Supplemental material for "Hierarchical Hidden Markov Models for Response Time Data"

Deborah Kunkel[1], **Zhifei Yan**[2],
Peter F. Craigmile[2], **Mario Peruggia**[2], **Trisha Van Zandt**[3,2].

[1] School of Mathematical and Statistical Sciences, Clemson University, Clemson, SC 29634, USA

[2] Department of Statistics, The Ohio State University, Columbus, OH 43210, USA

[3] Department of Psychology, The Ohio State University, Columbus, OH 43210, USA

Last updated January 29, 2020

1 The Markov chain Monte Carlo algorithm

Here we present the Markov chain Monte Carlo (MCMC) algorithm that we use to fit the Bayesian model presented in the main article.

Always assuming conditional independence, the full hierarchical model for participant $i, i = 1, \ldots, n$ is below.

$$
\begin{aligned}
Y_{i,t}|r_{i,t} &\sim \text{Weibull}(\beta_i, \lambda_i(r_{i,t})),\\
\log(\lambda_i(r_{i,t})) &\sim N(\mu_i^{(r)}, \sigma_\lambda^{(r)}), \quad r = 1,2,3,\\
\log(\beta_i) &\sim N(\mu_\beta, \sigma_\beta),\\
\boldsymbol{\mu}_i &\sim N_3(\boldsymbol{\mu}, \boldsymbol{V}_\mu) I(\mu_i^{(1)} < \mu_i^{(2)} < \mu_i^{(3)}),\\
\boldsymbol{\mu} &\sim N_3(\boldsymbol{\kappa}, \boldsymbol{V}) I(\mu^{(1)} < \mu^{(2)} < \mu^{(3)}),\\
\sigma_\mu^{(r)}, \sigma_\lambda^{(r)} &\sim \text{halfCauchy}(a), \quad r = 1,2,3,\\
P(r_{i,t+1} = s | r_{i,t} = r, e_{i,t+1} = e) &= \boldsymbol{P}_{rs}^{i,e},\\
P(e_{i,t+1} = e | e_{i,t} = d) &= \boldsymbol{Q}_{de}^{(i)},\\
\boldsymbol{q}_j^{(i)} &\sim \text{Dirichlet}(\boldsymbol{\pi}_j), \quad j = 1,2,3,\\
\boldsymbol{p}_j^{(i,e)} &\sim \text{Dirichlet}(\boldsymbol{\pi}_e), \quad j = 1,2,3, e = 1,2,3,
\end{aligned}
$$

where $\boldsymbol{\mu} = (\mu^{(1)}, \mu^{(2)}, \mu^{(3)})^T$, $\boldsymbol{\mu}_i = (\mu_i^{(1)}, \mu_i^{(2)}, \mu_i^{(3)})^T$, $\boldsymbol{\sigma}_\mu = (\sigma_\mu^{(1)}, \sigma_\mu^{(2)}, \sigma_\mu^{(3)})^T$, $\boldsymbol{\sigma}_\lambda = (\sigma_\lambda^{(1)}, \sigma_\lambda^{(2)}, \sigma_\lambda^{(3)})^T$, $\boldsymbol{V}_\mu = \text{diag}(\sigma_\mu^{(r)})$, and $\boldsymbol{V} = \text{diag}(v_r), r = 1,2,3$, and $\boldsymbol{a}_j$ denotes the jth row of matrix $\boldsymbol{A}$.

The following hyperparameters are fixed: $\mu_\beta = 2$, $\sigma_\beta = 0.3$, $\boldsymbol{\kappa} = (6,7,8)^T$, $v_1 = v_2 = v_3 = 0.2$, $a = 0.5$, $\boldsymbol{\pi}_1 = (9, 0.5, 0.5), \boldsymbol{\pi}_2 = (0.5, 9, 0.5)^T, \boldsymbol{\pi}_3 = (0.5, 0.5, 9)^T$.

We update the parameters according to the following steps.

1. Sample hyperparameters $\boldsymbol{\mu}, \boldsymbol{\sigma}_\mu, \boldsymbol{\sigma}_\lambda$.
2. For $i = 1, \ldots, n$, sample $\log(\beta_i), \log(\lambda_i(r)), \mu_i^{(r)}, r = 1,2,3$ from their full conditionals.
3. For $i = 1, \ldots, n$, sample the latent response modes and environments, $(r_{i,t}, e_{i,t}), t = 1, \ldots, T$, from their full conditionals using a forward-backward algorithm.
4. For $i = 1, \ldots, n$, sample $\boldsymbol{Q}^{(i)}$ and $\boldsymbol{P}^{(i,e)}$ from their full conditionals.

