Supplementary 4: Bayesian estimation of the correlation between game players' environmental perception and in-game behaviors

Quan-Hoang Vuong ^{1,2,*}, Manh-Toan Ho ^{1,3}, Minh-Hoang Nguyen ^{1,3}, Thanh-Hang Pham ^{4,5}, Hoang-Anh Ho ¹, Thu-Trang Vuong ⁶, Viet-Phuong La ^{1,3}

¹ Centre for Interdisciplinary Social Research, Phenikaa University, Yen Nghia Ward, Ha Dong District, Hanoi 100803, Vietnam

² Centre Emile Bernheim, Université Libre de Bruxelles, 1050 Brussels, Belgium

³ A.I. for Social Data Lab, Vuong & Associates, 3/161 Thinh Quang, Dong Da District, Hanoi, 100000, Viet Nam

⁴ Faculty of Management and Tourism, Hanoi University, Km9, Nguyen Trai Road, Thanh Xuan, Hanoi 100803, Vietnam

⁵ School of Business, RMIT Vietnam University, Hanoi, 100000, Viet Nam

⁶ Sciences Po Paris, 27 Rue Saint-Guillaume, 75007 Paris, France

*Contact: Quan-Hoang Vuong (quvuong@ulb.ac.be); Manh-Toan Ho (toan.homanh@phenikaa-uni.edu.vn)

1. Aim

Given the contradicted results between game players' real-life environmental perceptions and in-game environmental behaviors, we are driven to explore whether there is the existence of a correlational relationship. In this Supplementary file, we present the mathematical formulas, codes, results, and interpretation of the association between game players' environmental perceptions and in-game behaviors towards the environment.

2. Model construction and results

2.1. Model1: Association between perception and in-game behaviors towards animals and plants

The formula representing the relationship between game players' perceptions and their in-game behaviors towards animals and plants is as follows:

$C2 \sim \text{BinomialLogit}(1 \mid b_Animal * Animal + b_Tree * Tree + a_Sex[Sex]);$

Where:

- Sex[1] ~ Male;
- Sex[2] ~ Female

This formula can be coded and simulated with the prior being Normal (0,10) using the following commands of the **bayesvl** R package (La & Vuong, 2019; Q.-H. Vuong, La, Nguyen, Ho, Ho, et al., 2020; Q.-H. Vuong, La, Nguyen, Ho, Tran, et al., 2020). Similar applications of the bayesvl R package can be found in the following articles (Q.-H. Vuong et al., 2019; Q.-H. Vuong et al., 2018; Q.-H. Vuong, Ho, et al., 2020; Q.-H. Vuong, La, Nguyen, Ho, Ho, et al., 2020). The reporting of the results and limitations strictly follows the principle of Q. H. Vuong (2020).

```

# Design the model
model <- bayesvl()
model <- bvl_addNode(model, "C2_bin", "binom")
model <- bvl_addNode(model, "Animal", "binom")
model <- bvl_addNode(model, "Tree", "binom")
model <- bvl_addNode(model, "Sex", "cat")

model <- bvl_addArc(model, "Animal", "C2_bin", "slope")
model <- bvl_addArc(model, "Tree", "C2_bin", "slope")

model <- bvl_addArc(model, "Sex", "C2_bin", "varint")

model <- bvl_modelFix(model, dat)
model_string <- bvl_model2Stan(model)
cat(model_string)

# detect the number of cores of your CPU
options(mc.cores = parallel::detectCores())

# fit the model using appropriate technical parameters
model <- bvl_modelFit(model, filtered, warmup = 2000, iter = 5000, chains = 4)

# Visualize the model and its distribution of posterior probability
bvl_plotIntervals(model)+theme_classic()

```

Table S4_1 shows the posterior coefficients of Model 1 employing datasets [1] and [2]. In all circumstances, the values of sample size (n_eff) were greater than 1000, hinting at good signals of the strong correlation between dependent and independent variables. Meanwhile, the Gelman Shrink Factor values of all parameters (Rhat) were equal to 1, indicating good signals of posterior convergence. The posterior coefficients of Model 1 generated from two different datasets are illustrated in Figure S4_1.

Table S4_1: Posterior coefficients of Model 1 with the prior being Normal (0,10).

Dataset	Parameter	mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
Dataset [1]	b_Animal_C2_bin	2.22	1.39	-0.15	1.25	2.09	3.04	5.42	4904	1
	b_Tree_C2_bin	0.03	0.77	*1.46	-0.48	0.03	0.53	1.56	6129	1
	a_Sex[1]	-1.51	1.37	-4.65	-2.31	-1.37	-0.56	0.78	5817	1
	a_Sex[2]	-2.90	1.38	-6.12	-3.71	-2.76	-1.94	-0.58	5551	1
	a0_Sex	-2.04	3.40	-9.23	-3.71	-2.07	-0.42	5.28	1991	1
	sigma_Sex	4.17	3.78	0.55	1.47	2.90	5.68	14.43	3078	1
Dataset [2]	b_Animal_C2_bin	2.32	1.45	-0.15	1.31	2.22	3.20	5.50	5303	1
	b_Tree_C2_bin	-0.15	0.81	-1.73	-0.69	-0.16	0.38	1.50	6653	1
	a_Sex[1]	-1.23	1.36	-4.25	-2.03	-1.11	-0.28	1.07	6793	1
	a_Sex[2]	-2.85	1.38	-5.86	-3.68	-2.74	-1.89	-0.52	6821	1

	a0_Sex	-1.68	3.63	-8.64	-3.48	-1.79	-0.12	6.21	1646	1
	sigma_Sex	4.38	3.92	0.64	1.64	3.09	5.80	15.18	2666	1

a) Dataset [1]

b) Dataset [2]

Figure S4_1: The posterior coefficients of Model 1 with the prior being Normal (0,10).

