ARTICLE IN PRESS

Journal of Mathematical Psychology ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

Semiparametric Bayesian approaches to systems factorial technology

Joseph W. Houpt [a], Steven N. MacEachern [b], Mario Peruggia [b], James T. Townsend [c], Trisha Van Zandt [d,*]

[a] Department of Psychology, Wright State University, Dayton, OH, United States
[b] Department of Statistics, The Ohio State University, Columbus, OH, United States
[c] Psychological and Brain Sciences, Indiana University, Bloomington, IN, United States
[d] Department of Psychology, The Ohio State University, Columbus, OH, United States

HIGHLIGHTS

- A semiparametric piecewise exponential model is used to estimate response time hazard functions.
- This semiparametric model is fit to data in a Bayesian hierarchy.
- The resulting posterior estimates of the hazard functions can be used to compute the posterior estimates of systems factorial functionals.

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Reaction times
Hazard rates
Piecewise exponential
Capacity
Survivor function

ABSTRACT

Systems factorial technology (e.g., Townsend and Nozawa, 1995) is a toolbox of methods to test characteristics of multichannel processing systems. In this paper, we present a semiparametric Bayesian model of response times based on a piecewise-exponential model. This model allows us to compute the posterior estimates of the response time distribution hazard function and the systems factorial technology functionals that can be computed from it, including the survivor function interaction contrast and the capacity coefficient. We then compute posterior estimates of statistics that reflect different behaviors of the functionals, and so arrive at a fully Bayesian and theoretically-motivated approach to systems factorial technology. Simulation studies demonstrate the effectiveness of the approach, and we apply it to a published data set.

1. Introduction

Consider a set of independent and identically distributed response time (RT) measurements $\boldsymbol{T} = \{t_1, t_2, \ldots, t_n\}$ collected from an observer performing an experimental task. The researcher's goal in collecting such data is usually to test a model of the RT-generating process. This model will describe the behavior of the continuous random variable T_k for trial k by specifying how it is distributed, that is, by stating that it follows a distribution with (positive and Lebesgue-integrable) probability density function (pdf) $f(t \mid \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the vector of the distribution's parameters and $f(t \mid \boldsymbol{\theta})$ has support on the positive real line $[0, \infty)$.

Under these conditions we can define the cumulative distribution function (cdf) F of T_k as

$$F(t \mid \boldsymbol{\theta}) = P(T_k \leq t) = \int_0^t f(u \mid \boldsymbol{\theta}) du,$$

and its corresponding survivor function

$$S(t \mid \boldsymbol{\theta}) = P(T_k > t) = 1 - F(t \mid \boldsymbol{\theta}).$$

Then the hazard function or failure rate of T_k is

$$h(t \mid \boldsymbol{\theta}) = f(t \mid \boldsymbol{\theta}) / [1 - F(t \mid \boldsymbol{\theta})] = f(t \mid \boldsymbol{\theta}) / S(t \mid \boldsymbol{\theta}).$$

The hazard function can be viewed as the instantaneous rate of change in the likelihood that an event will occur at time t: For some small time interval Δt, $h(t \mid \boldsymbol{\theta})\Delta t$ is approximately the probability that an event will occur in the interval $[t, t + \Delta t)$ given that the event has not occurred before time t.

The functions in Table 1 are different ways of presenting the same information about T_k. The use of the hazard function in modeling RT data arises in part from the observation that, although empirical RT density and distribution functions all look quite similar (i.e., unimodal and positively skewed), the hazard function is much better for visually discriminating between distributional families. Luce (1986), for example, showed how two variables with nearly indistinguishable pdfs had very different and discriminable

* Corresponding author.
E-mail address: van-zandt.2@osu.edu (T. Van Zandt).

http://dx.doi.org/10.1016/j.jmp.2016.02.008
0022-2496/

Please cite this article in press as: Houpt, J. W., et al., Semiparametric Bayesian approaches to systems factorial technology. Journal of Mathematical Psychology (2016), http://dx.doi.org/10.1016/j.jmp.2016.02.008

Table 1
Relationships among functions describing the distribution of a random variable. Dependence on the distribution's parameters is suppressed.

Notation	Name	Relationship
$f(t)$	Density	$\frac{d}{dt}F(t) = h(t)S(t)$
$F(t)$	Distribution	$\int_0^t f(u)du$
$S(t)$	Survivor	$1 - F(t)$
$h(t)$	Hazard	$f(t)/S(t)$
$H(t)$	Integrated hazard	$\int_0^t h(u)du = -\ln S(t)$
$k(t)$	Reverse hazard	$f(t)/F(t)$
$K(t)$	Integrated reverse hazard	$\int_0^t k(u)du = \ln F(t)$

hazard functions. Others have described ways that the hazard function can be used to identify cognitive processing properties (Ashby, 1982; Chechile, 2003; Townsend & Ashby, 1983; Van Zandt & Ratcliff, 1995).

Systems factorial technology (SFT), developed by Townsend and colleagues, uses a number of "functionals" based on the cdf $F(t|\boldsymbol{\theta})$ and the survivor function $S(t|\boldsymbol{\theta})$ to determine different characteristics of multichannel cognitive processing systems (Houpt & Townsend, 2010, 2012; Schweickert, Fisher, & Sung, 2012; Townsend & Nozawa, 1995; Townsend & Wenger, 2004). These characteristics include whether different channels process information simultaneously (in parallel) or one at a time (in series), how information is combined over the channels (the "stopping rule"), and the capacity of the system (how performance is affected by increasing the number of signals to be processed).

The SFT approach to model testing requires that the functionals be estimated from the RT data. Therefore, the estimated functionals are random functions with considerable variability. Some methods have been proposed for accommodating this variability so that arguments about plausible and implausible models can be objectively supported (Houpt & Townsend, 2010, 2012). In this paper we suggest an alternative approach that uses a Bayesian hierarchical model to estimate the posterior distribution of the hazard function and the functionals that can be computed from it.

In the rest of this paper we first present the functional approach to model testing, followed by our new Bayesian approach. Then we apply this new approach to simulated data, and conclude by analyzing a data set published by Eidels, Townsend, Hughes, and Perry (2015).

2. Systems factorial technology

Consider the problem posed by a task requiring observers to process multiple sources of information, each source represented by a channel devoted to acquiring, processing and transmitting the information within it. In the simplest case, which is all we will consider here, the observers are asked to press a button whenever a stimulus is detected, no matter how many stimuli are presented, where or in what modality they are presented.

For example, an experiment might present simple stimuli (e.g., lights or tones) of varying intensities (e.g., bright or dim, or soft or loud) in two locations. Each location, by virtue of stimulating the receptive fields of different sets of neurons, corresponds to a processing channel. Across trials, two stimuli could be presented simultaneously by presenting one in each of the two locations, or only one stimulus could be presented in one of the two locations. (In some cases, catch trials in which no stimulus is presented could occur, but would not be included in an analysis.) The intensity of a stimulus results in faster or slower processing in the channel in which it was presented: low intensity stimuli, being more difficult to perceive, result in slower processing times.

This experiment is an example of a 2×2 factorial design: channel by processing speed. Under this design we obtain four samples of RTs, which we can designate by $\boldsymbol{T}_c$, where $c = ss$ (both channels slow), sf (Channel 1 slow and Channel 2 fast), fs (Channel 1 fast and Channel 2 slow), or ff (both channels fast). Under the assumption that the RTs within a condition are independent and identically distributed, the hazard function for T_{ck} on any trial k under condition c can be written as $h_c(t \mid \boldsymbol{\theta}_c)$ (and similarly for the other functions), where $\boldsymbol{\theta}_c$ is the vector of parameters defining the model.

As we introduced it above, SFT is concerned with three basic properties of the system of channels. The first property is the temporal arrangement or architecture[1] of the channels, that is, whether processing in the channels is serial or parallel. The second property is the stopping rule that determines when processing ends, that is, whether processing can stop when any channel finishes (self-terminating or OR processing) or whether all channels must process their stimuli to completion (exhaustive or AND processing). The third property is the capacity of the system, which describes the system's efficiency as more sources of information need to be processed. We briefly summarize first the capacity coefficient and then the survivor function interaction contrast; interested readers should consult Townsend and Nozawa (1995) for more details.

2.1. The capacity coefficient

Capacity refers to the ability of the processing system to perform its task under increases in the amount of work it is asked to perform. A super-capacity system is one that gets faster as the load on the system increases. A limited-capacity system, by contrast, slows down. In an unlimited-capacity system, the rate at which information is processed in each channel is not affected by the number of channels in operation.

Consider the 2×2 experimental design where the "slow" low-intensity stimulus is of zero intensity, or turned off. If a stimulus is presented to Channel 1 and no stimulus is presented to Channel 2 (Condition $c = u\cdot$, where $u = s$ or f, slow or fast) the processing time $T_{ck} = T_{u\cdot,k}$ follows a distribution with pdf $f_{u\cdot}(t \mid \boldsymbol{\theta}_{u\cdot})$. If a stimulus is presented to Channel 2 and no stimulus is presented to Channel 1 (Condition $c = \cdot v$ where $v = s$ or f, slow or fast), the processing time $T_{\cdot v,k}$ follows a distribution with pdf $f_{\cdot v}(t \mid \boldsymbol{\theta}_{\cdot v})$. If stimuli are presented to both Channel 1 and Channel 2, the processing time $T_{uv,k}$ follows a distribution with pdf $f_{uv}(t \mid \boldsymbol{\theta}_{uv})$.

The capacity coefficient compares performance to the predicted response time distributions of an unlimited-capacity, independent, parallel (UCIP) model. For an OR task, the UCIP model predicts that the survivor function of the processing time with two signals $S_{uv}(t \mid \boldsymbol{\theta}_{uv})$ is the product of the survivor functions $S_{u\cdot}(t \mid \boldsymbol{\theta}_{u\cdot})$ and $S_{\cdot v}(t \mid \boldsymbol{\theta}_{\cdot v})$ for the processing times for each signal presented alone. For an AND task, the UCIP model predicts that the cdf of the processing time with two signals $F_{uv}(t \mid \boldsymbol{\theta}_{uv})$ is the product of the cdfs $F_{u\cdot}(t \mid \boldsymbol{\theta}_{u\cdot})$ and $F_{\cdot v}(t \mid \boldsymbol{\theta}_{\cdot v})$ for the processing times of each signal presented alone. The capacity coefficients are expressed in terms of the natural logarithm of the predicted and observed survivor (or distribution) functions. Formally, with $H(t) = -\ln S(t)$ and $K(t) = \ln F(t)$,

$$C_{\text{OR}}(t) = \frac{H_{uv}(t \mid \boldsymbol{\theta}_{uv})}{H_{u\cdot}(t \mid \boldsymbol{\theta}_{u\cdot}) + H_{\cdot v}(t \mid \boldsymbol{\theta}_{\cdot v})} \quad \text{and}$$
$$C_{\text{AND}}(t) = \frac{K_{u\cdot}(t \mid \boldsymbol{\theta}_{u\cdot}) + K_{\cdot v}(t \mid \boldsymbol{\theta}_{\cdot v})}{K_{uv}(t \mid \boldsymbol{\theta}_{uv})}. \tag{1}$$

[1] In other cognitive modeling approaches, e.g., ACT-R (Anderson, 2007), the term "architecture" is reserved for fixed characteristics of the cognitive processing system that do not vary according to task, strategy or otherwise. Within SFT, architecture (e.g., parallel or serial temporal organization) may vary across tasks.

Please cite this article in press as: Houpt, J. W., et al., Semiparametric Bayesian approaches to systems factorial technology. Journal of Mathematical Psychology (2016), http://dx.doi.org/10.1016/j.jmp.2016.02.008

One approach to estimating the capacity coefficients is to use the natural logarithm of the empirical cdfs, and then use bootstrapped confidence intervals to evaluate differences between the estimates. (e.g., Van Zandt, 2002a). More recently, Houpt and Townsend (2012) proposed using the empirical cumulative hazard function (also known as the Nelson–Aalen estimator when generalized to include censored data). Houpt and Townsend's approach allows for a null-hypothesis test based on a z-score of the difference between the model-based predicted performance and the observed performance (see Houpt & Townsend, 2012, for details).

2.2. The survivor function interaction contrast

The survivor function interaction contrast (SIC) assesses the system's architecture and stopping rule. The SIC is defined as the contrast between the change in the survivor function when Channel 1 is slow and Channel 2 goes from slow to fast, and when Channel 1 is fast and Channel 2 goes from slow to fast, so

$$SIC(t) = \left[S_{ss}(t \mid \boldsymbol{\theta}_{ss}) - S_{sf}(t \mid \boldsymbol{\theta}_{sf})\right] - \left[S_{fs}(t \mid \boldsymbol{\theta}_{fs}) - S_{ff}(t \mid \boldsymbol{\theta}_{ff})\right].$$

Each combination of serial and parallel architectures with AND and OR stopping rules results in different shapes of the SIC (see Table 2). The SIC can also indicate whether information is shared across the two channels, that is, whether there is coactivation.

The SIC function is usually estimated with an interaction contrast of the empirical survivor functions (or equivalently with the negative of the interaction contrast of the empirical cdfs). As with the capacity coefficients, uncertainty in the estimated SIC has often been estimated using bootstrapping (e.g., Van Zandt, 2002a). Houpt and Townsend (2010) derived tests of the SIC function shape based on observed data using the supremum $\hat{D}^+$ and infimum $\hat{D}^-$ of the estimated SIC. The estimates $\hat{D}^+$ and $\hat{D}^-$ can be combined for inferences about the architecture and stopping rules of the system.

We take a quite different approach to estimation and inference for both the capacity coefficients and the SIC in this paper. We first develop a Bayesian model of RT data, which is presented in the next section, and use that model to estimate the posterior distributions of hazard functions across conditions. From those posteriors we can estimate the posterior distributions of the SIC and capacity coefficient directly. We then compute the posterior estimates of various statistics that describe the behavior of the functions. In this way we present a comprehensive approach to applying the functions to data, and a principled method for drawing conclusions from those applications.