1.1 Sample $\boldsymbol{\mu}$, $\boldsymbol{\sigma}_\mu$, $\boldsymbol{\sigma_\lambda}$

Sample $\boldsymbol{\mu}$. The full conditional for $\boldsymbol{\mu}$ satisfies

$$f(\boldsymbol{\mu}|-) \propto f(\mu|\boldsymbol{\kappa}, \boldsymbol{V}) \prod_{i=1}^{n} f(\boldsymbol{\mu}_i|\boldsymbol{\mu}, \boldsymbol{\sigma}_\mu),$$

with

$$\boldsymbol{\mu}_i|\boldsymbol{\mu}, \boldsymbol{\sigma}_\mu \sim N_3(\boldsymbol{\mu}, \boldsymbol{V}_\mu) I(\mu_i^{(1)} < \mu_i^{(2)} < \mu_i^{(3)}),$$

and

$$\boldsymbol{\mu}|\boldsymbol{\kappa}, \boldsymbol{V} \sim N_3(\boldsymbol{\kappa}, \boldsymbol{V}) I(\mu^{(1)} < \mu^{(2)} < \mu^{(3)}).$$

Thus, $(\boldsymbol{\mu}|-)$ is proportional to a constrained trivariate Normal:

$$f(\mu|-) \propto N_3(\boldsymbol{\mu}_n, \boldsymbol{V}_n) I(\mu_n^{(1)} < \mu_n^{(2)} < \mu_n^{(3)}),$$

with

$$\boldsymbol{\mu}_n = (\mu_n^{(1)}, \mu_n^{(2)}, \mu_n^{(3)})^T, \text{ where } \mu_n^{(r)} = \frac{nv_r \sum_{i=1}^{n} \mu_i^{(r)} + \kappa_r (\sigma_\mu^{(r)})^2}{nv_r + (\sigma_\mu^{(r)})^2},$$

and

$$\boldsymbol{V}_n = \text{diag}\{v_n^{(r)}\}, \text{ where } v_n^{(r)} = \frac{(\sigma_\mu^{(r)})^2 v_r}{(\sigma_\mu^{(r)})^2 + nv_r}, \quad r = 1, 2, 3.$$

Sample $\boldsymbol{\sigma}_\mu$. Conditional on $\boldsymbol{\mu}_i$, the $\sigma_\mu^{(r)}$ are independent of each other. For $r = 1, 2, 3$, the full conditional of $\sigma_\mu^{(r)}$ satisfies

$$f(\sigma_\mu^{(r)}|-) \propto f(\sigma_\mu^{(r)}|a) \prod_{i=1}^{n} f(\mu_i^{(r)}|\mu^{(r)}, \sigma_\mu^{(r)}) I(\sigma_\mu^{(r)} > 0),$$

where

$$\mu_i^{(r)}|\mu^{(r)}, \sigma_\mu^{(r)} \sim N(\mu^{(r)}, (\sigma_\mu^{(r)})^2), \quad \text{for all } i = 1, \ldots, n,$$

and

$$\sigma_\mu^{(r)} \sim \text{halfCauchy}(a).$$

We use a Metropolis-Hastings (MH) step with a random walk proposal to sample $\sigma_\mu^{(r)}$.

Sample $\boldsymbol{\sigma_\lambda}$. Conditional on $\lambda_i(r)$, the $\sigma_\lambda^{(r)}$ are independent of each other. For $r = 1, 2, 3$, the full conditional of $\sigma_\lambda^{(r)}$ satisfies

$$f(\sigma_\lambda^{(r)}|-) \propto f(\sigma_\lambda^{(r)}|a) \prod_{i=1}^{n} f(\lambda_i(r)|\mu_i^{(r)}, \sigma_\lambda^{(r)}) I(\sigma_\lambda^{(r)} > 0),$$

where

$$\log(\lambda_i(r))|\mu_i^{(r)}, \sigma_\lambda^{(r)} \sim N(\mu_i^{(r)}, (\sigma_\lambda^{(r)})^2), \quad \text{for all } i = 1, \ldots, n,$$

and

$$\sigma_\lambda^{(r)} \sim \text{halfCauchy}(a).$$

We use a Metropolis-Hastings (MH) step with a random walk proposal to sample $\sigma_\lambda^{(r)}$.

