The impact of prior preference and item popularity on social learning

Naeun Oh, Seongjae Park

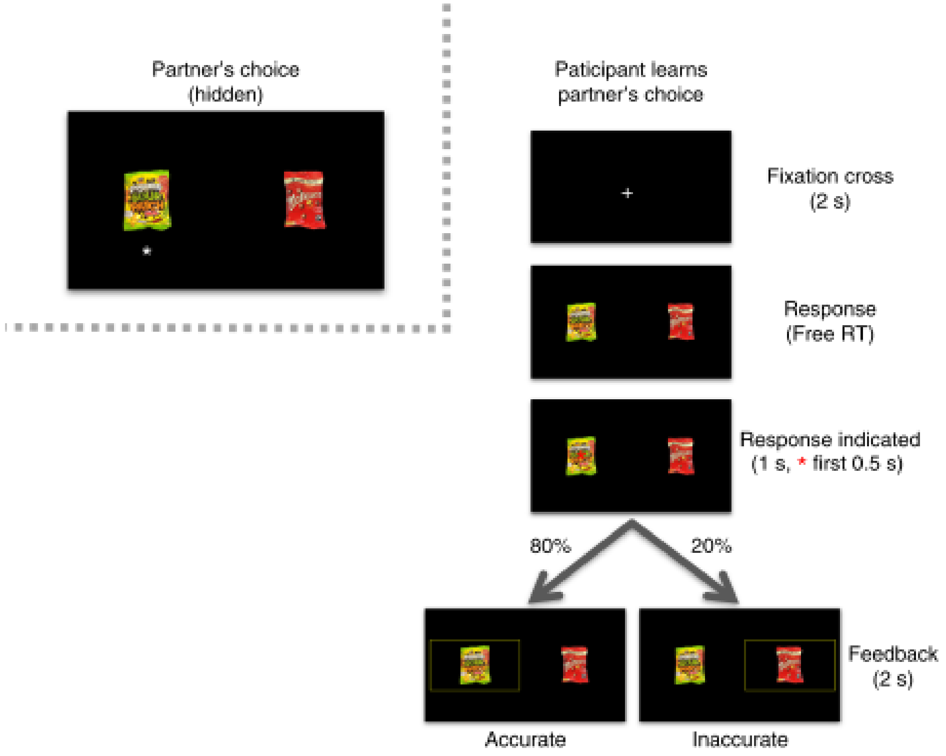
Abstract

Tarantola et al.(2017) found that our personal preferences affect the way we learn the preferences of other people. Through computational models combining inter-trial Bayesian learning and intra-trial choice process, they found the effects of participants' preferences on both the learning and choice process. The preferences were reflected on the learning process through the influence of priors and on the choice process through the influence of the choice bias. These effect generalized to non-social learning experiment.

When they modeled the influence of relative item popularity on the prior in the Bayesian learning process, they could find that only the participants in the social learning experiment additionally benefit by using their knowledge about the popularity of certain preferences.

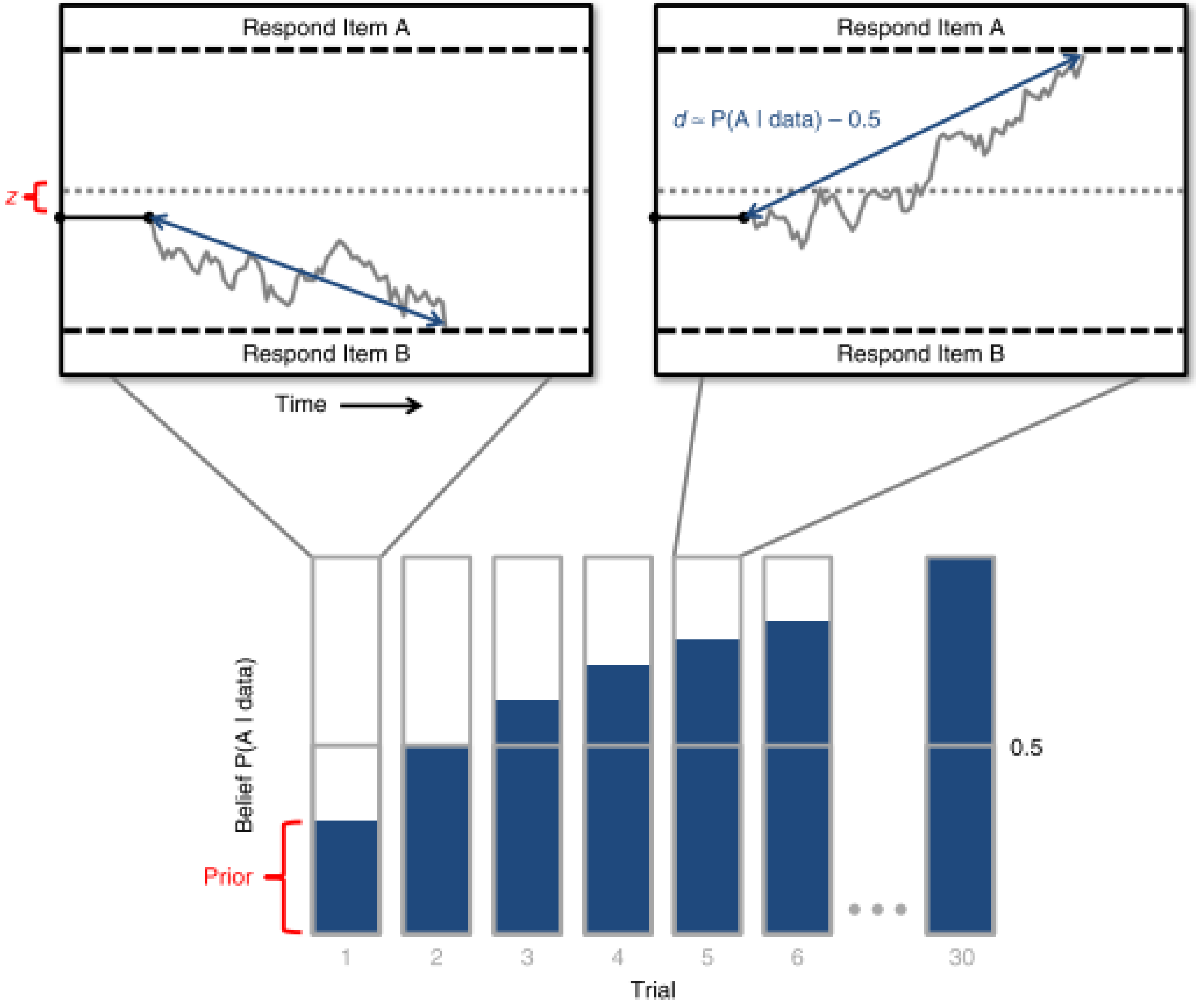
However, they didn't take into account the possible influence of item popularity on the choice bias. Therefore, we modified the author's code and fitted two new models.

