*, 56, 1-106.
- 1448 Grabenhorst, M., Maloney, L. T., Poeppel, D., & Michalareas, G. (2021). Two sources of uncertainty
1449 independently modulate temporal expectancy. *Proceedings of the National Academy of*
1450 *Sciences*, 118(16), e2019342118. <https://10.1073/pnas.2019342118>
- 1451 Grabenhorst, M., Michalareas, G., Maloney, L. T., & Poeppel, D. (2019). The anticipation of events in
1452 time. *Nature Communications*, 10(1), 1–13.
- 1453 Grabenhorst, M., Poeppel, D., & Michalareas, G. (2025). Neural signatures of temporal anticipation
1454 in human cortex represent event probability density. *Nature Communications*, 16(1), 2602.
- 1455 Grosjean, M., Rosenbaum, D. A., & Elsinger, C. (2001). Timing and reaction time. *Journal of*
1456 *Experimental Psychology: General*, 130(2), 256-272.

- 1457 Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*,
1458 4, 11-26.
- 1459 Houshmand Chatroudi, A., Mioni, G., & Yotsumoto, Y. (2024). On the nonlinearity of the foreperiod
1460 effect. *Scientific Reports*, 14(1), 2780.
- 1461 Janssen, P., & Shadlen, M. N. (2005). A representation of the hazard rate of elapsed time in macaque
1462 area LIP. *Nature Neuroscience*, 8(2), 234–241.
- 1463 JASP Team. (2025). *JASP (Version 0.95.4) [Computer software]*. <https://jasp-stats.org/>.
- 1464 Jazayeri, M., & Shadlen, M. N. (2010). Temporal context calibrates interval timing. *Nature*
1465 *Neuroscience*, 13(8), 1020–1026.
- 1466 Jepma, M., Wagenmakers, E., & Nieuwenhuis, S. (2012). Temporal expectation and information
1467 processing: A model-based analysis. *Cognition*, 122(3), 426–441.
- 1468 Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What’s new in
1469 Psychtoolbox-3. *Perception*, 36(14), 1-16.
- 1470 Kohl, C., Spieser, L., Forster, B., Bestmann, S., & Yarrow, K. (2019). The Neurodynamic Decision
1471 Variable in Human Multi-alternative Perceptual Choice. *Journal of Cognitive Neuroscience*,
1472 31(2), 262–277.
- 1473 Kohl, C., Spieser, L., Forster, B., Bestmann, S., & Yarrow, K. (2020). Centroparietal activity mirrors the
1474 decision variable when tracking biased and time-varying sensory evidence. *Cognitive*
1475 *Psychology*, 122, 101321.
- 1476 Lambert, B. (2018). *A student’s guide to Bayesian statistics*. Sage.

- 1477 Los, S. A., & van den Heuvel, C. E. (2001). Intentional and unintentional contributions to nonspecific
 1478 preparation during reactiontime foreperiods. *Journal of Experimental Psychology: Human*
 1479 *Perception and Performance*, 27(2), 370–386. <https://10.1037//0096-1523.27.2.370>
- 1480 Los, S. A. (2004). Inhibition of return and nonspecific preparation: separable inhibitory control
 1481 mechanisms in space and time. *Perception & Psychophysics*, 66, 119-130.
- 1482 Los, S. A. (2013). The role of response inhibition in temporal preparation: Evidence from a go/no-go
 1483 task. *Cognition*, 129(2), 328–344.
- 1484 Los, S. A., Kruijne, W., & Meeter, M. (2017). Hazard Versus History: Temporal Preparation Is Driven
 1485 by Past Experience. *Journal of Experimental Psychology. Human Perception and Performance*,
 1486 43(1), 78–88. <https://10.1037/xhp0000279>
- 1487 Luce, R. D. (1986). *Response Times: Their Role in Inferring Elementary Mental Organization*. Oxford
 1488 University Press.
- 1489 Mattiesing, R. M., Kruijne, W., Meeter, M., & Los, S. A. (2017). Timing a week later: The role of long-
 1490 term memory in temporal preparation. *Psychonomic Bulletin & Review*, 24, 1900–1905.
- 1491 Moon, J., Choe, S., Lee, S., & Kwon, O. (2019). Temporal dynamics of visual attention allocation.
 1492 *Scientific Reports*, 9(1), 1–11.
- 1493 Niemi, P., & Lehtonen, E. (1982). Foreperiod and visual stimulus intensity: A reappraisal. *Acta*
 1494 *Psychologica*, 50(1), 73–82.
- 1495 Niemi, P., & Näätänen, R. (1981). Foreperiod and simple reaction time. *Psychological Bulletin*, 89(1),
 1496 133-162.

- 1497 Nobre, A. C., & Van Ede, F. (2018). Anticipated moments: temporal structure in attention. *Nature*
1498 *Reviews Neuroscience*, 19(1), 34–48.
- 1499 Nobre, A. C., Correa, A., & Coull, J. T. (2007). The hazards of time. *Current Opinion in Neurobiology*,
1500 17(4), 465–470.
- 1501 R Core Team, 2021. R: A language and environment for statistical computing. R Foundation for
1502 Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- 1503 Ramayya, A. G., Buch, V., Richardson, A., Lucas, T., & Gold, J. I. (2025). Human response times are
1504 governed by dual anticipatory processes with distinct neural signatures. *Communications*
1505 *Biology*, 8(1), 124.
- 1506 Rigoux, L., Stephan, K. E., Friston, K. J., & Daunizeau, J. (2014). Bayesian model selection for group
1507 studies—revisited. *Neuroimage*, 84, 971–985.
- 1508 Rolke, B., & Hofmann, P. (2007). Temporal uncertainty degrades perceptual processing. *Psychonomic*
1509 *Bulletin & Review*, 14(3), 522–526.
- 1510 Rolke, B., & Ulrich, R. (2010). On the locus of temporal preparation: Enhancement of premotor
1511 processes. In A. C. Nobre, & J. T. Coull (Eds.), *Attention and time* (pp. 227–241). Oxford
1512 University Press.
- 1513 Salet, J. M., Kruijne, W., van Rijn, H., Los, S. A., & Meeter, M. (2022). FMTP: A unifying computational
1514 framework of temporal preparation across time scales. *Psychological Review*, 129(5), 911–948.
- 1515 Sharma, J., Sugihara, H., Katz, Y., Schummers, J., Tenenbaum, J., & Sur, M. (2015). Spatial attention
1516 and temporal expectation under timed uncertainty predictably modulate neuronal responses in
1517 monkey V1. *Cerebral Cortex*, 25(9), 2894–2906.