1.2 Sample $\lambda_i(r_{i,t})$ and β_i

The variables $\lambda_i(r_{i,t})$ and β_i are sampled separately for each participant i. For the scale parameter, $\lambda_i(r_{i,t})$, the full conditional satisfies

$$f(\log(\lambda_i(r))|-) \propto \prod_{t:r_{i,t}=r} f(Y_{i,t}|r_{i,t}, \lambda_i(r), \beta_i) f(\log(\lambda_i(r))|\mu_i^{(r)}, \sigma_\lambda^{(r)}),$$

where

$$Y_{i,t}|r_{i,t} = r, \log(\lambda_i(r)), \beta_i \sim \text{Weibull}(\beta, \lambda_i(r)),$$

and

$$\log(\lambda_i(r))|\mu_i^{(r)}, \sigma_\lambda^{(r)} \sim N(\mu_i^{(r)}, (\sigma_\lambda^{(r)})^2),$$

for $r = 1, 2, 3$. To sample from this distribution we use a Metropolis-Hastings step with a random walk proposal.

For the shape parameter, β_i, the full conditional satisfies

$$f(\beta_i|-) \propto \prod_{t=1}^{T} f(Y_{i,t}|r_{i,t}, \lambda_i(r_{i,t}), \beta_i) f(\log(\beta_i)),$$

where

$$Y_{i,t}|r_{i,t},\lambda_i(r_{i,t}),\beta_i \sim \text{Weibull}(\beta_i,\lambda_i(r_{i,t})),$$

and

$$\log(\beta_i) \sim N(\mu_\beta,\sigma_\beta^2).$$

To sample from this distribution we use a Metropolis-Hastings step with a random walk proposal.

Sample $\boldsymbol{\mu}_i$. The full conditional satisfies

$$f(\boldsymbol{\mu}_i|-) \propto \prod_{r=1}^{3} f(\log(\lambda_i(r))|\mu_i^{(r)},\sigma_\lambda^{(r)})f(\boldsymbol{\mu}_i|\boldsymbol{\mu},\boldsymbol{\sigma}_\mu),$$

where

$$\log(\lambda_i^{(r)})|\mu_i^{(r)},\sigma_\lambda^{(r)} \sim N(\mu_i^{(r)},(\sigma_\lambda^{(r)})^2), r=1,2,3,$$

and

$$\boldsymbol{\mu}_i|\boldsymbol{\mu},\boldsymbol{\sigma}_\mu \sim N_3(\boldsymbol{\mu},V_\mu)I(\mu_i^{(1)}<\mu_i^{(2)}<\mu_i^{(3)}).$$

The full conditional is proportional to a constrained trivariate Normal:

$$\boldsymbol{\mu}_i|- \sim N_3(\boldsymbol{\mu}_{ni},\boldsymbol{V}_{n\mu})I(\mu_{ni}^{(1)}<\mu_{ni}^{(2)}<\mu_{ni}^{(3)}),$$

with

$$\boldsymbol{\mu}_{ni}=(\mu_{ni}^{(1)},\mu_{ni}^{(2)},\mu_{ni}^{(3)})^T, \text{ where } \mu_{ni}^r = \frac{\log(\lambda_i^{(r)})(\sigma_\mu^{(r)})^2+\mu^{(r)}(\sigma_\lambda^{(r)})^2}{(\sigma_\lambda^{(r)})^2+(\sigma_\mu^{(r)})^2}, \ r=1,2,3,$$

and

$$\boldsymbol{V}_{n\mu}=\text{diag}\{v_{n\mu}^{(r)}\}, \text{ where } v_{n\mu}^{(r)} = \frac{(\sigma_\mu^{(r)})^2(\sigma_\lambda^{(r)})^2}{(\sigma_\lambda^{(r)})^2+(\sigma_\mu^{(r)})^2}, r=1,2,3.$$

1.3 Sample $\{r_{i,t}\}$ and $\{e_{i,t}\}$

To sample the latent Markov chain $\{(r_{i,t}, e_{i,t}) : t = 1, \ldots, T\}$, we derive a forward filtering backward sampling (FFBS) algorithm. Details of forward filtering backward sampling can be found in Frühwirth-Schnatter (2006). The latent variables are sampled separately for each participant i, and for notational convenience we suppress the subscript i in the subsequent steps. The notation Y^t denotes the sequence $\{Y_1, \ldots, Y_t\}$.

Let $f(Y_t|r_t)$ denote the Weibull probability density function evaluated at Y_t conditional on the latent response mode r_t.