Social Learning Task



- Social Group: learned the preference of virtual partner
- Non-Social Group: learned the correct item in each of the same 20 pairs of snacks.
- Preference: the amount of money the participant bid more for the correct item than the alternative in each pair

Intra-trial: Drift Diffusion Model + Inter-trial: Bayesian Learning



- Inter-trial component:
  - The prior belief about the correct answer in the first trial is influenced by the participant's preference and item popularity.
  - Prior belief in each trial, except the first trial, is updated by given feedback so far
- Intra-trial component:
  - For each participant and each item pair, the prior belief in each trial is converted into the drift rate by the drift weight parameter.
  - For each participant, the choice bias(z) in each pair is influenced by the participant's preference and item popularity in each participant.
  - The non-decision time is different for each participant
  - The threshold is different for each participant

Hierarchical Bayesian Parameter Estimation – 5 Models

- **Null model** (Social + Non-social)
  - (w/o the influence of preference and popularity)

- **Dual influence model** (Only Social)
  - (only w/ the influence of preference on both prior and choice bias)

$$P(A)_{n=1} = \frac{1}{1 + e^{-\beta_{\Delta\nu}\Delta\nu}}$$

$$z = \frac{1}{1 + e^{-\kappa\Delta\nu}}$$

Hierarchical Bayesian Parameter Estimation – 5 Models

- **Dual influence + popularity prior** (Social + Non-social)
  - (w/ the influence of both preference and item popularity on prior, and only the influence of preference on the choice bias)

$$P(A)_{n=1} = \frac{1}{1 + e^{-(\beta_{\Delta v}\Delta v + \beta_{\rho}\rho)}}$$

- **Dual influence + popularity bias** (Only Social)
  - (w/ the influence of both preference and item popularity on choice bias, and only the influence of preference on the prior)

$$z = \frac{1}{1 + e^{-(\kappa \Delta v + \beta_{\rho}\rho)}}$$

- **Dual influence + popularity dual** (Only Social)
  - (w/ the influence of both preference and item popularity on both prior and choice bias)

Hierarchical Bayesian Parameter Estimation – 7 Parameters

- **Non_decision_time_int**
  - Individual non-decision-time parameter
- **Threshold_int_mean**
  - Group mean threshold parameter
- **Drift_rate_learning_mean**
  - Group mean drift weight parameter
- **Cong_weight_prior_mean**
  - Group mean weight parameter used to convert preference into the prior
- **Cong_weight_drift_bias_mean**
  - Group mean weight parameter used to convert preference into the choice bias
- **Insight_mean**
  - Group mean weight parameter used to convert item popularity into the prior
- **Insight_bias_mean**
  - Group mean weight parameter used to convert item popularity into the choice bias

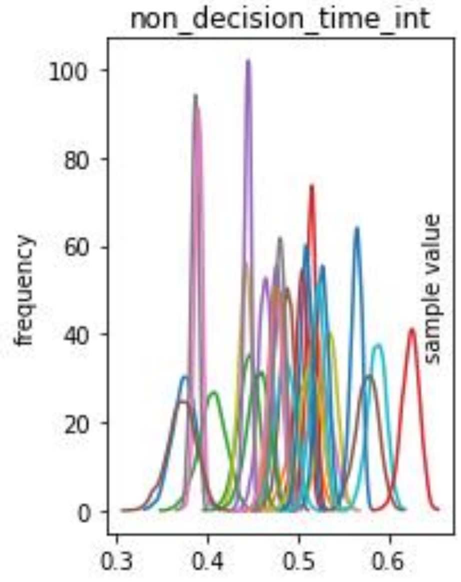
Initial Value Setting for Iteration

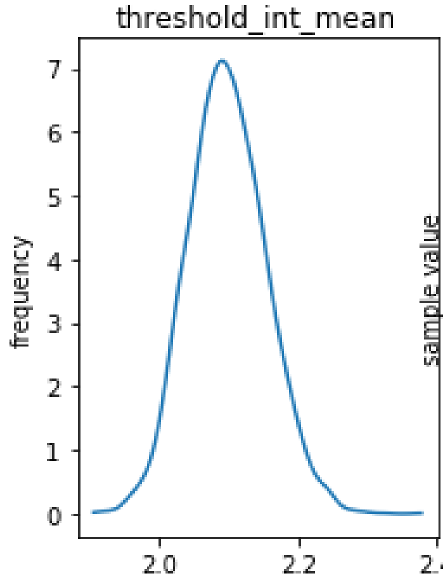
- In the case of dual influence model, we constrained the initial values of parameters based on range of prior distributions.
- When the initial value of parameter ‘threshold’ was randomly selected, the sampling process did not converge allowing sampled value to asymptotically approach infinity.
- For all other models, we used random initial values for all parameters.

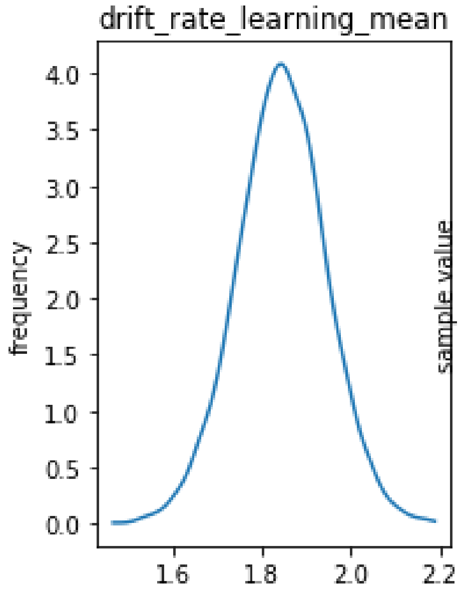
Null social

```
Inference for Stan model: model_dual_insight_priorbias_bayesian_3f4b0e8861fe35442899151acfda9b9d.
4 chains, each with iter=2000; warmup=1000; thin=1;
post-warmup draws per chain=1000, total post-warmup draws=4000.
```

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
threshold_int_mean	2.13	7.8e-4	0.05	2.03	2.09	2.13	2.17	2.24	4956	1.0
threshold_int_sd	0.3	6.0e-4	0.04	0.23	0.27	0.3	0.33	0.4	5096	1.0
drift_rate_learning_mean	1.86	4.8e-3	0.1	1.66	1.79	1.86	1.93	2.05	425	1.01
drift_rate_learning_sd	0.54	2.6e-3	0.07	0.41	0.49	0.53	0.58	0.71	852	1.0



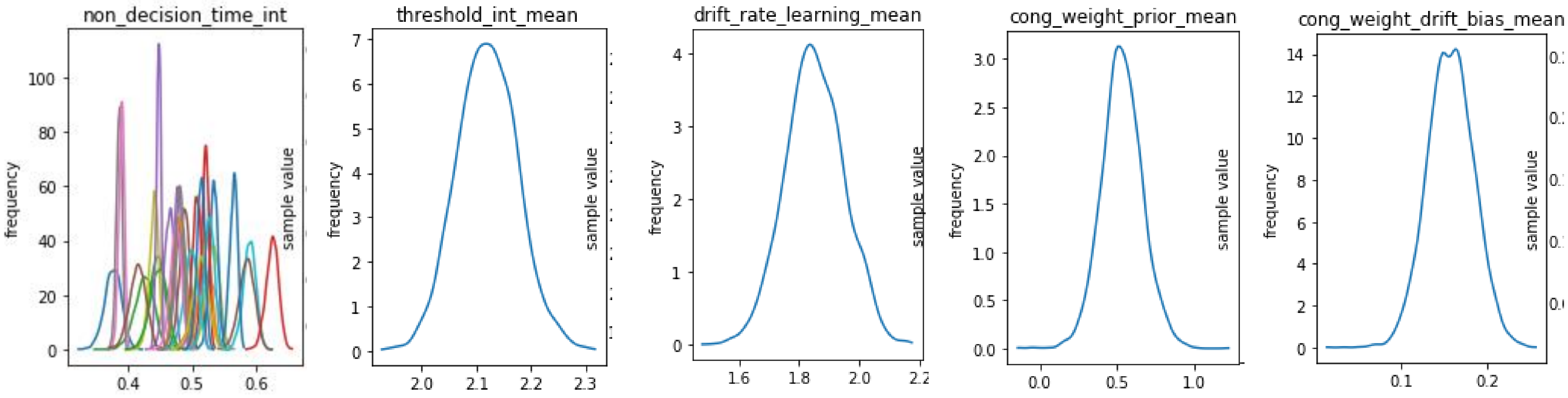




Dual influence social

Inference for Stan model: model_dual_bayesian_0e83ec402f54bbe4b85063383c9d295c.
4 chains, each with iter=2000; warmup=1000; thin=1;
post-warmup draws per chain=1000, total post-warmup draws=4000.