3. The Bayesian model

To begin, recall that, for a vector of parameters $\boldsymbol{\theta}$, the likelihood of data $\boldsymbol{T} = \{t_1, t_2, \ldots, t_n\}$ is given by

$$L(\boldsymbol{\theta} \mid \boldsymbol{T}) = \prod_{k=1}^{n} f(t_k \mid \boldsymbol{\theta}),$$

when $\boldsymbol{T}$ is assumed to be an independent and identically distributed sample drawn from a distribution with pdf f. As we defined above, the hazard function for the variable T_k is

$$h(t \mid \boldsymbol{\theta}) = f(t \mid \boldsymbol{\theta})\left(S(t \mid \boldsymbol{\theta})\right)^{-1},$$

where we will assume implicitly that the support of T_k is the positive real line, and hence f takes on the value 0 for all non-positive values of t. Noting that

$$f(t \mid \boldsymbol{\theta}) = h(t \mid \boldsymbol{\theta}) S(t \mid \boldsymbol{\theta}),$$

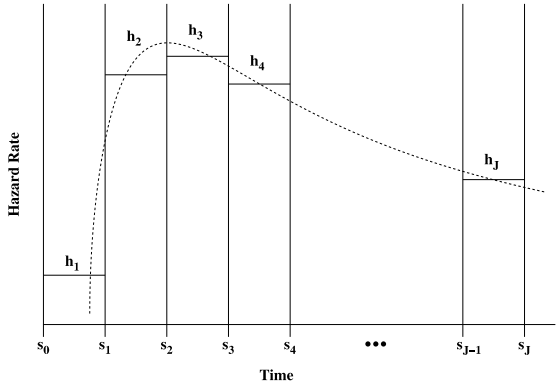

Fig. 1. The piecewise exponential approximation of a continuous hazard function (dotted line). The approximation is represented by the horizontal line segments with heights h_j between the vertical boundaries at $s_j, j = 0, \ldots, J$.

we can rewrite the likelihood of the data $\boldsymbol{T}$ as

$$L(\boldsymbol{\theta} \mid \boldsymbol{T}) = \prod_{k=1}^{n} h(t_k \mid \boldsymbol{\theta}) S(t_k \mid \boldsymbol{\theta}).$$

We want to estimate values of the function $h(t \mid \boldsymbol{\theta})$ at a finite number J of points over the observed range of T. To do this we will use a semiparametric approach in which we assume that the behavior of the hazard function at these points can be well described by a piecewise exponential model (Ibrahim, Chen, & Sinha, 2005), depicted in Fig. 1. We first place $J+1$ points $\{s_0, s_1, \ldots, s_J\}$ along the support of T_k, where $s_0 = 0$ and $s_J > \max_k\{T_t\}$. The hazard function $h^*(t)$ of the piecewise exponential model is piecewise constant, i.e.,

$$h^*(t) = h_j \quad \text{for } s_{j-1} < t \le s_j,\ j = 1, \ldots, J.$$

Given the values of the hazard function, $\boldsymbol{h} = \{h_1, h_2, \ldots, h_J\}$, the piecewise exponential pdf is

$$f(t \mid \boldsymbol{h}) = \begin{cases} h_j \exp\left\{-h_j(t - s_{j-1}) - \sum_{m=1}^{j-1} h_m(s_m - s_{m-1})\right\}, \\ \qquad \text{for } s_{j-1} < t \le s_j, j = 1, \ldots, J, \text{ and} \\ 0, \quad \text{for } t \notin [0, s_J). \end{cases} \tag{2}$$

Note that the parameters of the model are the hazard rates $\{h_1, h_2, \ldots, h_J\}$ that define the exponential rates within each bin. The likelihood can then be rewritten as

$$\begin{aligned} L(\boldsymbol{h} \mid \boldsymbol{T}) &= \prod_{k=1}^{n} f(t_k \mid \boldsymbol{h}) \\ &= \prod_{k=1}^{n} \prod_{j=1}^{J} \left[h_j \exp\left\{ - h_j(t_k - s_{j-1}) \right.\right. \\ &\qquad \left.\left. - \sum_{m=1}^{j-1} h_m(s_m - s_{m-1}) \right\} \right]^{I(s_{j-1} < t_k \le s_j)}, \end{aligned} \tag{3}$$

where $I(x)$ is an indicator function that equals 1 when x is true and 0 otherwise. Over conditions (c) and observers (i), our goal is to estimate the posteriors of $h_{ic,j}, j = 1, \ldots, J$, and thus estimate the posteriors of the hazard functions of a hierarchical piecewise exponential model.

The semiparametric approach represented by the piecewise exponential model is common in Bayesian survival analysis. The hazard rates $\boldsymbol{h}_{ic}$ are often modeled as a stochastic process (see Sinha & Dey, 1997, for a review). Without such a process, the hazard rates over the intervals $[s_{ic,j-1}, s_{ic,j}), j = 1, \ldots, J$ are independent. In particular, the hazard rates in adjacent intervals

Please cite this article in press as: Houpt, J. W., et al., Semiparametric Bayesian approaches to systems factorial technology. Journal of Mathematical Psychology (2016), http://dx.doi.org/10.1016/j.jmp.2016.02.008

Table 2
The behavior of the survivor function interaction contrast $SIC(t)$ for different cognitive architectures and stopping rules. The time t^* is a constant that is not necessarily the same for all models; it only marks the point at which the behavior of the function changes.

Stopping rule		
Architecture	Self-terminating (OR)	Exhaustive (AND)
Serial	$SIC(t) = 0$ for all t	$SIC(t) < 0$ for $t < t^*$ and $SIC(t) > 0$ for $t > t^*$
Parallel	$SIC(t) > 0$ for all t	$SIC(t) < 0$ for all t
Coactivation	$SIC(t) < 0$ for $t < t^*$ and $SIC(t) > 0$ for $t > t^*$	

are independent. This is problematic for continuous measurements such as RTs, for which we know the hazard rate varies smoothly over the real line.

Following Gamerman (1991), we introduce an autoregressive process of order 1 (AR(1)) over the elements of $\boldsymbol{h}_{ic}$. One way to view the AR(1) process over the rates is as a smoothing process that ensures the changes from one interval to the next are not too irregular. More importantly, the AR(1) process allows the estimation of the hazard rates in each interval to be informed by the rates in adjacent intervals.

To fully specify the model, note that an RT T_{ick} from Observer i in Condition c on Trial k is distributed as a piecewise exponential with hazard rates $\boldsymbol{h}_{ic} = \{h_{ic,1}, h_{ic,2}, h_{ic,3}, \ldots, h_{ic,J}\}$. Given $h_{ic,1}$, the hazard rates $h_{ic,j}$ for $j = 2, \ldots, J$ are defined as a stationary AR(1) process on the log scale, so

$$\ln h_{ic,j} = \mu_{ic} + \phi_{ic}\left(\ln h_{ic,j-1} - \mu_{ic}\right) + \epsilon_{ic,j}. \tag{4}$$

The parameter $\mu_{ic} \in \mathbb{R}$ is the mean of the AR(1) process, $\phi_{ic} \in \mathbb{R}$ is the autoregressive coefficient, and the $\epsilon_{ic,j}$s are normal errors with mean 0 and variance $\sigma^2_{h,ic}$.

We modeled the autoregressive parameter ϕ_{ic} as a rescaled inverse logit transformation of a parameter α_{ic}, so that

$$\phi_{ic} = 2\left[1 + \exp(-\alpha_{ic})\right]^{-1} - 1,$$

which restricts the autoregressive parameter ϕ_{ic} to the interval $(-1, 1)$, which in turn guarantees stationarity of the AR(1) process. Model stationarity and normality of the $\epsilon_{ic,j}$s imply that

$$\ln h_{ic,j} \mid \ln h_{ic,j-1} \sim \mathcal{N}\left[\mu_{ic} + \phi_{ic}(\ln h_{j-1} - \mu_{ic}), \sigma^2_{h,ic}\right]$$

for $j \geq 2$, and

$$\ln h_{ic,1} \sim \mathcal{N}\left[\mu_{ic}, \sigma^2_{h,ic}/(1 - \phi^2_{ic})\right].$$

Marginally, all the $h_{ic,j}$s have the same mean μ_{ic} and variance $\sigma^2_{h,ic}/(1 - \phi^2_{ic})$.

To construct the full hierarchical model over observers and conditions, we gave μ_{ic} and α_{ic} normal priors so

$$\mu_{ic} \sim \mathcal{N}(\mu_{\mu,i}, \sigma^2_{\mu,i}) \quad \text{and} \quad \alpha_{ic} \sim \mathcal{N}(\mu_{\alpha,i}, \sigma^2_{\alpha,i}),$$

where

$$\mu_{\mu,i} \sim \mathcal{N}(\mu_0, \sigma^2_{0,\mu}) \quad \text{and} \quad \mu_{\alpha,i} \sim \mathcal{N}(\alpha_0, \sigma^2_{0,\alpha}).$$

To complete the model, we gave the hyperparameters normal and gamma priors,[2] so

$$\mu_0, \alpha_0 \sim \mathcal{N}(0, 1),$$
$$\sigma^2_{\mu,i}, \sigma^2_{\alpha,i}, \sigma^2_{h,ic} \sim \Gamma(1, 1),$$

and

$$\sigma^2_{0,\mu}, \sigma^2_{0,\alpha} \sim \Gamma(1, 1).$$

[2] Because the hazard rates generally varied over a narrow range, depending on the unit of measurement for T_{ick}, hyperpriors with high precision resulted in priors for $\boldsymbol{h}$ with low precision.

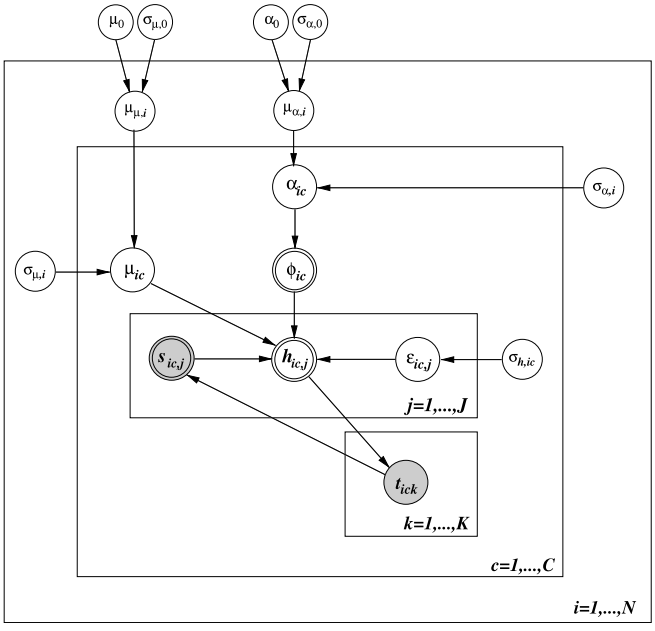

Fig. 2. The graphical model of the semiparametric hazard function estimate. The index j runs over the J bins in which the hazard rates are estimated, k represents trials, c represents experimental conditions and i represents individuals. Continuous parameters to be estimated are shown in single circles and continuous computed values in double circles. Both response times t_{ick} and bin boundaries s_{icj} are observable variables represented by shaded circles.

Fig. 2 shows a graphical representation of this model. With this model structure in place, it is important to note that the prior mean for $\boldsymbol{h}_{ic}$ is $\mathbf{1}$, or that the prior mean shape of the hazard is that of the exponential distribution with rate parameter 1. We present the prior distributions of $\boldsymbol{h}_{ic}$ as well as the prior distributions of the capacity coefficients and the SIC in Section 4.2.

It is an open question as to the best way to select the bin boundaries $s_{ic,j}$. There are three possibilities that we considered: (1) using fixed, equidistant points between $s_{ic,0} = 0$ and some reasonable choice for $s_{ic,J} > \max_k T_{ick}$; (2) using the observed quantiles of the samples $\boldsymbol{T}_{ic}$; and (3) incorporating the values of $\boldsymbol{s}_{ic}$ into the structure of the model and estimating their posteriors along with the posteriors of $\boldsymbol{h}_{ic}$.

Using fixed, equidistant points has much to recommend it. Unfortunately, for distributions with heavy tails, often observed for RT data, the value of $s_{ic,J} > \max_k T_{ick}$ is likely to be distant from most of the values in the sample. Slicing the range from 0 to $s_{ic,J}$ into J bins of equal width results in a potentially large number of bins in the tail of the distribution within which no observations were sampled. There seemed to be no practical way around this problem, even for small numbers of observers in small numbers of conditions.

We explored the inclusion of $\boldsymbol{s}_{ic}$ among the model parameters by defining a new parameter vector $\boldsymbol{p}_{ic} \in (0, 1)^J$ with a Dirichlet prior, and setting $s_{ic,q} = (1 + \delta) \max_k T_{ick} \sum_{m=1}^{q} p_{ic,m}$ for small values of δ. The increase in the dimensionality of the model

Please cite this article in press as: Houpt, J. W., et al., Semiparametric Bayesian approaches to systems factorial technology. Journal of Mathematical Psychology (2016), http://dx.doi.org/10.1016/j.jmp.2016.02.008

Table 3
The four models simulated to produce data for Study 1. Gamma distributions are parameterized as (shape, rate); log-normal distributions are parameterized as (mean, standard deviation), inverse-normal distributions are parameterized as (threshold, drift rate) with drift coefficient = 1.0, and exponential distributions are parameterized as (rate).

	Condition		
	$T_{u\cdot}$	$T_{\cdot v}$	T_{uv}
Model 1	$\Gamma(2, 1)$	$\Gamma(2, 1)$	$\min\{T_{u\cdot}, T_{\cdot v}\}$
Model 2	$\ln\mathcal{N}(0, 1)$	$\ln\mathcal{N}(0, 1)$	$T_{u\cdot} \times T_{\cdot v} \sim \ln\mathcal{N}(0, \sqrt{2})$
Model 3	$\mathcal{N}^{-1}(20.0, 1.5)$	$\mathcal{N}^{-1}(20.0, 1.5)$	$\mathcal{N}^{-1}(20.0, 3.0)$
Model 4	$\mathcal{E}(0.5)$	$\mathcal{E}(0.4)$	$\max\{T_{u\cdot}, T_{\cdot v}\}$

resulted in unstable estimates which led to an inability to fit hierarchical structures in a practical way.

The failure of these two methods left us with the quantile solution. For J hazard rates, we used the J/N quantiles of the sample which assured that there were N/J observations within each bin. This approach also has drawbacks, namely that the estimates of $\boldsymbol{h}_{ic}$ reflect in part the possible variance in the bin boundaries $\boldsymbol{s}_{ic}$, and that we are using the data twice, once to estimate the bins and then to estimate the hazard rates. Despite these drawbacks, the procedure worked well, and permitted us to obtain accurate posterior estimates of $\boldsymbol{h}_{ic}$ in reasonable amounts of time. In the next section, we demonstrate this accuracy by recovering the hazard rates of a number of different heavy-tailed distributions.

4. Simulation studies

We undertook simulation studies to examine the feasibility of using Bayesian methods to estimate the hazard functions, capacity coefficients and SICs. In particular, we looked at the influence of sample size on the accuracy of the posteriors for data generated from a number of models consistent with systems factorial technology. We first simulated data for examining system capacity by way of the capacity coefficient (three conditions corresponding to two single-stimulus conditions and one dual-stimulus condition), and then we simulated data for examining system architecture (four conditions corresponding to fast and slow processing in each of two channels).

4.1. Description of the studies

Recall that the systems factorial experimental design must present stimuli to two processing channels. The RT T_c in condition c will be determined by the processing times for each channel, and the precise relationship between processing and response time depends on the system architecture and capacity.

To investigate system capacity, a single stimulus may be presented to Channel 1 or Channel 2 alone, or single stimuli may be presented to both Channels 1 and 2 simultaneously. Recalling the notation defined in Section 2.1, let $T_{u\cdot}$ be the RT when a stimulus is presented to Channel 1 alone, $T_{\cdot v}$ be the RT when it is presented to Channel 2 alone, and T_{uv} be the RT when single stimuli are presented in each channel simultaneously. For the first simulation study, we simulated finishing times under each condition for four models.