- 1518 Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017).
 1519 Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience*,
 1520 40, 99–124.
- 1521 Smith, P. L. & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in*
 1522 *Neuroscience*, 27, 161-168.
- 1523 Smith, P. L. (1995). Psychophysically Principled Models of Visual Simple Reaction Time. *Psychological*
 1524 *Review*, 102(3), 567–593. <https://10.1037/0033-295X.102.3.567>
- 1525 Spieser, L., Kohl, C., Forster, B., Bestmann, S., & Yarrow, K. (2018). Neurodynamic evidence supports
 1526 a forced-excursion model of decision-making under speed/accuracy instructions. *Eneuro*, 5(3),
 1527 ENEURO.0159-18.2018
- 1528 Stan Development Team (2022). RStan: the R interface to Stan. R package version 2.21.5. [https://mc-](https://mc-stan.org/)
 1529 [stan.org/](https://mc-stan.org/).
- 1530 Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., & Friston, K. J. (2009). Bayesian model
 1531 selection for group studies. *Neuroimage*, 46(4), 1004-1017.
- 1532 Stevens, C. F., & Zador, S. (1996). When is an integrate-and-fire neuron like a Poisson neuron? In
 1533 *Advances in neural information processing systems* (Vol. 8, pp. 103-109).
- 1534 Tal-Perry, N., & Yuval-Greenberg, S. (2020). Pre-target oculomotor inhibition reflects temporal
 1535 orienting rather than certainty. *Scientific Reports*, 10(1), 21478.
- 1536 Trillenberg, P., Verleger, R., Wascher, E., Wauschkuhn, B., & Wessel, K. (2000). CNV and temporal
 1537 uncertainty with ‘ageing’ and ‘non-ageing’ intervals. *Clinical Neurophysiology*, 111(7), 1216–
 1538 1226.

- 1539 Tsunoda, Y., & Kakei, S. (2008). Reaction time changes with the hazard rate for a behaviorally
 1540 relevant event when monkeys perform a delayed wrist movement task. *Neuroscience Letters*,
 1541 433(2), 152–157.
- 1542 Vallesi, A., Arbula, S., & Bernardis, P. (2014). Functional dissociations in temporal preparation:
 1543 Evidence from dual-task performance. *Cognition*, 130, 141-151.
- 1544 Vallesi, A., Lozano, V. N., & Correa, Á. (2013). Dissociating temporal preparation processes as a
 1545 function of the inter-trial interval duration. *Cognition*, 127(1), 22–30.
- 1546 Vallesi, A., Mussoni, A., Mondani, M., Budai, R., Skrap, M., & Shallice, T. (2007a). The neural basis of
 1547 temporal preparation: Insights from brain tumor patients. *Neuropsychologia*, 45(12), 2755–
 1548 2763.
- 1549 Vallesi, A., & Shallice, T. (2007). Developmental dissociations of preparation over time:
 1550 deconstructing the variable foreperiod phenomena. *Journal of Experimental Psychology:*
 1551 *Human Perception and Performance*, 33(6), 1377.
- 1552 Vallesi, A., Shallice, T., & Walsh, V. (2007b). Role of the prefrontal cortex in the foreperiod effect:
 1553 TMS evidence for dual mechanisms in temporal preparation. *Cerebral Cortex*, 17(2), 466–474.
- 1554 van den Brink, R. L., Murphy, P. R., Desender, K., de Ru, N., & Nieuwenhuis, S. (2021). Temporal
 1555 expectation hastens decision onset but does not affect evidence quality. *Journal of*
 1556 *Neuroscience*, 41(1), 130–143.
- 1557 van Elswijk, G., Kleine, B. U., Overeem, S., & Stegeman, D. F. (2007). Expectancy induces dynamic
 1558 modulation of corticospinal excitability. *Journal of Cognitive Neuroscience*, 19(1), 121–131.

- 1559 Vangkilde, S., Petersen, A., & Bundesen, C. (2013). Temporal expectancy in the context of a theory of
 1560 visual attention. *Philosophical Transactions of the Royal Society B: Biological Sciences*,
 1561 368(1628), 20130054.
- 1562 Van Lambalgen, R. M., & Los, S. A. (2008). The role of attention in nonspecific preparation. In B. C.
 1563 Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the*
 1564 *Cognitive Science Society* (pp. 1525-1530). Cognitive Science Society.
- 1565 Vehtari, A., Gabry, J., Magnusson, M., Yao, Y., Bürkner, P., Paananen, T., Gelman, A. (2022). "loo:
 1566 Efficient leave-one-out cross-validation and WAIC for Bayesian models." R package version
 1567 2.5.1, <https://mc-stan.org/loo/>.
- 1568 Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out
 1569 cross-validation and WAIC. *Statistics and Computing*, 27(5), 1413–1432.
- 1570 Visalli, A., Capizzi, M., Ambrosini, E., Kopp, B., & Vallesi, A. (2021). Electroencephalographic
 1571 correlates of temporal Bayesian belief updating and surprise. *NeuroImage*, 231, 117867.
 1572 <https://10.1016/j.neuroimage.2021.117867>
- 1573 Visalli, A., Capizzi, M., Ambrosini, E., Kopp, B., & Vallesi, A. (2023). P3-like signatures of temporal
 1574 predictions: a computational EEG study. *Experimental Brain Research*, 241(7), 1919–1930.
- 1575 Vishne, G., Deouell, L. Y., & Landau, A. N. (2025). Computational Mechanisms of Temporal
 1576 Anticipation in Perception and Action. *bioRxiv*, 2025.03. 28.646008.
- 1577 Wang, M., Huang, Y., Luo, H., & Zhang, H. (2020). Sustained Visual Priming Effects Can Emerge from
 1578 Attentional Oscillation and Temporal Expectation. *The Journal of Neuroscience*, 40(18), 3657–
 1579 3674. <https://10.1523/JNEUROSCI.2539-19.2020>