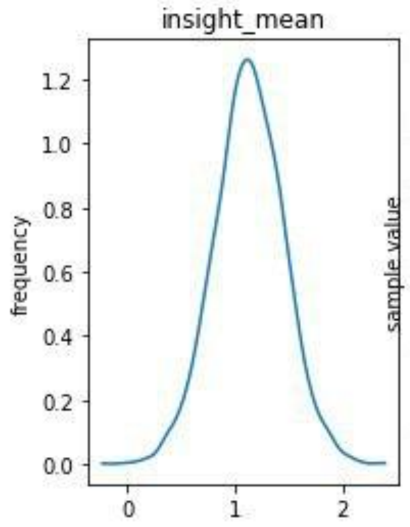
	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
threshold_int_mean	2.12	7.2e-4	0.06	2.01	2.08	2.12	2.16	2.23	5877	1.0
threshold_int_sd	0.3	6.4e-4	0.04	0.23	0.27	0.3	0.33	0.4	4733	1.0
drift_rate_learning_mean	1.85	4.0e-3	0.1	1.66	1.79	1.85	1.92	2.04	581	1.01
drift_rate_learning_sd	0.53	2.0e-3	0.07	0.41	0.48	0.52	0.57	0.69	1251	1.0
cong_weight_prior_mean	0.53	3.4e-3	0.13	0.28	0.45	0.53	0.62	0.81	1557	1.0
cong_weight_prior_sd	0.64	4.3e-3	0.15	0.4	0.54	0.63	0.74	0.98	1234	1.0
cong_weight_drift_bias_mean	0.16	6.5e-4	0.03	0.1	0.14	0.16	0.17	0.21	1728	1.0
cong_weight_drift_bias_sd	0.13	6.3e-4	0.02	0.09	0.12	0.13	0.15	0.19	1428	1.0

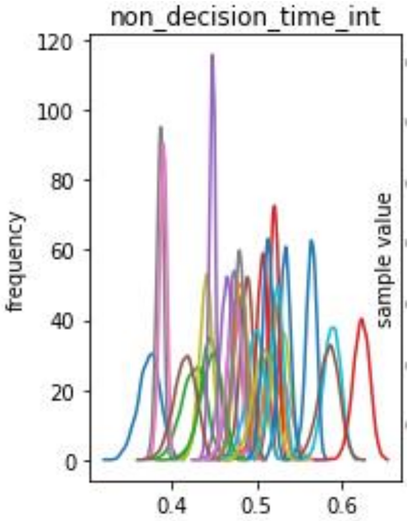


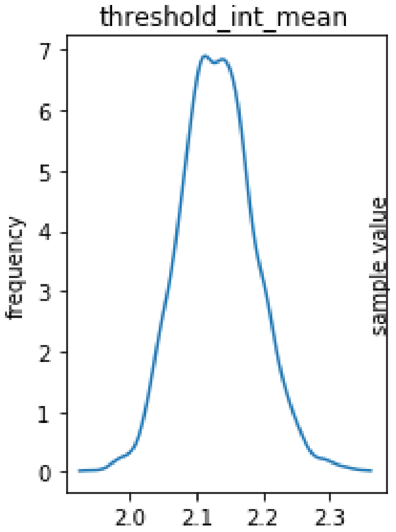
Dual influence social w/ popularity prior

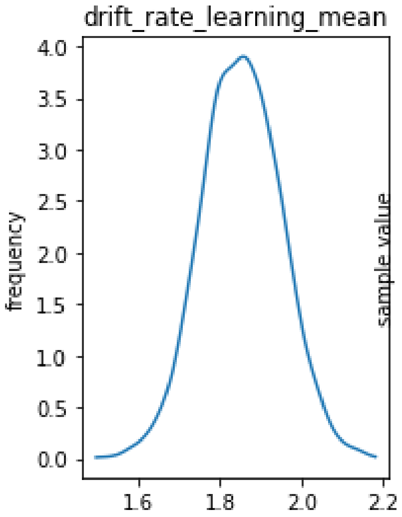
```
Inference for Stan model: model_dual_insight_prior_bayesian_a5f068290117c5bb88cda0180a2f3099.
4 chains, each with iter=2000; warmup=1000; thin=1;
post-warmup draws per chain=1000, total post-warmup draws=4000.
```

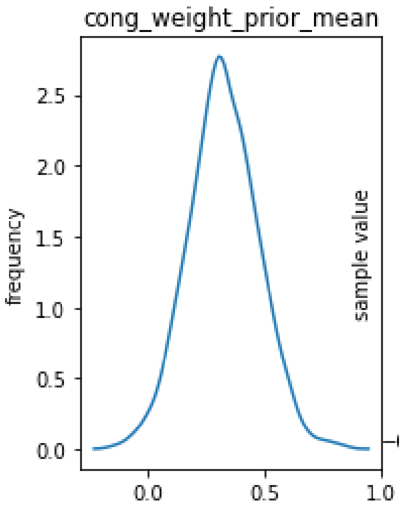
	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
threshold_int_mean	2.13	7.0e-4	0.06	2.03	2.09	2.13	2.17	2.25	6504	1.0
threshold_int_sd	0.3	5.9e-4	0.04	0.23	0.27	0.3	0.33	0.4	5552	1.0
drift_rate_learning_mean	1.85	4.1e-3	0.1	1.66	1.79	1.85	1.92	2.04	566	1.01
drift_rate_learning_sd	0.54	1.9e-3	0.07	0.42	0.49	0.53	0.58	0.7	1435	1.0
cong_weight_prior_mean	0.33	3.5e-3	0.15	0.03	0.23	0.32	0.42	0.62	1873	1.0
cong_weight_prior_sd	0.74	3.7e-3	0.15	0.49	0.63	0.72	0.83	1.08	1661	1.0
cong_weight_drift_bias_mean	0.16	6.1e-4	0.03	0.1	0.14	0.16	0.17	0.21	2055	1.0
cong_weight_drift_bias_sd	0.13	5.2e-4	0.02	0.09	0.12	0.13	0.15	0.19	2024	1.0
insight_mean	1.13	6.5e-3	0.32	0.49	0.93	1.13	1.35	1.77	2454	1.0
insight_sd	1.54	6.2e-3	0.28	1.05	1.34	1.51	1.72	2.16	2109	1.0