Table 3 shows the models simulated in the first study. In the first model, $T_{u\cdot}$ and $T_{\cdot v}$ both followed gamma distributions (parameterized by shape and rate) and T_{uv} was modeled as the minimum of the two channels' finishing times (a parallel OR structure). In the second model $T_{u\cdot}$ and $T_{\cdot v}$ followed log-normal distributions (parameterized by mean and standard deviation) and T_{uv} was modeled as the product of the two channels' finishing times. This model describes T_{uv} as a sum of independent normal finishing times on the log scale. In the third model $T_{u\cdot}$ and $T_{\cdot v}$ followed inverse-normal distributions characterizing the first passage time of a diffusion process (written as $\mathcal{N}^{-1}$(threshold, drift rate); the diffusion coefficients, not shown in the table, equaled 1.0, 1.0 and $\sqrt{2}$ for $T_{u\cdot}$, $T_{\cdot v}$ and T_{uv}, respectively). The variable T_{uv} was also the first passage time of a diffusion process with drift rate equal to the sum of the drift rates for each single channel (coactivation). The fourth model assumed exponential finishing times (parameterized by rate) for $T_{u\cdot}$ and $T_{\cdot v}$, and T_{uv} was modeled as the maximum finishing time of the two channels (a parallel AND structure).

To investigate system architecture, the experimental design must present stimuli of varying intensities to each of the two channels. As the intensities vary, the stimuli must evoke faster or slower processing time in the channel to which they are presented. Recalling the notation defined in Section 2.2, let T_{ss} be the processing time when "slow" stimuli are presented to each channel, T_{sf} and T_{fs} be the processing times when a "slow" stimulus is presented to one channel and a "fast" stimulus is presented to the other, and T_{ff} be the processing time when "fast" stimuli are presented to each channel. For the second simulation study, we simulated finishing times under each condition for five models, each model representing an architecture as shown in Table 4.

The first model was a serial AND architecture, where finishing times in each channel followed gamma distributions. The finishing times for each channel were summed to give the overall RTs. The second model was a serial OR architecture, where the finishing time in one of the two channels, randomly chosen, determined the RT. The RTs in each of the four conditions were then distributed as mixtures of log-normals.

The third model was a parallel OR architecture, where finishing times in each channel followed a Weibull distribution (parameterized by shape and scale, with mean equal to scale $\times$ $\Gamma(1 + 1/\text{shape})$). The processing time for each condition was then the minimum of the finishing times for both channels. The fourth model was a parallel AND architecture, where finishing times in each channel followed a gamma distribution. The processing time for each condition was then the maximum of the finishing times for both channels.

The fifth model was a parallel coactivation model, where finishing times in each channel followed an inverse-normal distribution representing the first passage time of a diffusion process. The processing times in each condition were then modeled as inverse-normal distributions with drift rates equal to the sum of the drift rates in the two channels.

4.2. Procedure

From each model we generated 50, 100 or 500 samples of finishing times for each experimental condition (three conditions for the first study and four conditions for the second study). We subtracted the minimum from each sample and computed bin boundaries based on both the shifted sample deciles (10%, 20%, . . . , 100%) and the sample quintiles (20%, 40%, . . . , 100%). We found that using the deciles resulted in finer resolution and better estimates of the tail behavior of the hazard rates, even at the smallest sample size, so we present only results using deciles here.

It was necessary to shift the samples because the first bin boundary must start at 0. Samples arising from right-shifted distributions would otherwise have a very wide initial bin within which the observations were clustered close to the right boundary, a pattern that the piecewise-exponential distribution cannot accommodate. We fit both shifted and unshifted samples and found that the model fits much better to the shifted samples, so we present only these results here.

We used the simulated, shifted data and the decile bin boundaries to estimate the posterior hazard functions for each model and condition using the hierarchical structure shown in

Please cite this article in press as: Houpt, J. W., et al., Semiparametric Bayesian approaches to systems factorial technology. Journal of Mathematical Psychology (2016), http://dx.doi.org/10.1016/j.jmp.2016.02.008

Table 4
The five models simulated to produce data for Study 2. Parameterizations are the same as for Study 1, and in addition the Weibull distribution is parameterized as (shape, scale).

	Condition			
	T_{ss}	T_{sf}	T_{fs}	T_{ff}
Serial-AND	$\Gamma(4, 1.5) + \Gamma(4, 1.5)$	$\Gamma(4, 1.5) + \Gamma(1, 1.5)$	$\Gamma(1, 1.5) + \Gamma(4, 1.5)$	$\Gamma(1, 1.5) + \Gamma(1, 1.5)$
Serial-OR	$0.5 \ln \mathcal{N}(0.5, 1) + 0.5 \ln \mathcal{N}(1, 1)$	$0.5 \ln \mathcal{N}(.5, 1) + 0.5 \ln \mathcal{N}(-.5, 1)$	$0.5 \ln \mathcal{N}(-1.5, 1) + 0.5 \ln \mathcal{N}(1, 1)$	$0.5 \ln \mathcal{N}(-1.5, 1) + 0.5 \ln \mathcal{N}(-0.5, 1)$
Parallel-OR	$\min\{\mathcal{W}(3, 6), \mathcal{W}(3, 6)\}$	$\min\{\mathcal{W}(3, 6), \mathcal{W}(3, 3)\}$	$\min\{\mathcal{W}(3, 3), \mathcal{W}(3, 6)\}$	$\min\{\mathcal{W}(3, 3), \mathcal{W}(3, 3)\}$
Parallel-AND	$\max\{\Gamma(3, 0.5), \Gamma(3, 0.5)\}$	$\max\{\Gamma(3, 0.5), \Gamma(3, 1.5)\}$	$\max\{\Gamma(3, 1.5), \Gamma(3, 0.5)\}$	$\max\{\Gamma(3, 1.5), \Gamma(3, 1.5)\}$
Coactive	$\mathcal{N}^{-1}(20, 1.5 + 1.5)$	$\mathcal{N}^{-1}(20, 1.5 + 3.0)$	$\mathcal{N}^{-1}(20, 3.0 + 1.5)$	$\mathcal{N}^{-1}(20, 3.0 + 3.0)$

Fig. 2. We fit the model in Gelman et al.'s Stan (Stan Development Team, 2014), which uses Homan and Gelman's (2014) No U-Turn Sampler (NUTS), a modification of Hamiltonian Monte Carlo sampling.

For each sample size we generated three chains of 10,000 samples following a burn-in period of 1000 samples. The NUTS sampler produced very stable, well-mixing chains at all levels of the hierarchy. For larger sample sizes, we needed to increase the value of the delta parameter, one of the adaptive control parameters responsible for acceptance rate and step size, from its default value of 0.8 to 0.99 to prevent the chains from getting stuck in low-probability regions of the parameter space. We therefore used values of delta equal to 0.99 for all sample sizes. We evaluated convergence of the chains by visual inspection of the trace plots and by confirming that all values of the Gelman–Rubin statistic $\hat{R}$ (Gelman & Rubin, 1992) for all parameters at all levels of the hierarchy were very close to 1, and the effective sample sizes for all parameters were very large.

Fig. 3 shows the thinned trace plots of the log posterior likelihood at all three sample sizes for the first simulation study with the burn-in periods removed. (Trace plots for the second simulation study were similar.) The 10,000 samples from each chain were concatenated to produce a single chain of length 30,000, and then for purposes of this figure thinned to every 50th sample. The log posterior likelihood in this figure is the likelihood computed over the simultaneous fit to all data models and conditions. Convergence of the chains is demonstrated by the lack of change in the mean and variance over samples and the lack of strong autocorrelation (which is also demonstrated in the autocorrelation functions, which are not shown). Similar convergence is reflected in the trace plots for all parameters (which are not shown).

We also verified that the posterior estimates were sensitive to the data by contrasting the posteriors of all parameters to their priors. The priors were quite diffuse, while the posteriors were narrow and usually shifted far from the prior mean. Even with sample sizes of 50 (5 observations per bin) there was considerable learning for all models. (Contrasts between the priors and posteriors for the second simulation study showed similar patterns.)

Fig. 4 shows boxplots of the prior distributions of the hazard rates for each of the 10 bins, together with the posterior distributions for each data model and condition for sample size $N = 50$. (The placement of the bin boundaries, generated by computing the quintiles of data generated by simulating draws from the hierarchical model shown in Fig. 2 and described in Section 3, had no effect on the shape of the prior distributions.) Fig. 6 shows boxplots of the prior distributions of the parameters μ_{ic} and ϕ_{ic} for each condition, together with the posterior distributions for each model and condition for sample size $N = 50$. As sample size increased, the location and variance of the posteriors for μ_{ic} changed little. The posteriors for the hazard rates also shifted little as sample size increased (except in the tail), although the variance decreased greatly. Similarly, the location of the posteriors for ϕ_{ic} shifted toward one and the variance decreased greatly. The tendency for ϕ_{ic} to increase with sample size suggests a lack of fit of the model, which we will address in Section 6.

We note that the prior distributions over the CC and the SIC, constructed as transformations of samples from the prior distributions of the hazard rates, are well defined and reasonably behaved (see Fig. 5). This reassures us that we can transform samples from the posterior distributions of the hazard rates into samples from the posterior distributions of the CCs and SIC without worrying about the prior distributions of the CCs and SIC being in some way unreasonable. Although these priors may appear to be concentrated within narrow regions, it is important to remember that the values of the CCs and the SIC are constrained within these regions. The SIC, for example, can only take values between −2 and 2.

In addition to the full hierarchical model shown in Fig. 2, only for the first simulation study we also fit a "null" model in which the hazard rate for T_{uv} was constrained to be equal to the sum of the hazard rates for $T_{u\cdot}$ and $T_{\cdot v}$. This null model has capacity coefficient everywhere equal to 1, and is consistent with parallel OR processing (the first minimum gamma model described above for the first simulation study). The chains of samples from the posteriors converged just as well as for the full model and the Gelman–Rubin statistic $\hat{R}$ was just as close to one.

The Stan code used to fit the model and the simulated data can be downloaded from https://github.com/trishvz/BayesHazard.git.

4.3. Results

The procedure yielded 30,000 posterior draws of the hazard rates estimated for each condition in each data model for a given sample size. With these posterior draws in hand, we can estimate the posteriors of the functions shown in Table 1 as well as of the capacity coefficients and SICs.

The piecewise exponential density function is given in Eq. (2), and noting that

$$F(t \mid \boldsymbol{h}) = \int_0^t f(u \mid \boldsymbol{h}) du = 1 - \exp\left\{-h_j(t - s_{j-1}) - \sum_{m=1}^{j-1} h_m(s_m - s_{m-1})\right\}, \quad \text{for } s_{j-1} < t < s_j,\ j = 1, \ldots, J,$$

for J bins, we can compute, most importantly, estimates of the posterior of the survivor function $S(t \mid \boldsymbol{h}) = 1 - F(t \mid \boldsymbol{h})$ and the integrated hazard function $H(t \mid \boldsymbol{h}) = -\ln S(t \mid \boldsymbol{h})$, leading directly to the posterior estimates of the capacity coefficient and the SIC. The estimates of the posteriors of the functions shown in Table 1 and of the capacity coefficients and SICs are obtained by transforming the samples of the hazard functions through the piecewise exponential density and distribution functions.

We first examined the estimates of the hazard functions themselves. Then we looked at the accuracy of the estimates of the capacity coefficient and the SIC (computed after shifting the fitted distributions to their appropriate locations by adding back the minima). Finally we explored the goodness of fit of the model to the data from the first simulation study (the three-condition case), contrasting the full model to a null model in which capacity is unlimited. There was no equivalent null model for the second simulation study.

Please cite this article in press as: Houpt, J. W., et al., Semiparametric Bayesian approaches to systems factorial technology. Journal of Mathematical Psychology (2016), http://dx.doi.org/10.1016/j.jmp.2016.02.008

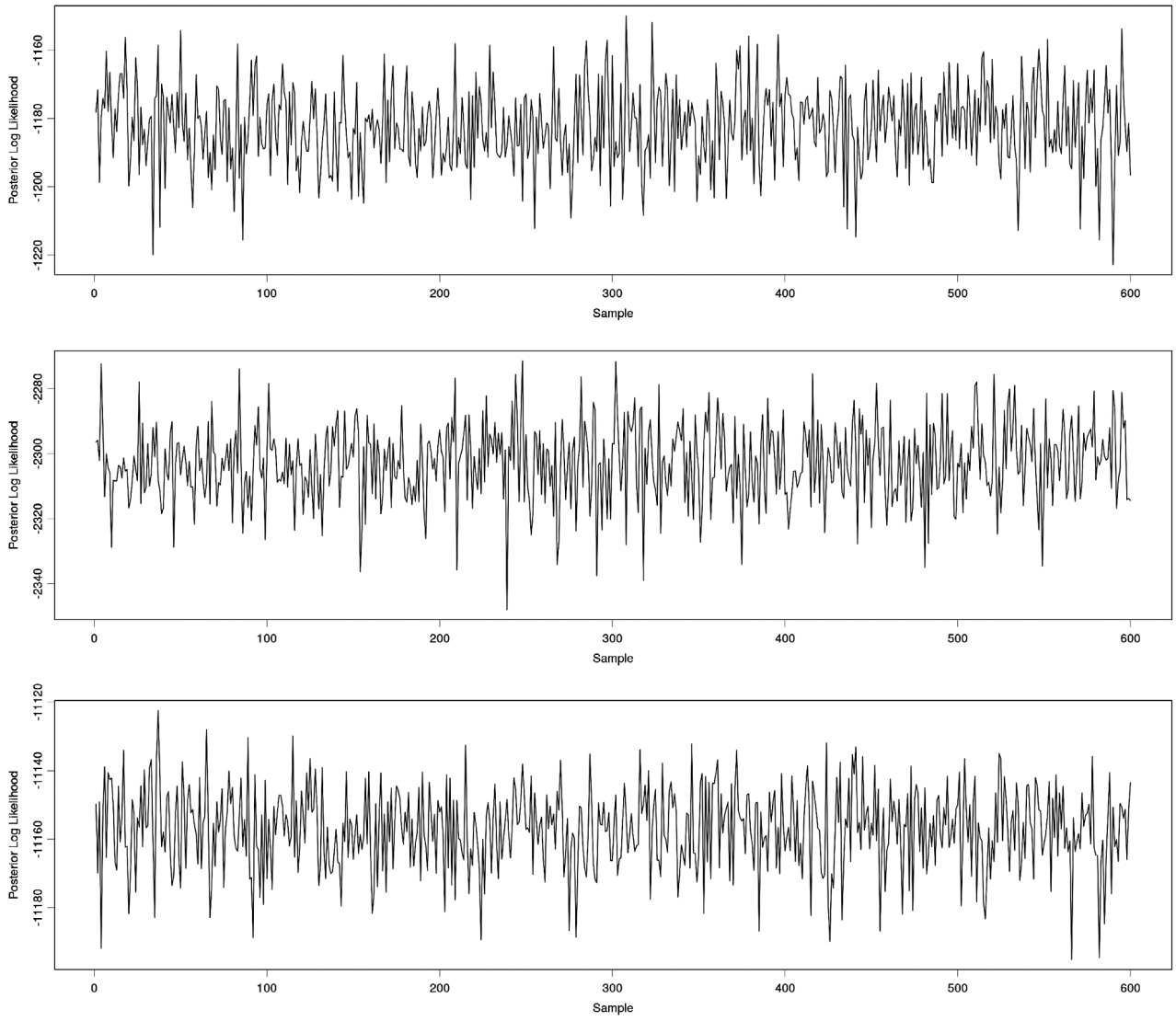

Fig. 3. Thinned (1/50) trace plots of log posterior likelihood of the complete model to the data generated for the first simulation study for sample sizes of $N = 50$ (top panel), $N = 100$ (middle panel) and $N = 500$ (bottom panel). The decreasing values in the posterior log likelihood over panels are due to the increase in sample size.