1.3.1 Forward filtering

We first derive the following filters

$$F_t = \begin{cases} f(r_1|Y_1), & t = 0; \\ f(e_t, r_{t+1}|Y^{t+1}), & t = 1, \ldots, T-1; \\ f(e_T|Y^T), & t = T, \end{cases}$$

that will be used in the backward sampling algorithm defined in Section 1.3.2. For F_0, we have

$$f(r_1|Y_1) \propto f(Y_1|r_1)f(r_1),$$

where $f(r_1)$ is the stationary distribution of the latent response modes, $\{r_t\}$. Given the initial F_0, we are able to derive the expression of F_{t+1} from F_t, for all $t = 0, \ldots, T-1$.

From F_0 to F_1. We have

$$f(e_1, r_2|Y_1) = \sum_{r_1} f(r_2|e_1, r_1)f(e_1)f(r_1|Y_1),$$

where $f(e_1)$ is the stationary distribution of the latent environments, $\{e_t\}$, and we have that $f(r_1|Y_1) = F_0$. The expression of F_1 can be obtained by noting that

$$F_1 = f(e_1, r_2|Y^2) \propto f(Y_2|r_2)f(e_1, r_2|Y_1).$$

From F_t to F_{t+1}, for $t = 1, \ldots, T-2$. We have

$$f(e_{t+1}, r_{t+2}|Y^{t+1}) = \sum_{e_t, r_{t+1}} f(r_{t+2}|e_{t+1}, r_{t+1}) f(e_{t+1}|e_t) f(e_t, r_{t+1}|Y^{t+1}),$$

where $f(e_t, r_{t+1}|Y^{t+1}) = F_t$. The expression of F_{t+1} can be obtained by noting that

$$F_{t+1} = f(e_{t+1}, r_{t+2}|Y^{t+2}) \propto f(Y_{t+2}|r_{t+2}) f(e_{t+1}, r_{t+2}|Y^{t+1}).$$

From F_{T-1} to F_T. We can obtain F_T from F_{T-1} by noting that

$$F_T = f(e_T|Y^T) = \sum_{e_{T-1}, r_T} f(e_T|e_{T-1}) f(e_{T-1}, r_T|Y^T) = \sum_{e_{T-1}, r_T} f(e_T|e_{T-1}) F_{T-1}.$$

1.3.2 Backward sampling

The key to backward sampling lies in the following factorization of $f(e^T, r^T|Y^T)$:

$$f(e^T, r^T|Y^T) = f(e_T|Y^T) f(e_{T-1}, r_T|e_T, Y^T) \prod_{t=T-2}^{1} f(e_t, r_{t+1}|e_{t+1}, r_{t+2}, Y^{t+1}) f(r_1|e_1, r_2, Y_1). \tag{1}$$

Each piece in the factorization may be found using the forward filters derived in Section 1.3.1, noting that

$$\begin{aligned} f(e_T|Y^T) &= F_T, \\ f(e_{T-1}, r_T|e_T, Y^T) &\propto f(e_T|e_{T-1}) F_{T-1}, \\ f(e_t, r_{t+1}|e_{t+1}, r_{t+2}, Y^{t+1}) &\propto f(r_{t+2}|e_{t+1}, r_{t+1}) f(e_{t+1}|e_t) F_t, \quad \text{for all } t = 1, \ldots, T-2, \end{aligned}$$

and

$$f(r_1|e_1, r_2, Y_1) \propto f(r_2|r_1, e_1) f(e_1) F_0.$$

Hence, the factorization (1) together with the filters obtained from the forward pass enable us to draw samples of the $\{e_t\}$ and $\{r_t\}$ chains backward from e_T to r_1.

1.4 Sample Q, P^1, P^2, and P^3

The matrices $\boldsymbol{Q}^{(i)}$, $\boldsymbol{P}^{(i,1)}$, $\boldsymbol{P}^{(i,2)}$, and $\boldsymbol{P}^{(i,3)}$, $i = 1, \ldots, n$ are sampled separately for each participant i. For notational convenience we suppress the subscript i in the subsequent steps.