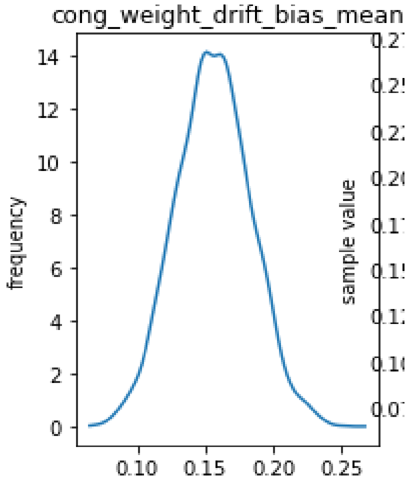






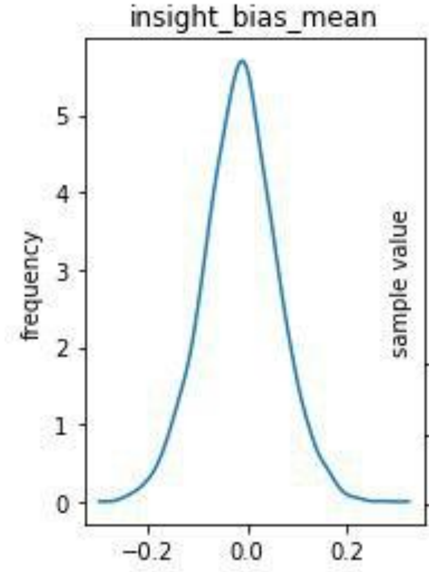


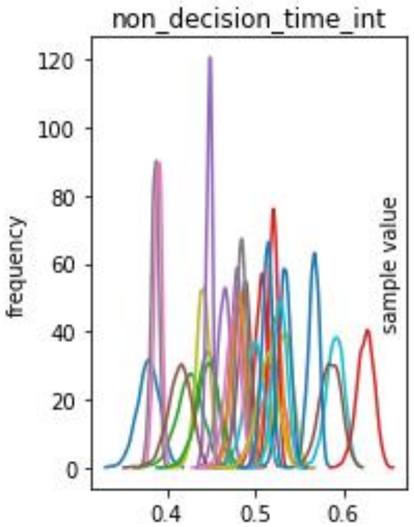


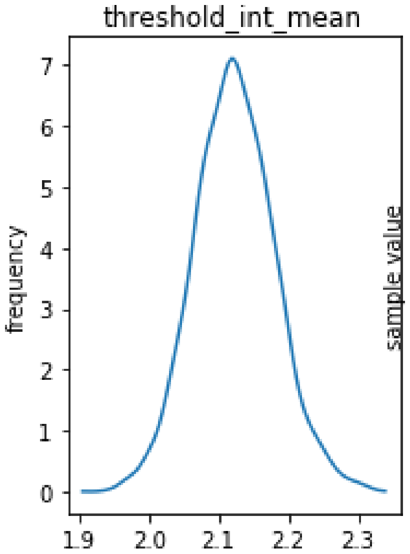


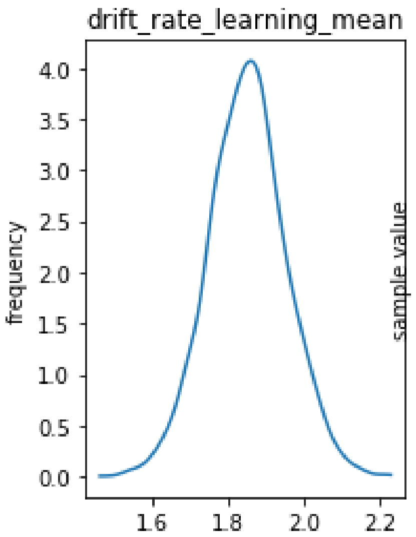
Dual influence social w/ popularity bias

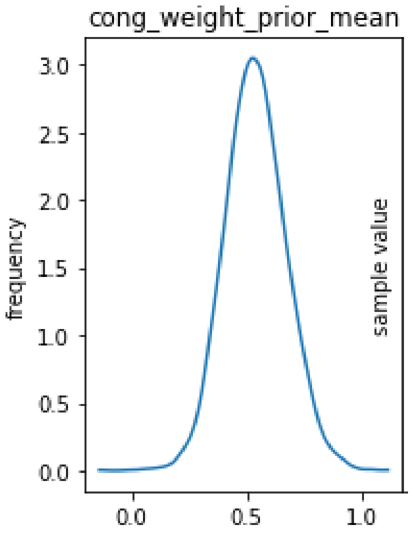
```
Inference for Stan model: model_dual_insight_bias_bayesian_ecadd60a72b9b5c8acc8ce907155f599.
4 chains, each with iter=2000; warmup=1000; thin=1;
post-warmup draws per chain=1000, total post-warmup draws=4000.

                          mean se_mean     sd   2.5%    25%    50%    75%  97.5%  n_eff   Rhat
threshold_int_mean        2.12  8.2e-4   0.06   2.01   2.08   2.12   2.16   2.24   4994    1.0
threshold_int_sd           0.3  7.3e-4   0.04   0.23   0.27    0.3   0.33   0.41   3671    1.0
drift_rate_learning_mean  1.85  5.1e-3    0.1   1.65   1.78   1.85   1.92   2.05    403   1.01
drift_rate_learning_sd    0.53  2.5e-3   0.07   0.41   0.48   0.53   0.58    0.7    888    1.0
cong_weight_prior_mean    0.54  4.1e-3   0.13   0.29   0.45   0.54   0.63    0.8   1032    1.0
cong_weight_prior_sd      0.63  4.4e-3   0.14    0.4   0.53   0.61   0.71   0.94   1046    1.0
cong_weight_drift_bias_mean 0.16  8.9e-4   0.03    0.1   0.14   0.16   0.18   0.22   1293    1.0
cong_weight_drift_bias_sd 0.15  7.2e-4   0.03    0.1   0.13   0.14   0.16   0.21   1505    1.0
insight_bias_mean        -0.01  1.7e-3   0.08  -0.16  -0.06  -0.01   0.03   0.14   1885    1.0
insight_bias_sd           0.35  1.4e-3   0.06   0.25   0.31   0.35   0.39   0.49   1957    1.0
```