4.3.1. The posterior hazard functions

For each posterior draw of $\boldsymbol{h}_{ic}$, we can plot the sampled values as the heights of the piecewise exponential hazard function changing at the bin boundaries $\boldsymbol{s}_{ic}$ (see Fig. 1). In the following figures, we plotted these line segments in gray with opacity, so the darkness of the gray shows the density of the posterior distribution. We plotted a subset (every 100th sample) of the obtained draws. We focus only on the posterior hazard functions obtained for the first simulation study. The results for the second study were similar.

Fig. 7 shows a subset of the posterior samples for all models and conditions for sample size $N = 50$ in gray. Superimposed on the posterior distributions are the true hazard rates (solid lines), and the posterior medians and 95% (pointwise) credible sets (circles and error bars extending from the 2.5 to the 97.5% posterior quantiles). Even for this small sample size, the posteriors are quite accurate, containing the true hazard rates within the credible sets.

Fig. 8 shows the effect of sample size ($N = 50$, 100 and 500 in the left, middle and right panels, respectively) on the posterior estimates for Model 2, Condition 1. Model 2 was perhaps the least well-fitting of those shown in Fig. 7, but even for this model the sample size increases, the posterior variance of the estimates decreases dramatically around the true hazard rate. However, even at the largest sample size, there is a systematic overestimate of the hazard rate in the last bin. This overestimate could be due to the AR(1) process defined over $\boldsymbol{h}_{ic}$, an issue that we revisit in Section 6.

The dotted lines in Figs. 7 and 8 are the 95% (pointwise) confidence intervals around the median bootstrapped kernel density estimates of the hazard function, a procedure advocated by Van Zandt (2002b) for evaluating the variability of the hazard function and other functionals derived from it. The procedure is as follows: (1) Sample n observations with replacement from the data set of size n to obtain a bootstrap sample; (2) Compute a kernel density estimate of the hazard function at a fixed number of points for the bootstrapped sample; (3) Repeat steps 1 and 2 a large number of times; (4) For each fixed point at which the hazard function was estimated, compute the 2.5%, 50% and 97.5% quantiles over all the repetitions performed in step 3 (shown as dashed lines). Keeping in mind that the variance of the hazard function indicated by the error bars is the variance of the posterior hazard rate while the variance indicated by the dotted lines is the variance of the kernel density hazard estimate, two very different things, the estimate based on the posterior median is considerably more accurate than the bootstrapped estimate, especially in the critical right tails of the distribution.

The bootstrapped medians show undulations over the support of the distribution that depend on the value of the bandwidth

Please cite this article in press as: Houpt, J. W., et al., Semiparametric Bayesian approaches to systems factorial technology. Journal of Mathematical Psychology (2016), http://dx.doi.org/10.1016/j.jmp.2016.02.008

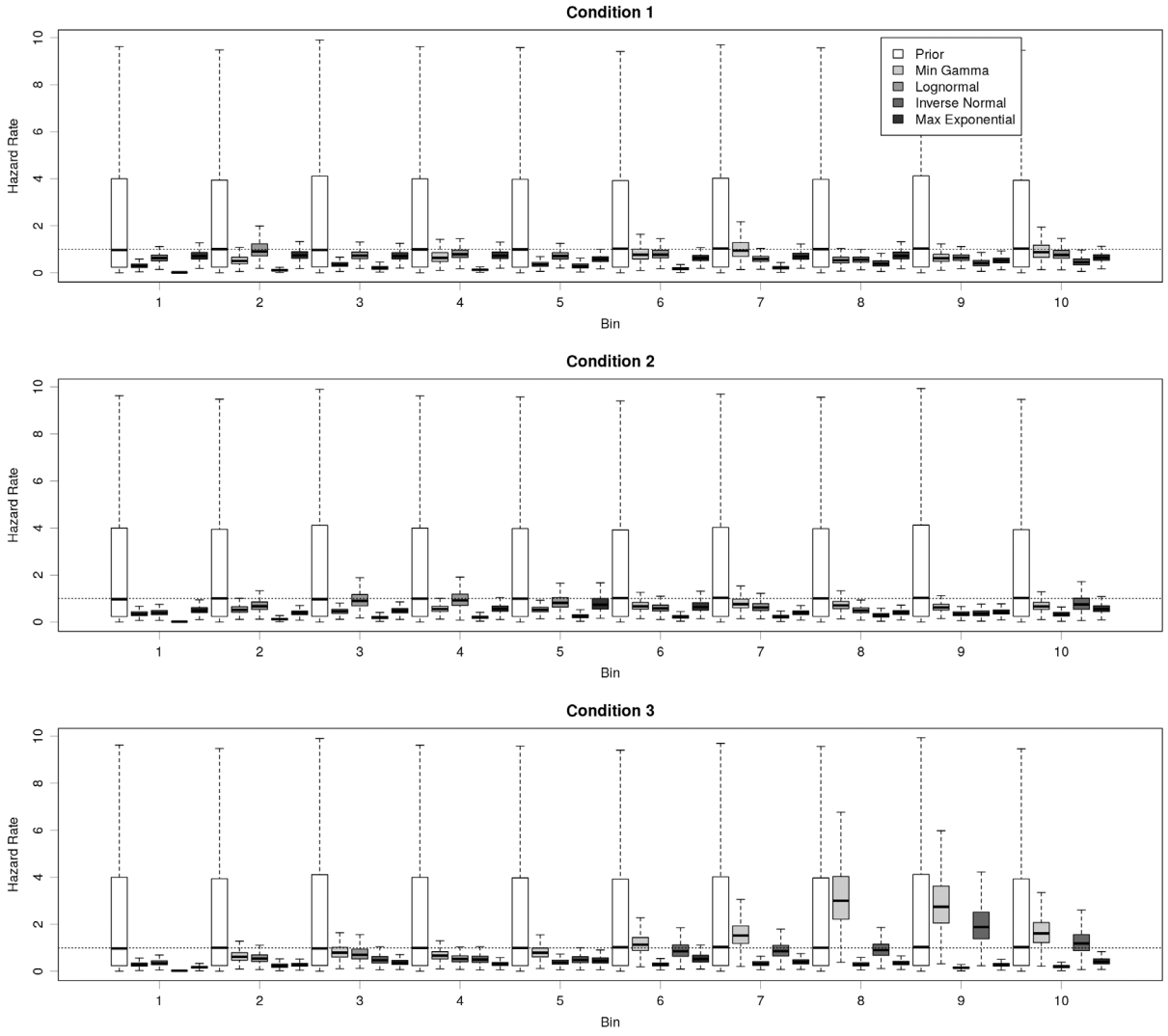

Fig. 4. Prior vs. posterior distributions of the hazard rates across bins for each condition ($T_{u\cdot}$, $T_{\cdot v}$ and T_{uv} are shown in the top, middle and bottom panels, respectively), for sample size $N = 50$. Each model is identified by grayscale color and the models are ordered along the x-axis as follows: the minimum gamma (lightest gray), log-normal, inverse-normal, and the maximum exponential (darkest gray). The horizontal dotted line is equal to one, the prior mean.

parameter used to smooth the kernel density estimates. While frequently very close to the posterior median, these bootstrapped estimates are difficult to interpret, and their accuracy does not improve as dramatically as that of the posterior median with increases in sample size. We therefore recommend that, for the estimation of the hazard function, the procedures recommended by Van Zandt (2002b) be abandoned in favor of this Bayesian approach.

4.3.2. The capacity coefficient

Before looking at the capacity coefficients $C_{\text{OR}}(t)$ and $C_{\text{AND}}(t)$, we computed the posterior distributions of $F(t \mid \boldsymbol{h}_{ic})$ and $H(t \mid \boldsymbol{h}_{ic})$ for each model and condition. These are the functions from which the two capacity coefficients are derived (Eq. (1)). Figs. 9 and 10 show the posterior distributions of the cdfs $F(t \mid \boldsymbol{h}_{ic})$ and integrated hazard $H(t \mid \boldsymbol{h}_{ic})$, respectively, for sample size $N = 50$ for each model i (rows) and condition c (columns). A subset of the posterior samples is shown in gray, and superimposed on the posteriors are the posterior median and 95% credible interval as dotted lines. The solid line indicates the true values of the cdf or integrated hazard for each model and condition.

Figs. 9 and 10 show good recovery of the true cdfs and integrated hazards by the posterior medians. For most models and conditions the true function is contained within the 95% credible interval even for $N = 50$. As sample size increases, the credible interval shrinks around the true function value.

We used the posteriors of $F(t \mid \boldsymbol{h}_{ic})$ and $H(t \mid \boldsymbol{h}_{ic})$ to estimate the posteriors of $C_{\text{OR}}(t \mid \boldsymbol{h}_{i,1}, \boldsymbol{h}_{i,2}, \boldsymbol{h}_{i,3})$ and $C_{\text{AND}}(t \mid \boldsymbol{h}_{i,1}, \boldsymbol{h}_{i,2}, \boldsymbol{h}_{i,3})$ (see Eq. (1)). Fig. 11 shows the estimated posteriors of the capacity coefficients for all models with sample size $N = 50$ and Fig. 12 shows the estimated posteriors of the capacity coefficients for Model 2 for sample sizes $N = 50,\ 100$and 500. Recall that Model 1 (top row of Fig. 11) is the parallel OR model and that Model 4 (bottom row of Fig. 11) is the parallel AND model. Therefore, Model 1 should have a flat OR capacity coefficient equal to 1.0 and Model 4 should have a flat AND capacity coefficient equal to 1.0.

These two figures show that the smaller sample size $N = 50$ does not provide posterior estimates of the capacity coefficients that are as accurate as we might have hoped, given the good accuracy of the hazard rates observed for that sample size. For most models, the values of the true capacity coefficient are contained within the 95% credible intervals, with some deviations. As sample size increases, however, the accuracy of the posterior estimates improves. One point to remember is that the capacity coefficient is diagnostic by being either less than, greater than, or equal to 1. Even in situations where some misfit is evident, and even for the

Please cite this article in press as: Houpt, J. W., et al., Semiparametric Bayesian approaches to systems factorial technology. Journal of Mathematical Psychology (2016), http://dx.doi.org/10.1016/j.jmp.2016.02.008

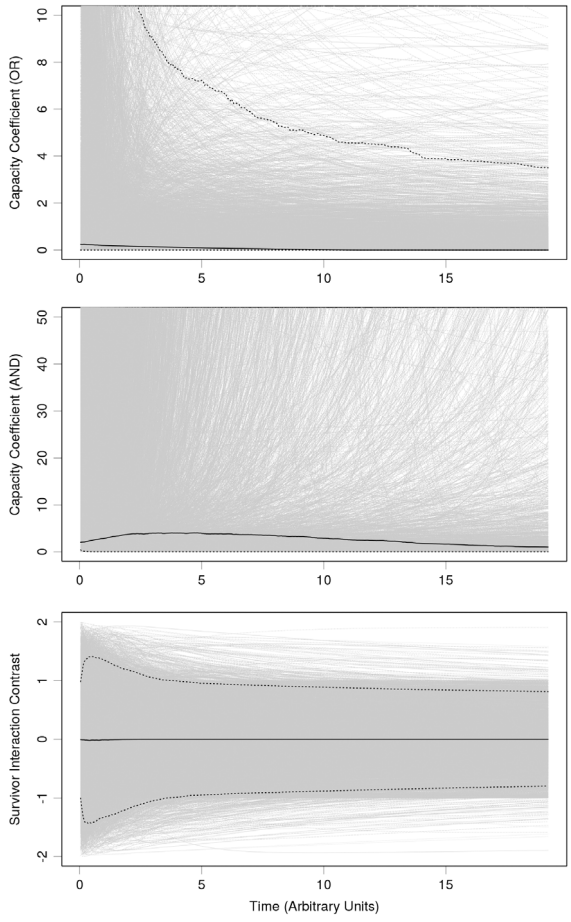

Fig. 5. Prior distributions of the AND and OR capacity coefficients (top two panels) and the survivor function interaction contrast (SIC; bottom panel). Each sample from the prior is shown as a gray line, and the prior median and 95% credible intervals are shown as solid black and dotted black lines, respectively.

smaller sample sizes, this diagnostic is appropriately preserved for all the models. The most difficult models to fit are those for which the capacity coefficient is everywhere equal to one (e.g., Models 1 and 4). For these models, or for data that indicate that the capacity coefficient may be equal to one, a large number of observations will be required to conclusively demonstrate unlimited capacity processing.

4.4. Model fit

To evaluate model fit, we examined the posterior predictive distributions for the first study and compared them to the distributions from which the simulated data were obtained. In addition, for the first study, we fit a null model and computed the Watanabe–Akaike Information Criterion (WAIC) to contrast the null and full models.

4.4.1. Posterior predictive distributions

For each sample obtained from the posterior distributions of $\boldsymbol{h}_{ic}$, we drew a finishing time for each model i and condition c. With 30,000 posterior samples, this generated 30,000 draws from the posterior predictive distributions of $\boldsymbol{T}_{ic}$. We constructed quantile–quantile plots of the draws from the posterior predictive distribution against the quantiles of the distribution from which the data were sampled. While overall the model is able to fit the data well, for $N = 50$, there were some clear misfits in the tails of the distributions (see Fig. 13). Fig. 14 shows the influence of sample size on the quantile–quantile plots for Model 2, Condition 1. As sample size increased, the fits generally improved, although some problems remained.

The posterior predictive distributions suggest that the model, which is simultaneously fitting data generated from 12 different distributions, is overall doing a very good job. These results were similar to those obtained for the second study. Depending on how an acceptable level of model fit is determined, a sample size as large as $N = 500$ may be required. Methods for improving model fits are discussed in Section 6.

Table 5
WAIC for the unrestricted hazard model (Full) and the null model (Null; unlimited capacity) for the simulated data.

	50 samples		100 samples		500 samples	
	Full	Null	Full	Null	Full	Null
Model 1	**423**	451	**833**	846	**4 242**	4 255
Model 2	**505**	596	**943**	1041	**4 809**	5 333
Model 3	**671**	688	**1378**	1428	**6 998**	7 178
Model 4	**527**	602	**1162**	1272	**5 699**	6 153
Group	**2126**	2336	**4316**	4588	**21 748**	22 919

4.4.2. Model comparisons

To contrast the full and null models we computed the Watanabe–Akaike Information Criterion (WAIC Watanabe, 2010). Like the deviance information criterion (DIC), the WAIC is based on an estimation of the effective number of parameters rather than a simple count of the number of parameters and is thus more appropriate for hierarchical models than the Bayesian information criterion (BIC) or the Akaike information criterion (AIC). In addition, the WAIC is more stable in practice (Gelman, Hwang, & Vehtari, 2014) than these other criteria. Because the data from each model were embedded in a more complex hierarchy, we computed; a partial WAIC based on the partial likelihoods for each model separately. The results are shown in Table 5.

