Implementation of

Reinforcement-Learning Drift Diffusion

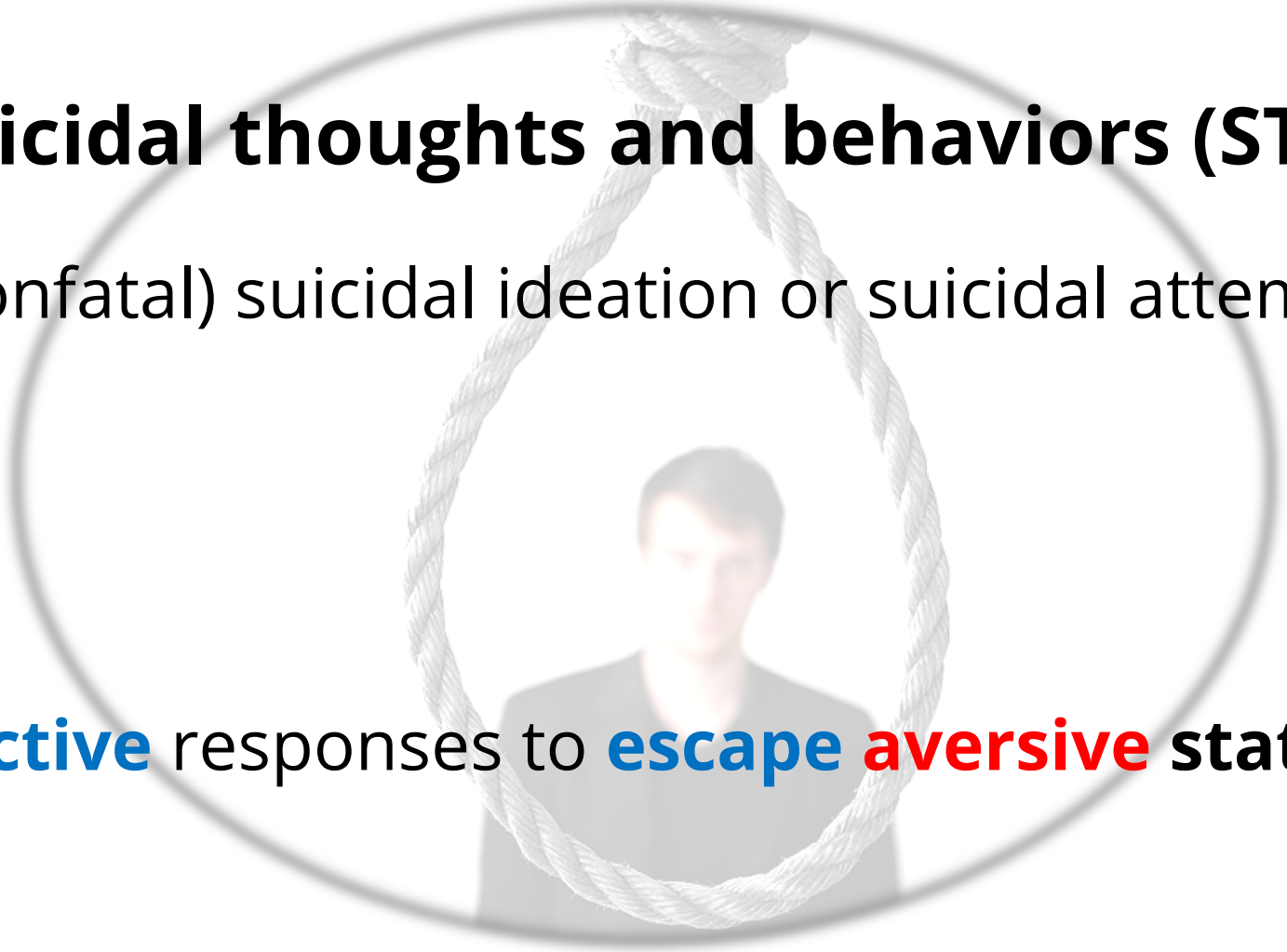
Model in rStan

Computational Modeling

Jihyun Hur

Contents

Introduction



Suicidal thoughts and behaviors (STB)

: (nonfatal) suicidal ideation or suicidal attempts

: **active** responses to **escape** **aversive** **states**

Research Paper

American Psychological Association

Journal of Abnormal Psychology

0021-843X/19/$12.00

2019, Vol. 128, No. 2, 106–118
http://dx.doi.org/10.1037/abn0000395

Suicidal Thoughts and Behaviors Are Associated With an Increased Decision-Making Bias for Active Responses to Escape Aversive States

Alexander J. Millner
Harvard University

Hanneke E. M. den Ouden
Radboud University

85 military veterans with Lifetime STB
44 with Psychiatric disorders, no STB

Research Question

Would these two groups of military veterans have different parameter values, which account for response bias, in escape condition?

Task Go/No-go

Model RL-DDM (Matlab)

Hierarchical Expectation Maximization

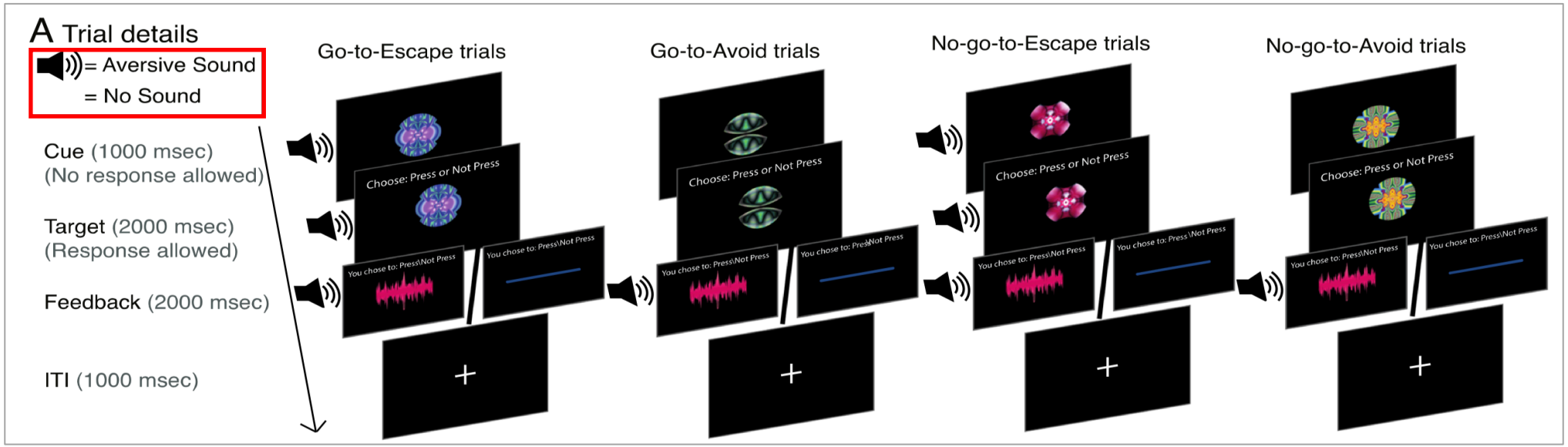
1. **Starting points**
2. Drift rates

	Escape		Avoid	
	Starting point	Drift rate	Starting point	Drift rate
Suicidal group	w1	β_1	w2	β_2
Non-suicidal group	w1	β_1	w2	β_2

Note The authors compared two main models: one with different starting points for each condition and one with two drift rates, each for one condition. As they calculated point estimates, BIC was the tool for model comparison.

Task: Go/No-go

Go/No-go Task with Aversive Sound Cues



Millner et al. (2019)

Go/no-go task with aversive sound cues includes two conditions, escape and avoid, which are distinguished by the existence of aversive sound cues (fork scraping a plate) at the beginning of a trial. A participant makes either go or no-go response for each trial, and the optimal response has to be learned throughout the trials by relating the sound cue and presented image with the corresponding feedback. There are two feedback types: aversive sound or silence.

Modeling in the Paper

Reinforcement Learning

$$Q_{t+1}(s_t, a_t) = Q_t(s_t, a_t) + \alpha[r_t - Q_t(s_t, a_t)]$$

Drift Rate based on Q-value

$$\mu_t = \beta_0 + \beta_1[Q_t(s_t, go) - Q_t(s_t, nogo)]$$

Drift Diffusion Model

$$Wiener(y|\alpha, \tau, \beta, \delta) = \frac{\alpha^3}{(y-\tau)^{3/2}} exp\left(-\delta\alpha\beta - \frac{\delta^2(y-\tau)}{2}\right) \sum_{k=-\infty}^{\infty} (2k+\beta)\phi \frac{2k\alpha+\beta}{\sqrt{y-\tau}}$$

α: learning rate, s_t: stimulus, a_t: action, r_t: reward, β_0: constant go bias, β_1: shared go bias

α: boundary separation, τ: non − decision time, β: drift rate, δ: starting point

Modeling in the Paper

Hierarchical Expectation Maximization

Learning rate (alpha)

Starting point for escape condition (w1)

Starting point for avoid condition (w2)

Non-decision time (tau)

Decision boundary (omega)

Constant go bias and shared go bias (β_0, β_1)

Replication

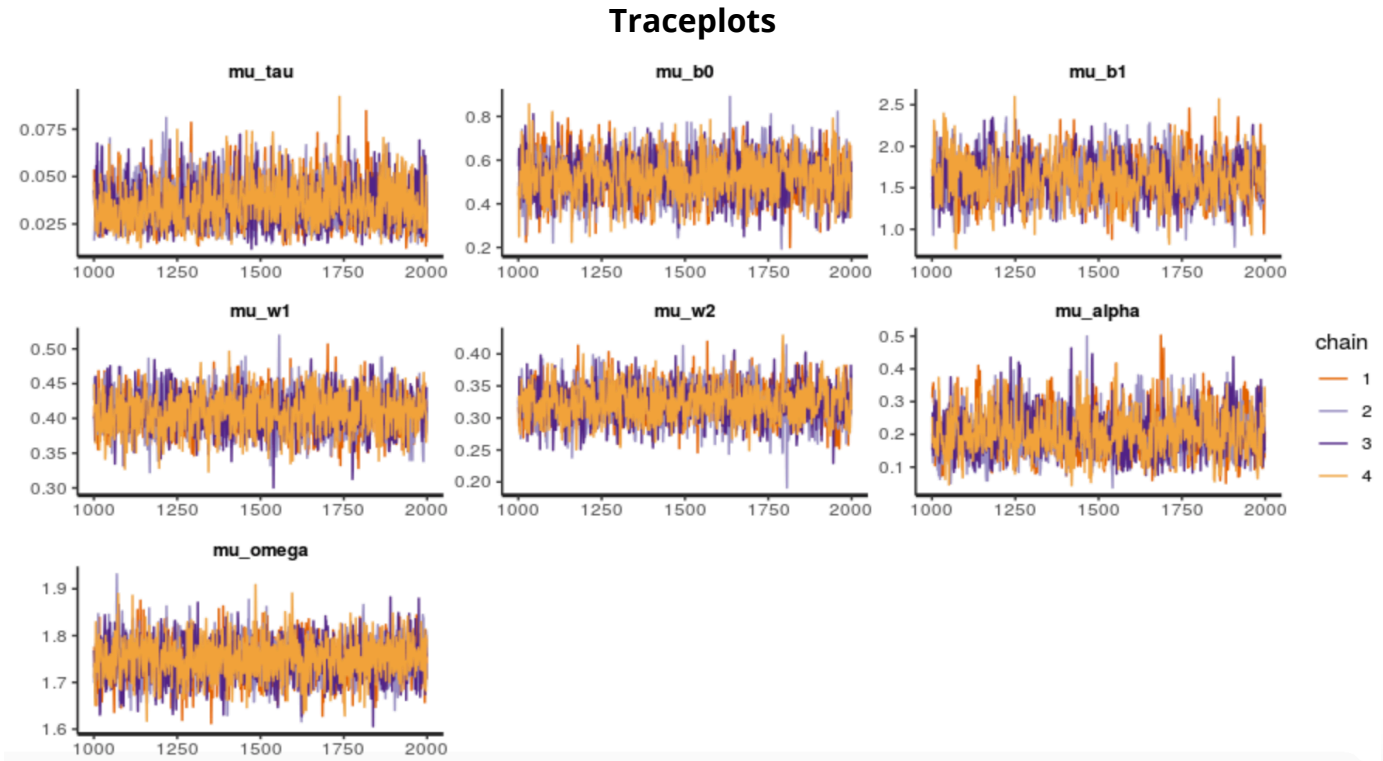
Implementation

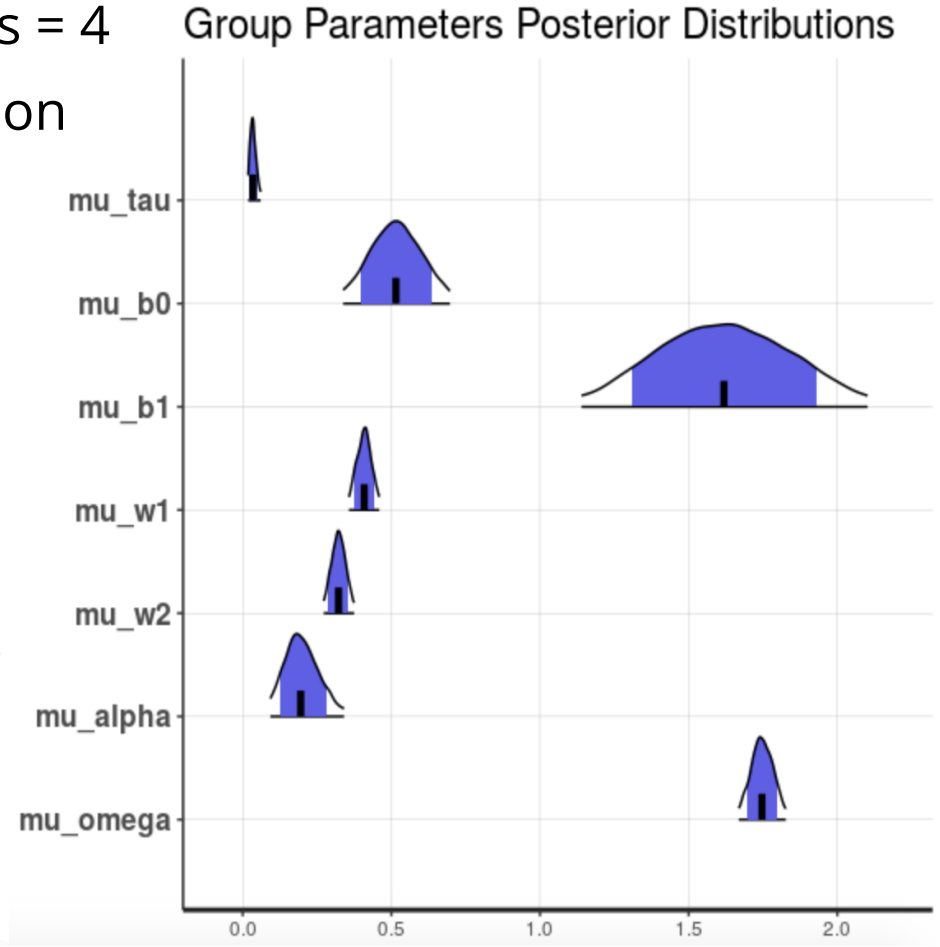
Model Block

```
// subject loop
for (i in 1:N) {
  // assign individual variables
  real w;
  int c;
  int s;
  matrix[4, 2] Q;

  // initial values for state-action values
  Q = rep_matrix(0.0, 4, 2);

  for (t in 1:T) {
    real v;
    c = C[t, i];
    s = cond[t, i];

    w = (E[t, i] == 1) ? w1[i] : w2[i];
    v = b0[i] + (b1[i] * (Q[s, 2] - Q[s, 1]));

    // increment log probability density and update posterior
    if (c == 1) {
      rt[t, i] ~ wiener(omega[i], tau[i], w, v);
    } else {
      nogo_lik = 0.0;
      for (n in 1:Steps) {
        step_size = (n * 1.0) / Steps;
        temp_rt = step_size * (RTmax - tau[i]) + tau[i];
        prob_density = exp(wiener_lpdf(temp_rt | omega[i], tau[i], 1-w, -v));
        nogo_lik += ((RTmax - tau[i]) / Steps) * prob_density;
      }
    if (is_nan(nogo_lik) == 1) {
     nogo_lik = 2.220446e-16;
    }
    target += log(nogo_lik);
    }

    // update Q-value
    Q[s, c+1] += alpha[i]*(fd[t, i] - Q[s, c+1]);
  }
 }
}
```