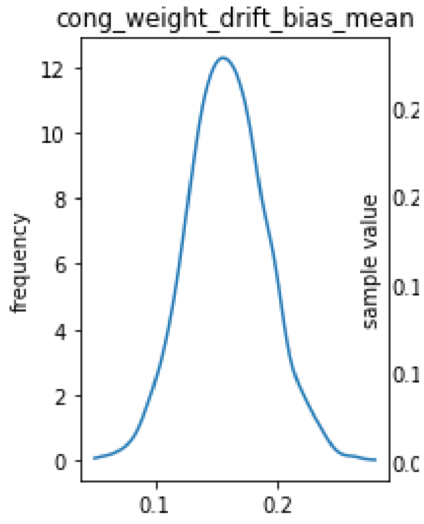








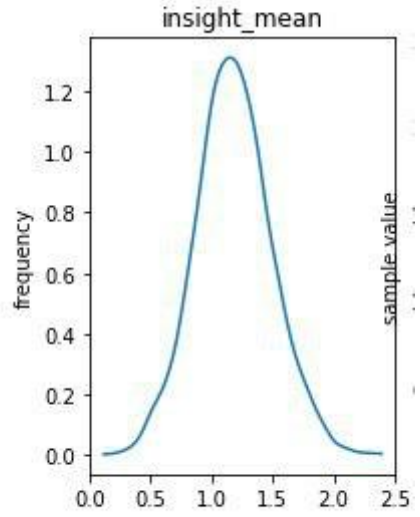


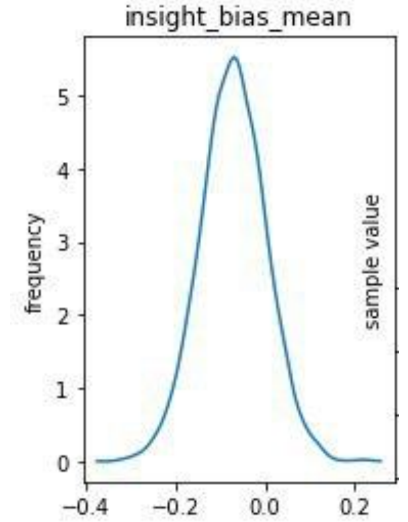


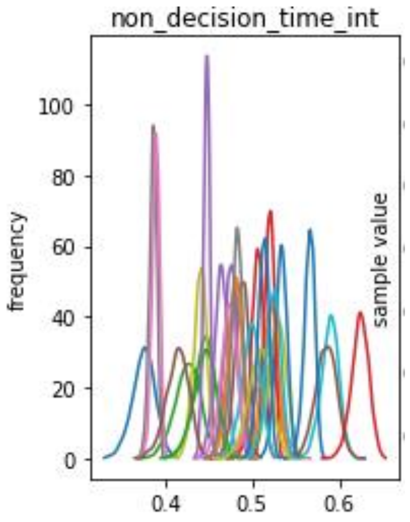
Dual influence w/ popularity dual

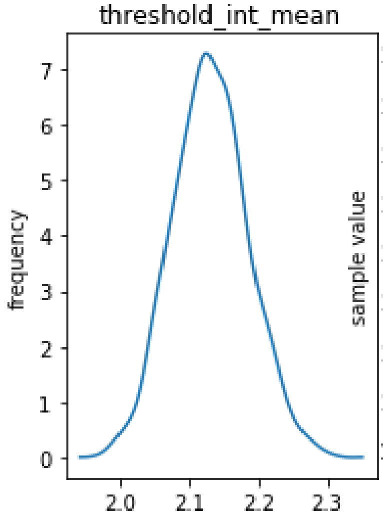
```
Inference for Stan model: model_dual_insight_priorbias_bayesian_3f4b0e8861fe35442899151acfda9b9d.
4 chains, each with iter=2000; warmup=1000; thin=1;
post-warmup draws per chain=1000, total post-warmup draws=4000.
```

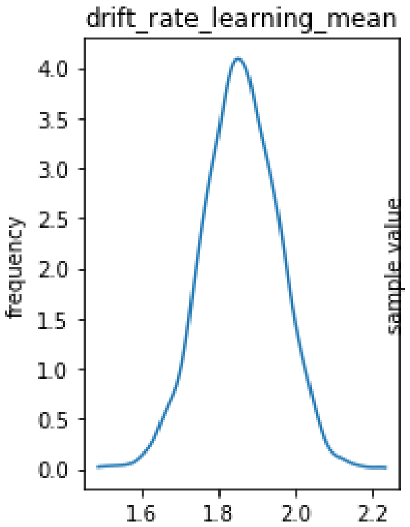
	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
threshold_int_mean	2.13	7.8e-4	0.05	2.03	2.09	2.13	2.17	2.24	4956	1.0
threshold_int_sd	0.3	6.0e-4	0.04	0.23	0.27	0.3	0.33	0.4	5096	1.0
drift_rate_learning_mean	1.86	4.8e-3	0.1	1.66	1.79	1.86	1.93	2.05	425	1.01
drift_rate_learning_sd	0.54	2.6e-3	0.07	0.41	0.49	0.53	0.58	0.71	852	1.0
cong_weight_prior_mean	0.31	3.9e-3	0.15	0.02	0.21	0.31	0.41	0.61	1508	1.0
cong_weight_prior_sd	0.72	4.0e-3	0.15	0.47	0.62	0.71	0.81	1.06	1398	1.0
cong_weight_drift_bias_mean	0.17	6.9e-4	0.03	0.11	0.15	0.17	0.19	0.23	1906	1.0
cong_weight_drift_bias_sd	0.14	6.4e-4	0.03	0.1	0.13	0.14	0.16	0.2	1762	1.0
insight_mean	1.18	6.4e-3	0.31	0.57	0.98	1.17	1.38	1.81	2354	1.0
insight_sd	1.46	6.6e-3	0.28	0.99	1.26	1.43	1.63	2.08	1778	1.0
insight_bias_mean	-0.07	1.5e-3	0.07	-0.22	-0.12	-0.07	-0.02	0.08	2573	1.0
insight_bias_sd	0.33	1.4e-3	0.06	0.22	0.28	0.32	0.36	0.47	1916	1.0