Fits of the null model restricted the hazard rates in Condition 3 to equal the sum of the hazard rates in Conditions 1 and 2, so

$$\boldsymbol{h}_{i,3} = \boldsymbol{h}_{i,1} + \boldsymbol{h}_{i,2}.$$

Recall that only Model 1, a race between channels operating in parallel, is consistent with this null model. Therefore, information criterion comparison statistics should be smaller for the null model for Model 1, but larger for the null model for Models 2, 3 and 4. However, the WAIC was lower for the full, unrestricted model for all generating models and at the group level.

To test whether the model misclassified the data from Model 1 due to the influence on the posterior (shrinkage) from the other models, which did not conform to the null model, we ran a follow-up simulation study in which all four data models conformed to the null model assumption. Even in this case, the WAIC indicated better performance by the full model. This result indicates that the model comparison between the full and null model may not be sufficient for inferring whether the generating process conformed to the baseline unlimited capacity model or at least that some modification of the WAIC may be appropriate. However, as sample size increased, the advantage of the full model decreased for Model 1 but increased for all the other models, suggesting in addition that a larger sample size might be required to accurately distinguish between limited and unlimited capacity models on the basis of the WAIC.

Considering the entire hierarchy, values of the log posterior likelihood were smaller for the null model (medians −1174.6, −2336.6 and −10,941.0 for sample sizes 50, 100 and 500, respectively) than for the full model (medians −1136.4, −2233.6 and −11,601.0 for sample sizes 50, 100 and 500, respectively), consistent with the fact that the hierarchy was fit to three models that were inconsistent with the null model. In addition, visual

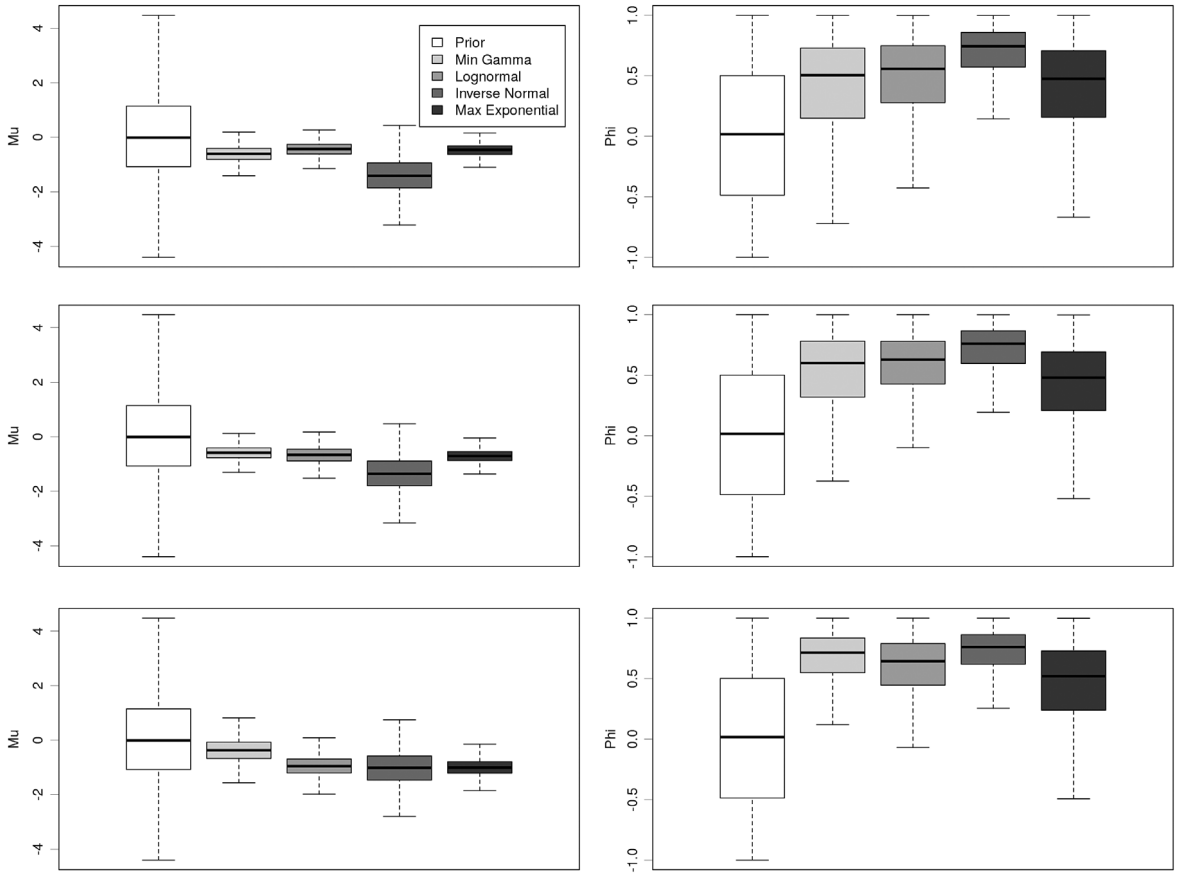

Fig. 6. Prior vs. posterior distributions of μ_{ic} (left panels) and ϕ_{ic} (right panels) for each condition ($T_{u\cdot}$, $T_{\cdot v}$ and T_{uv} are shown in the top, middle and bottom rows, respectively) for sample size $N = 50$. Each model is identified by grayscale color and the models are ordered along the x-axis as follows: the minimum gamma (lightest gray), log-normal, inverse-normal, and the maximum exponential (darkest gray).

inspection of the posterior median hazard rate obtained by fits of the null model showed large, systematic differences from the true hazard rate that increased with sample size for all models except the minimum gamma.

4.5. Survivor function interaction contrast

Following the same procedures as before, we used the posterior samples of $\boldsymbol{h}_{ic}$ for each data model i to compute the posteriors of the survivor functions, from which we then computed the posteriors of the SICs. Figs. 15 and 16 show the results.

Each of the five models selected for the second simulation study represented an architecture that produces a different behavior of the SIC (see Table 2). Taking Model 5, the coactive model, as an example, coactive models should show SICs with an initial negative trajectory moving to positive, a crossing point, and the positive trajectory moving to negative with a final asymptote at zero. The bottom panel in Fig. 15 shows the posterior of the SIC for this model for $N = 50$ and, despite some misfits in the tails of the four distributions contributing to the SIC, the posterior median very accurately captures this behavior even for small sample sizes.

To summarize the posterior distribution over SIC shapes, we use the maximum and minimum values of the SIC according to the test derived by Houpt and Townsend (2010). Finite data samples from a set of distributions with a flat SIC will still have some negative and positive values of the SIC, so it is not appropriate to compare the posteriors to a point null hypothesis. Instead, we compare the posteriors to distribution implied by a flat SIC, which was derived in Houpt and Townsend (2010).

Table 6
Posterior probabilities that the maximum (D^+) and minimum values (D^-) of each model's SIC have a larger magnitude than predicted by a flat SIC. Values in bold indicate those which are predicted to be larger than the null based on the generating model.

	$P(X > D^+_{\text{null}})$			$P(X < D^-_{\text{null}})$		
	50	100	500	50	100	500
Serial-AND	**0.9285**	**0.9981**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
Serial-OR	0.8605	0.7060	0.9072	0.3665	0.5375	0.4376
Parallel-OR	**0.9994**	**1.0000**	**1.0000**	0.8245	0.5314	0.2748
Parallel-AND	0.2020	0.4271	0.0473	**1.0000**	**1.0000**	**1.0000**
Coactive	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**

Table 6 gives the probabilities, for each model and sample size, that the posterior maximum and minimum SIC values were of a larger magnitude than that which would be predicted by a flat SIC. The probabilities in bold font are those that are predicted to be larger, and consequently, those probabilities should be very high (greater than 0.50). Probabilities that are not in bold font should be smaller (around 0.50). Without specifying upper and lower criteria around 0.5, or a region of practical equivalence (ROPE), the posteriors reflect the expected patterns for the data generated by the models. The higher probabilities associated with D^+ for the Serial-OR data and (possibly) those associated with D^- for the Parallel-OR data, probabilities that do not steadily decrease or approach 0.5 with increasing sample size, indicate that different model structures may be more or less problematic when using the posterior maxima and minima to perform model evaluations. We will revisit this issue again when we fit the piecewise exponential model to real data in the next section.

Please cite this article in press as: Houpt, J. W., et al., Semiparametric Bayesian approaches to systems factorial technology. Journal of Mathematical Psychology (2016), http://dx.doi.org/10.1016/j.jmp.2016.02.008

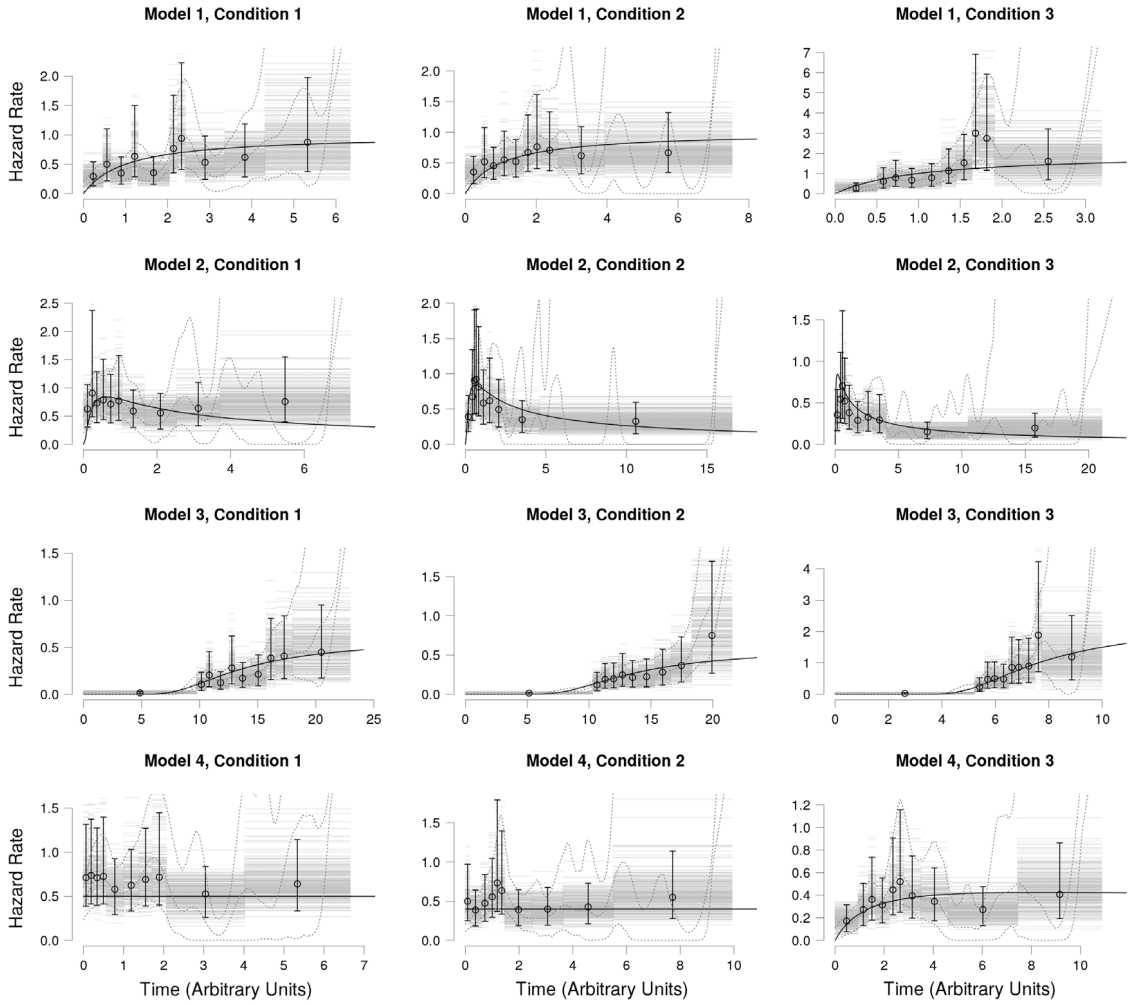

Fig. 7. Posterior estimates of the hazard functions for the three conditions for four models with sample size equal to $N = 50$. The posterior medians and 95% credible intervals for each bin of the fitted piecewise exponential hazard are shown as circles with error bars. The solid line is the true hazard function given the model that generated the data. The gray curves show a subset of the sampled posteriors. The dotted curves show the median estimate of the hazard function obtained by bootstrap sampling, together with the estimated (pointwise) 95% confidence interval.

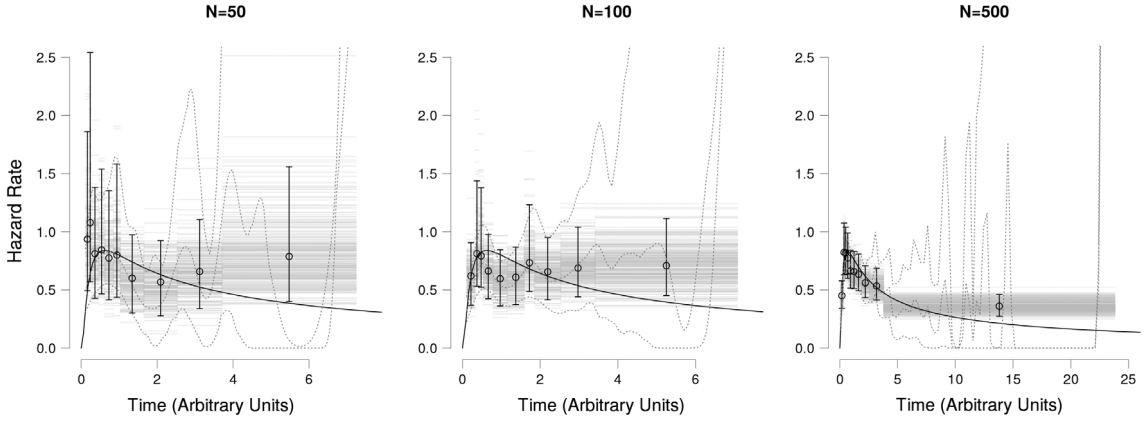

Fig. 8. Effects of sample size (left panel $N = 50$; middle panel $N = 100$; right panel $N = 500$) on the posterior estimates of the hazard functions for Model 2, Condition 1. The posterior medians and 95% credible intervals for each bin of the fitted piecewise exponential hazard are shown as circles with error bars. The solid line is the true hazard function given the model that simulated the data. The gray curves show a subset of the sampled posteriors. The dotted curves show the median estimate of the hazard function obtained by bootstrap sampling, together with the estimated (pointwise) 95% confidence interval.