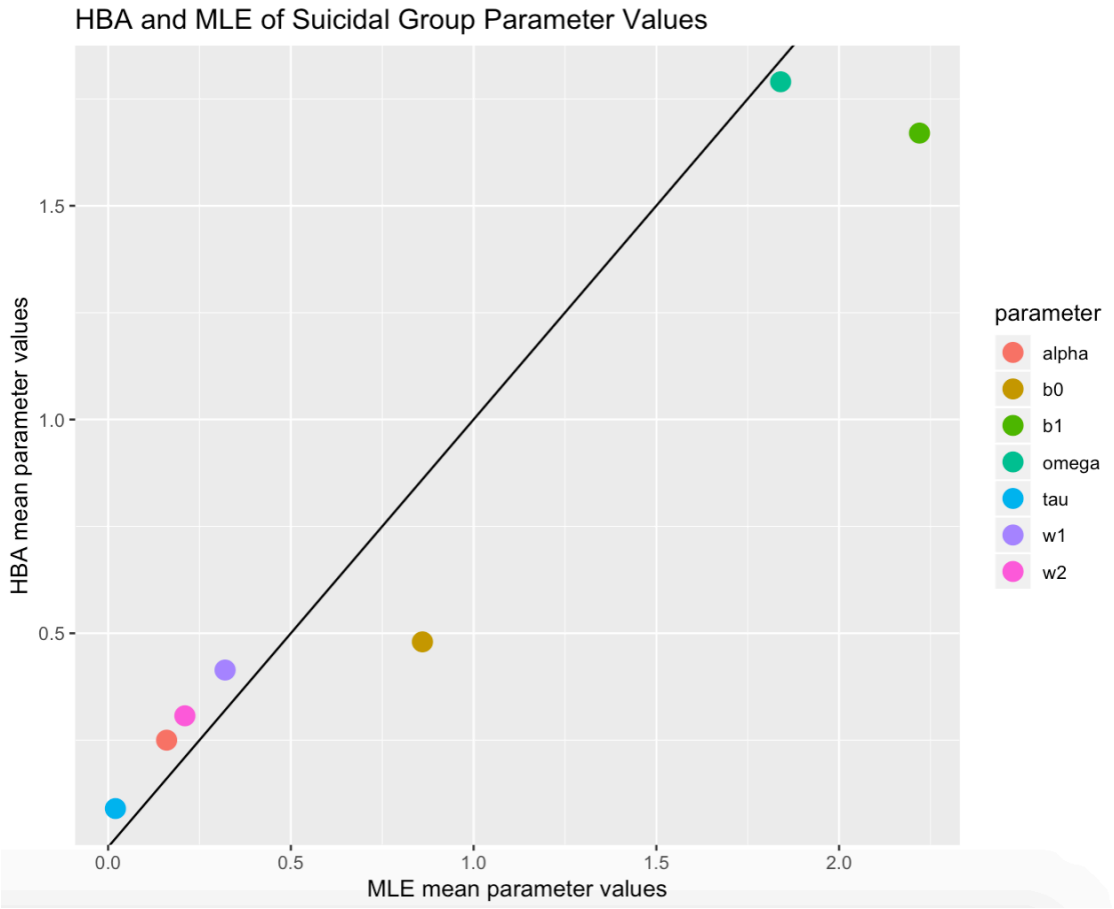
Replication

- Number of subjects = **30** (20 suicidal, 10 non-suicidal)
- Step size = 0.04, iters = 2000, warmup = 1000, chains = 4
- Hierarchical Bayesian, two starting points by condition

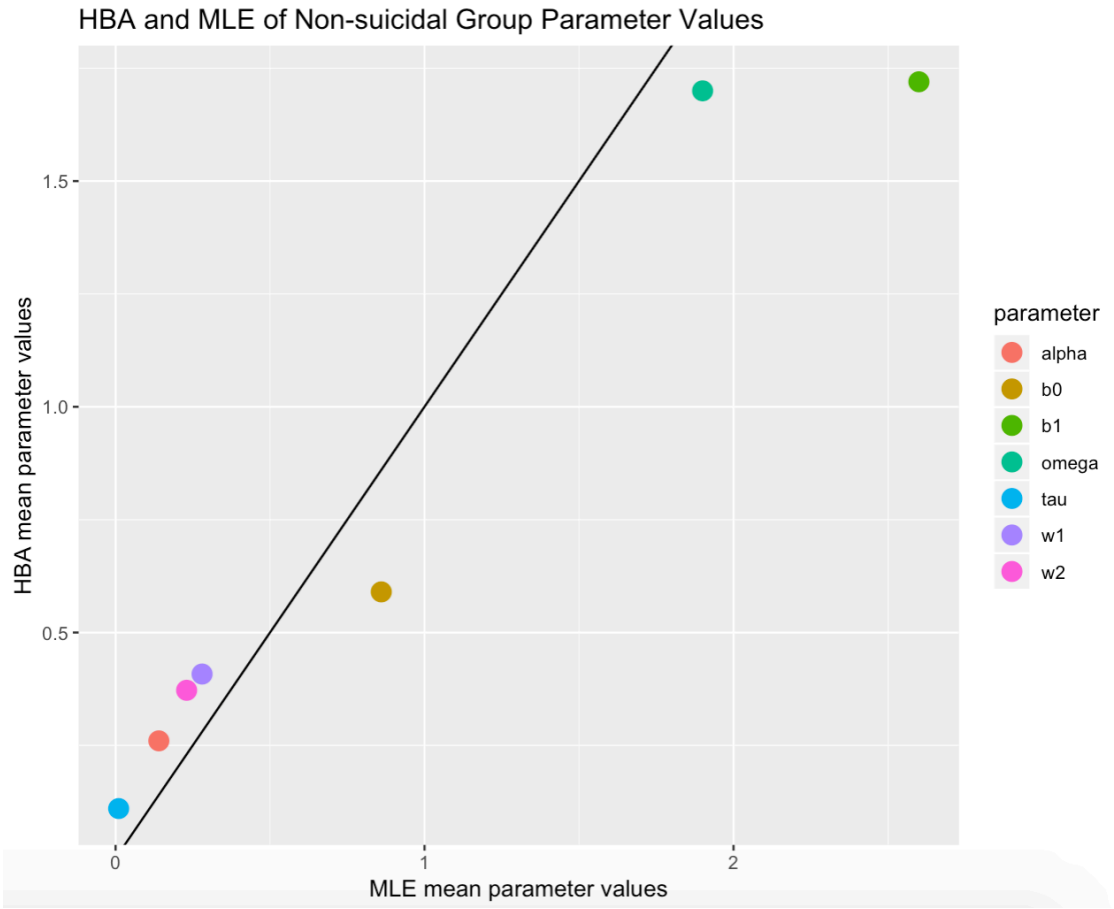




Correlation with MLE Estimates

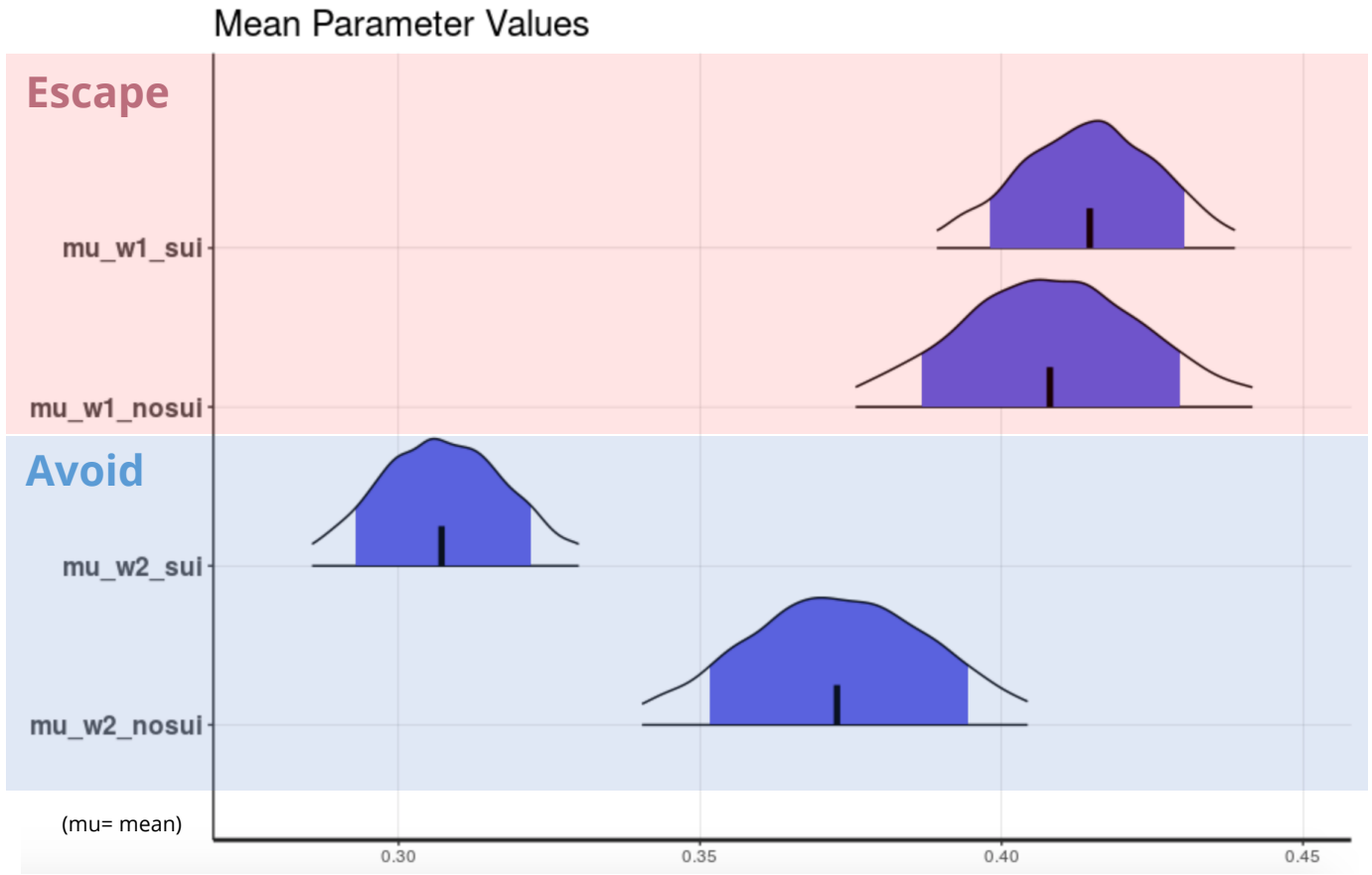


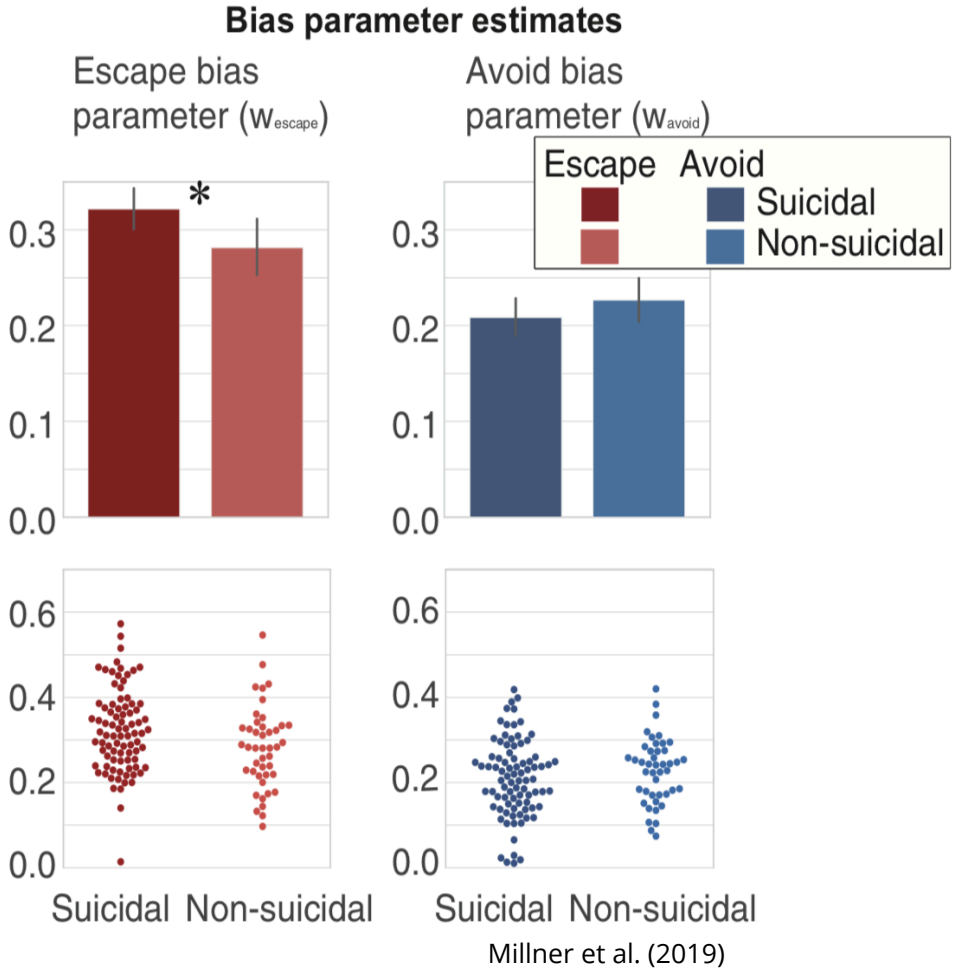
HBA: **20** suicidal participants



HBA: **10** non-suicidal participants

Results





Millner et al. (2019)

Original Model Discussions

1. Sampling using Model 1 has been converged well
2. In escape condition, the suicidal group starts their drifting closer to the go boundary, compared to the non-suicidal group
3. In avoid condition, the suicidal group starts their drifting closer to the no-go boundary, compared to the non-suicidal group
4. The non-suicidal group shows a higher constant go-bias across conditions

Model Comparison

Model Comparison

- Number of subjects = **30** (20 suicidal, 10 non-suicidal)
- Step size = 0.04, iters = 2000, warmup = 1000, chains = 4

Shared parameters: non-decision time, decision threshold, learning rate

Models (hierarchical)	**Added parameters**	**LOOIC**
1. Original	**w1, w2,** β_0	6580.8
2. Two drift rates	w, $\boldsymbol{\beta_1}$, $\boldsymbol{\beta_2}$	6496.0
3. Both	**w1, w2,** $\boldsymbol{\beta_1}$, $\boldsymbol{\beta_2}$	6404.4

Two Drift Rates Model (Model 2)

Model 2: Two drift rates + One starting point

- For escape trials:

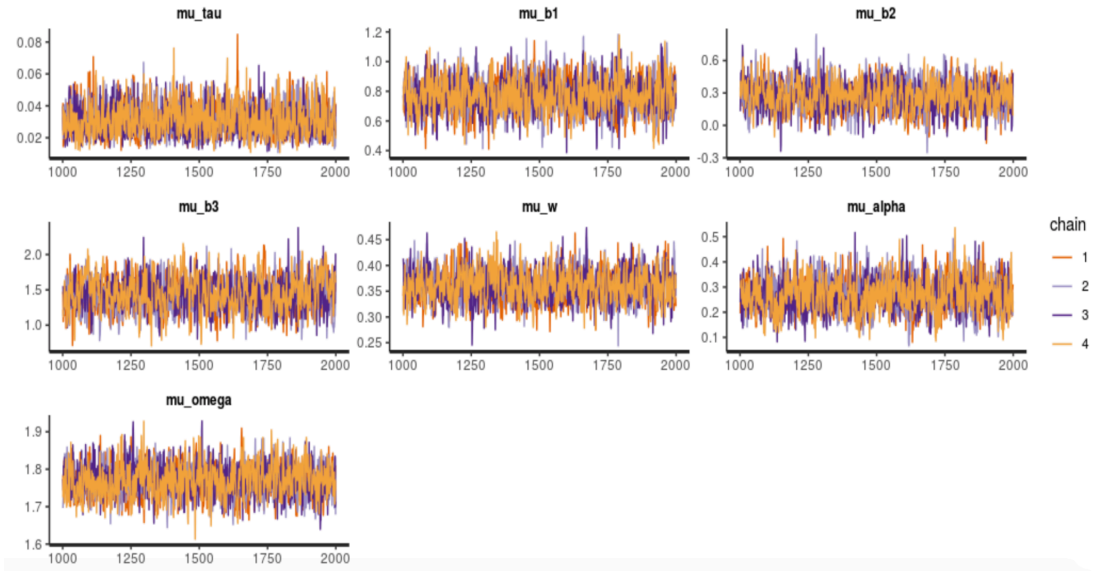
$$\mu_t = \boldsymbol{\beta_1} + \beta_3[Q_t(s_t, go) - Q_t(s_t, nogo)]$$

- For avoid trials:

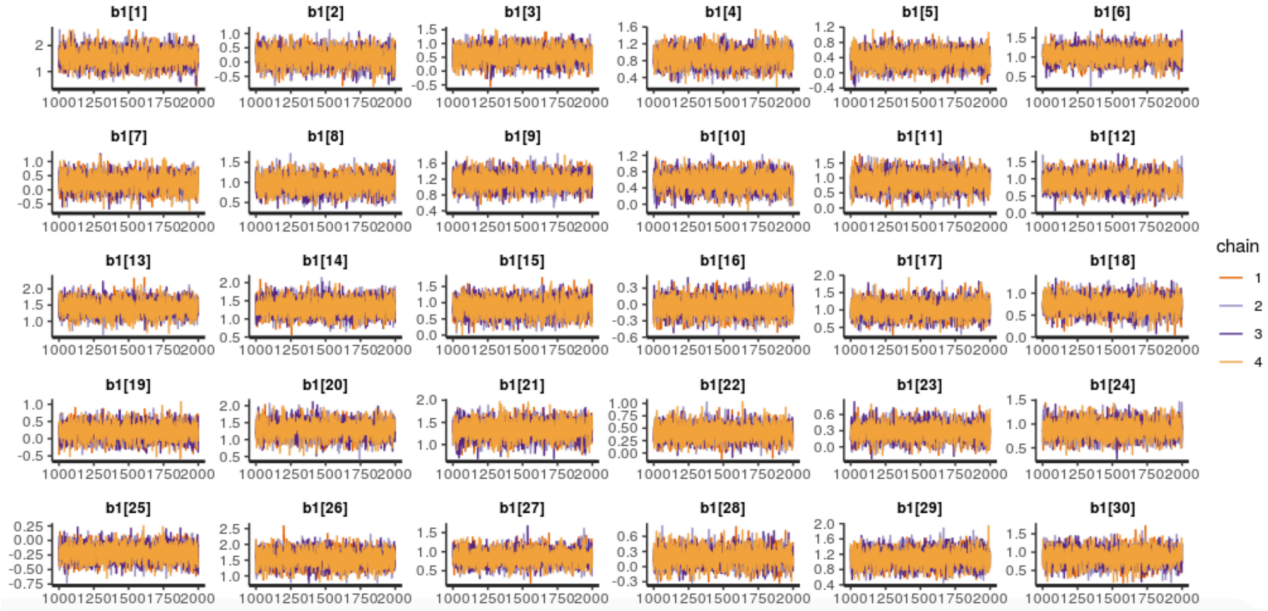
$$\mu_t = \boldsymbol{\beta_2} + \beta_3[Q_t(s_t, go) - Q_t(s_t, nogo)]$$

Model 2 Traceplots

Group Parameters Traceplots



Individual Parameter (b1) Traceplots



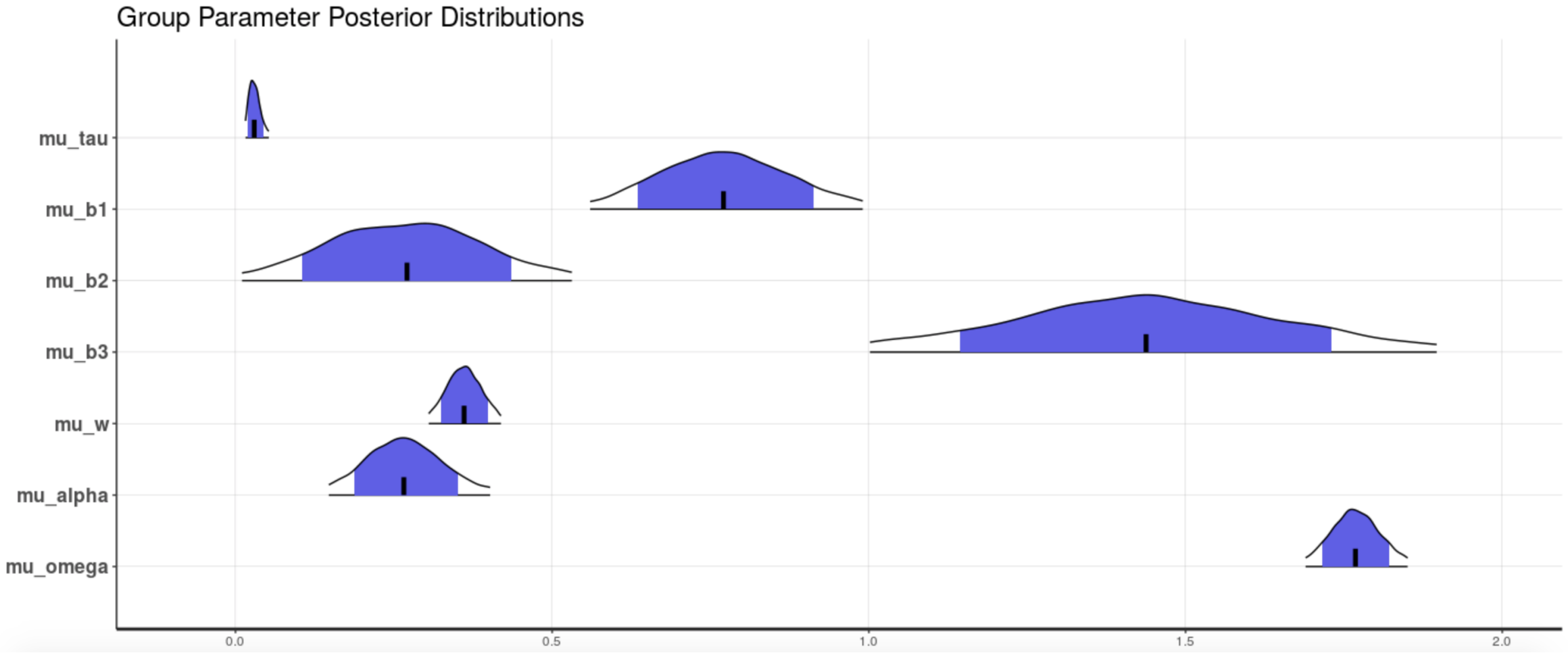
Group Parameters Rhat

mu_tau	mu_b1	mu_b2	mu_b3	mu_w	mu_alpha	mu_omega
1.001166	1.001342	1.004477	1.005206	1.002314	1.005245	1.005193

Individual Parameter (b1) Rhat

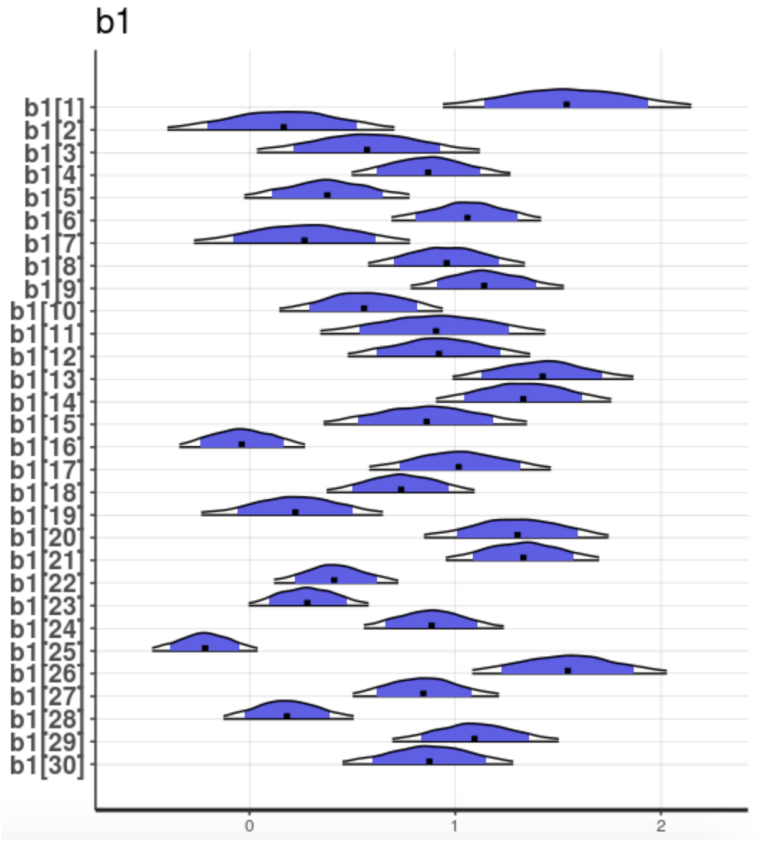
0.9998846	0.9993839	0.9991189	0.9993771	1.0003322	1.0001786
1.0005688	0.9996061	0.9999021	1.0000047	0.9998642	0.9998884
0.9996462	1.0007733	1.0001590	0.9995579	0.9999377	0.9993409
0.9995668	0.9994451	0.9993031	1.0001534	1.0003226	0.9996304
0.9999373	0.9997360	0.9998385	1.0002911	0.9994443	0.9992494

Model 2 Posterior Distributions

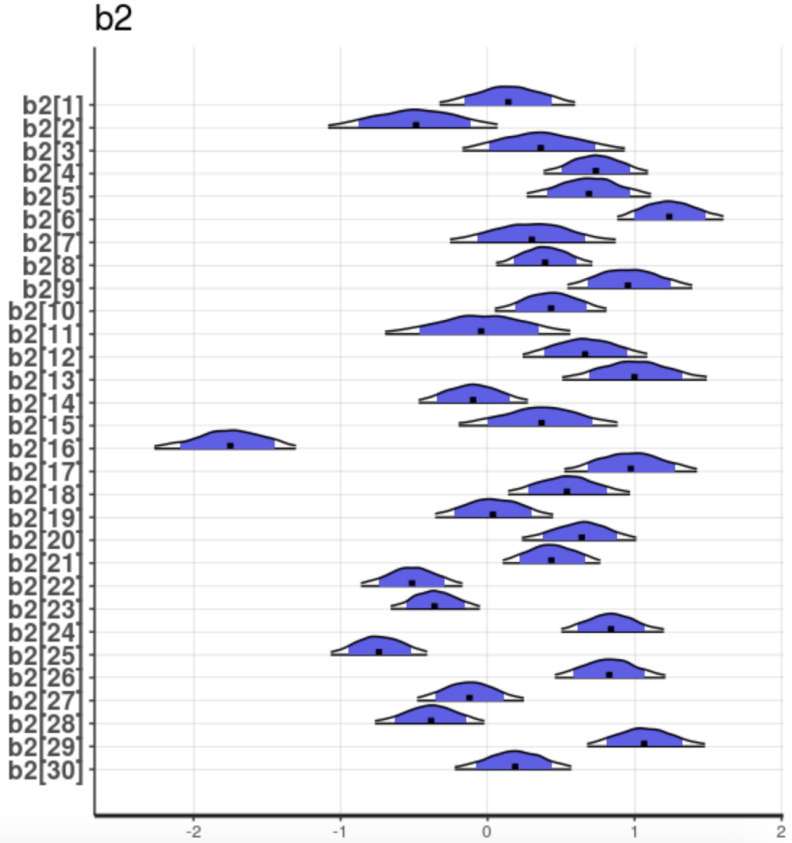


Model 2 Posterior Distributions

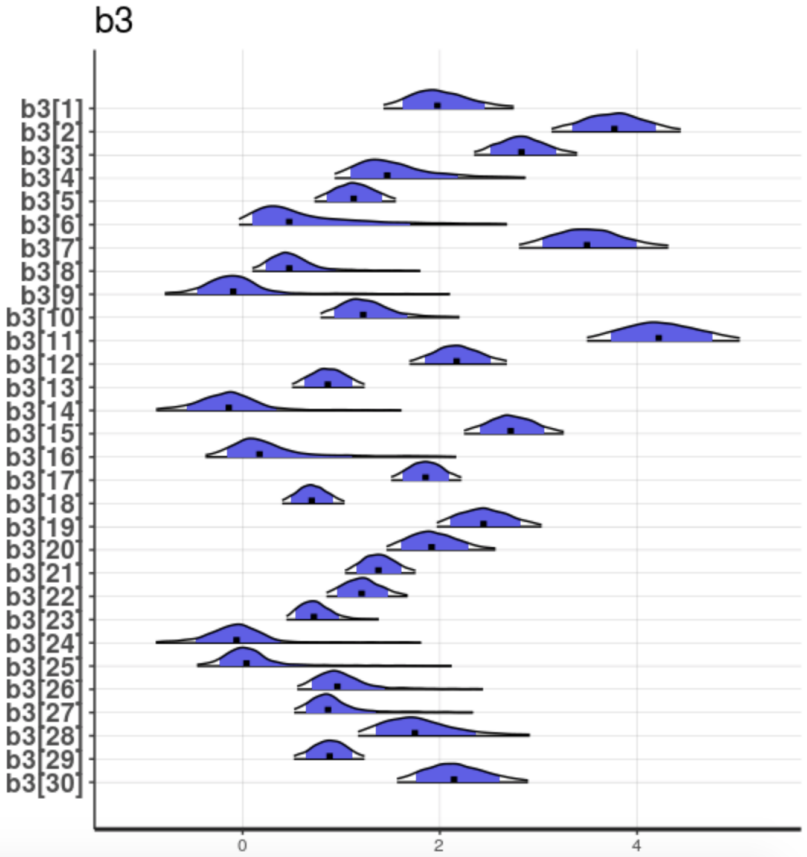
$\boldsymbol{\beta_1}$: Intercept for drift rate in escape