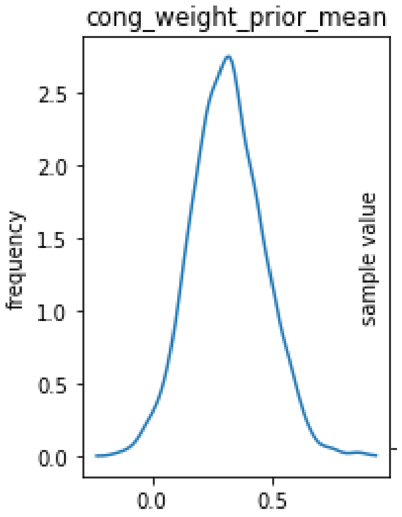


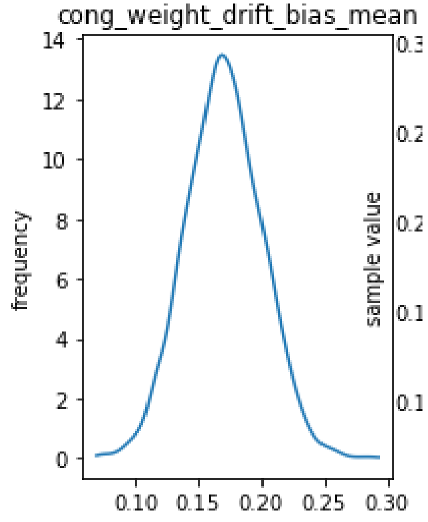






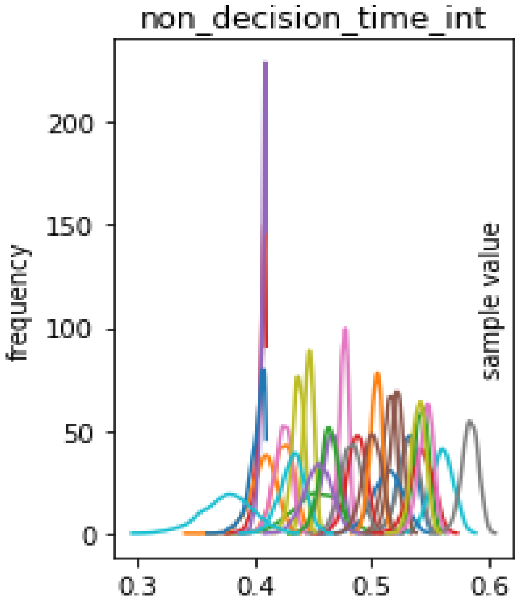


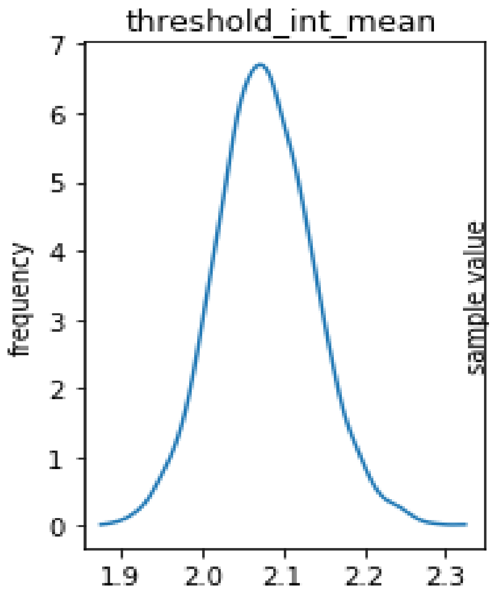


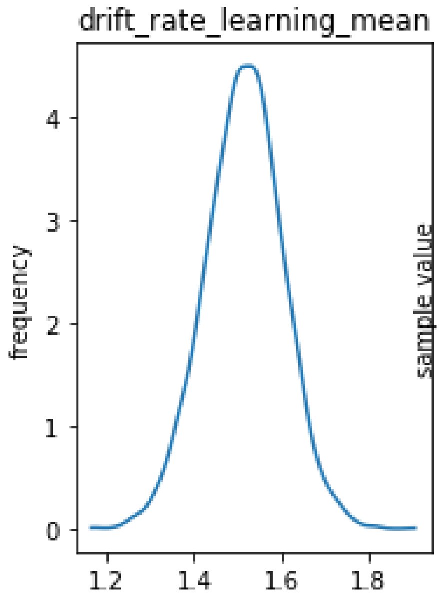


Null non-social

```
Inference for Stan model: model_null_bayesian_non_social_1649474275fde6026e721c32820ce796.
4 chains, each with iter=2000; warmup=1000; thin=1;
post-warmup draws per chain=1000, total post-warmup draws=4000.

                           mean se_mean     sd   2.5%    25%    50%    75%  97.5%  n_eff   Rhat
threshold_int_mean         2.08  8.2e-4   0.06   1.96   2.04   2.08   2.12    2.2   5422    1.0
threshold_int_sd           0.32  7.1e-4   0.05   0.24   0.28   0.31   0.34   0.42   4305    1.0
drift_rate_learning_mean   1.52  3.5e-3   0.09   1.34   1.46   1.52   1.57   1.69    632    1.0
drift_rate_learning_sd     0.48  2.0e-3   0.07   0.37   0.44   0.47   0.52   0.63   1044    1.0
```



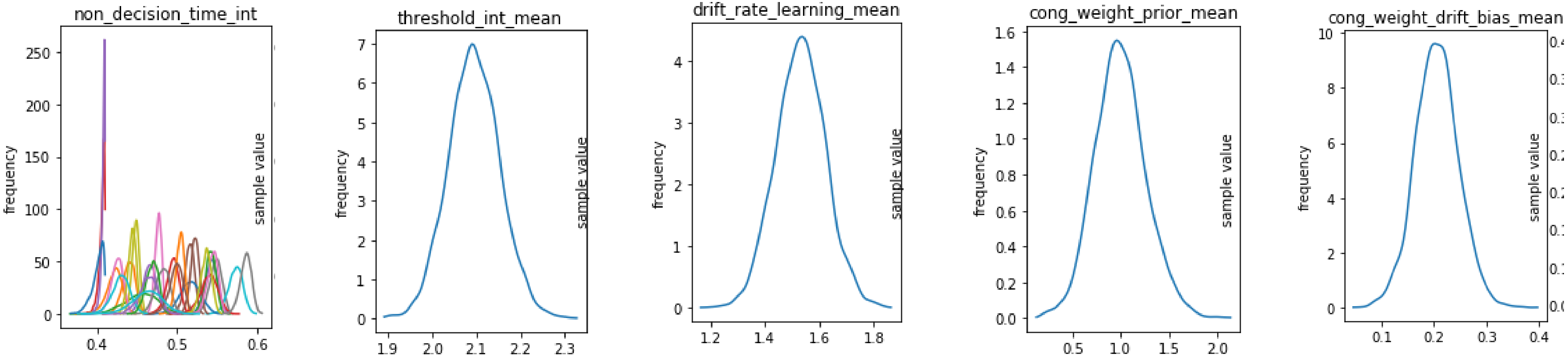




Dual influence non-social

```
Inference for Stan model: model_dual_bayesian_non_social_9b17a80e51ff1e772158dac3f05e48fa.
4 chains, each with iter=2000; warmup=1000; thin=1;
post-warmup draws per chain=1000, total post-warmup draws=4000.
```

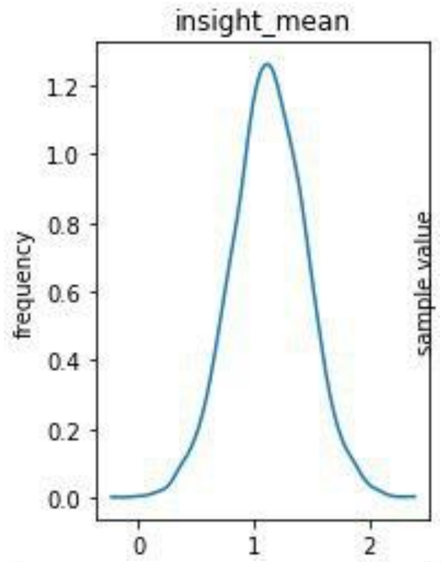
	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
threshold_int_mean	2.1	8.7e-4	0.06	1.98	2.06	2.09	2.13	2.21	4553	1.0
threshold_int_sd	0.31	7.6e-4	0.04	0.24	0.28	0.31	0.34	0.41	3425	1.0
drift_rate_learning_mean	1.53	4.5e-3	0.09	1.35	1.47	1.53	1.59	1.71	417	1.01
drift_rate_learning_sd	0.49	2.4e-3	0.07	0.38	0.44	0.49	0.54	0.65	882	1.01
cong_weight_prior_mean	1.0	9.2e-3	0.27	0.5	0.82	0.99	1.16	1.55	843	1.01
cong_weight_prior_sd	1.34	8.2e-3	0.25	0.93	1.17	1.31	1.49	1.91	912	1.0
cong_weight_drift_bias_mean	0.21	1.4e-3	0.04	0.12	0.18	0.21	0.23	0.29	836	1.01
cong_weight_drift_bias_sd	0.2	1.1e-3	0.04	0.14	0.18	0.2	0.23	0.28	1174	1.0