5. Application to a published data set

In this section we apply the semiparametric Bayesian analysis to the dot detection data reported in Eidels et al. (2015) and included in the R package "sft" (Houpt, Blaha, McIntire, Havig, & Townsend, 2013). In this experiment, nine observers were asked to detect a dot presented on a computer monitor by pressing a key. On each trial, a dot appeared in one of two locations, above or below the midline of the monitor, both positions, or neither. Observers performed under two blocked conditions. In one

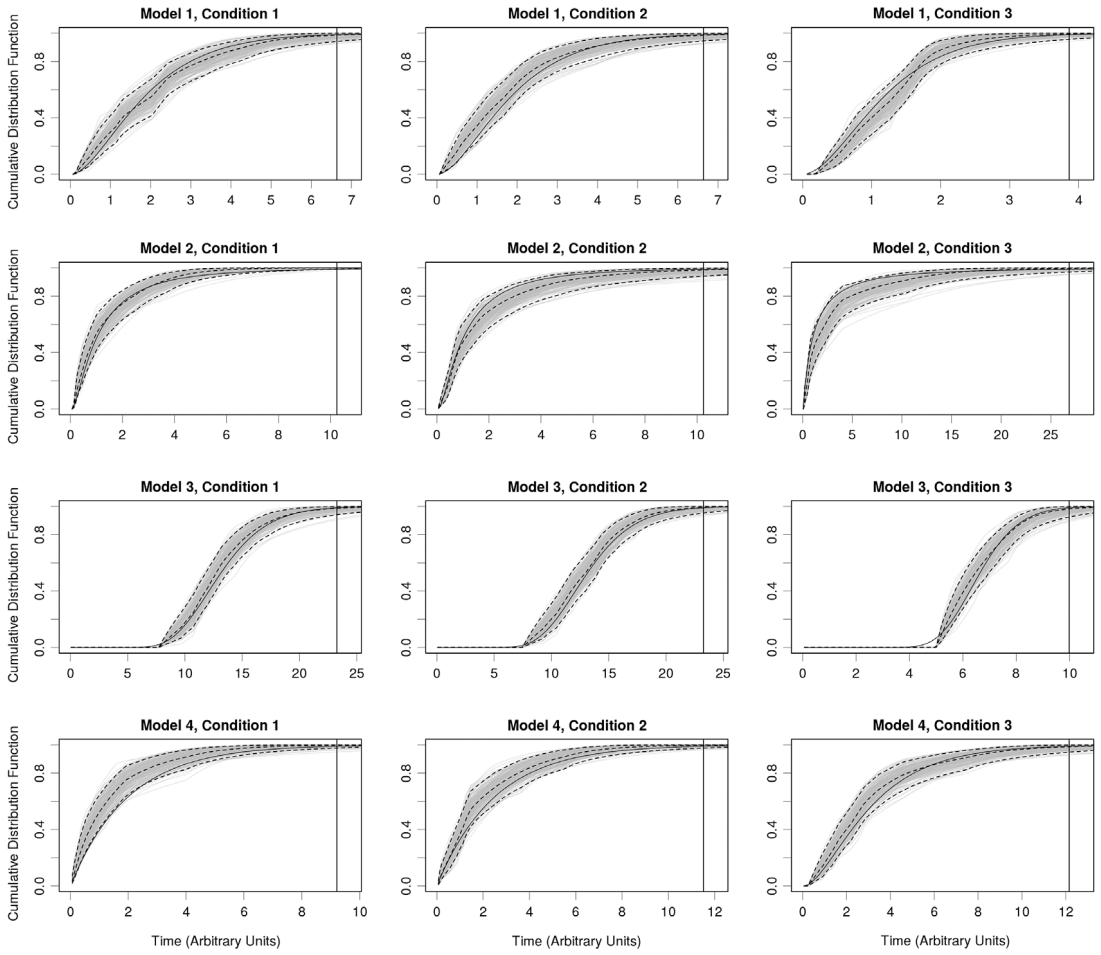

Fig. 9. Posterior estimates of the cumulative distribution function $F(t \mid \boldsymbol{h}_{ic})$ for the three conditions c (columns) for four models i (rows) with sample size equal to $N = 50$. A subset of the samples from the posterior is shown in gray, and the posterior medians and 95% credible intervals are shown as dashed lines. The solid line is the true cumulative distribution function given the model that generated the data. The vertical lines mark the 99% quantiles of the distributions.

Table 7
WAIC for the unrestricted hazard model and the null model (unlimited capacity) for the Dots OR Task Data. The lowest WAIC is bolded, showing that, for all observers except Observer 3 and for the group overall, the full model had a lower WAIC indicating better performance than the null model.

Observer	Full	Null
1	**26,126**	26,230
2	**26,488**	26,540
3	26,496	**26,494**
4	**29,095**	29,176
5	**29,451**	29,512
6	**27,516**	27,575
7	**28,376**	28,552
8	**27,402**	27,652
9	**28,259**	28,463
Group	**249,208**	250,194

condition, observers were asked to indicate whether any dot was present (the "OR" task). In the other condition, they were asked to indicate whether both dots were present (the "AND" task). Dots were presented on the screen at one of two levels of contrast leading to two levels of detection difficulty. When both dots were presented, observers saw each 2×2 factorial combination of high and low contrast in both locations with equal probability. We will focus only on data from the OR task because the analyses are analogous for the AND task.

To analyze the capacity coefficient, we estimated the cumulative hazard functions when both dots were present, when only the upper dot was present and when only the lower dot was present—the conditions analogous to the three conditions of simulation Study 1, with sample sizes of approximately 800 RTs per condition. The posterior distributions of the capacity coefficients for two arbitrarily-selected observers are shown in the left panels of Fig. 17. As for all the observers, the posterior medians of these two observers' capacity coefficients, being almost entirely less than one, demonstrate limited capacity processing.

To further compare whether the observers' data were of unlimited capacity, we fit a model in which the cumulative hazard of the response times when both dots were present was fixed at the sum of the cumulative hazard functions from each of the single-dot-present conditions, as in simulation Study 1. We then compared the unrestricted model and the restricted model using both the WAIC calculated on the full data set and separately for the WAIC calculated only on each individual's data.

In agreement with the conclusions based on the posterior estimates of the capacity function, all observers showed significantly limited capacity using the null hypothesis test from Houpt and Townsend (2012). The WAIC results, reported in Table 7, lead to the same conclusion, save those of Observer 3 whose results are consistent with unlimited capacity processing. However, using what we learned in the simulation study, we cannot yet support the limited-capacity conclusion based on the WAIC alone, because it may not have indicated support for the null model even if observers were actually performing with unlimited capacity.

To analyze the survivor function interaction contrast, we estimated the survivor functions for response times for each of

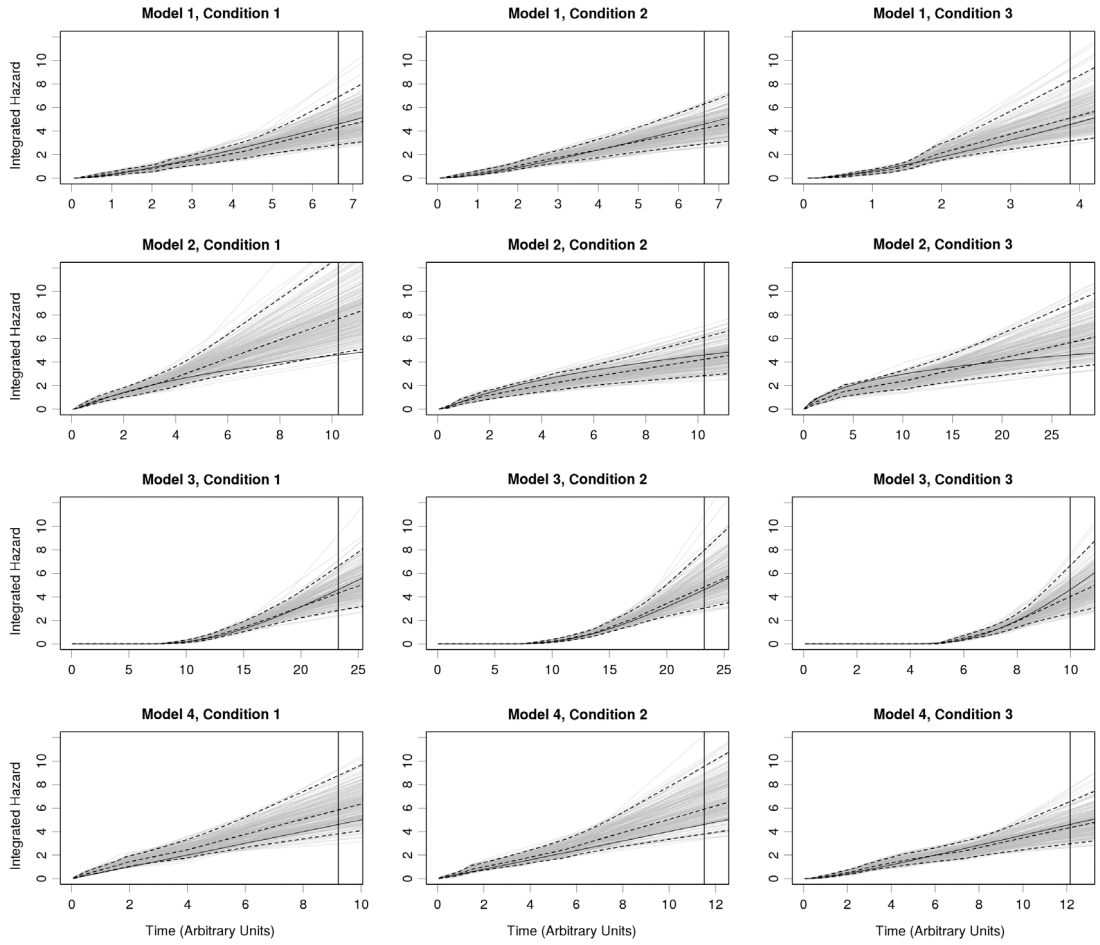

Fig. 10. Posterior estimates of the integrated hazard function $H(t \mid \boldsymbol{h}_{ic})$ for the three conditions c (columns) for four models i (rows) with sample size equal to $N = 50$. A subset of the samples from the posterior is shown in gray, and the posterior medians and 95% credible intervals are shown as dashed lines. The solid line is the true integrated hazard function given the model that generated the data. The vertical lines mark the 99% quantiles of the distributions.

the 2×2 contrast level by location conditions, each of which had approximately 200 RTs. The posterior distributions of the SIC from the same two observers as before are shown in the right panels of Fig. 17. The shapes of the SICs for these and the other observers are consistent with Parallel-OR processing.

In addition to calculating the full posterior over SIC shapes, we also calculated the posterior distributions of the maximum (D^+) and minimum (D^-) values of each observer's SIC across time. The posteriors of the maximum and minimum values of the SIC assuming a flat SIC can be used as a basis for evaluating the Parallel-OR conclusion as we did for the simulation study (Table 6). The results are given in Table 8 for the dots-OR data.

For all observers, the probability that the maximum of their SIC was larger than the maximum predicted by a flat SIC was high, larger than 0.999 for all but two observers. Observer 2 had the weakest evidence for a larger maximum SIC than that predicted by the null, and this observer also did not have a significant D^+ using the Houpt and Townsend (2010) test. All other observers' D^+s were significant using the null hypothesis test. Observers 4, 5 and 9 have the strongest evidence for a smaller minimum SIC than that predicted by the null, with probabilities greater than 0.5. This conflicts with the results of Houpt and Townsend (2010), who presented nonsignificant negative excursions of the SIC for all observers.

To explain the difference between our results and those of Houpt and Townsend (2010), consider that the simulated Study 2 showed that the left tails of the Parallel-OR SIC are consistently underestimated by the median posterior (see Fig. 15, third panel)

Table 8
Posterior probabilities that the maximum (D^+) and minimum values (D^-) of each observer's SIC have a larger magnitude than predicted by a flat SIC.

Observer	$P(X > D^+_{\text{null}})$	$P(X < D^-_{\text{null}})$
1	>0.9999	0.0360
2	0.9819	0.5419
3	>0.9999	0.0100
4	0.9999	0.6850
5	0.9990	0.3676
6	>0.9999	0.0725
7	0.9991	0.1104
8	>0.9999	0.0982
9	>0.9999	0.8864

for smaller sample sizes. The data we analyzed here had sample sizes of $N = 200$, which may not be sufficient to provide an accurate picture of these early negative trends in the SIC. It is also the case that real data is always noisier than simulated data. Other processes could be contributing to these fast responses, creating a mixture distribution that does not conform to the assumptions of either the D^- statistic or the piecewise exponential model (Craigmile, Peruggia, & Van Zandt, 2010).

Without disregarding the early negative excursions of the SIC, we may conclude that the results confirm either Parallel-OR or coactive processing (see Table 2). Most observers could detect the dots in parallel and responded as soon as they detected either the top or the bottom dot. Observers 4, 5 and 9 may have been engaging coactive processing.

Please cite this article in press as: Houpt, J. W., et al., Semiparametric Bayesian approaches to systems factorial technology. Journal of Mathematical Psychology (2016), http://dx.doi.org/10.1016/j.jmp.2016.02.008

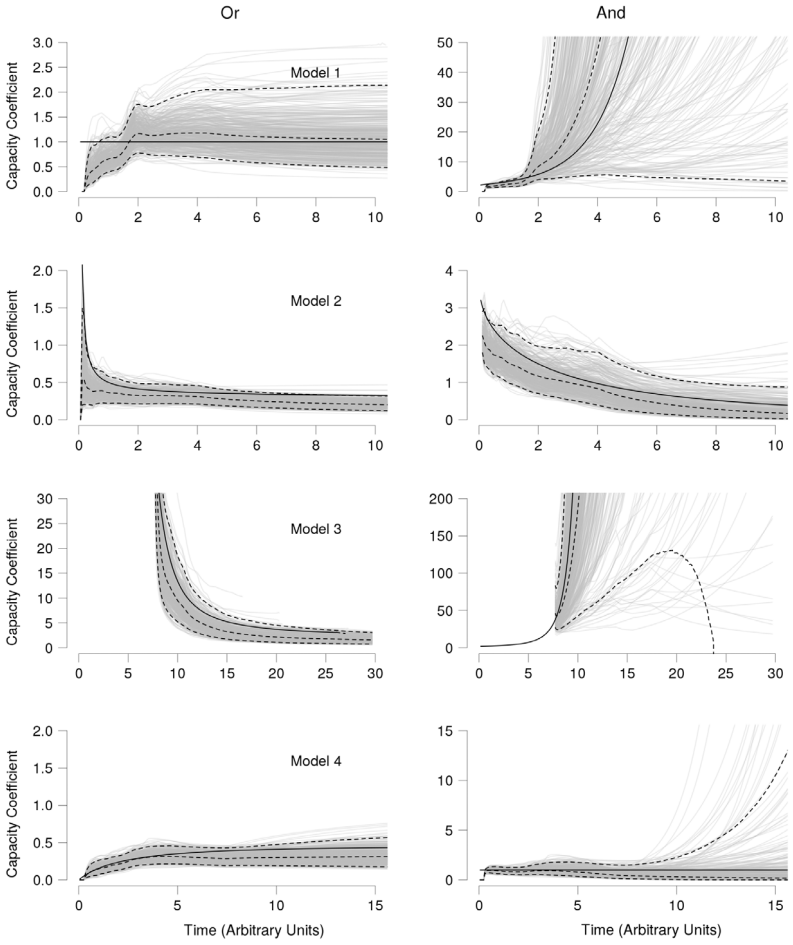

Fig. 11. Posterior estimates of the capacity coefficients $C_{OR}(t \mid \boldsymbol{h}_{i,1}, \boldsymbol{h}_{i,2}, \boldsymbol{h}_{i,3})$ (left column) and $C_{AND}(t \mid \boldsymbol{h}_{i,1}, \boldsymbol{h}_{i,2}, \boldsymbol{h}_{i,3})$ (right column) for the four models, $i = 1, 4$ (rows), of the first study with sample size equal to $N = 50$. A subset of the samples from the posterior is shown in gray together with the posterior medians and 95% credible intervals (dashed lines). The solid line is the true capacity coefficient given the model that generated the data.