Let $\boldsymbol{q}_j, j = 1, 2, 3$, denote the jth row of $\boldsymbol{Q}$. Let $n_{jk}, j, k = 1, 2, 3$, denote the number of transitions in the Markov chain $\{e_t\}$ from environment j to k. Use $f()$ to denote the prior distribution of $\boldsymbol{q}_j$, a Dirichlet distribution with parameter $\boldsymbol{\pi}$. The full conditional of $\boldsymbol{q}_j$ satisfies

$$
\begin{aligned}
f(\boldsymbol{q}_j|-) &\propto f(e_1|\boldsymbol{Q}) \prod_{t\in\{1,\ldots,T-1\}:e_t=j} f(e_{t+1}|e_t, \boldsymbol{q}_j) f(\boldsymbol{q}_j), \\
&\propto f(e_1|\boldsymbol{Q}) \mathrm{Dir}(n_{j1} + \pi_1, n_{j2} + \pi_2, n_{j3} + \pi_3).
\end{aligned}
$$

To sample from $f(\boldsymbol{q}_j|-)$ we use a Metropolis-Hastings (MH) algorithm with proposal distribution $\mathrm{Dir}(n_{j1} + \pi_1, n_{j2} + \pi_2, n_{j3} + \pi_3)$.

Similarly, we sample each row of each of the transition matrices $\boldsymbol{P}^1$, $\boldsymbol{P}^2$, and $\boldsymbol{P}^3$, separately.