- 1580 Wearden, J. H., & Lejeune, H. (2008). Scalar properties in human timing: Conformity and violations.
 1581 *Quarterly Journal of Experimental Psychology*, 61(4), 569–587.
- 1582 Woodrow, H. (1914). The measurement of attention. *The Psychological Monographs*, 17(5), i-158.
- 1583 Yarrow, K. (2022) Foreperiod x FlippedVExp x Opportunity cost, simple RT (October 2022).
 1584 AsPredicted. <https://aspredicted.org/kt3qu.pdf>
- 1585 Yarrow, Kielan (2025). Data and code accompanying a paper entitled "How beliefs about the
 1586 warning-target foreperiod inform temporal preparation to determine simple reaction times".
 1587 City St George's, University of London Dataset. <https://doi.org/10.25383/city.30715439.v1>
- 1588 Yarrow, K., Saurels, B. W., Arnold, D., Tal-Perry, N., & Solomon, J. (2026, April 19). How beliefs about
 1589 the warning-target foreperiod inform temporal preparation to determine simple reaction times.
 1590 Retrieved from osf.io/preprints/psyarxiv/69xvk_v1
- 1591 Zhang, S., Huang, H. C., & Yu, A. J. (2014). Sequential effects: A Bayesian analysis of prior bias on
 1592 reaction time and behavioral choice. Paper presented at the *Proceedings of the Annual Meeting*
 1593 *of the Cognitive Science Society*, 36.
- 1594
- 1595

Appendix A: Detailed methods for Experiments 1 and 3

Transparency and openness

Either here or in the main text, we report how we determined our sample size, all data exclusions, all manipulations, and all measures. All raw (trial-by-trial) data and analysis scripts are available at <https://doi.org/10.25383/city.30715439>. Experimental scripts are available on request from the corresponding author. We cite all program code developed by others that was used in our analyses. We indicate whether experiments were preregistered.

Experiment 1

Participants

Participant recruitment followed a pre-registered schedule (<https://aspredicted.org/kt3qu.pdf>).¹⁹ Initially, 48 participants, all recruited via the Prolific Academic online panel, completed the experiment (some further participants were rejected during practice and hence did not complete the experiment; see below). Because data from some completed participants failed to meet pre-registered criteria (also outlined below) recruitment was re-opened until a further 3 participants had successfully met these criteria, such that equal numbers of participants were retained for analysis in each of four groups (used to counterbalance the order of completion for experimental blocks). This yielded a final sample for analysis of 36 participants (21 male, 14 female, one non-binary, mean age = 30.8, sd = 9.7). Participants who completed the experiment (which took a median 31 minutes) were paid a guaranteed £5 plus a performance-dependent bonus (mean £1.93). Experimental procedures were approved by the City, University of

¹⁹ Developments and practical realisations arising during the modelling process mean that the analysis presented here is not exactly as pre-registered, although many aspects (and the overall spirit) have been retained.

1617 London Department of Psychology Research Ethics Committee (ETH2021-2016 Modelling foreperiod
1618 effects).

1619 Apparatus, Stimuli, Design and Procedure

1620 The experiment was programmed using PsychoPy3 (v2021.1.4), which created Javascript
1621 code hosted on the Pavlovia.org platform (except for consent and age/sex demographics,
1622 implemented and hosted via Qualtrics). Because participants completed the task on their own
1623 desktop or laptop devices, with variable screen-refresh rates and display characteristics, all timings
1624 are approximate, and exact stimulus sizes and luminances cannot be reported.

1625 The participant's bonus reward (£2.00 at the start of the experiment) was displayed
1626 continuously as a fixation stimulus, against a mid-grey background. A peripheral track, consisting of
1627 two white ovals (drawn at 62% and 78% of the screen's size) was also presented continuously. Each
1628 trial began with the fixation stimulus changing colour to white, after which a variable foreperiod
1629 preceded the onset of an imperative "go" stimulus (a small white oval surrounding fixation). The
1630 imperative stimulus was displayed for a maximum of 200 ms. Responses were made using the space
1631 key, which was monitored continuously from the beginning of the foreperiod. Following a response,
1632 the imperative stimulus (if still present) was removed, and the fixation stimulus changed colour for
1633 500 ms to signal feedback. It could change to red (for an incorrect anticipation, $RT < 50$ ms,
1634 accompanied by a 1 pence loss of bonus reward), amber (slow response, no change in bonus reward)
1635 or green (fast response, 0.5 pence increase in bonus reward). Unbeknownst to participants, the
1636 threshold between slow and fast responses used to determine rewards was adaptive (62.5th
1637 percentile of RTs in the previous 16-trial "miniblock") to equate speed/accuracy pressure for
1638 respondents with different individual RTs and hardware delays.

1639 Throughout each trial, additional white stimuli could be flashed within the peripheral track
1640 for a duration of one display frame. The chance of a peripheral stimulus occurring was 1/600 on each
1641 frame. This corresponds to an objectively flat hazard function for the time between peripheral

1642 stimuli. If such a stimulus occurred, its position and type (either a target “+” or a non-target “x”) was
 1643 allocated at random. In some blocks participants were instructed to count the number of target “+”
 1644 stimuli that had occurred since their last report (while ignoring non-targets).