$\boldsymbol{\beta_2}$: Intercept for drift rate in avoid

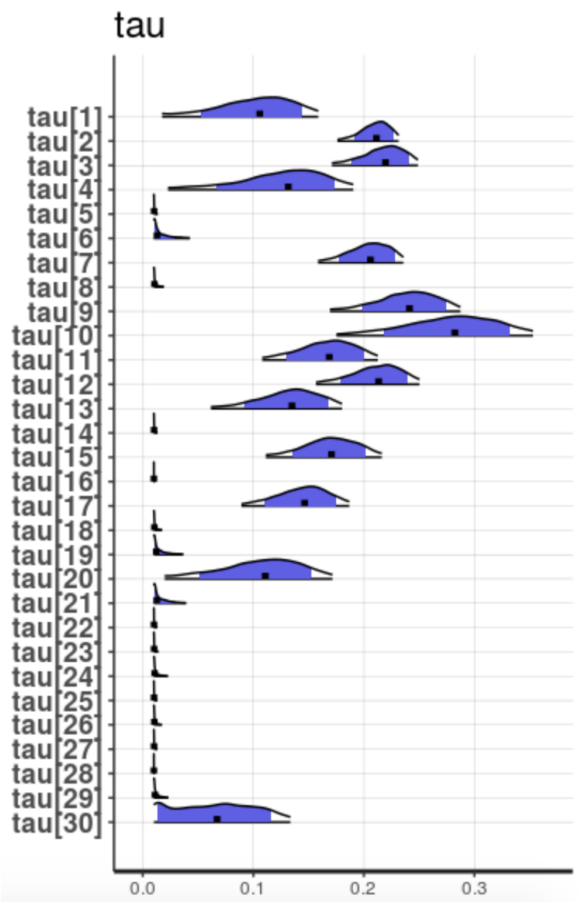


$\boldsymbol{\beta_3}$: Differential weight

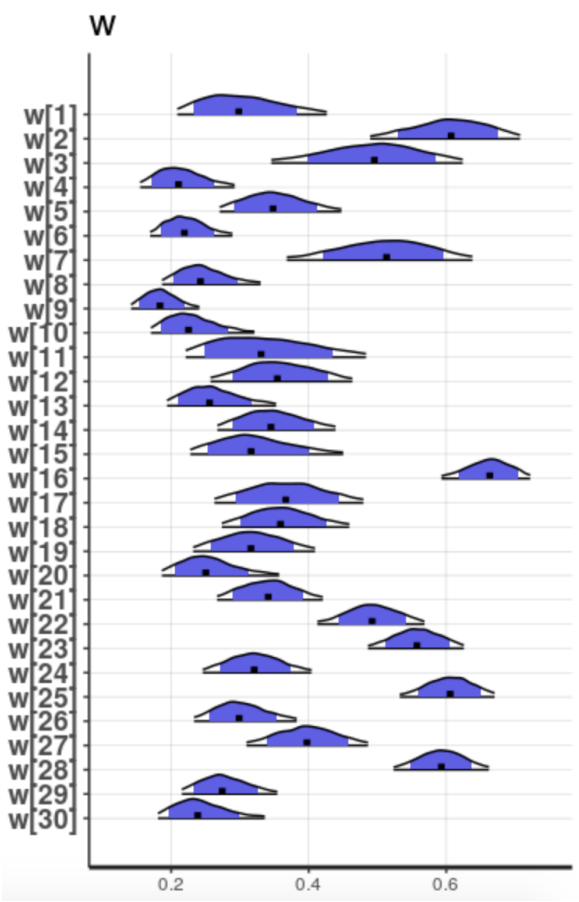


Model 2 Posterior Distributions

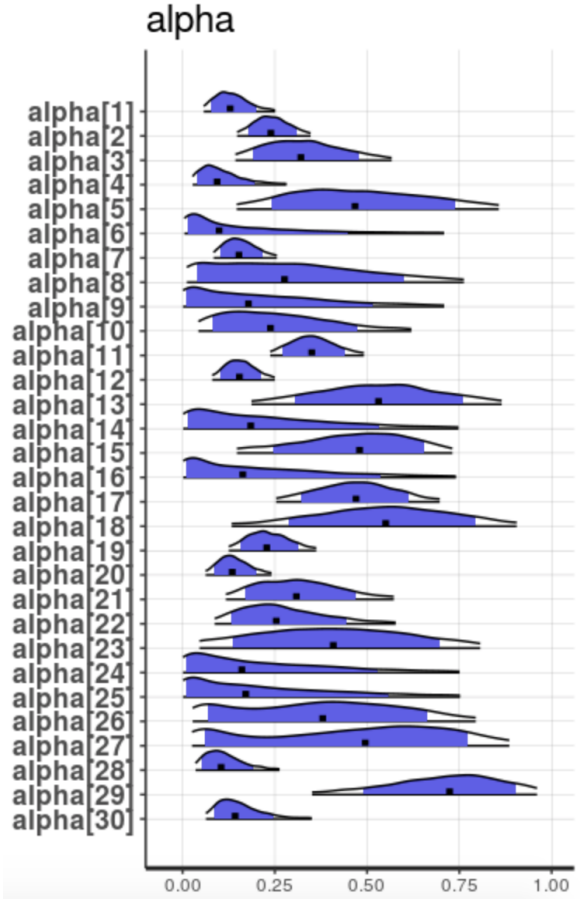
τ : Non-decision time



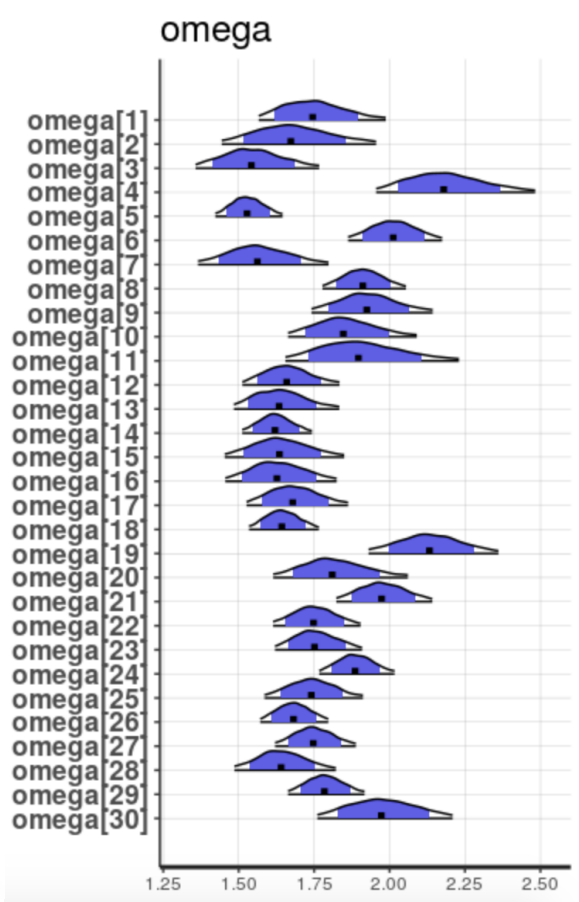
w : Starting point for both conditions



α : Learning rate



ω : Boundary separation



Model 2 Suicidal versus Non-suicidal

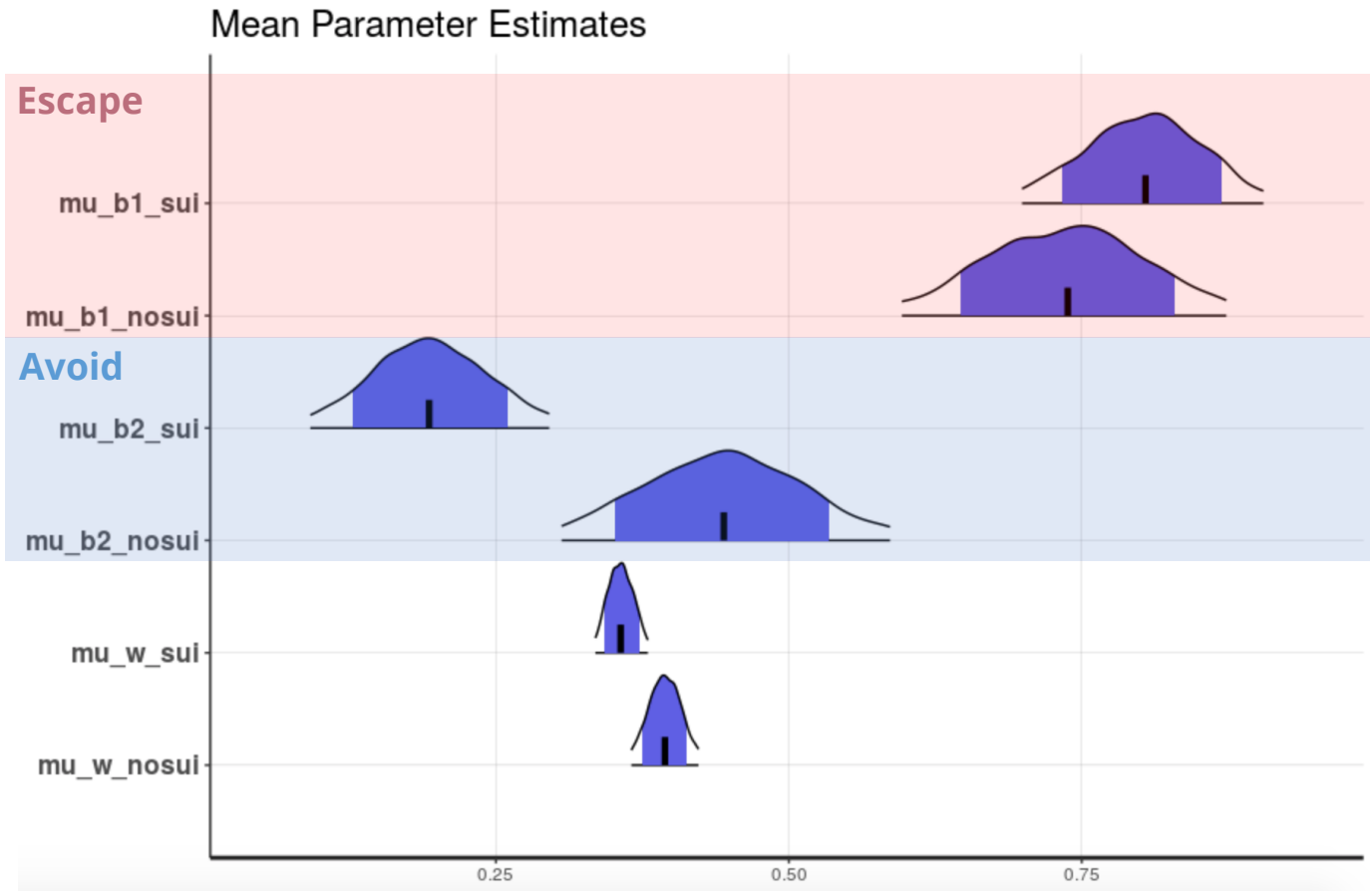


Table 2: Mean Estimated Parameter Values by Group

	mean	sd	Rhat
mu_b1_sui	0.8029664	0.0528404	0.9994440
mu_b1_nosui	0.7364864	0.0715878	0.9999966
mu_b2_sui	0.1934772	0.0517712	0.9994596
mu_b2_nosui	0.4440975	0.0719152	1.0005187
mu_b3_sui	1.4629532	0.1125108	1.0049626
mu_b3_nosui	1.5828252	0.1509983	1.0060614
mu_w_sui	0.3566904	0.0115995	0.9995703
mu_w_nosui	0.3942289	0.0146981	1.0002459

(mu = mean)

Model 2 Discussions

1. Sampling using Model 2 has been converged well
2. Model 2 is the second best model among three based on LOOIC with 30 subjects
3. In escape condition, the suicidal group has a higher drift rate on average, compared to the non-suicidal group
4. In avoid condition, the average drift rate of the suicidal group is lower than that of the non-suicidal group
5. The non-suicidal group shows a higher constant go-bias across conditions

Both Model (Model 3)

Model 3: Two drift rates + Two starting points

- For escape trials:

$$\mu_t = \boldsymbol{\beta_1} + \beta_3[Q_t(s_t, go) - Q_t(s_t, nogo)]$$

$$\mathrm{w} = \mathrm{w1}$$

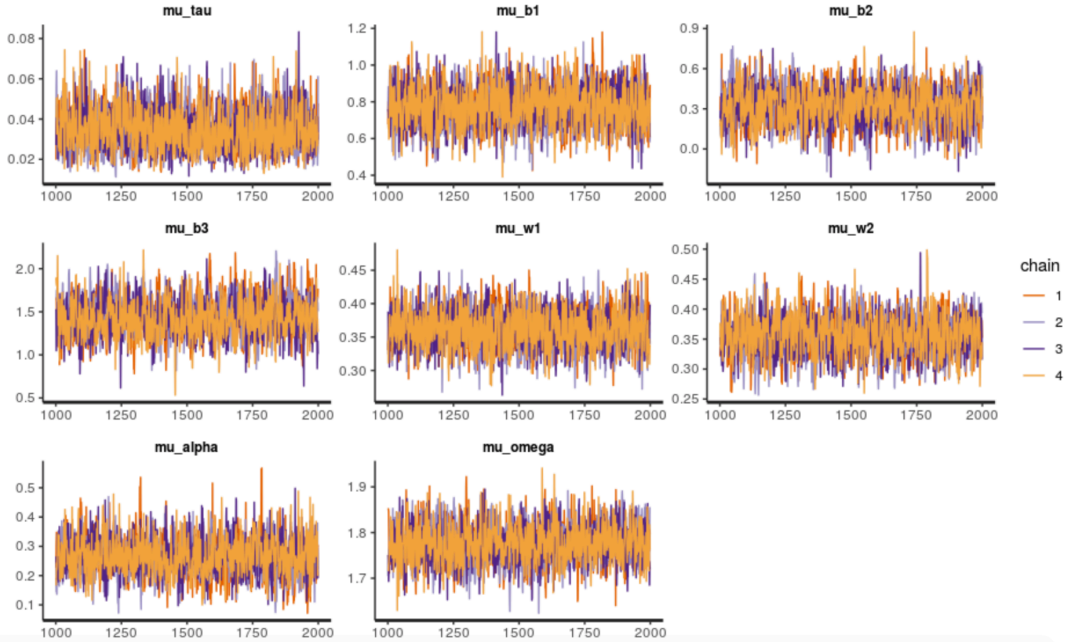
- For avoid trials:

$$\mu_t = \boldsymbol{\beta_2} + \beta_3[Q_t(s_t, go) - Q_t(s_t, nogo)]$$

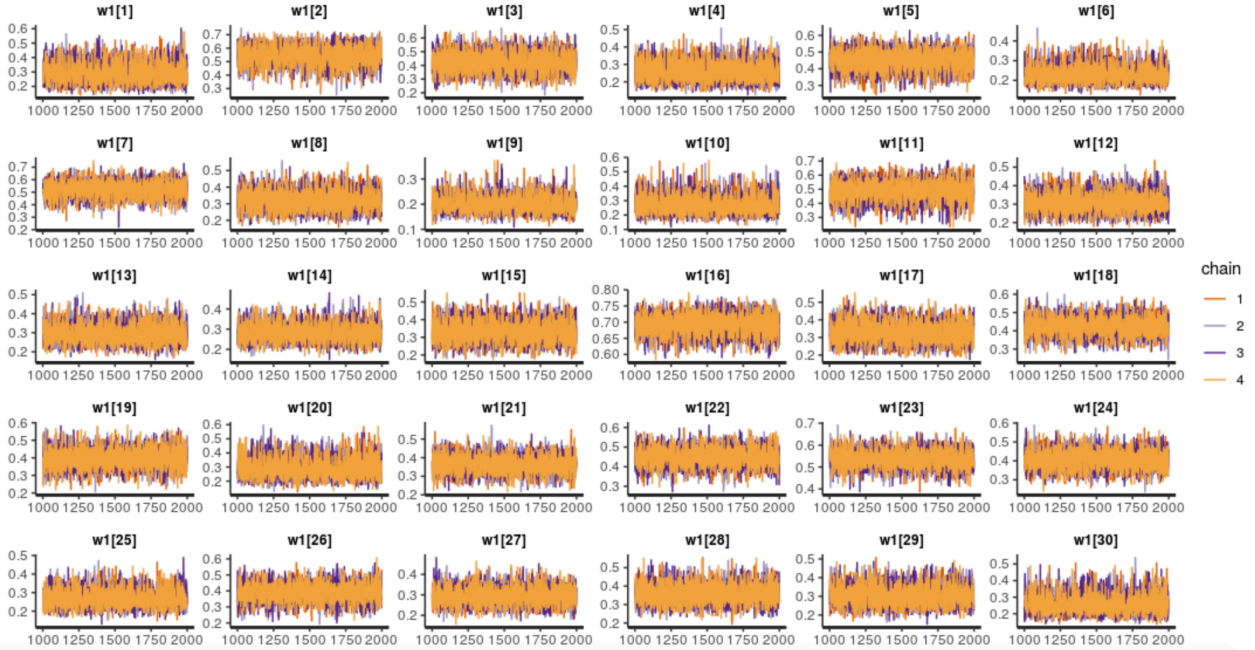
$$\mathrm{w} = \mathrm{w2}$$

Model 3 Traceplots

Group Parameters Traceplots



Individual Parameter (w1) Traceplots



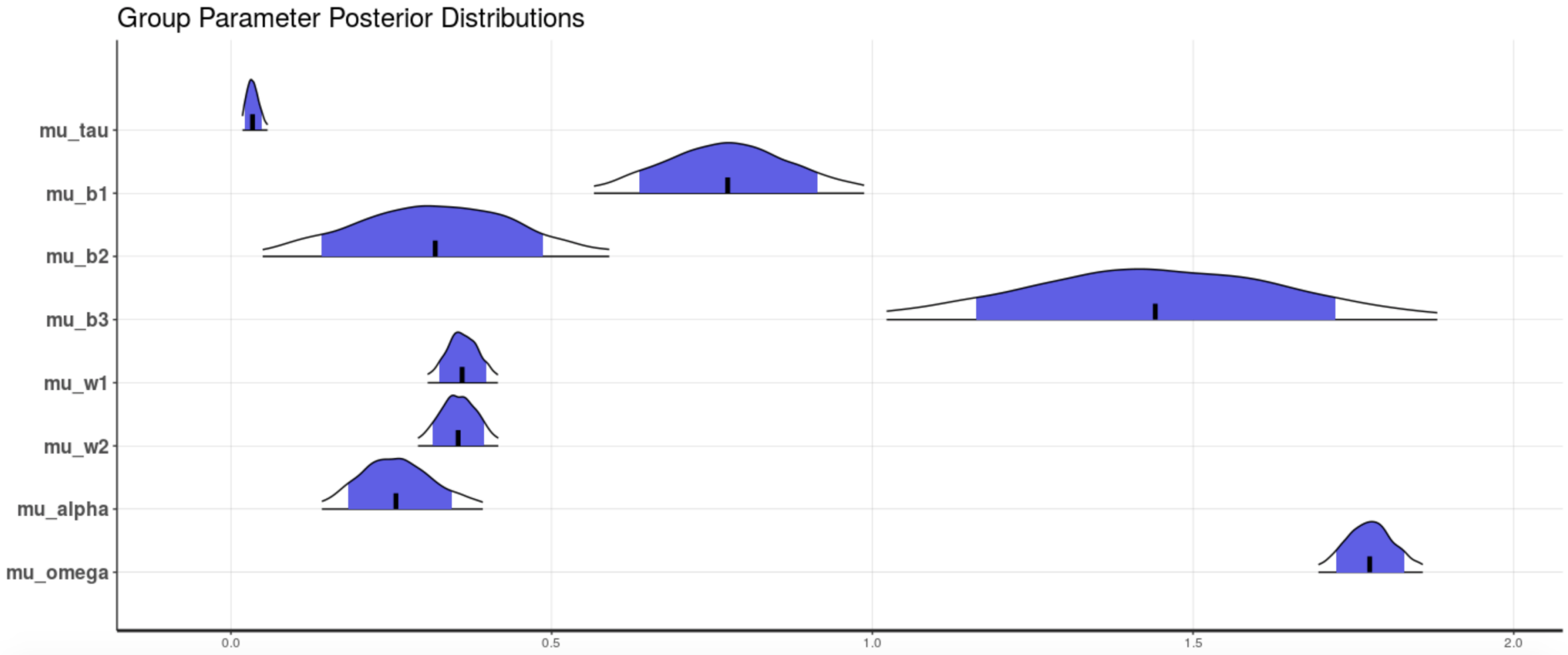
Group Parameters Rhat

mu_tau	mu_alpha	mu_omega	mu_b1	mu_b2	mu_b3	mu_w1	mu_w2
1.004038	1.010854	1.002062	1.001119	1.000251	1.012955	1.003401	1.003091

Individual Parameter (w1) Rhat

1.0022604	1.0006801	1.001479	1.0010816	1.000591	0.9995712	0.9996424	0.9995426	0.9994502	0.9996790
1.0003806	0.9995791	1.000540	0.9999075	1.000334	0.9998915	1.0006931	0.9997029	0.9992349	0.9999627
0.9993025	1.0006932	1.000011	0.9996544	1.000278	0.9994667	0.9995979	0.9998168	0.9996939	1.0003782

Model 3 Posterior Distributions



Model 3 Posterior Distributions

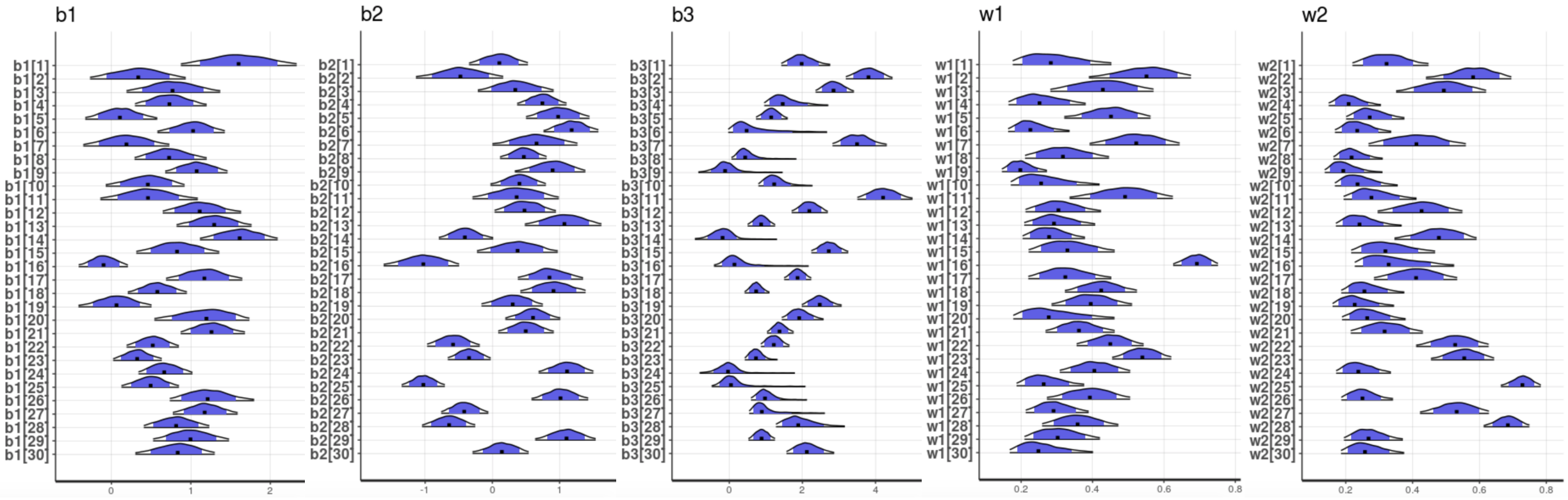
$\boldsymbol{\beta_1}$: Intercept of drift rate in escape condition