Let $\boldsymbol{p}^e_j, j = 1, 2, 3$ denote the jth row of $\boldsymbol{P}^e$, $e = 1, 2, 3$. Let n^e_{jk}, $j, k, e = 1, 2, 3$, denote the number of transitions in the sequence $\{r_t\}$ from response mode j to response mode k, environment e; that is, $n^e_{jk} = \sum_{t=1,\ldots,T-1} I(r_{t+1} = k, r_t = j, e_{t+1} = e)$. Use $f()$ to denote the prior distribution of $\boldsymbol{p}^e_j$, a Dirichlet distribution with parameter $\boldsymbol{\pi}$. We have

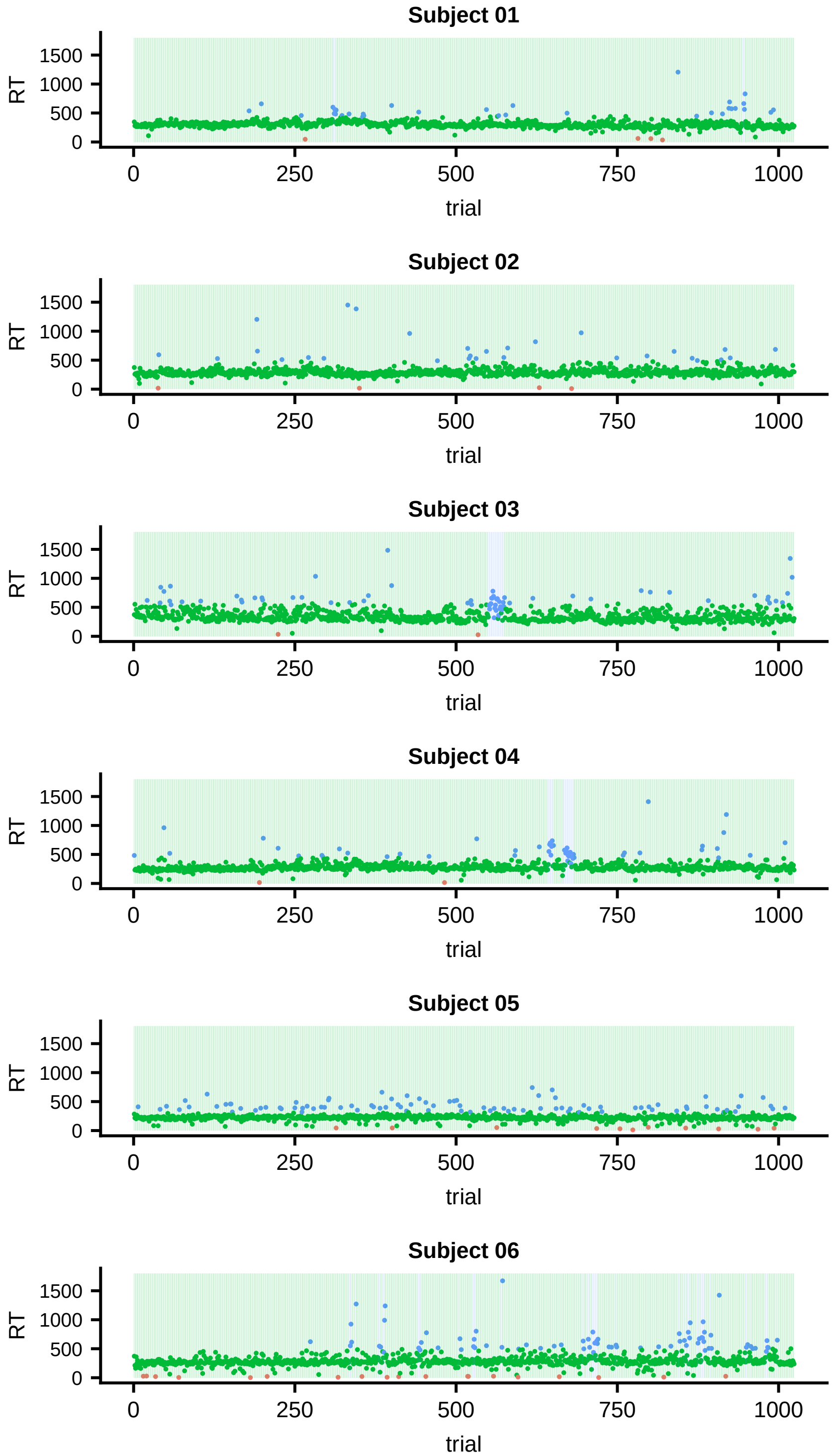
$$
\begin{aligned}
f(\boldsymbol{p}^e_j|-) &\propto f(r_1|\boldsymbol{P}^1, \boldsymbol{P}^2, \boldsymbol{P}^3) \prod_{t\in\{1,\ldots,T-1\}:e_{t+1}=e, r_t=j} f(r_{t+1}|r_t, e_{t+1}, \boldsymbol{p}^e_j) f(\boldsymbol{p}^e_j), \\
&\propto f(r_1|\boldsymbol{P}^1, \boldsymbol{P}^2, \boldsymbol{P}^3) \mathrm{Dir}(n^e_{j1} + \pi_1, n^e_{j2} + \pi_2, n^e_{j3} + \pi_3).
\end{aligned}
$$