1645 The experiment began with practice sessions, first for the main task alone, then for the
 1646 peripheral task alone. Participants needed to respond correctly (RT >50 and <1000 ms 10 times in a
 1647 row) or count peripheral stimuli correctly (across a 10-second interval, three times in a row) to
 1648 proceed to the next stage and were screened out if they could not accomplish this within a
 1649 reasonable period of time. During practice, a fixation cross replaced the display showing the current
 1650 bonus reward, and the foreperiod was constant (1.14 s).

1651 Following practice a 2x2x9 repeated-measures design was used. The first factor, *distribution*,
 1652 manipulated the foreperiod distribution from which foreperiods were being drawn (exponential or
 1653 flipped, in separate blocks). The second factor, *cost*, manipulated the relevance of the secondary
 1654 “peripheral track” task (to yield either a cost, resulting from failures to accurately monitor this track
 1655 with peripheral visual attention, or no such cost, in separate blocks). Participants completed the four
 1656 (2x2) blocks relating to these factors in one of four (nested counterbalanced) orders.

1657 The third factor randomised the exact foreperiod experienced on each trial. Initially, each
 1658 block was programmed to contain a series of nine mini-blocks, each of which contained a random-
 1659 without-replacement ordering of eight possible foreperiods each presented twice. In exponential-
 1660 distribution blocks, these eight foreperiods were 0.77, 0.8, 0.84, 0.9, 0.98, 1.1, 1.27 and 1.5 s. In
 1661 flipped blocks, they were 0.77, 1, 1.17, 1.29, 1.37, 1.43, 1.47, and 1.5 s. However, the final (i.e. 144th)
 1662 trial of each block was replaced with a surprise foreperiod of 0.5 s (cf. Grosjean et al., 2001).

1663 Participants received feedback (median RT, % correct, and current level of bonus reward)
 1664 and an opportunity to rest after each miniblock. Before this feedback was presented, they reported
 1665 the number of times that a “+” stimulus had flashed in the peripheral track during that miniblock
 1666 (selecting from options: 0, 1, 2, 3, 4+). They made this report even in no-cost blocks, but between

blocks they were informed about the relevance of the peripheral track for the next set of miniblocks (either that it should be ignored entirely as it did not affect the bonus reward, or that it should be attended because it did). This instruction was reinforced after each miniblock as follows: In cost blocks, feedback indicated a correct or incorrect report about the number of “+” stimuli and its effect on the bonus reward (a 2-pence gain or a 20-pence loss, respectively). In no-cost blocks, no information on peripheral-task accuracy was provided, with a message instead stating that the peripheral track should be being ignored. After the experiment, to ensure that there had been adequate adherence to the cost manipulation, participants were excluded if the correlation between the number of “+” stimuli in each miniblock and a participant’s report about this number was 1) not significant ($p > 0.05$) in cost blocks and 2) not greater in cost blocks than in no-cost blocks. To ensure adequate performance on the primary task, participants were also excluded if they had >20% of either anticipatory errors or RTs exceeding one second.

Inferential statistics

For inferential statistics relating to trends across conditions we made use of Matlab v2021a (The Mathworks), in particular the `fitrm` and `ranova` commands for ANOVAs, and JASP v0.95.4 (JASP Team, 2025).

Experiment 3

Participants

Eight participants completed the experiment (five male, three female, mean age = 26.9, sd = 10.4). The sample size was arbitrary, and the study was not pre-registered. Most participants were naive to the hypotheses and completed the experiment in two sessions (which each took 1-2 hours), receiving £40 for their time. The exception was one author participant (KNY) who provided twice as much data (and was not paid). Experimental procedures were approved by the City, University of

1691 London Department of Psychology Research Ethics Committee (ETH2324-2257 Modelling hand and
1692 foot responses).

1693 Apparatus, Stimuli, Design and Procedure

1694 A PC running Windows and Matlab R2013b (MathWorks, Natick, MA, USA) with functions
1695 from the Psychtoolbox extension v3 (Brainard, 1997; Kleiner et al., 2007) controlled the experiment.
1696 Instructions were presented via a Display ++ M0623 monitor (Cambridge Research Systems,
1697 Rochester, UK; vertical refresh rate 100 Hz) in white text on a black background. Warning and go
1698 stimuli were 10 ms pure tones with frequencies of 100 Hz and 500 Hz yielding a “thud” and a “beep”
1699 respectively. They were played through PX360 headphones (Sennheiser, Wedemark, Germany).
1700 Responses were recorded using a digital button (pinched between forefinger and thumb) and a
1701 digital pedal, both interfaced via a national instrument X-series PCIe-6323 A/D card (National
1702 Instruments, Austin, TX, USA) sampling at 10,000 Hz.

1703 A 2x2x2x2x24 repeated-measures design was used. Most participants completed 16 blocks,
1704 each of at least 192 trials. The first factor, *distribution*, manipulated the foreperiod distribution from
1705 which foreperiods were being drawn (either Gaussian or uniform, in separate blocks, switching over
1706 after half of the blocks were completed). The second factor, *rate*, manipulated the rate at which
1707 catch trials occurred, and switched every four blocks. There were either no catch trials, or the
1708 addition of 24 catch trials (implying an 11.1% rate) at random positions within a block.²⁰ The third
1709 factor, *effector*, manipulated the response (either a pinch, or a foot press), alternating every block.
1710 Each participant completed blocks relating to these factors in one of eight possible (nested
1711 counterbalanced) orders. The fourth factor, *intensity*, randomised the intensity of the imperative
1712 stimulus within blocks (with strong and weak stimuli having a loudness of 71 vs 51 dBa respectively).
1713 The fifth and final factor randomised the exact foreperiod experienced on each (non-catch) trial,

²⁰ The author participant completed further blocks with 4% and 7.7% catch trials after completing the first two levels of this factor, such that they only swapped to a new distribution after 16 blocks.