$\boldsymbol{\beta_2}$: Intercept of drift rate in avoid condition

$\boldsymbol{\beta_3}$: Differential weight

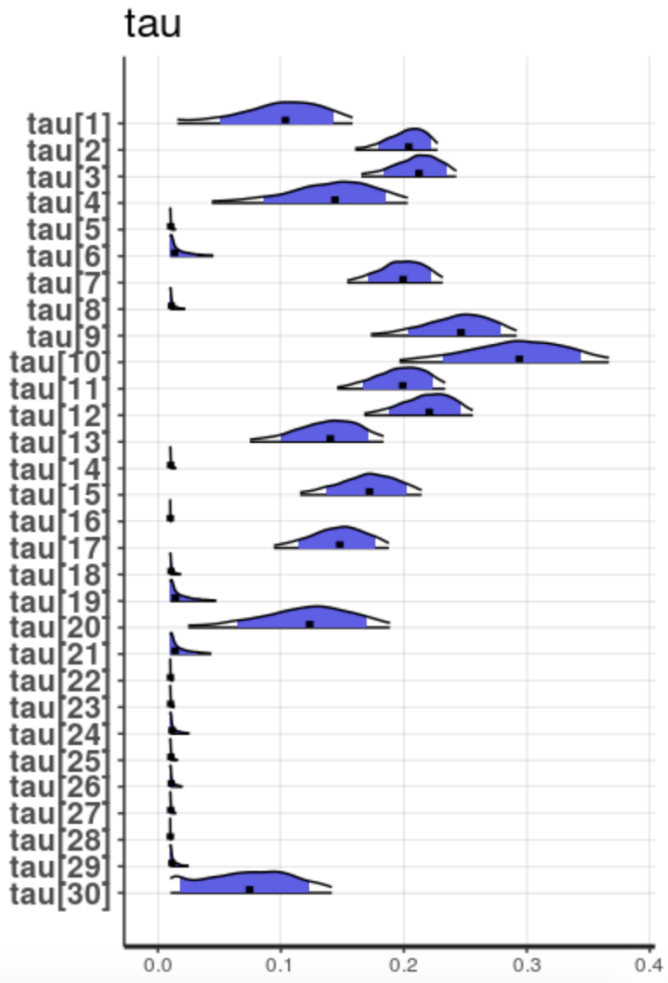
w1 : Starting point in escape condition

w2 : Starting point in avoid condition

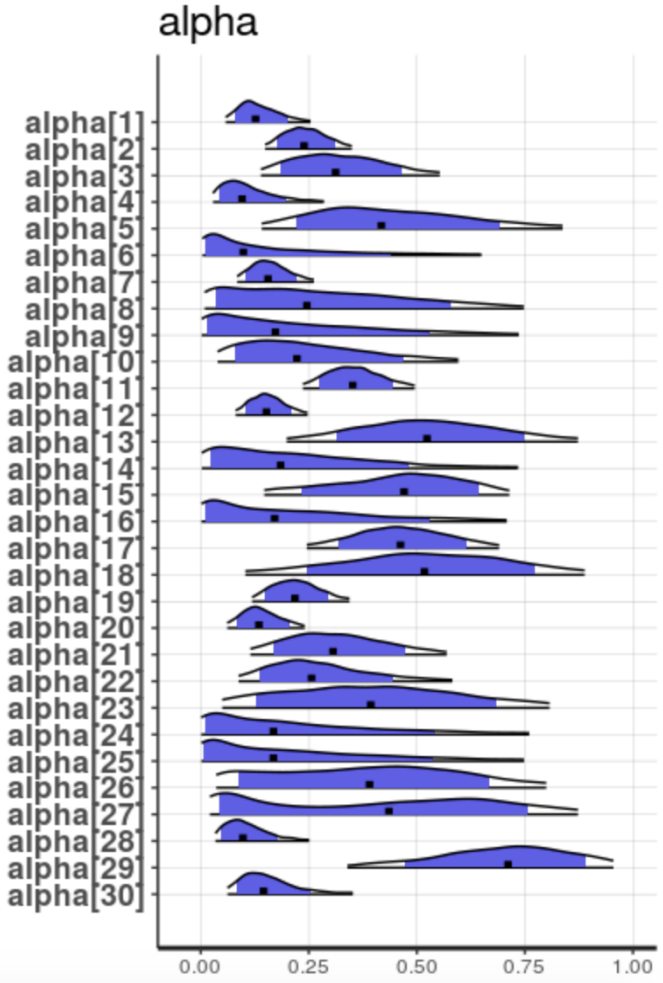


Model 3 Posterior Distributions

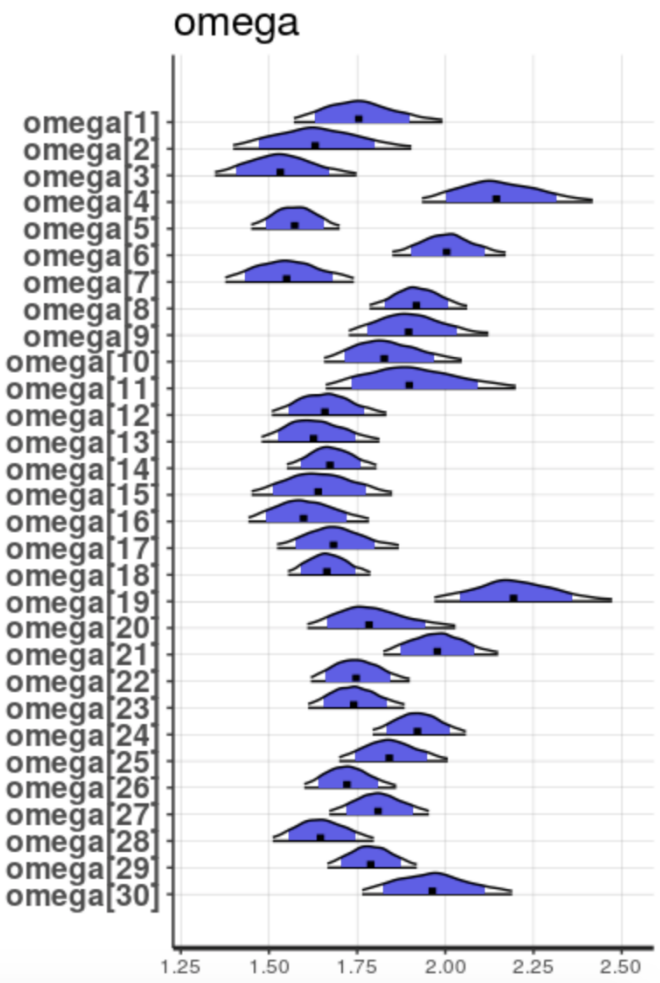
τ : Non-decision time



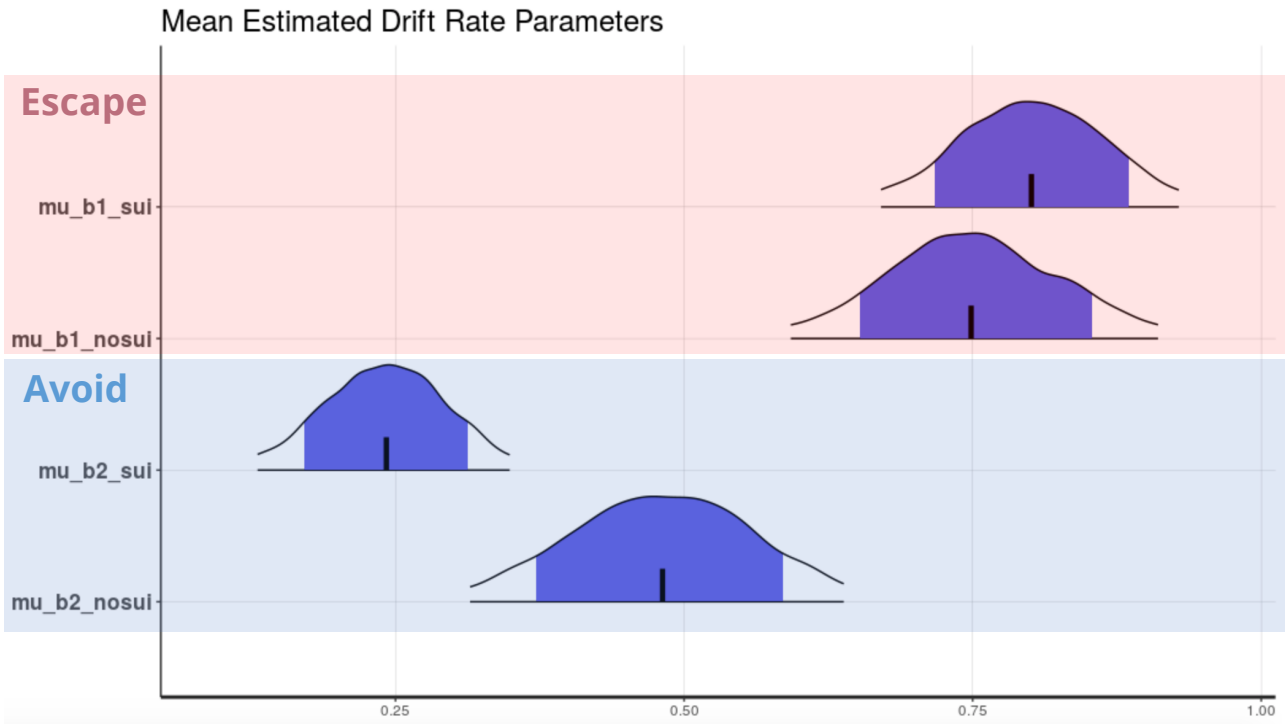
α : Learning rate

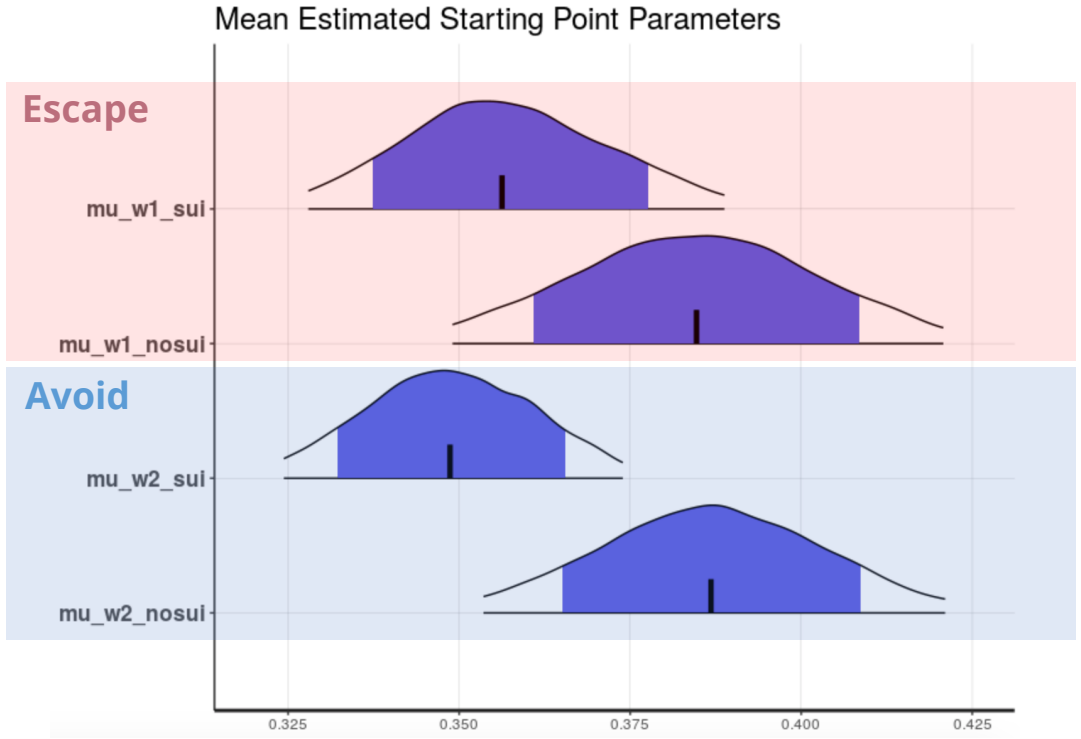


ω : Boundary separation



Model 3 Suicidal versus Non-suicidal





	mu_b1_sui	mu_b1_nosui	mu_b2_sui	mu_b2_nosui	mu_b3_sui	mu_b3_nosui
mean	0.8008688	0.7504390	0.2413617	0.4803858	1.4633480	1.598060
sd	0.0656118	0.0791603	0.0552296	0.0830168	0.1052431	0.137212

(mu = mean)

	mu_w1_sui	mu_w1_nosui	mu_w2_sui	mu_w2_nosui
mean	0.3570474	0.3846383	0.3488559	0.3869334
sd	0.0155943	0.0184014	0.0128842	0.0171818

(mu = mean)

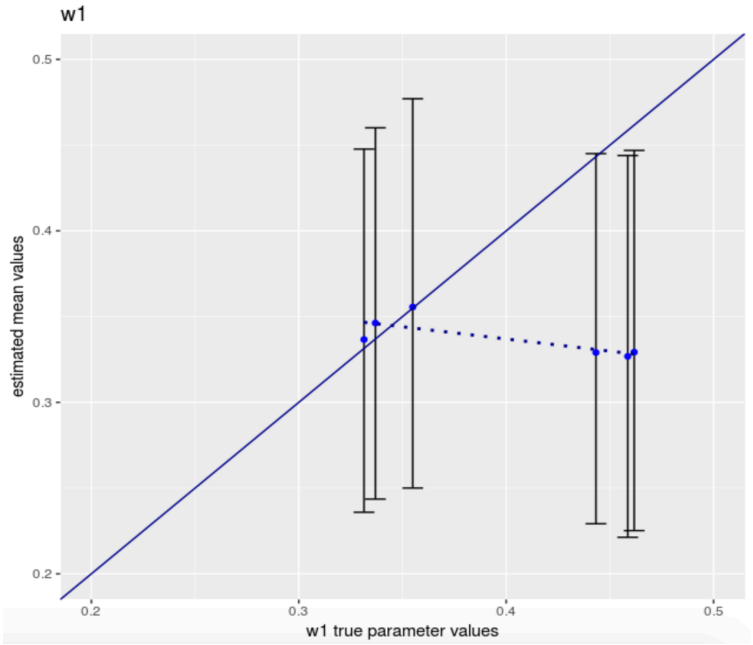
Model 3 Discussions

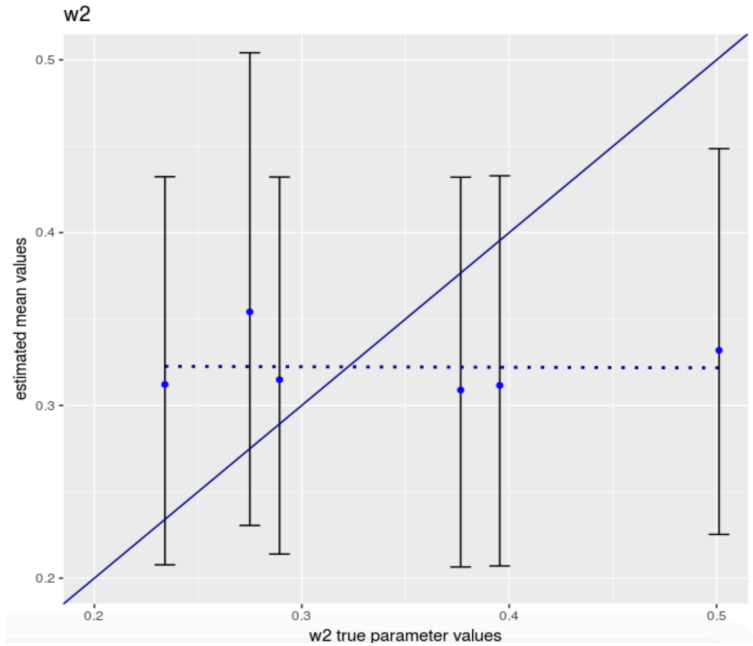
1. Sampling using Model 3 has been converged well
2. Model 3 is the best model among three based on LOOIC with 30 subjects
3. In escape condition, the drift rate of the suicidal group is higher than that of the non-suicidal group, whereas the trend is flipped in the avoid condition
4. ★ In both escape and avoid conditions, the non-suicidal group starts its drifting at a point closer to the go-boundary than the suicidal group does

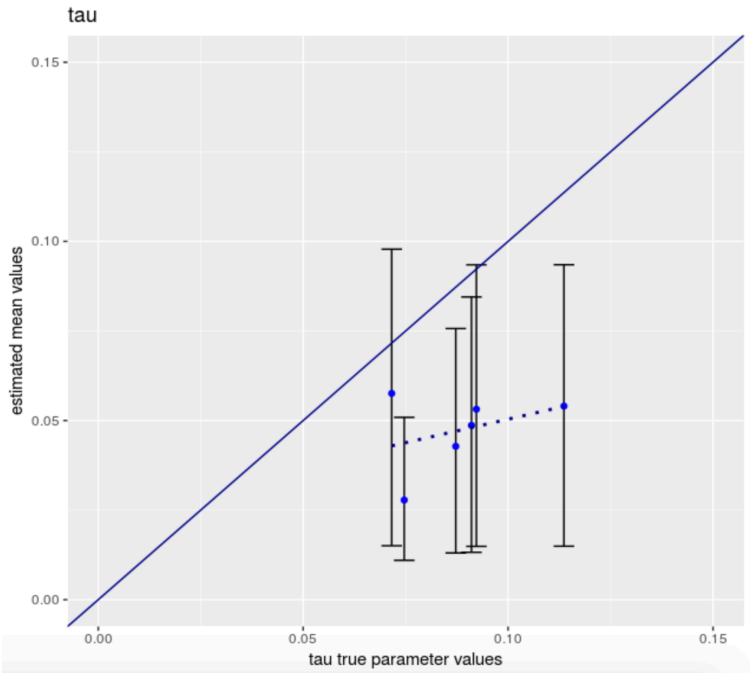
Parameter Recovery

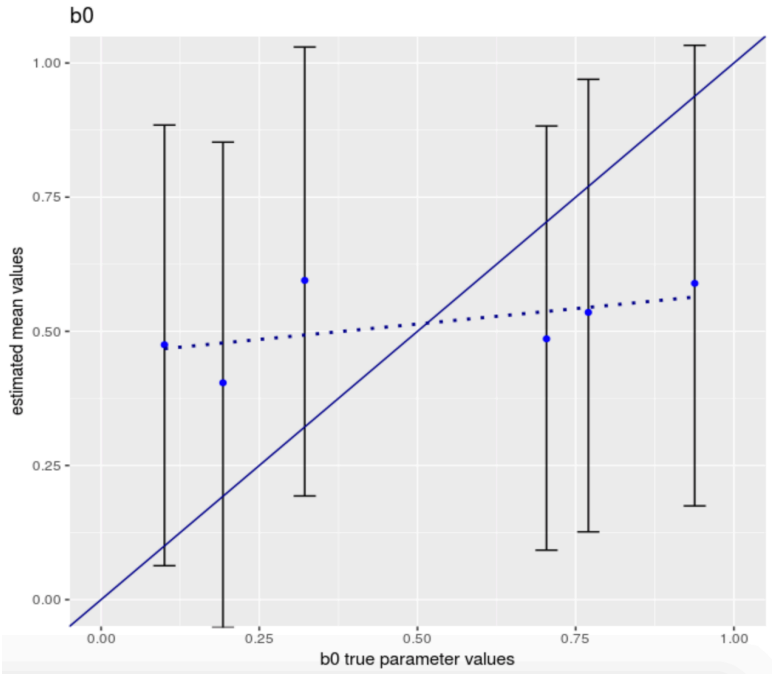
Original Model

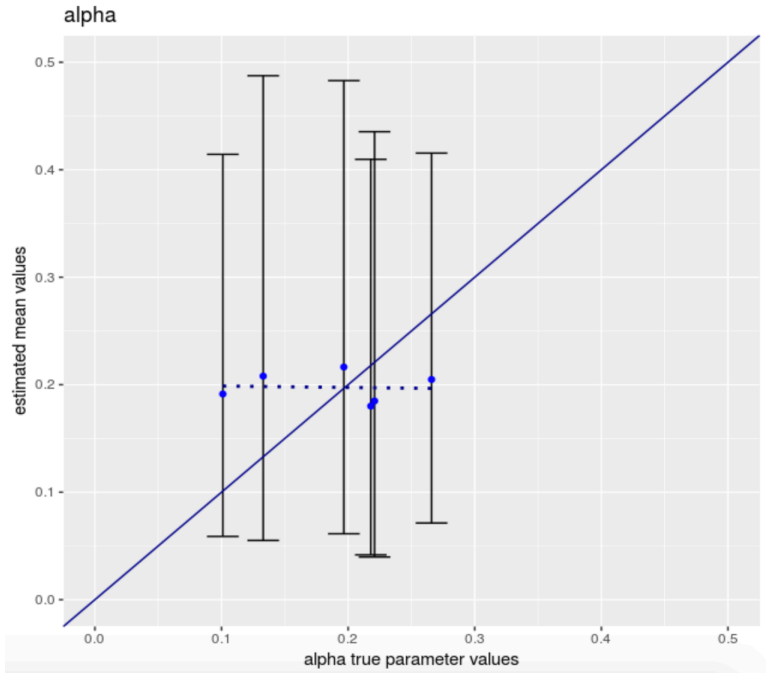
- Number of subjects = 6
- Step size = 0.02
- Iterations = 2000
- Warmup = 1000
- Chains = 4