Figure S4_1a) and b) represent the posterior coefficients generated from the datasets [1] and [2], respectively. Collecting animal behavior has a positive association with the perception supporting the exploitation of the environment, while the impact of collecting plants on the environmental perceptions is negligible. The difference between males and females in terms of the agreement with perception C2 is consistent with the result in Supplementary 2. The sensitivity when applying the model on different datasets is negligible.

The probability of agreeing with the perception exploiting natural environment can be calculated by the **boot** R package (Canty & Ripley, 2020):

```
# Launch the package
library(boot)
anim = 1
tree = 1

# Calculate the probability of male and female, respectively
inv.logit(2.22 * anim + 0.03 * tree - 1.51)

inv.logit(2.22 * anim + 0.03 * tree - 2.90)
```

The estimated probabilities agreeing with the perception C2 of males and females are 67.69% and 34.29%, respectively, given they are likely to collect all in-game animals and plants. The probability density of males and females is shown in Figure S4_2.

Figure S4_2: The comparative densities between male and female game players estimated with the prior being Normal (0,10).

Figure S4_2 show relatively different probabilistic trends of the agreement with perception C2 between males and females, with females less likely to agree than males. $a_Sex[1]$ and $a_Sex[2]$ represents male and female game players, respectively.

2.2. Model 2: Association between perception and in-game behaviors towards animals, plants, and their interaction

The formula representing the relationship between game players' perceptions and their in-game behaviors towards animals, plants, and the interaction (between animals and plants) is as follows:

$$C2 \sim \text{BinomialLogit}(1 \mid b_Animal * Animal + b_Tree * Tree + b_Animal_Tree * Animal * Tree + a_Sex[Sex])$$

Where:

- $Sex[1] \sim \text{Male}$;
- $Sex[2] \sim \text{Female}$

This formula can be coded and estimated by the following commands (Stan Development Team, 2020):

```
# Design the model
model_string <-
"data{
  // Define variables in data
  int<lower=1> Nobs; // Number of observations (an integer)
  int<lower=0,upper=1> C2_bin[Nobs]; // outcome variable
  int<lower=0,upper=1> Animal[Nobs];
  int<lower=0,upper=1> Tree[Nobs];
  int NSex;
  int<lower=1,upper=NSex> Sex[Nobs];
}
parameters{
  // Define parameters to estimate
  real b_Animal;
  real b_Tree;
  real b_Animal_Tree;
  real a0_Sex;
  real<lower=0> sigma_Sex;
  vector[NSex] u_Sex;
}
transformed parameters{
  // Transform parameters
  real theta_C2_bin[Nobs];
  vector[NSex] a_Sex;
  // Varying intercepts definition
  for(k in 1:NSex) {
    a_Sex[k] = a0_Sex + u_Sex[k];
  }

  for (i in 1:Nobs) {
    theta_C2_bin[i] = b_Animal * Animal[i] + b_Tree * Tree[i] + b_Animal_Tree * Animal[i] *
    Tree[i] + a_Sex[Sex[i]];
  }
}
model{
```

```

// Priors
b_Animal ~ normal( 0, 10 );
b_Tree ~ normal( 0, 10 );
b_Animal_Tree ~ normal( 0, 10 );
a0_Sex ~ normal(0,10);
sigma_Sex ~ normal(0,10);
u_Sex ~ normal(0, sigma_Sex);

// Likelihoods
C2_bin ~ binomial_logit(1, theta_C2_bin);
}
generated quantities {
  // log-likelihood posterior
  vector[Nobs] log_lik_C2_bin;
  for (i in 1:Nobs) {
    log_lik_C2_bin[i] = bernoulli_logit_lpmf(C2_bin[i] | theta_C2_bin[i]);
  }
}
}"

# fit the model using appropriate technical parameters
datEnv <- with(filtered,
  list(Nobs      = length(A1),
       Animal    = as.numeric(Animal),
       Tree      = as.numeric(Tree),
       C2_bin    = as.numeric(C2_bin),
       NSex = length(unique(Sex)),
       Sex       = as.numeric(Sex)))

mstanInter <- rstan::stan(model_code = model_string, data = datEnv,
  warmup=2000 , iter = 5000, chains = 4)

# Visualize the distribution of posterior probability
bvl_plotParams(model, row=1,col=1,params=c("b_Animal_Tree"))

posterior <- as.data.frame(mstanInter)

mcmcMat = as.matrix(posterior[par_names], chains=TRUE)

exclude <- c("lp__","theta_C2_bin")
par_names <- c("b_Animal","b_Tree","b_Animal_Tree","a0_Sex","sigma_Sex")

bayesplot::color_scheme_set("blue")
posterior <- as.data.frame(mstanInter)
posterior <- posterior[par_names]
bayesplot::mcmc_intervals(posterior, point_est = "mean", prob = 0.8, prob_outer =
0.95)+theme_minimal()

bayesplot::mcmc_pairs(posterior, pars = par_names, off_diag_args = list(size = 1))