To sample from $f(\boldsymbol{p}^e_j|-)$ we use a Metropolis-Hastings (MH) algorithm with proposal distribution $\mathrm{Dir}(n^e_{j1} + \pi_1, n^e_{j2} + \pi_2, n^e_{j3} + \pi_3)$.

2 Analysis of Wagenmakers data

The main manuscript presents an analysis of the data of Wagenmakers et al. (2004) under the short RSI condition. We performed the same analysis on the data from the long RSI condition. Figure 1 shows the RTs for each participant colored by their estimated response mode and with background color corresponding to the estimated environment as well as the posterior means of the TPMs.

There are fewer estimated transitions among environments in this condition than in the short RSI condition; most participants are estimated to remain in environment 2 for nearly the entire duration of the experiment, with occasional shifts into environment 3. In fact, response mode 1 is very seldom used and environment 1 is never used by any participant, perhaps because the prior on the scale parameters puts high probability on fast responses in response mode 1 and the RTs in this condition tend to be slower.



Subject 01

Q

0.89	0.06	0.05
0.00	1.00	0.00
0.03	0.16	0.81

P1

0.90	0.05	0.05
0.89	0.05	0.05
0.89	0.06	0.05

P2

0.04	0.93	0.04
0.00	0.97	0.02
0.01	0.89	0.10

P3

0.05	0.05	0.90
0.04	0.06	0.90
0.04	0.07	0.90

Subject 02

Q

0.90	0.05	0.05
0.00	1.00	0.00
0.06	0.08	0.87

P1

0.90	0.05	0.05
0.91	0.04	0.05
0.90	0.05	0.05

P2

0.03	0.86	0.11
0.00	0.97	0.03
0.01	0.95	0.04

P3

0.05	0.05	0.90
0.05	0.06	0.89
0.05	0.05	0.90

Subject 03

Q

0.89	0.06	0.06
0.00	1.00	0.00
0.02	0.09	0.90

P1

0.90	0.05	0.05
0.90	0.05	0.05
0.89	0.06	0.05

P2

0.04	0.89	0.07
0.00	0.95	0.05
0.01	0.80	0.19

P3

0.05	0.04	0.90
0.04	0.05	0.91
0.02	0.08	0.90

Subject 04

Q

0.89	0.06	0.05
0.00	0.99	0.00
0.01	0.12	0.87

P1

0.90	0.05	0.06
0.90	0.05	0.06
0.90	0.04	0.05

P2

0.05	0.90	0.04
0.00	0.97	0.03
0.01	0.93	0.06

P3

0.06	0.05	0.89
0.03	0.07	0.91
0.02	0.14	0.85

Subject 05

Q

0.89	0.06	0.05
0.00	1.00	0.00
0.05	0.07	0.87

P1

0.91	0.05	0.04
0.90	0.05	0.05
0.90	0.05	0.05

P2

0.02	0.88	0.10
0.01	0.89	0.10
0.02	0.91	0.07

P3

0.05	0.05	0.91
0.06	0.04	0.90
0.05	0.05	0.90

Subject 06

Q

0.87	0.07	0.06
0.00	0.97	0.03
0.01	0.29	0.70

P1

0.89	0.06	0.05
0.91	0.05	0.05
0.91	0.04	0.05

P2

0.05	0.90	0.06
0.02	0.96	0.02
0.02	0.90	0.07

P3

0.05	0.05	0.90
0.02	0.10	0.88
0.03	0.25	0.73

Figure 1: RTs and estimated environments and response modes for the long RSI condition

3 Sensitivity analysis

To assess the sensitivity of our analysis to the prior assumptions on the TPMs we consider three possible alternative priors resulting from using different values for the Dirichlet parameters $\boldsymbol{\pi}_1, \boldsymbol{\pi}_2$, and $\boldsymbol{\pi}_3$. The prior in the main manuscript used the following values:

$$\begin{aligned} \boldsymbol{\pi}_1 &= (9, 0.5, 0.5); \\ \boldsymbol{\pi}_2 &= (0.5, 9, 0.5); \\ \boldsymbol{\pi}_3 &= (0.5, 0.5, 9), \end{aligned} \tag{2}$$

so that 0.90 is the expected probability of remaining in a given environment and of transitioning to response mode e while in environment e. The three additional priors we considered are as follows:

$$\begin{array}{lllllllll} \text{Prior 1:} & & & \text{Prior 2:} & & & \text{Prior 3:} & & \\ \boldsymbol{\pi}_1 & = & (8, 1, 1); & \boldsymbol{\pi}_1 & = & (6, 2, 2); & \boldsymbol{\pi}_1 & = & (4.5, 0.25, 0.25); \\ \boldsymbol{\pi}_2 & = & (1, 8, 1); & \boldsymbol{\pi}_2 & = & (2, 6, 2); & \boldsymbol{\pi}_2 & = & (0.25, 4.5, 0.25); \\ \boldsymbol{\pi}_3 & = & (1, 1, 8). & \boldsymbol{\pi}_3 & = & (2, 2, 6). & \boldsymbol{\pi}_3 & = & (0.25, 0.25, 4.5). \end{array} \tag{3}$$

Priors 1 and 2 modify the expected TPMs so that the expected probability of remaining in a given environment is 0.80 and 0.60, respectively. Prior 3 leaves the expected TPMs unchanged, but produces a less-informative prior by reducing the concentration about the prior means.

Figure 2 summarizes the impact of these changes on the posterior estimated environments and response modes. For each prior, we calculated the the overall proportion of trials for which the estimated environment and response mode matched the estimate obtained under the original prior. These proportions are shown for the environments in the left panel of Figure 2 and for the response modes in the right panel. The figures indicate that with high probability the estimated response modes will be similar to those under the original prior across the three priors, with slightly higher variability under Prior 2. The estimated environments, on the other hand, differ considerably under Prior 2 from those under the