1714 with 8 repetitions of 24 possible foreperiods per block. The foreperiods were 0.793, 0.893, 0.948,
1715 0.989, 1.023, 1.052, 1.078, 1.102, 1.125, 1.147, 1.169, 1.190, 1.210, 1.231, 1.253, 1.275, 1.298,
1716 1.322, 1.348, 1.377, 1.411, 1.452, 1.507, and 1.607s in Gaussian blocks and 0.793, 0.828, 0.864,
1717 0.899, 0.935, 0.97, 1.005, 1.041, 1.076, 1.112, 1.147, 1.182, 1.218, 1.253, 1.288, 1.324, 1.359, 1.395,
1718 1.430, 1.465, 1.501, 1.536, 1.572 and 1.607s in uniform blocks.

1719 The experimental task is schematised in Figure 11A. Each trial began with a warning thud,
1720 after which a variable foreperiod preceded either an imperative beep instructing a speeded
1721 response or, on catch trials, the end of the trial after 2s. Following the response or trial end, there
1722 was a 0.5s gap before the start of the next trial. Breaks with written summary feedback (median RT
1723 and, when they occurred, % anticipation errors) were provided after every 20 trials. If there were no
1724 anticipation errors, the feedback was accompanied by the exhortation, "Can you go faster?"

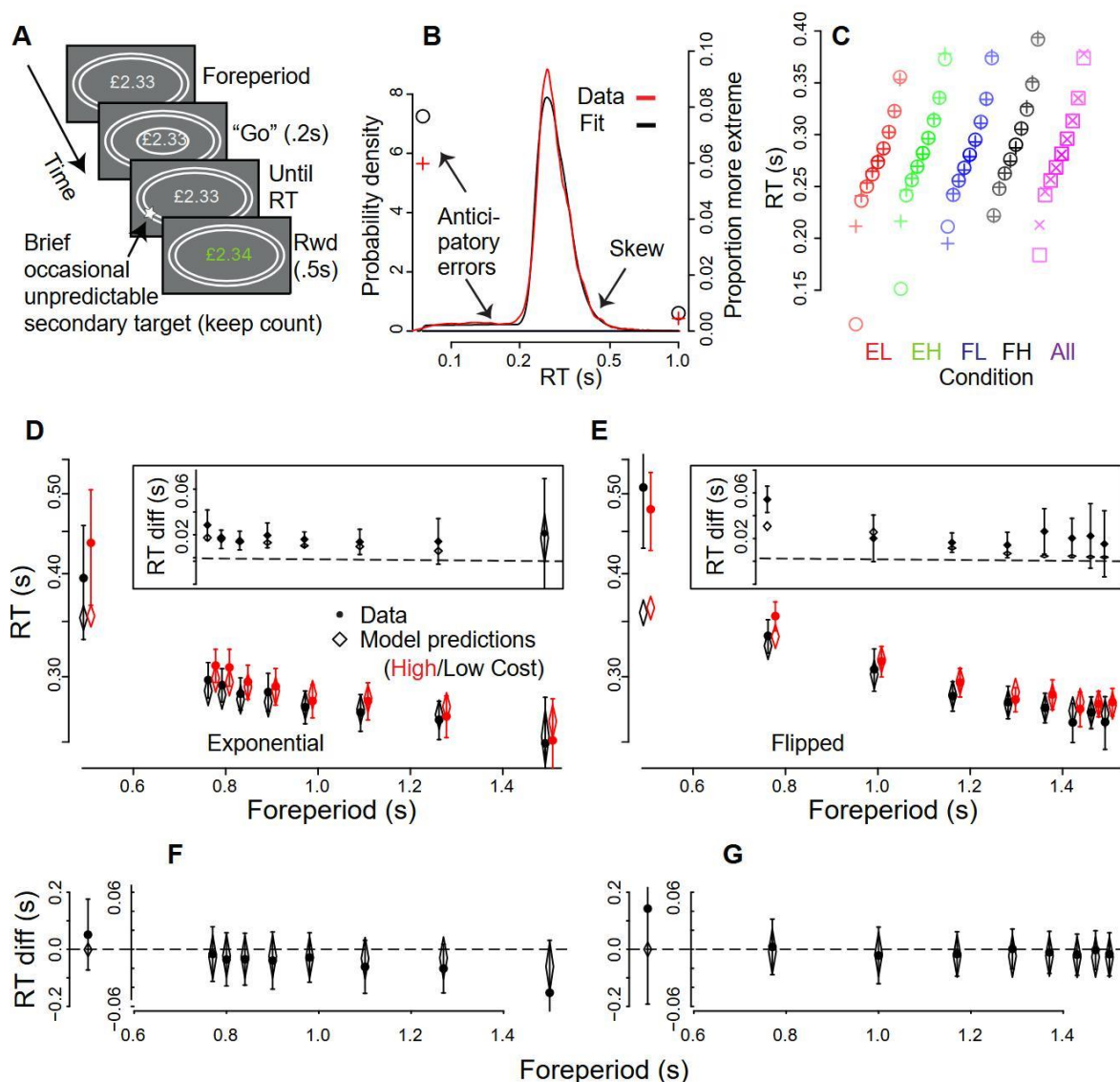
1725

1726

Appendix B: simplified model fit for Experiment 1

To conclude our analysis of data from Experiment 1, we created a final model variant, based on the winning saturating variant with hazard-based preparation shown at the top of Table 2. In this final variant, we reduced all parameters that are not being modified as a function of elapsed time within a trial (i.e. the Weber ratio, φ , the learning rate, α , and the rate of decay for stimulus-dependent drive, δ) to each be just a single parameter for the entire group (rather than one parameter per participant). We tested this reduced model against data from Experiment 1 in a full Bayesian fit. The new, reduced variant has 364 individual-level parameters (for 36 participants) compared to the original 469, so a reduction of over 100 parameters.