6. Summary and conclusions

In this paper we have presented a semiparametric hierarchical Bayesian approach to estimating the hazard functions of a set of RT data collected over multiple observers and conditions. The model that we presented is very flexible, able to accommodate data across a range of data-generating distributions and conditions simultaneously. Using the posterior distributions of hazard rates obtained by fitting the model, we have computed the posteriors of the systems factorial technology functionals, the capacity coefficients and the survivor function interaction contrast, that can be used for fully Bayesian explorations of cognitive architectures.

We demonstrated the feasibility and accuracy of the approach in two simulation studies involving four "observers" and three conditions in Study 1 and five "observers" and four conditions in Study 2. Posterior estimates of the hazard function obtained from the posterior median were accurate even for small sample sizes, far more accurate than bootstrapping kernel density estimates of the hazard function (Van Zandt, 2002b). Despite some misfits across the models, the capacity coefficients and survivor function interaction contrasts were also accurately estimated, though larger samples were necessary to achieve results as nice as for the hazard functions themselves. This is not surprising, as the hazard rates were sampled directly, while the SFT functionals were derived from combinations of them. The sample sizes we explored were all well within the range of a typical RT experimental design.

We applied these methods to data from a simple two-dot detection task previously published (Eidels et al., 2015). The posterior results indicated limited-capacity, parallel, first-terminating (OR) detection of the dots. These results generally conform to existing models of visual perception and match the inferences drawn from the existing null-hypothesis significance tests for the SIC and the capacity coefficient (Houpt & Townsend, 2010, 2012), although there is some ambiguity about early, negative excursions in the SIC for some observers. This ambiguity may be due to too-small sample sizes or mixtures of different kinds of fast responses.

The overall consistency of the conclusions across the two methods is reassuring. Our new approach complements the existing analyses from Houpt and Townsend by providing posterior distributions as the basis for inference rather than deviation from a null hypothesis. Furthermore, the new approach extends the focus of inference to function-wide statistics rather than single-point summaries.

We chose the piecewise-exponential model because of its flexibility and not for theoretical reasons; it is a descriptive model that, as we have demonstrated, can accommodate a wide range of hazard function shapes. The choices of the priors were therefore driven primarily by considerations of convenience. It is important to notice that we have used the identical model structure, including priors and hyperparameters, for all the work presented in this paper. This work encompassed 8 different models generating 32 different data sets at each sample size, data sets that represented RTs varying over different scales. The posteriors estimated for each of these data sets were far more accurate than any other estimator proposed to date, and were highly accurate at sample sizes that

Please cite this article in press as: Houpt, J. W., et al., Semiparametric Bayesian approaches to systems factorial technology. Journal of Mathematical Psychology (2016), http://dx.doi.org/10.1016/j.jmp.2016.02.008

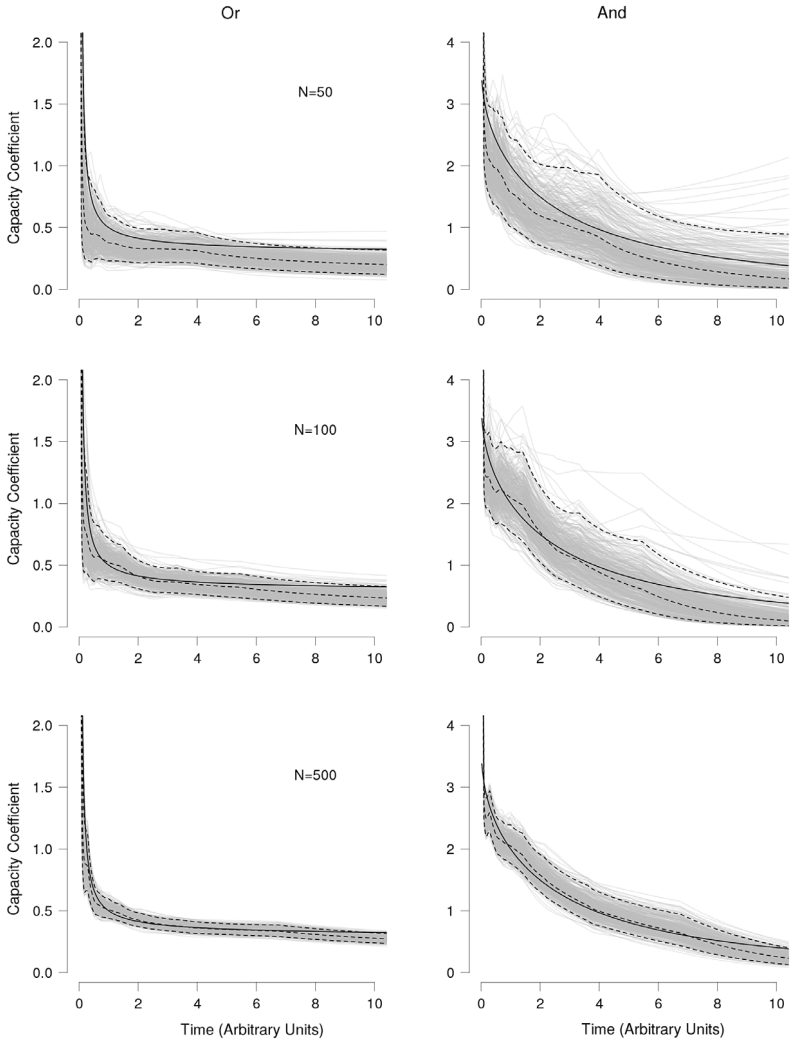

Fig. 12. Effects of sample size ($N = 50$, $N = 100$ and $N = 500$ in the top, middle and bottom rows, respectively) on the posterior estimates of the capacity coefficients for Model 2 $C_{\text{OR}}(t \mid \boldsymbol{h}_{2,1}, \boldsymbol{h}_{2,2}, \boldsymbol{h}_{2,3})$ (left column) and $C_{\text{AND}}(t \mid \boldsymbol{h}_{2,1}, \boldsymbol{h}_{2,2}, \boldsymbol{h}_{2,3})$ (right column). A subset of the samples from the posterior is shown in gray together with the posterior medians and 95% credible intervals (dashed lines). The solid line is the true capacity coefficient given the model that generated the data.

reflect the sample sizes typical of RT experiments. Therefore, while the specific choices we made might have influenced in some small way the posteriors estimated at smaller sample sizes, we have no reason to believe that these specific priors could have systematically biased the estimates, especially for larger sample sizes. We would never argue that our model could be applied to data in a "black box" fashion, but we believe it is easily generalizable to other kinds of data.

6.1. Evaluation of model fit

We discussed the issue of model fit, commenting on the good recovery of the hazard rates and the relative fit of the piecewise exponential model to the individual models that generated the data. We were quite pleased with how well the model performed, given that there were four (Study 1) or five (Study 2) very different structures that the piecewise exponential model had to fit simultaneously, and over three (Study 1) or four (Study 2) conditions with different model parameters. The fit of the model overall was quite impressive but at least three issues remain.

First, it is evident that the placement of the second bin boundary $s_{ic,1}$ can be problematic. For Model 3 in Study 1, the distribution of RTs was shifted relative to those of the other data sets. The piecewise exponential model predicts observations within this bin to be close to $s_{ic,0} = 0$, whereas the observations are more likely to be close to $s_{ic,1}$. If the samples are not shifted by their minima, this misfit produces hazard estimates that artificially elevate the left-hand tails of the distribution, resulting in systematic errors of all the other function estimates. For this reason, we subtracted the minimum of each sample, and then added it back after the shifted hazard rate posteriors were estimated.

A similar issue likely exists for the last bin boundary $s_{ic,J} > \max_k T_{ick}$. If the distribution is heavy-tailed, such as the log-normal for Model 2, then the last bin boundary $s_{ic,J}$ will be large but potentially not large enough, leading to a bin that is too narrow. The model assumes exponentially-distributed observations within the bin, but a heavy-tailed distribution will have more observations near $s_{ic,J}$, again resulting in an overestimate of $h_{ic,J}$. Such an overestimate is evident for Model 2 in Fig. 8.

A second source of misfit is likely to be shrinkage in the estimates of $\boldsymbol{h}_{ic}$ resulting from the attempts to fit a number of different models of varying location and scale within the same hierarchical structure. That is, because the $\boldsymbol{h}_{ic}$s are drawn from a common hyperdistribution, the values of the hazard rates across observers i will be pulled toward the center of that common hyperdistribution. Part of the misfit we observed for the first and last bins could be due to shrinkage. One way to minimize effects of individual differences in location and variance could be to normalize each observer's data

Please cite this article in press as: Houpt, J. W., et al., Semiparametric Bayesian approaches to systems factorial technology. Journal of Mathematical Psychology (2016), http://dx.doi.org/10.1016/j.jmp.2016.02.008

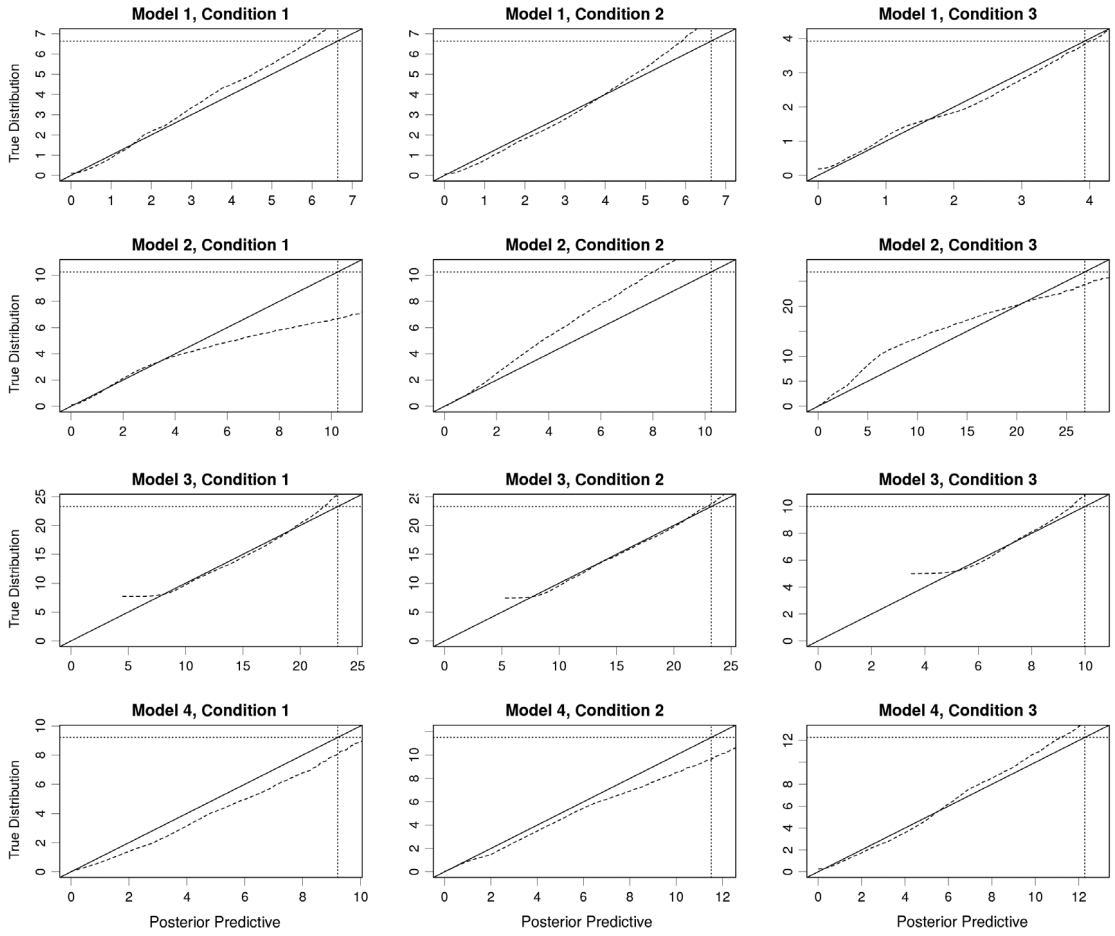

Fig. 13. Quantile–quantile plots (dashed lines) of the posterior predictive distributions and the true distributions from which the data were simulated for sample size $N = 50$ in the first study. The 99% quantiles are shown as dotted vertical and horizontal lines, and the solid lines are the identity. Columns 1, 2, and 3 contain the panels for Conditions 1, 2, and 3, respectively, while rows 1–4 contain the panels for Models 1–4, respectively.

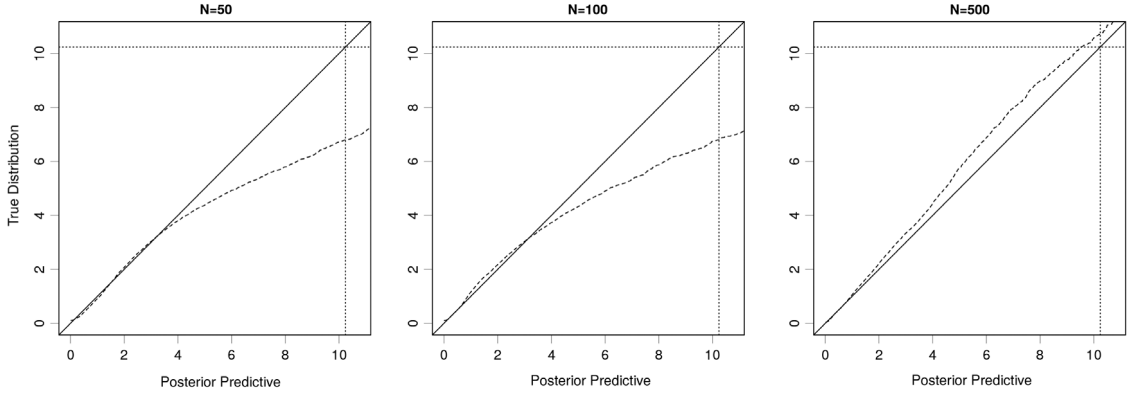

Fig. 14. Effects of sample size (left panel $N = 50$; middle panel $N = 100$; right panel $N = 500$) on model fit for Model 2, Condition 1. The 99% quantiles are shown as dotted vertical and horizontal lines, and the solid lines are the identity.

before fitting, and then transform the resulting posteriors back to the original scale of the data, by appropriate transformations of the piecewise exponential mean and variance.