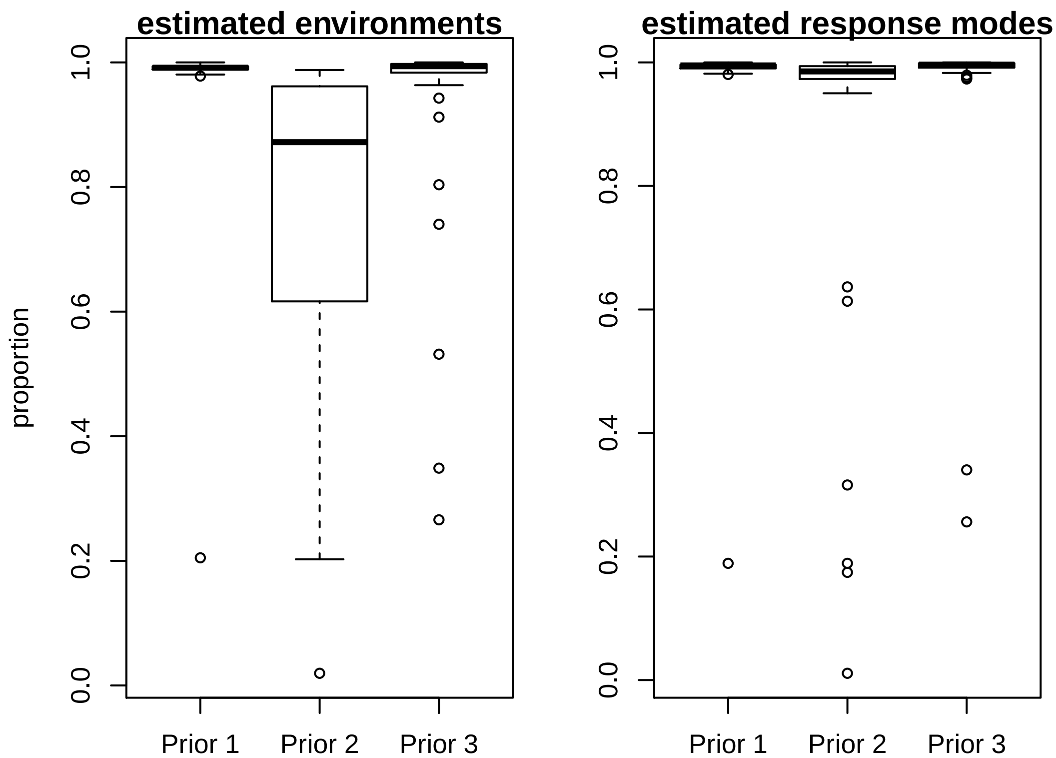


Figure 2: Across-participant average proportion of estimated environments (response modes) that match those estimated by the original prior.

original prior and exhibit slightly higher variability under Prior 3. These results suggest that the estimated latent states in this analysis are robust to small perturbations of the expected TPMs and to use of less-informative priors; however, the prior TPMs for Q do need to give each environment with a clear preference for one response mode in order to obtain results similar to those in our original analysis.

Figure 3 gives the across-participant average TPMs, $\bar{Q}$, $\bar{P}^1$, $\bar{P}^2$, and $\bar{P}^3$ for the original prior and priors 1-3. In $\bar{Q}$, the posterior estimates are fairly similar for all priors, especially among the original prior and priors 1 and 3. In $\bar{P}^1$, $\bar{P}^2$, and $\bar{P}^3$, the estimates for prior 3 are similar to those of the original prior, but reflect strong influence of priors 1 and 2.

Q

original

0.90	0.06	0.05
0.07	0.89	0.04
0.06	0.06	0.88

prior 1

0.90	0.05	0.05
0.07	0.89	0.04
0.07	0.07	0.85

prior 2

0.84	0.08	0.08
0.11	0.79	0.10
0.13	0.12	0.74

prior 3

0.90	0.06	0.04
0.07	0.89	0.04
0.06	0.06	0.87

P1

original

0.95	0.04	0.01
0.90	0.06	0.04
0.87	0.06	0.07

prior 1

0.94	0.05	0.02
0.82	0.12	0.06
0.77	0.10	0.13

prior 2

0.86	0.08	0.06
0.60	0.28	0.11
0.53	0.16	0.30

prior 3

0.95	0.04	0.01
0.88	0.09	0.04
0.84	0.06	0.09

P2

original

0.06	0.89	0.05
0.02	0.97	0.02
0.04	0.90	0.06

prior 1

0.12	0.79	0.08
0.03	0.95	0.02
0.07	0.80	0.13

prior 2

0.31	0.55	0.14
0.09	0.84	0.07
0.15	0.58	0.27

prior 3

0.07	0.88	0.05
0.02	0.97	0.01
0.04	0.88	0.08

P3

original

0.06	0.04	0.89
0.04	0.08	0.88
0.02	0.03	0.95

prior 1

0.13	0.08	0.79
0.08	0.12	0.80
0.04	0.04	0.92

prior 2

0.33	0.17	0.50
0.16	0.34	0.49
0.10	0.11	0.79

prior 3

0.08	0.04	0.88
0.04	0.08	0.88
0.02	0.03	0.95

Figure 3: Across-participant posterior mean TPMs under the original prior (top left) and priors 1-3.

References

Frühwirth-Schnatter, S. (2006). *Finite mixture and Markov switching models*. Springer, New York, NY.

Wagenmakers, E.-J., Farrell, S., and Ratcliff, R. (2004). Estimation and interpretation of $1/f$ noise in human cognition. *Psychonomic Bulletin & Review*, 11:579–615.