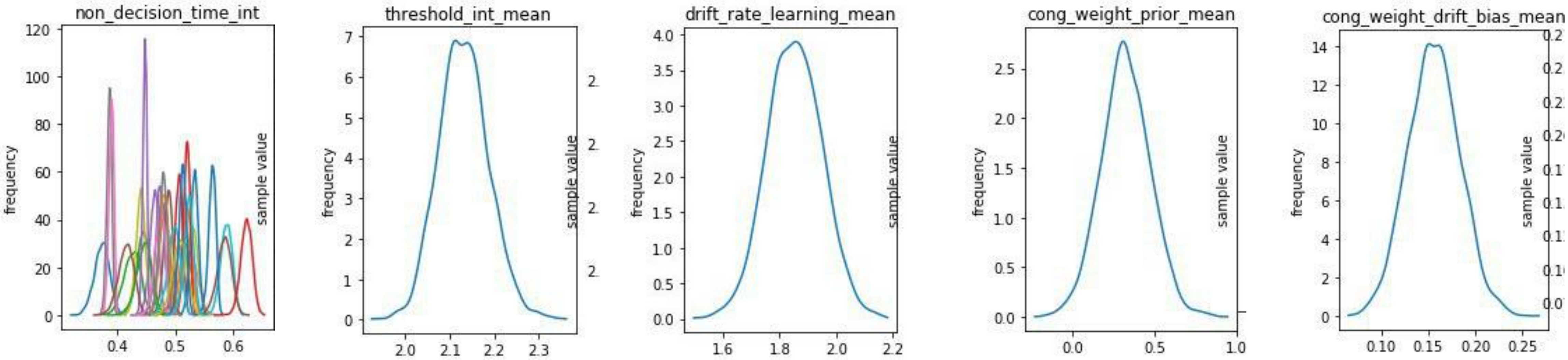


Dual influence w/ popularity prior non-social

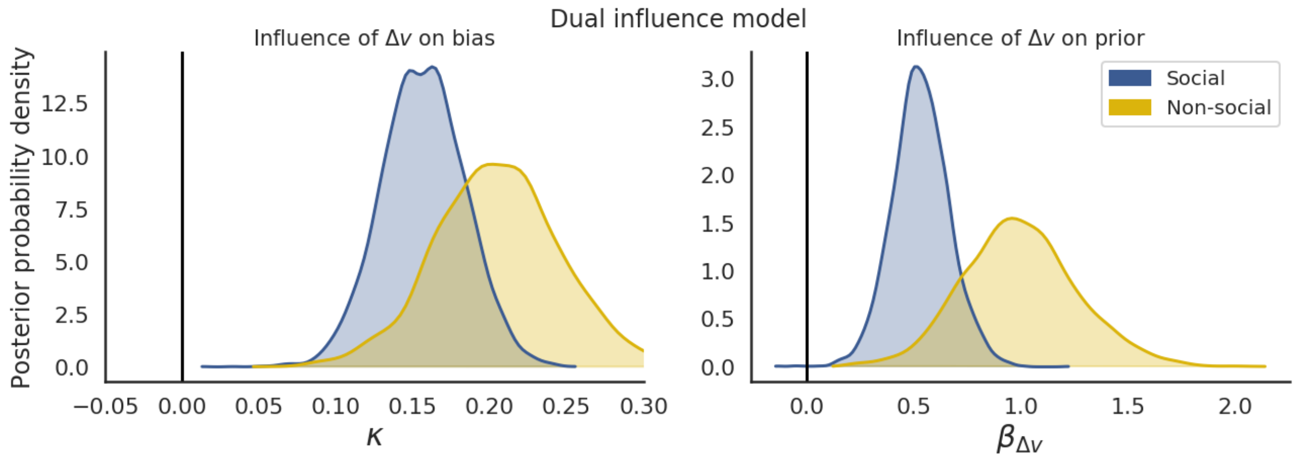
```
Inference for Stan model: model_dual_insight_prior_bayesian_a5f068290117c5bb88cda0180a2f3099.
4 chains, each with iter=2000; warmup=1000; thin=1;
post-warmup draws per chain=1000, total post-warmup draws=4000.
```

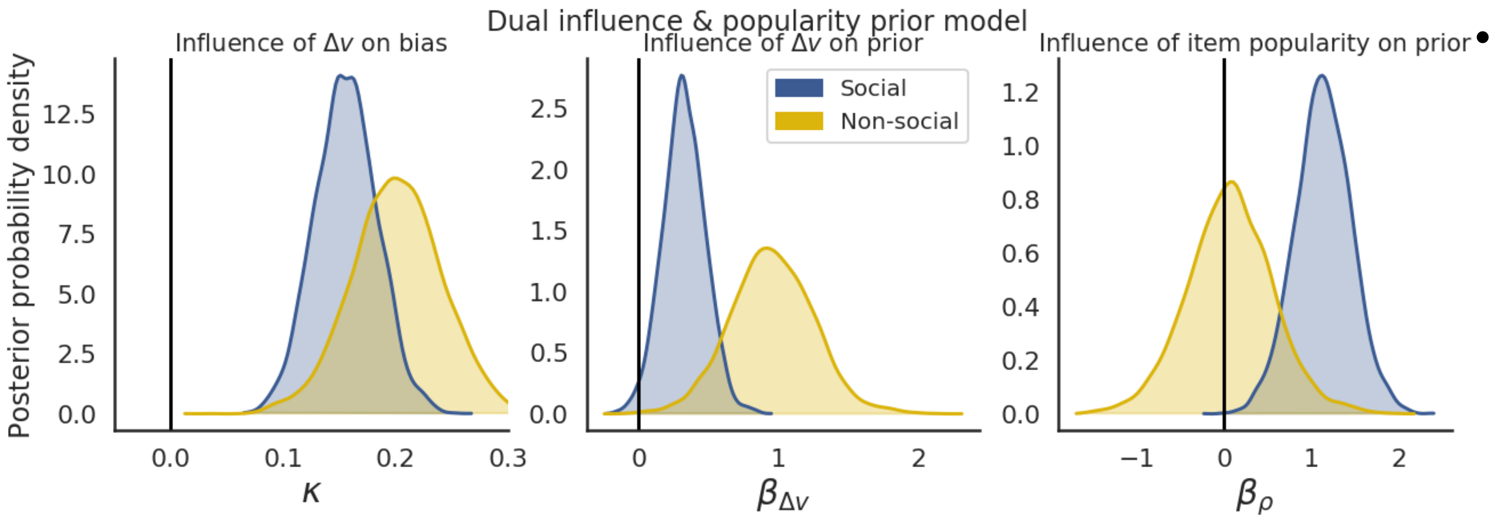
	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
threshold_int_mean	2.13	7.0e-4	0.06	2.03	2.09	2.13	2.17	2.25	6504	1.0
threshold_int_sd	0.3	5.9e-4	0.04	0.23	0.27	0.3	0.33	0.4	5552	1.0
drift_rate_learning_mean	1.85	4.1e-3	0.1	1.66	1.79	1.85	1.92	2.04	566	1.01
drift_rate_learning_sd	0.54	1.9e-3	0.07	0.42	0.49	0.53	0.58	0.7	1435	1.0
cong_weight_prior_mean	0.33	3.5e-3	0.15	0.03	0.23	0.32	0.42	0.62	1873	1.0
cong_weight_prior_sd	0.74	3.7e-3	0.15	0.49	0.63	0.72	0.83	1.08	1661	1.0
cong_weight_drift_bias_mean	0.16	6.1e-4	0.03	0.1	0.14	0.16	0.17	0.21	2055	1.0
cong_weight_drift_bias_sd	0.13	5.2e-4	0.02	0.09	0.12	0.13	0.15	0.19	2024	1.0
insight_mean	1.13	6.5e-3	0.32	0.49	0.93	1.13	1.35	1.77	2454	1.0
insight_sd	1.54	6.2e-3	0.28	1.05	1.34	1.51	1.72	2.16	2109	1.0





Social vs Non-Social





- The weights of preference on both the prior and bias are positive in both social and non-social group
  - People are biased(through both the prior and choice bias) toward their own preference both when learning other person's preference(social learning) and when just learning correct answers(non-social learning).
- The weight of item popularity on the prior is mainly positive only in social group.
  - People are biased toward(through the prior) the correct item's relative popularity only when learning other person's preference(social learning), but not when just learning correct answers(non-social learning).