The reduced model fitted significantly less well than the winning model upon which it had been based (via a method that takes account of its reduced parametric complexity) yielding an increase in PSIS-LOOIC of 419.6 (SE 73.2, $z = 7.96$, $p < 0.001$). This poorer fit justifies our original choice to model these parameters at the individual level. However, it also allows us to illustrate how a large reduction in parametric complexity has only a minor effect on the model's capacity to explain key trends in the data. For that purpose, we present Figures B1 and B2 below, which illustrate the same data trends shown in Figures 6 and 7 respectively against the fit of the reduced model. The model still captures key effects (of foreperiod, distribution, cost, and prior experience), albeit very slightly less well than a model varying more parameters at the individual level. Hence, our model's ability to capture the data patterns we have highlighted does not appear particularly sensitive to the parameter count.

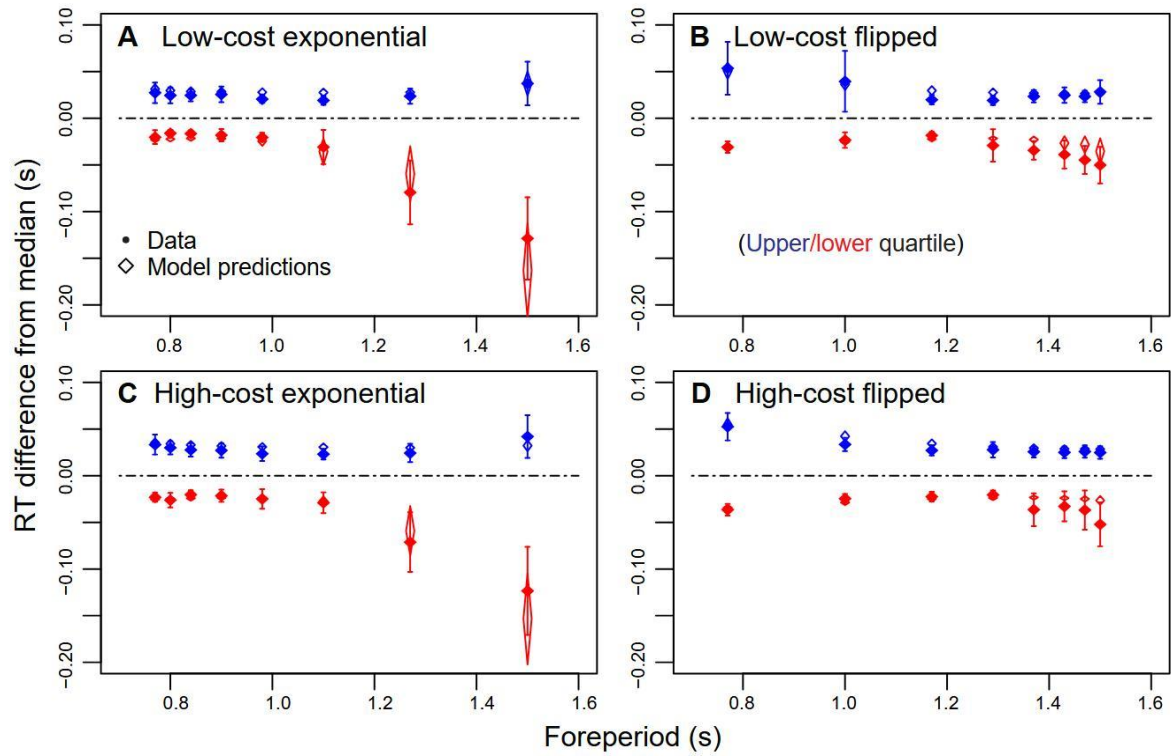


1748

1749 Figure B1. Experiment 1 procedure and results revisited. Figure shows whole-group data along with
 1750 equivalently averaged or collated reduced saturating hazard-based preparation SIP-then-SPuRT
 1751 model predictions. **A.** Schematic of procedure (single trial from high-cost condition, rewarded for a
 1752 fast and accurate response). **B.** Kernel-density estimate of the distribution of RTs (relative to go
 1753 stimulus) collated across all levels of experimental factors (red) along with equivalent model
 1754 prediction (black). Left axis relates to density for RTs both > 0.075 and < 1 second; right axis relates to
 1755 discrete probability for RTs < 0.075 or > 1 second (shown as crosses and circles at RT = 0.075 and 1,
 1756 respectively). Note that the x axis uses a logarithmic scale that masks the heavy right tail
 1757 characteristic of RT distributions. **C.** Quantiles of distributions of RT (from 10th to 90th percentile, left

1758 to right offset within each condition) for trials collated across participants and foreperiods but
 1759 separated by block (E = exponential distribution, F = flipped distribution, H denotes a high
 1760 opportunity-cost block, L a low-cost condition). Open symbols are model predictions, crosses are
 1761 data. **D-E.** Main plots show average effect of foreperiod on mean of median RTs in blocks utilising the
 1762 exponential foreperiod distribution (D) or the flipped foreperiod distribution (E). Solid symbols are
 1763 data, open diamonds are model predictions. The effect is shown separately for high (red) and low
 1764 (black) cost blocks (x-axis offset added for clarity). Insets illustrate $n - 1$ sequential effects as the
 1765 mean difference in median RTs (collated across high/low-cost blocks) between trials preceded by one
 1766 of the two longest foreperiods and trials preceded by one of the two shortest foreperiods. Error bars
 1767 (around data points) and top/bottom vertices (for model predictions) indicate 95% confidence
 1768 intervals across the group. **F-G.** Prior-block effects. Mean difference in median RTs between two
 1769 subgroups of participants who, as a result of counterbalancing, experienced the flipped vs.
 1770 exponential blocks first. The surprise (0.5 s) foreperiod data point is shown at a different axis scale
 1771 (far left) relative to more frequently experienced foreperiods. Error bars (around data points) and
 1772 top/bottom vertices (for model predictions) indicate 95% confidence intervals around between-group
 1773 differences.