Transformations of location are easily accomplished (as we did with the minimum), because the bin boundaries, computed outside of the sampling process, are the component of the model that determines location. However, transformations of scale are not. The variance of the piecewise exponential is not invertible as a function of the parameters $\boldsymbol{h}_{ic}$. This means that a scale parameter would need to be added to the model and its posterior estimated along with the others.

A final source of misfit could be the AR(1) structure imposed on the values of $\boldsymbol{h}_{ic}$. The contribution of the AR(1) structure to misfit is evident in the posterior distributions of ϕ_{ic} shown in Fig. 6. As sample size increases, the posterior medians of ϕ_{ic} move closer to the upper bound of 1.0, suggesting that the mean of the AR(1) process is not adequately specified (Peruggia, 2007).

6.2. Further directions

The issues of model misfit suggest a number of directions in which this work could be extended, at the cost, of course, of more

Please cite this article in press as: Houpt, J. W., et al., Semiparametric Bayesian approaches to systems factorial technology. Journal of Mathematical Psychology (2016), http://dx.doi.org/10.1016/j.jmp.2016.02.008

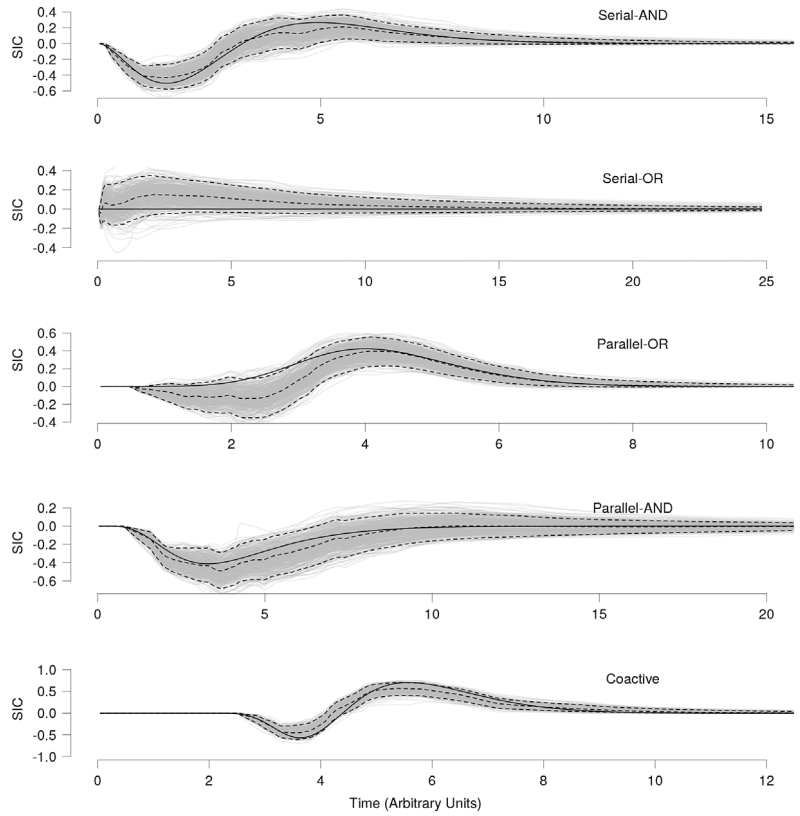

Fig. 15. Posterior estimates of the survivor function interaction contrast $SIC(t \mid \boldsymbol{h}_{i,1}, \boldsymbol{h}_{i,2}, \boldsymbol{h}_{i,3})$ for the five models $i = 1, \ldots, 5$ of the second study with sample size equal to $N = 50$. A subset of the samples from the posterior is shown in gray together with the posterior medians and 95% credible intervals (dashed lines). The solid line is the true survivor function contrast given the model that simulated the data.

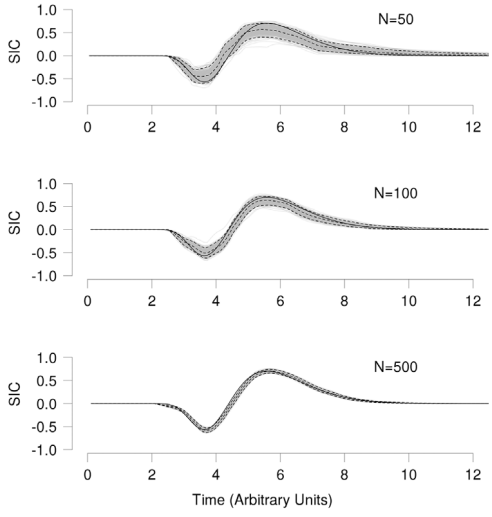

Fig. 16. Effects of sample size (top panel $N = 50$; middle panel $N = 100$; bottom panel $N = 500$) on the survivor function interaction contrast $SIC(t \mid \boldsymbol{h}_{5,1}, \boldsymbol{h}_{5,2}, \boldsymbol{h}_{5,3})$ for Model 5. A subset of the samples from the posterior is shown in gray together with the posterior medians and 95% credible intervals (dashed lines). The solid line is the true survivor function contrast given the model that generated the data.

complex computation. First, to address the problem of selection of bin boundaries, we might consider using a much finer grid. Rather than an AR(1) process over a small number of bins determined by the sample quantiles, we could place a Gaussian process on the log-hazard at the finest (continuous time) scale. The Gaussian process induces dependence in the hazard rates over time, say $h(t)$ and $h(t')$, through its covariance function at times t and t'. The dependence over bins modeled by the AR(1) process in the present model is in fact an approximation of this kind of dependence structure. Moving to a continuous Gaussian process would reduce the sensitivity of the analysis to the choice of bins and stabilize the behavior of the hazard estimates in the right tail.

A second direction involves the prior structure of the model, specifically for the means μ_{ic} of the AR(1) process. These means need not be centered on the constant hazard arising from an exponential distribution. Recall that μ_{ic} is the center about which the autoregressive log-hazard rate varies for Observer i under Condition c (see Eq. (4)), and we specified a $N(\mu_{\mu,i}, \sigma_{\mu}^2)$ prior for it. The expected value of the log hazard is therefore constant across bins, and, in our model, equal to 1.0. This is a quite restrictive model, which could be replaced with a richer mean structure.

As an example, we could write $\boldsymbol{\mu}_{ic} = \{\mu_{ic,1}, \mu_{ic,2}, \ldots, \mu_{ic,J}\}$, where

$$\mu_{ic,j} = \frac{\kappa_{ic}}{\lambda_{ic}^{\kappa_{ic}}} t_j^{\kappa_{ic}-1} \exp\left\{-(t_j/\lambda_{ic})^{\kappa_{ic}}\right\},$$

and t_j is equal to the bin midpoint $\left(s_j + s_{j-1}\right)/2$. That is, the bin-wise means for the log hazard may arise from a Weibull distribution's pdf, indexed by shape κ_{ic} and scale λ_{ic}, leading to different means for different bins. Depending on the shape parameter κ_{ic}, these means might be constant, monotonically increasing, or monotonically decreasing. Placing prior distributions on the Weibull shape and scale parameters results in a semiparametric hazard centered on a data-driven parametric form.

6.3. Conclusions

In conclusion, systems factorial technology can be applied in a fully Bayesian fashion, using a general model that is flexible enough

Please cite this article in press as: Houpt, J. W., et al., Semiparametric Bayesian approaches to systems factorial technology. Journal of Mathematical Psychology (2016), http://dx.doi.org/10.1016/j.jmp.2016.02.008

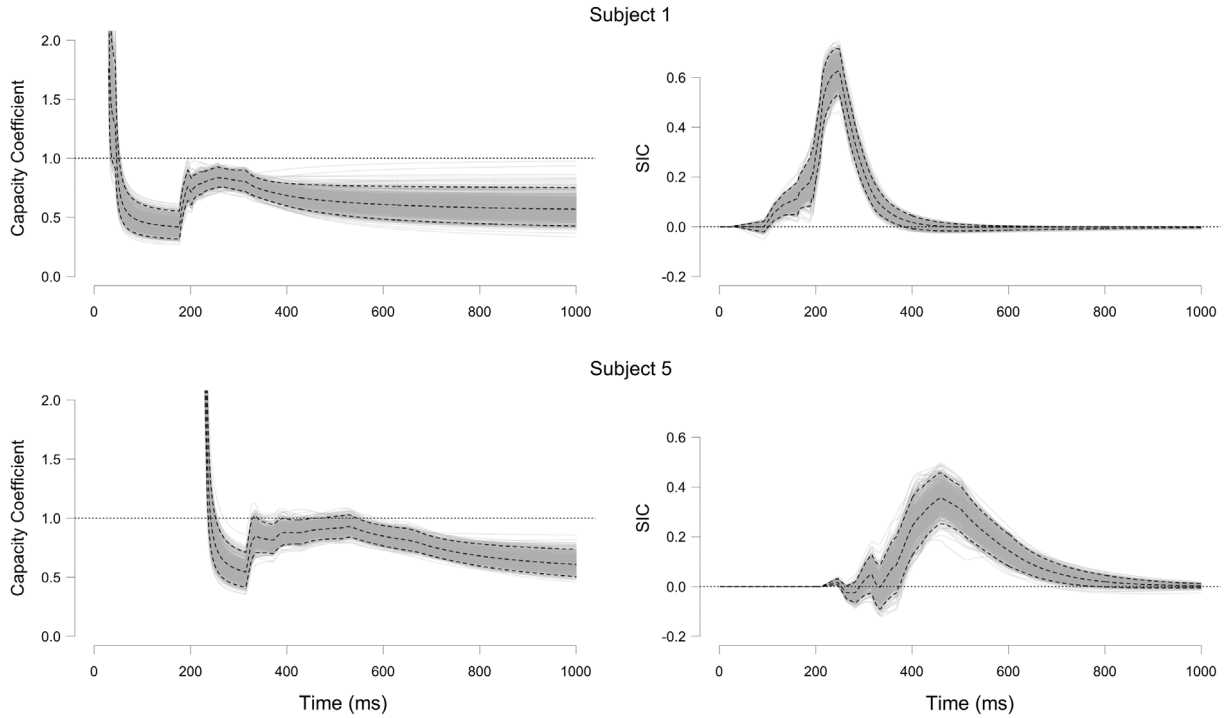

Fig. 17. Posterior capacity coefficients (left panels) and SICs (right panels) for Observer 1 (left panels) and Observer 5 (right panels) from Eidels et al. (2015). A sample of the posterior distributions of the capacity coefficients is shown as gray lines, with the (point-wise) posterior medians and 95% highest density intervals as dashed lines.

to accommodate a wide range of data-generating mechanisms. This is a significant advance, opening the door to new applications of systems factorial technology to more complex data. In addition, inferences about cognitive architecture can now be made on the basis of the posterior distributions of the systems factorial technology functionals, improving statistical decision making and better integrating theoretical questions with analytic techniques.

Acknowledgments

The authors would like to thank Peter Craigmile for helpful advice on model implementation. Portions of this material were presented at the 2013 symposium *Mathematical Models of Perception and Cognition: Essays in Honor of James T. Townsend*, Bloomington, IN, and at the 48th annual meeting of the Society for Mathematical Psychology, Newport Beach, CA. The research was made possible through grants from the National Science Foundation (BCS-1331047, SES 1024709, SES 1424481, and DMS 1209194), and the Air Force Office of Scientific Research (FA9550-12-1-0172 and FA9550-13-1-087).

References

Anderson, J. R. (2007). *How can the mind occur in the physical universe?*. New York: Oxford University Press.

Ashby, F. G. (1982). Testing the assumptions of exponential, additive reaction time models. *Memory and Cognition*, *10*, 125–134.

Chechile, R. A. (2003). Mathematical tools for hazard function analysis. *Journal of Mathematical Psychology*, *47*, 478–494.

Craigmile, P. F., Peruggia, M., & Van Zandt, T. (2010). An autocorrelated mixture model for sequences of response time data. *Psychometrika*, *75*, 613–632.

Eidels, A., Townsend, J. T., Hughes, H. C., & Perry, L. A. (2015). Evaluating perceptual integration: uniting response-time-and accuracy-based methodologies. *Attention, Perception, & Psychophysics*, *77*, 659–680.

Gamerman, D. (1991). Dynamic Bayesian models for survival data. *Applied Statistics*, *40*, 63–79.

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, *24*, 997–1016. http://dx.doi.org/10.1007/s11222–013-9416–2.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 457–472.

Homan, M. D., & Gelman, A. (2014). The no-u-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *The Journal of Machine Learning Research*, *15*, 1593–1623.

Houpt, J. W., Blaha, L. M., McIntire, J. P., Havig, P. R., & Townsend, J. T. (2013). Systems factorial technology with R. *Behavior Research Methods*, *46*, 307–330.

Houpt, J. W., & Townsend, J. T. (2010). The statistical properties of the survivor interaction contrast. *Journal of Mathematical Psychology*, *54*, 446–453.

Houpt, J. W., & Townsend, J. T. (2012). Statistical measures for workload capacity analysis. *Journal of Mathematical Psychology*, *56*, 341–355.

Ibrahim, J. G., Chen, M.-H., & Sinha, D. (2005). *Bayesian survival analysis*. Wiley Online Library.

Luce, R. D. (1986). *Response times: their role in inferring elementary mental organization*. New York: Oxford University Press.

Peruggia, M. (2007). Bayesian model diagnostics based on artificial autoregressive errors. *Bayesian Analysis*, *2*, 817–841.

Schweickert, R., Fisher, D. L., & Sung, K. (2012). *Discovering cognitive architecture by selectively influencing mental processes, Vol. 4*. World Scientific.

Sinha, D., & Dey, D. K. (1997). Semiparametric Bayesian analysis of survival data. *Journal of the American Statistical Association*, *92*, 1195–1212.

Stan Development Team. (2014). Stan: A c++ library for probability and sampling, version 2.5.0. URL: http://mc-stan.org/.

Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes*. Cambridge: Cambridge University Press.

Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial and coactive theories. *Journal of Mathematical Psychology*, *39*, 321–360.

Townsend, J. T., & Wenger, M. J. (2004). A theory of interactive parallel processing: New capacity measures and predictions for a response time inequality series. *Psychological Review*, *111*, 1003–1035.

Van Zandt, T. (2002a). Analysis of response time distributions. In H. Pashler, & J. Wixted (Eds.), *Stevens' handbook of experimental psychology* (3rd ed.) (pp. 461–516). New York: John Wiley and Sons.

Van Zandt, T. (2002b). Analysis of response time distributions. In J. T. Wixted, & H. Pashler (Eds.), *Stevens' handbook of experimental psychology, Vol. 4* (3rd ed.) (pp. 461–516). New York: Wiley Press.

Van Zandt, T., & Ratcliff, R. (1995). Statistical mimicking of reaction time distributions: Mixtures and parameter variability. *Psychonomic Bulletin and Review*, *2*, 20–54.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *The Journal of Machine Learning Research*, *11*, 3571–3594.