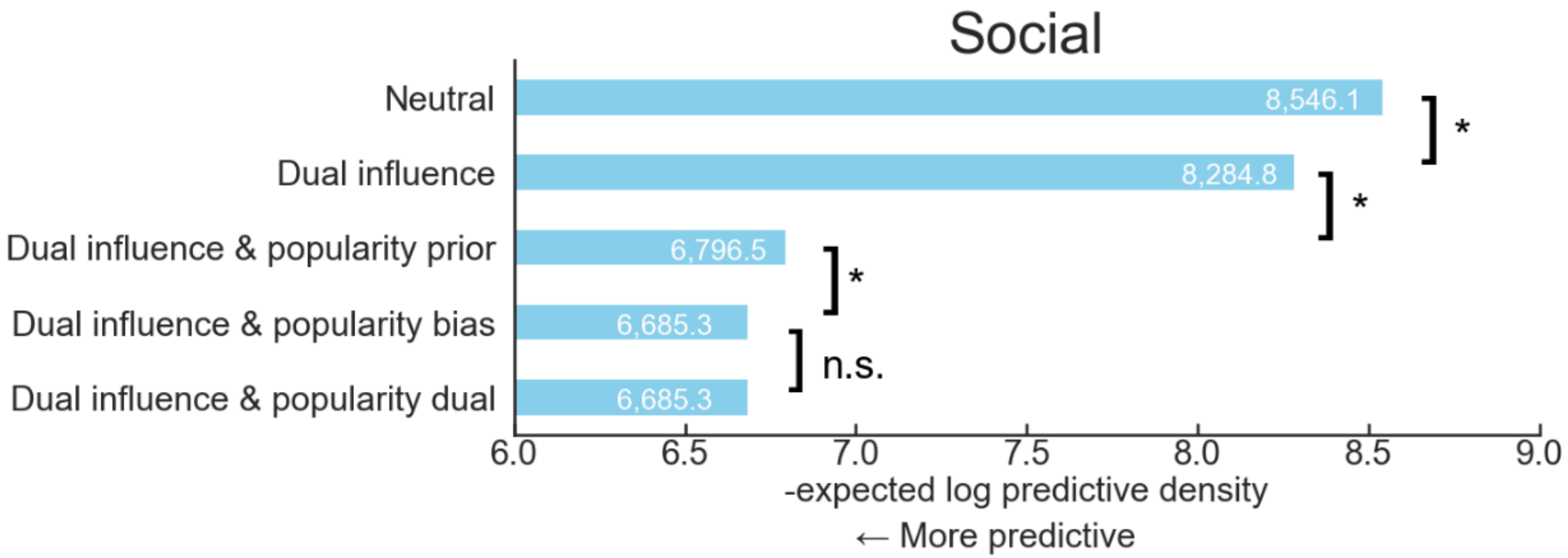
Model Comparison
Leave One Out Cross Validation

- **LOOCV** is a method for estimating pointwise out-of-sample prediction accuracy from a fitted Bayesian model using log-likelihood evaluated at the posterior simulations of parameter values.

- **Loo Package (Stan function)**
- From existing posterior simulation draws, **loo** computes approximate LOO-CV using Pareto smoothed importance sampling(PSIS), a procedure for regularizing importance weights.

- As a byproduct of the calculations, **loo** also obtains approximate standard errors for estimated predictive errors and for comparing predictive errors between two models.

Model Comparison
Leave One Out Cross Validation



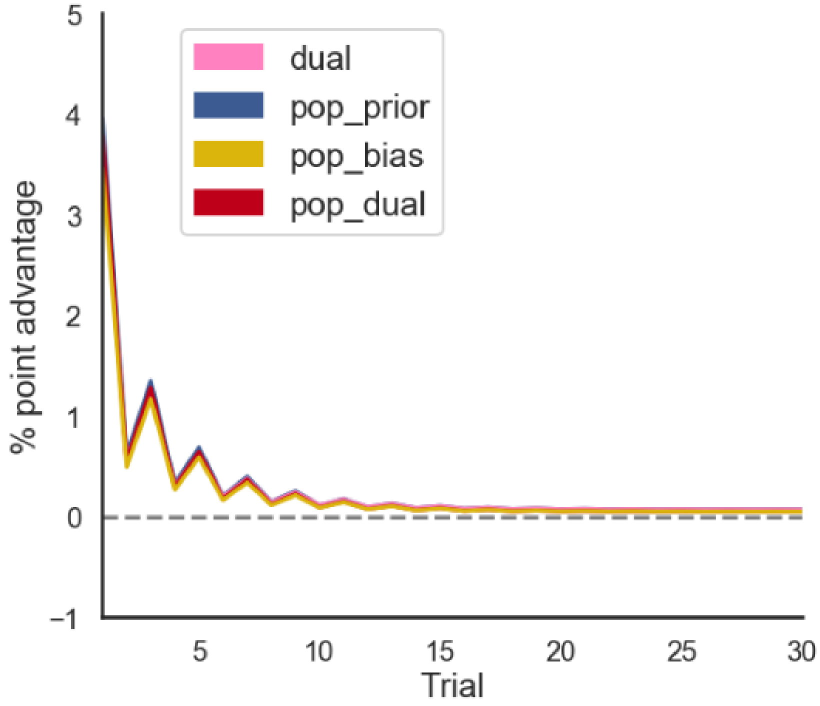
Minus ELPD score means minus log likelihood, thus the lowest minus ELPD score will have the highest posterior probability.

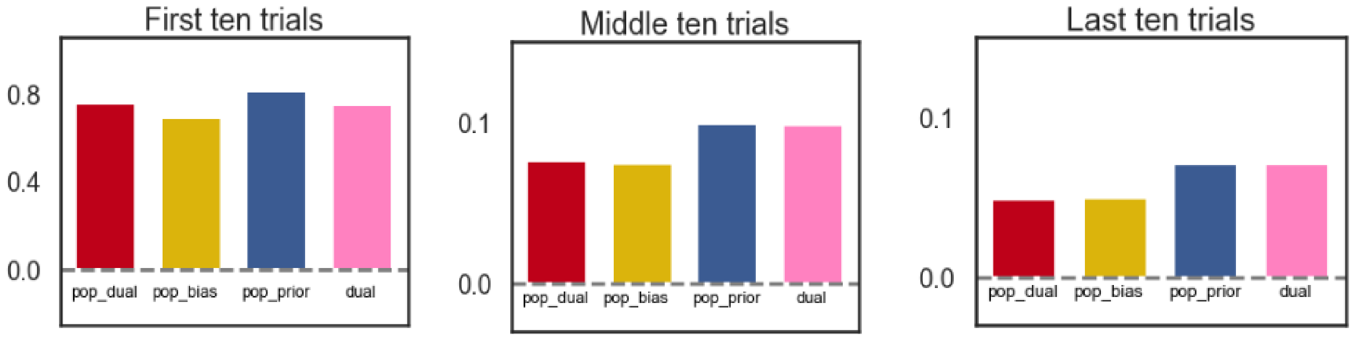
Dual influence & popularity bias model and Dual influence & popularity dual model got the lowest minus ELPD score.

Performance Simulation

- Tested whether the influence of preference with or without item popularity might provide a performance advantage(ratio of correct answers) over hypothetical actors who are not influenced by preference and item popularity.
- Used the group mean parameters extracted from the model with dual influence(preference) + popularity dual.
- Centered our simulations on the social group's parameter values in order to evaluate the efficiency of behavior in a social learning context.

% Point Advantage in each model





- Percentage point advantage in each model didn't differ much through entire process of learning.
- Benefits of all models compared to the null model decreased as the learning processed.
- On later 20 trials, dual influence & popularity prior model dual influence model showed slightly better benefit than other two models.

Conclusion

- People seem to be biased toward their preference in the context of both social learning and non-social learning
- People seem to also consider other people's preference(relative item popularity) which can be different from their own preference in the context of social learning, but not in non-social learning
- Models taking into account the influence of item popularity were found to have bigger likelihood.
- Performance Simulation showed that all the 4 models reflecting the influence of preference had benefit in the ratio of correct answer compared to null model.

References

- Tarantola, T., Kumaran, D., Dayan, P., & De Martino, B. (2017). Prior preferences beneficially influence social and non-social learning. *Nature communications*, *8*(1), 817.
- Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and computing*, *24*(6), 997-1016.