1774



1775

1776 Figure B2. Variable-foreperiod effects on the shape of the RT distribution (Experiment 1, revisited).

1777 Blue symbols show data (solid points) and reduced saturating hazard-based model predictions (open

1778 symbols) for the 50th to 75th percentile of RTs. Red symbols denote the 50th to 25th percentile. Error

1779 bars show the 95% confidence interval on the data, while upper and lower diamond vertices show the

1780 equivalent confidence interval around model predictions. **A.** Exponential distribution low-cost block.

1781 **B.** Flipped distribution low-cost block. **C.** Exponential distribution high-cost block. **D.** Flipped

1782 distribution high-cost block.

1783

Appendix C: Further results for Experiment 3

For the data presented in Figure 11, we ran further inferential tests to confirm that our positive-control manipulations were effective. As in Experiment 1, only the two most extreme foreperiods (0.793 and 1.607 s) appeared in blocks with both kinds of distribution (Gaussian and uniform), so we entered them into a 2 (distribution) x 2 (Intensity) x 2 (effector) x 2 (foreperiod) repeated-measures ANOVA. This yielded an interaction between distribution and foreperiod ($F(1,7) = 19.49, p = 0.003, \omega^2 = 0.042$) showing that preparation differed between the two kinds of foreperiod distribution. There were also main effects of effector ($F(1,7) = 114.35, p < 0.01, \omega^2 = 0.768$), intensity ($F(1,7) = 57.06, p < 0.01, \omega^2 = 0.144$) and foreperiod ($F(1,7) = 75.32, p < 0.001, \omega^2 = 0.408$). There was only one further significant effect, a three-way interaction between intensity, effector and foreperiod ($F(1,7) = 9.55, p = 0.018, \omega^2 = 0.006$). Two further separate 2x2x24 ANOVAs using all foreperiods confirmed that significant main effects of intensity (Gaussian: $F(1,7) = 110.12, p < 0.001, \omega^2 = 0.101$; uniform: $F(1,7) = 49.03, p < 0.001, \omega^2 = 0.160$), effector (Gaussian: $F(1,7) = 100.91, p < 0.001, \omega^2 = 0.716$; uniform: $F(1,7) = 117.00, p < 0.001, \omega^2 = 0.757$) and foreperiod (Gaussian: $F(23,161) = 20.27, p < 0.001, \omega^2 = 0.217$; uniform: $F(23,161) = 16.54, p < 0.001, \omega^2 = 0.164$) were present in each distribution. For the Gaussian-distributed blocks, there was also an interaction between intensity and foreperiod ($F(23,161) = 1.93, p = 0.006, \omega^2 = 0.101$), echoing a classic result from Niemi and Lehtonen (1982) but with a wider range of foreperiods. There were no further significant effects.

For concision we don't illustrate effects of the previous foreperiod for Experiment 3, but they were robustly present when considering the difference between trials preceded by the most or least extreme 25% of foreperiods, in both Gaussian ($F(1,7) = 21.94, p = 0.002, \omega^2 = 0.044$) and uniform ($F(1,7) = 17.32, p = 0.004, \omega^2 = 0.06$) distribution blocks. In uniform blocks this effect of prior foreperiod also interacted significantly with current foreperiod ($F(23,161) = 3.32, p = 0.015, \omega^2 =$

0.11). There were also effects of the current foreperiod in these ANOVAs ($F(23,161) = 4.83, p = 0.003$; $F(23,161) = 3.92, p = 0.006$).

The SIP-theN-SpuRT models were given capacity to explain the significant differences in RT between pinch and foot-pedal responses by allowing changes to the intercept, pre-saturation gain and post-saturation gain of the preparation function for residual RT. The relevant parameter values are shown in Figure C1 panels A-C respectively. Only the intercept (β_{Ebt}) changes reliably for the group, with a credible interval that excludes zero. The positive value indicates a larger residual RT at the onset of the trial for foot compared to hand responses (by around 80 ms). The post-saturation gain parameter (β_{Eat}) shows a non-significant trend towards negative values, but its interpretation is fairly complex, as for residual RT the post-saturation gain represents roughly the maximum *proportional* reduction seen across foreperiods (see footnote 6). Hence the trend towards a lower proportional reduction in the case of foot-pedal responses, which start with a higher residual RT, actually implies a fairly similar absolute reduction in RT from an initial relaxed state to the maximum point of preparation (in the region of 45 milliseconds for both pinch and foot-pedal responses). With inconsistent effects of the effector on pre-saturation gain (β_{Egt}) the overall group-level interpretation is that foot responses had longer residual RTs but a quite similar temporal preparation profile relative to pinch responses.