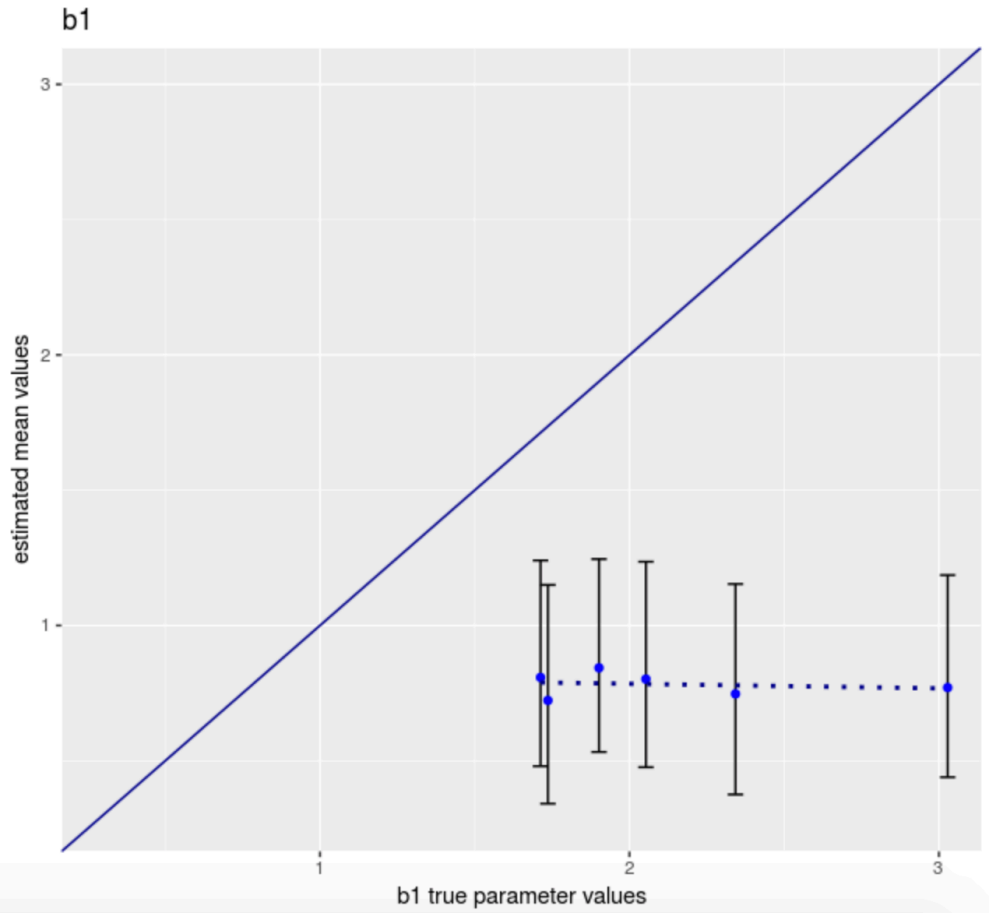


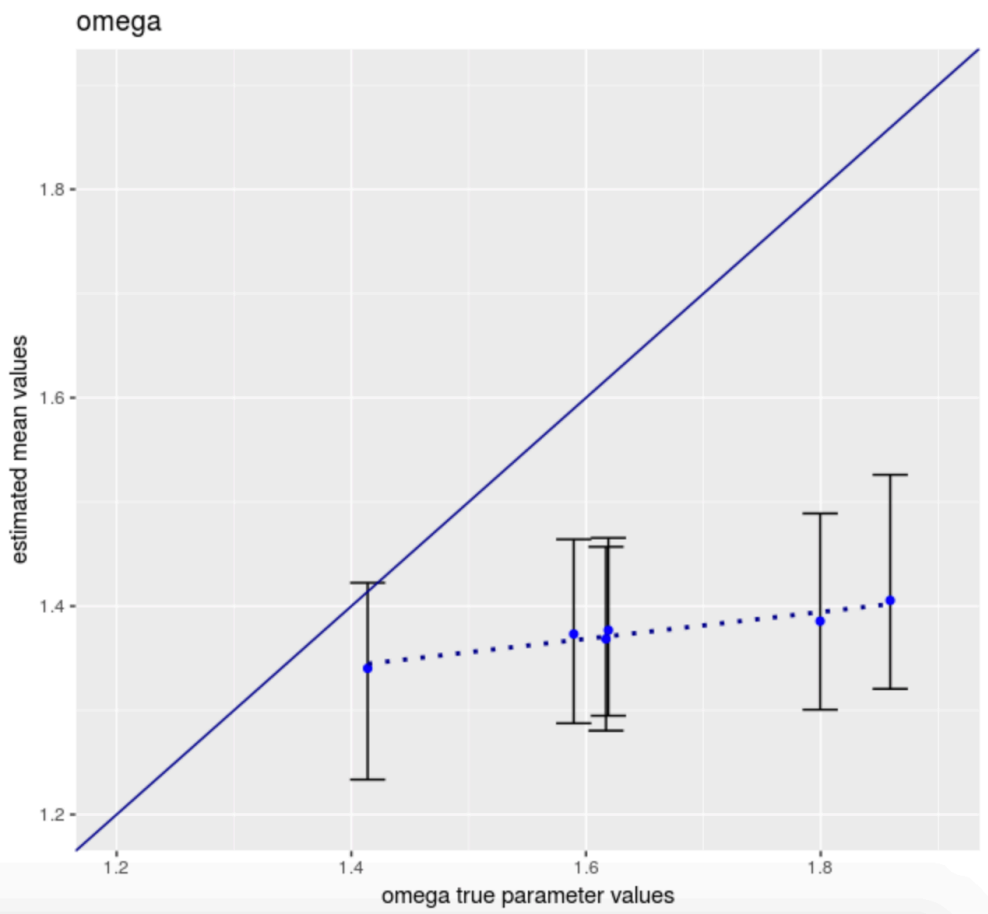






Original Model



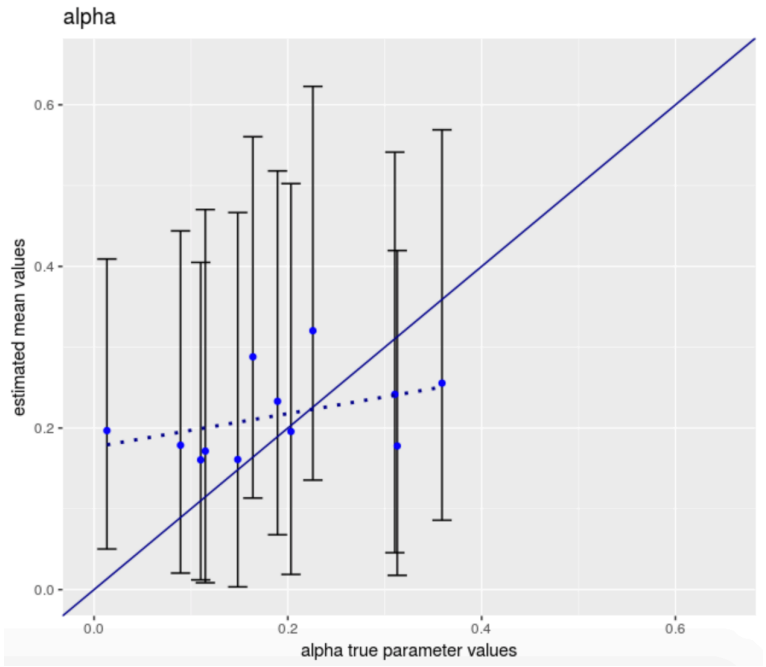


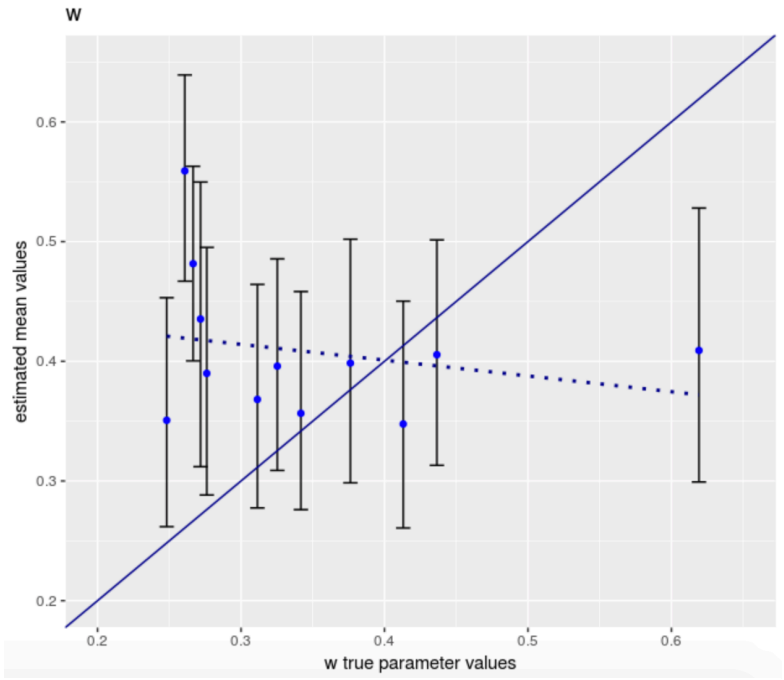
<u>Possible reasons</u> :

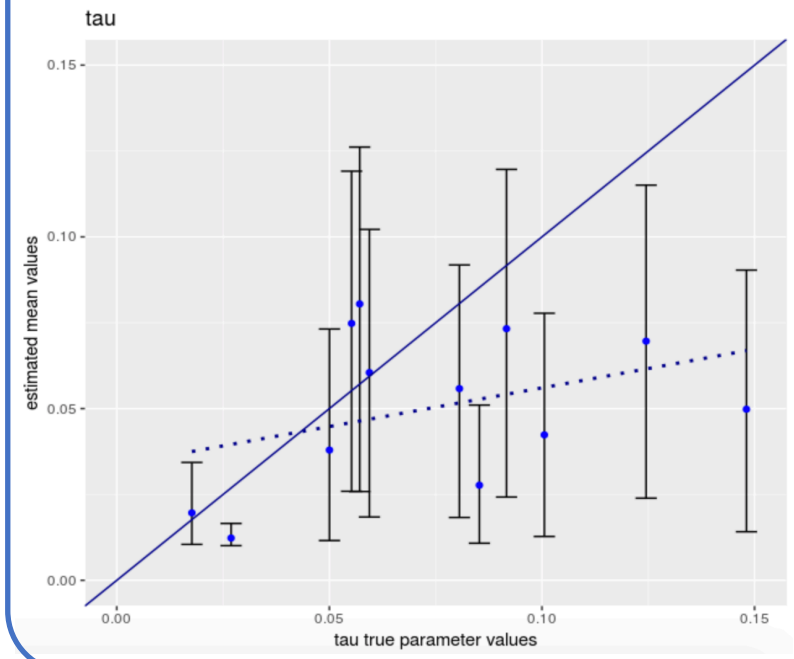
- No inverse temperatures when generating the choice dataset
- Model is too flexible

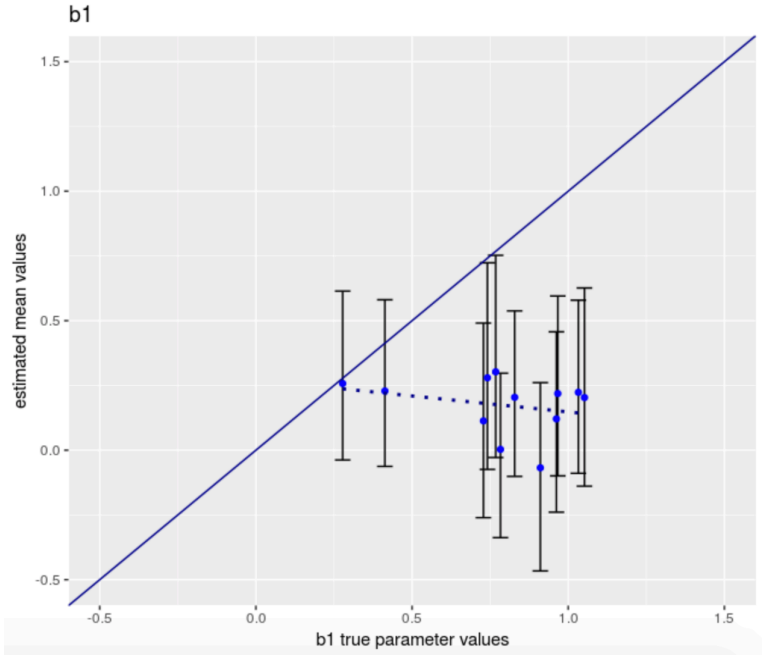
Model 2

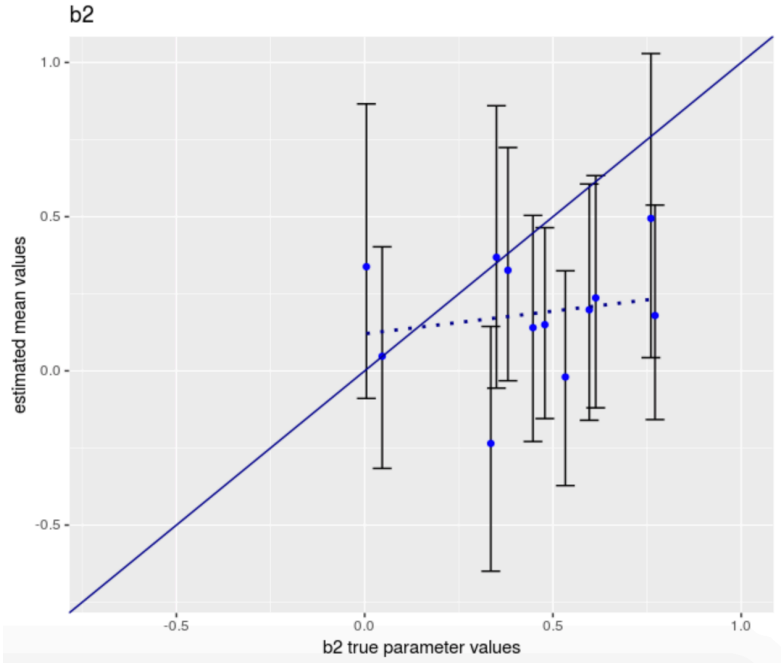
- Number of subjects = 12
- Step size = 0.04
- Iterations = 2000
- Warmup = 1000
- Chains = 4



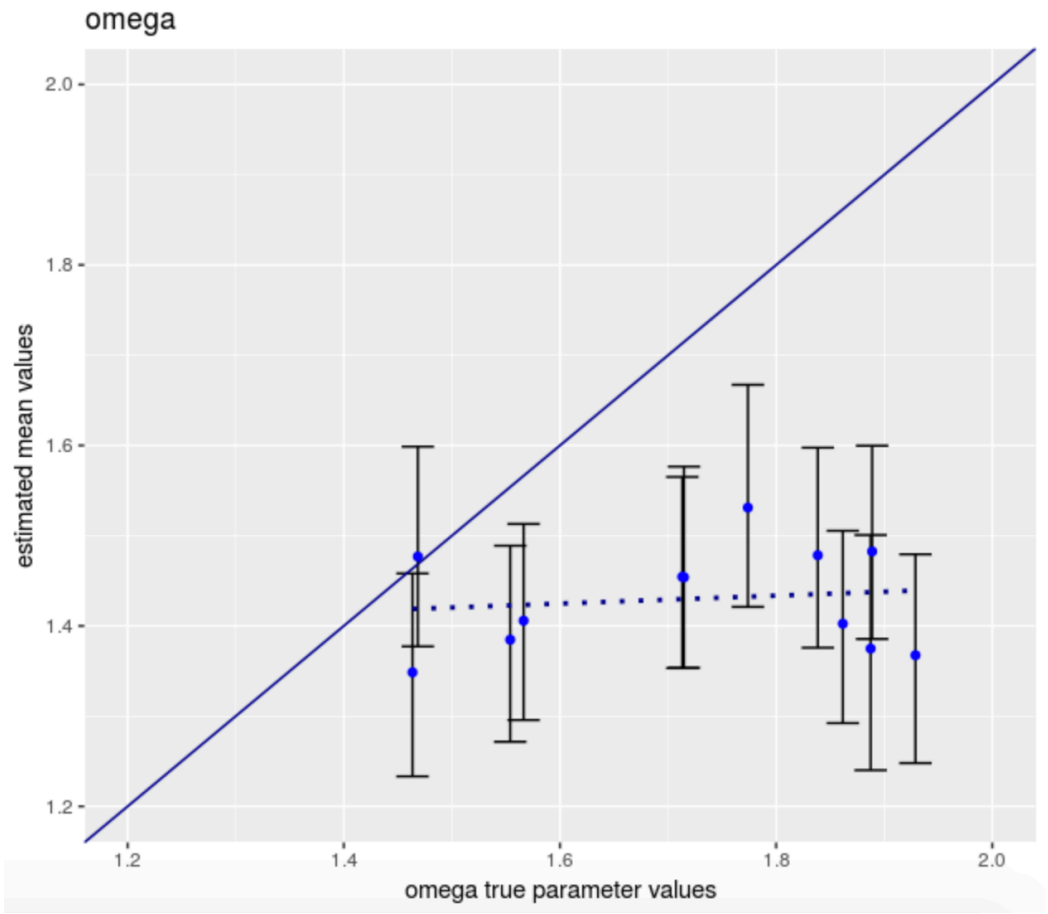


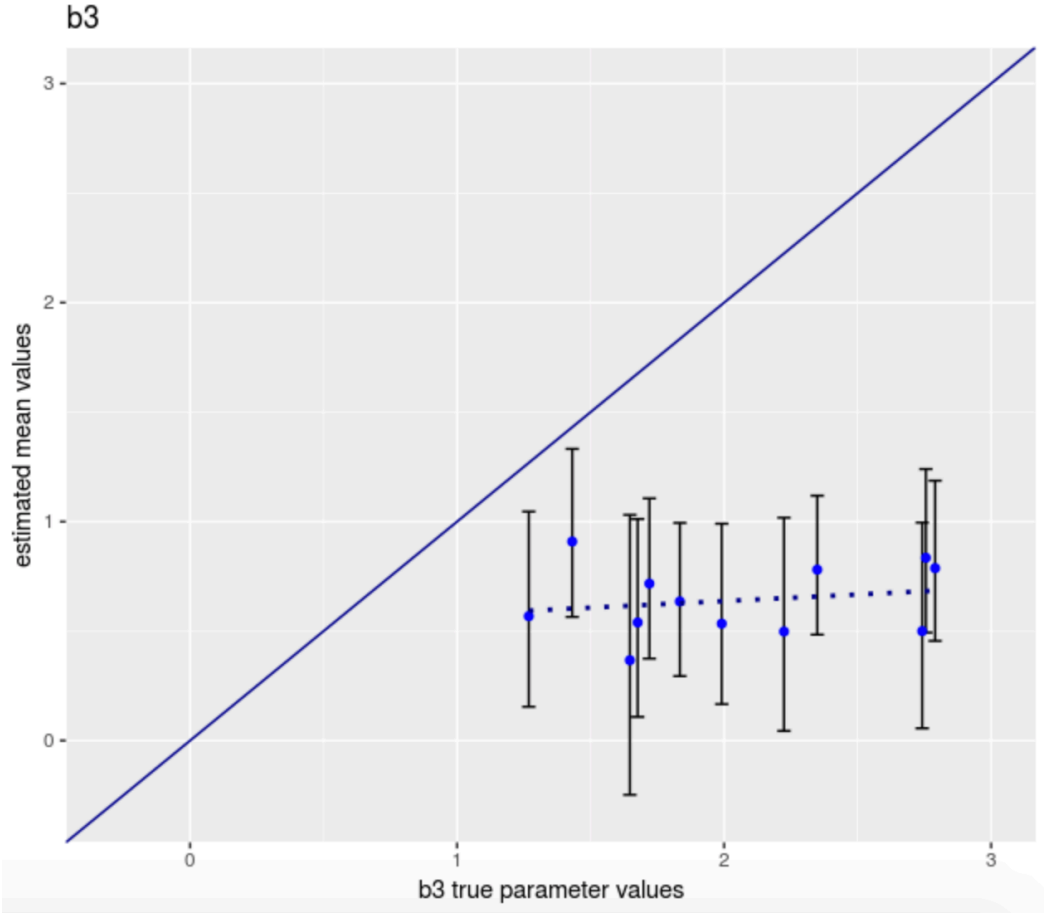






Model 2



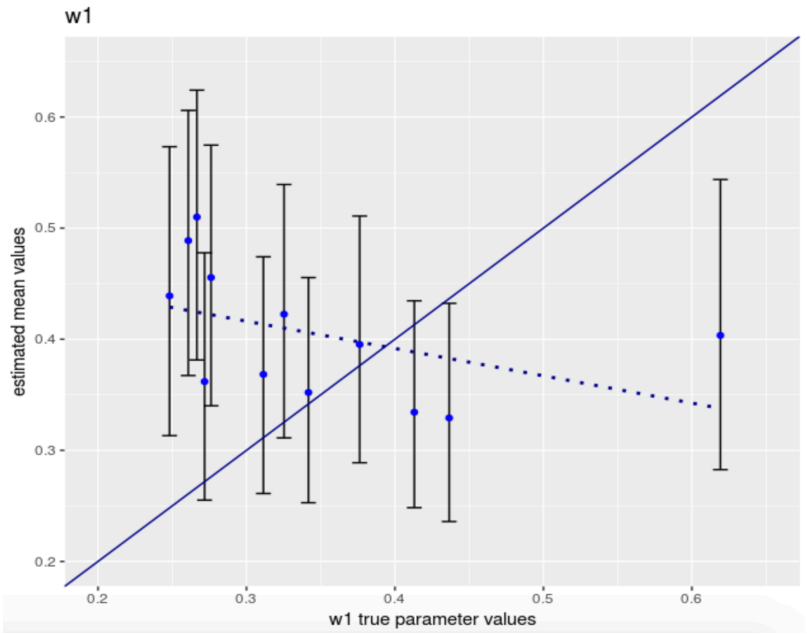


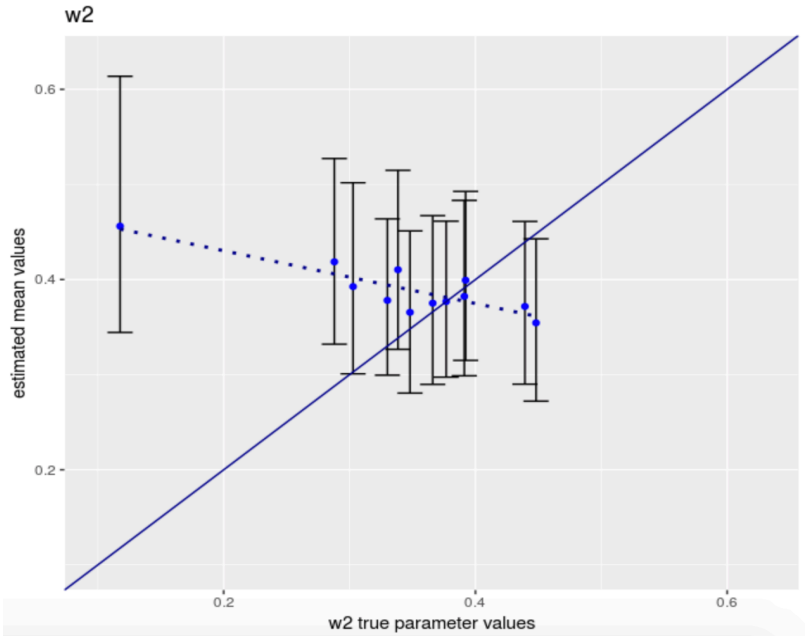
<u>Possible reasons</u> :

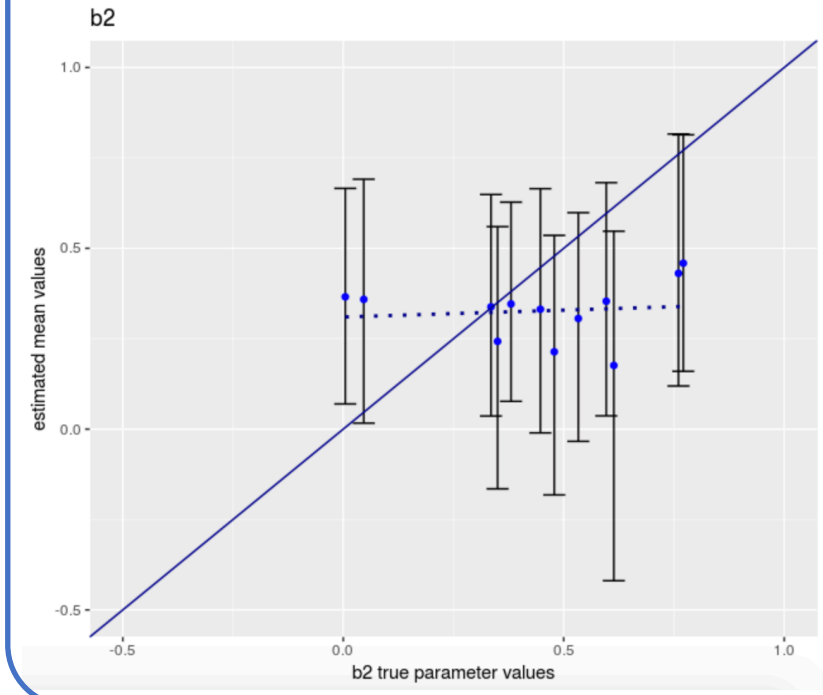
- No inverse temperatures when generating the choice dataset
- Model is too flexible

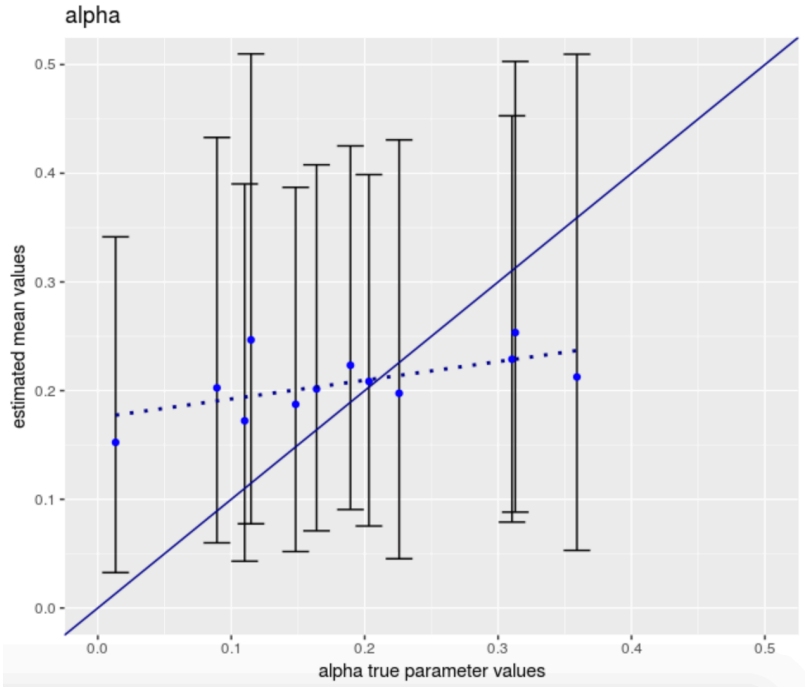
Model 3

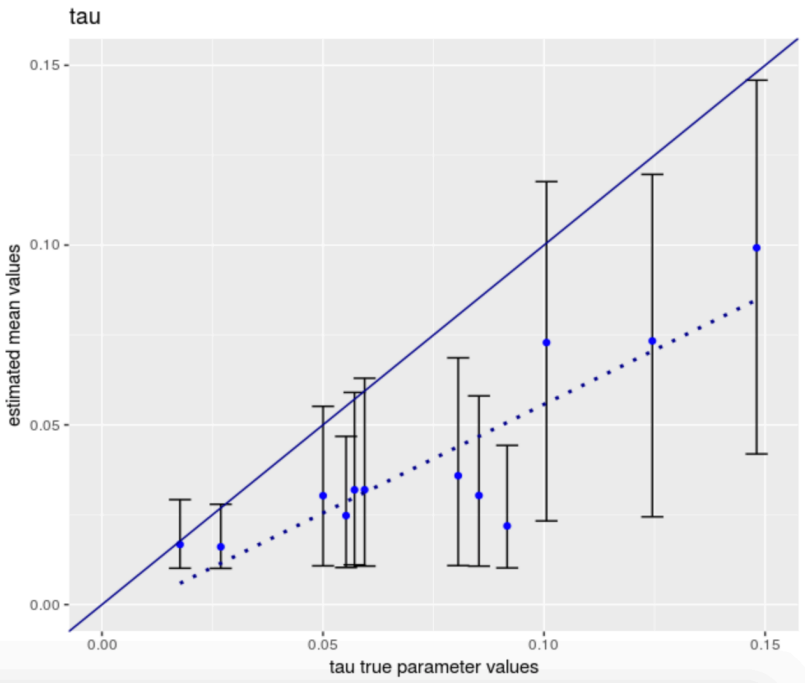
- Number of subjects = 12
- Step size = 0.04
- Iterations = 2000
- Warmup = 1000
- Chains = 4



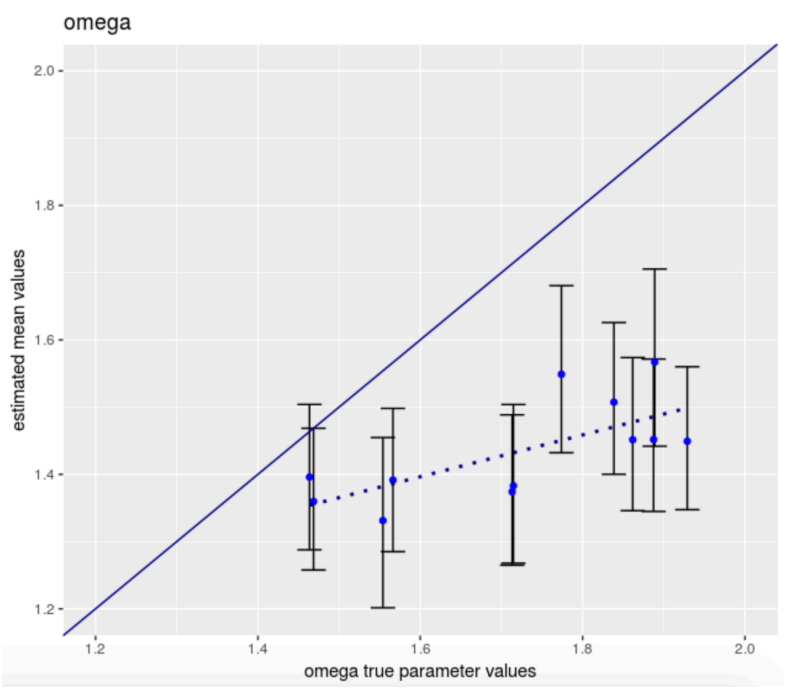


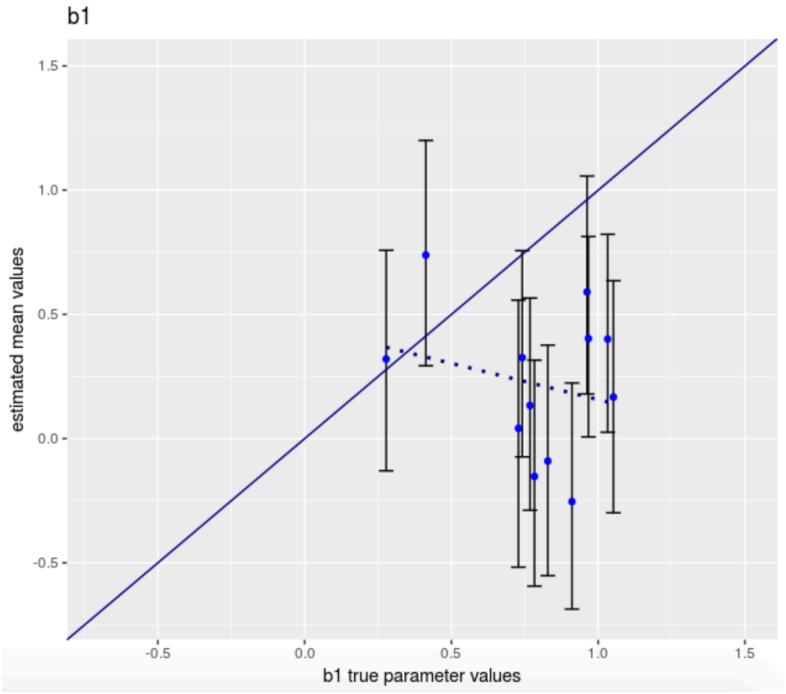


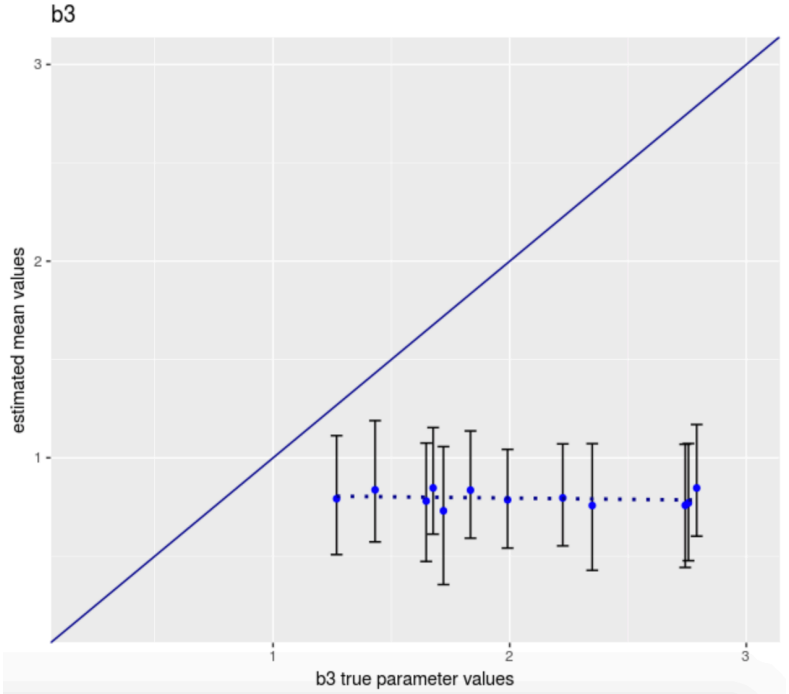




Model 3







<u>Possible reasons</u> :

- No inverse temperatures when generating the choice dataset
- Model misspecification for drift rate
- Badly simulated data (using an inappropriate seed for random number generator)

Conclusions

- The suicidal group tends to drift faster to the go-boundary in escape condition, but the non-suicidal group tends to have a constant go bias in both conditions
- ★ The suicidal group also shows a stronger no-go to avoid response bias in avoid condition
- Model 3 with two drift rates and two starting points is the best among three based on LOOIC with 30 subjects
- Not successful parameter recovery of the differential weight parameter (β_1 or β_3) in all three models
- Fair parameter recovery of the rest of the parameters in all three models

References

- Millner, A. J., den Ouden, H. E., Gershman, S. J., Glenn, C. R., Kearns, J. C., Bornstein, A. M., Marx, B. P., Keane, T. M., & Nock, M. K. (2019). Suicidal thoughts and behaviors are associated with an increased decision-making bias for active responses to escape aversive states. *Journal of Abnormal Psychology,* 128(2):106-118.

- Millner, A. J., Gershman, S. J., Nock, M. K., & den Ouden, H. E. (2018). Pavlovian control of escape and avoidance. *Journal of Cognitive Neuroscience,* 1–12.