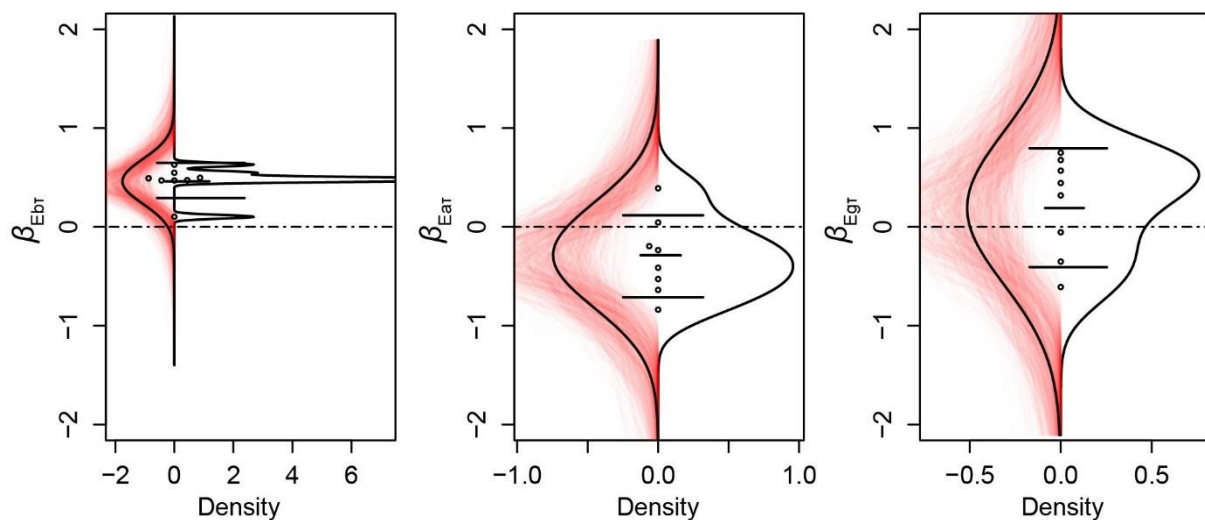


Figure C1. Saturating hazard-dependent SIP-theN-SPuRT model parameters quantifying changes across effector conditions in Experiment 3. Circles indicate parameter estimates for individual participants. Shorter central solid horizontal lines indicate the group mean. Longer central solid horizontal lines indicate the 95% Bayesian credible interval around the group mean. The left-hand butterfly wings describe the model's estimate of the group-level distribution, with the black line illustrating the mean posterior estimate while red shading is derived from the entire posterior to illustrate uncertainty. Right-hand butterfly wings denote kernel-density estimates of individual parameter values. **A.** β_{Ebt} , the (logarithmic) change in the intercept of residual RT. **B.** β_{Eat} , the (logarithmic) change in the rate with which preparation affects residual RT (subsequent to saturation). **C.** β_{Egt} , the (logarithmic) change in the rate with which preparation affects residual RT (prior to saturation).

For differences in RT between high and low intensity stimuli, six parameters (shown in Figure C2 panels A-F) were free to vary, but none did so reliably. Here, effects were available on both residual RT and sensory gain. In both cases, intercept parameters came closest to a reliable difference for the group (panels A and D). Sensory gain intercept values ($\beta_{\text{Ib}\lambda}$) were negative, consistent with an average 18% lower stimulus-dependent drive arising from the weaker stimulus. However, residual RT intercept values (β_{Ibt}) actually trended positive by 8%. These non-significant results must be interpreted cautiously, but as sensory drive affects both the spread and the central tendency of the RT distribution, it may be that changing its intercept to best capture the greater spread of RT for low-intensity stimuli (see Figure 11C) led to slightly too much of a concomitant reduction in central tendency, which was then countered via an (unrealistic) change to residual RT. In this way, excessive parametric flexibility allowed the model to perform slightly better than it really should have done, masking an underlying deficit in its architecture.

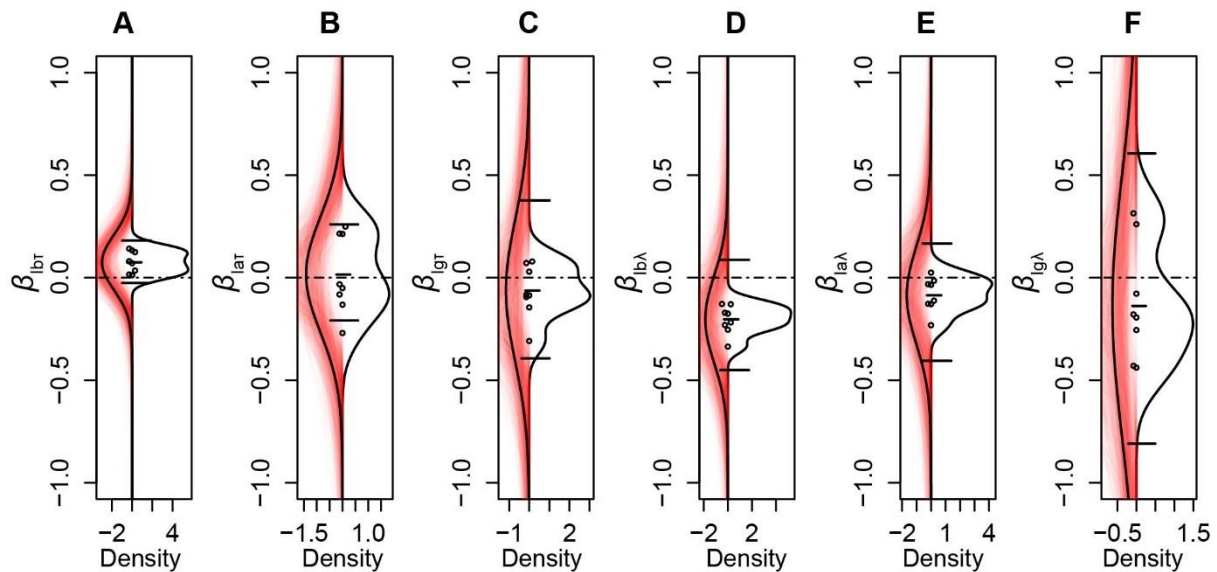


Figure C2. Saturating hazard-dependent SIP-theN-SPuRT model parameters quantifying changes across intensity conditions in Experiment 3. Circles indicate parameter estimates for individual participants. Shorter central solid horizontal lines indicate the group mean. Longer central solid horizontal lines indicate the 95% Bayesian credible interval around the group mean. The left-hand butterfly wings describe the model's estimate of the group-level distribution, with the black line illustrating the mean posterior estimate while red shading is derived from the entire posterior to illustrate uncertainty. Right-hand butterfly wings denote kernel-density estimates of individual parameter values. **A.** β_{lbr} , the (logarithmic) change in the intercept of residual RT. **B.** β_{lat} , the (logarithmic) change in the rate with which preparation affects residual RT (subsequent to saturation). **C.** β_{lgr} , the (logarithmic) change in the rate with which preparation affects residual RT (prior to saturation). **D.** β_{lba} , the (logarithmic) change in the intercept of stimulus-dependent drive. **E.** β_{lal} , the (logarithmic) change in the rate with which preparation affects stimulus-dependent drive (subsequent to saturation). **F.** β_{lgl} , the (logarithmic) change in the rate with which preparation affects stimulus-dependent drive (prior to saturation).