```

Table S4_2 shows the posterior coefficients of Model 2 employing datasets [1] and [2]. In all circumstances, the values of sample size (n_{eff}) were greater than 1000, hinting at good signals of the strong correlation between dependent and independent variables (McElreath, 2020). Meanwhile, the Gelman Shrink Factor values of all parameters (R_{hat}) were equal to 1, indicating

good signals of posterior convergence (Brooks & Gelman, 1998). The posterior coefficients of Model 2 generated from two different datasets are illustrated in Figure S4_3.

Table S4_2: Posterior coefficients of Model 1 with the prior being Normal (0,10).

Dataset	Parameter	mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
Dataset [1]	b_Animal	1.25	1.72	-1.88	0.11	1.14	2.30	4.97	5821	1
	b_Tree	-5.45	4.40	-15.54	-8.05	-4.81	-2.24	-1.49	3130	1
	b_Animal_Tree	5.75	4.44	-1.46	2.52	5.16	8.39	1.59	3151	1
	a_Sex[1]	0.47	3.09	-6.99	-0.60	0.57	1.70	7.09	1784	1
	a_Sex[2]	-0.92	3.09	-8.46	-2.00	-0.77	0.32	5.69	1784	1
	a0_Sex	-1.27	3.37	-8.16	-3.04	-1.27	4.18	6.10	2007	1
	sigma_Sex	4.14	3.82	0.53	1.44	2.83	5.62	1.44	2593	1
Dataset [2]	b_Animal	1.55	1.75	-1.69	0.39	1.47	2.62	5.26	5255	1
	b_Tree	-5.34	4.63	-15.78	-8.13	-4.70	-1.96	-1.90	3137	1
	b_Animal_Tree	5.44	4.69	-1.99	2.00	4.85	8.29	1.60	3133	1
	a_Sex[1]	0.62	3.29	-7.10	-0.56	0.71	1.95	7.53	1746	1
	a_Sex[2]	-1.01	3.29	-8.82	-2.22	-0.87	3.23	5.84	1781	1
	a0_Sex	-1.32	3.56	-8.64	-3.16	-1.32	4.20	6.58	1938	1
	sigma_Sex	4.36	3.89	0.65	1.63	3.04	5.84	0.15	3008	1

Figure S4_3: The posterior coefficients of Model 2 with the prior being Normal (0,10).

Figure S4_3a) and b) represent the posterior coefficients generated from the datasets [1] and [2], respectively. Similar to Model 1, collecting animal behavior has a positive association with the perception of exploiting the environment. However, when we consider the interaction of collecting animals and plants in this model as well, the impact of collecting plants on the environmental perception become negative. Interestingly, the effect of interaction of collecting animals and plants on the environmental perception is even stronger than the collecting animal behavior alone. The difference between males and females in terms of the agreement with perception C2 is consistent with the result in Supplementary 2 and the prior model. The sensitivity when applying the model on different datasets is negligible.

Figure S4_4: Bivariate scatterplot of posterior coefficients of Model 2 with the prior being Normal (0,10).

2.2. Models' weight comparison

To estimate and compare the weight of model 1 and model 2 (interaction model) for model specification, we employ the **loo** R package by the following commands (Vehtari & Gabry, 2019):

```
library("loo")

log_lik_1 <- extract_log_lik(model@stanfit, parameter_name = "log_lik_C2_bin", merge_chains = FALSE)
r_eff <- relative_eff(exp(log_lik_1))
loo_1 <- loo(log_lik_1, r_eff = r_eff, cores = 2)

log_lik_2 <- extract_log_lik(mstanInter, parameter_name = "log_lik_C2_bin", merge_chains = FALSE)
r_eff <- relative_eff(exp(log_lik_2))
loo_2 <- loo(log_lik_2, r_eff = r_eff, cores = 2)

loo_compare(list(model1=loo_1, model_interaction=loo_2))

loo_list <- list(model1=loo_1, model_interaction=loo_2)
```

```

loo_model_weights(loo_list) # stacking weight
loo_model_weights(loo_list, method = "pseudobma") # pseudo-BMA+ weight
loo_model_weights(loo_list, method = "pseudobma", BB = FALSE) # pseudo-BMA weight

(waic1 <- waic(log_lik_1))
(waic2 <- waic(log_lik_2))

waics <- c(
  waic1$estimates["elpd_waic", 1],
  waic2$estimates["elpd_waic", 1]
)

waic_wts <- exp(waics) / sum(exp(waics)) # WAIC weight

```

Table S4_3 presents the weight of model 1 and 2 measured by four different methods: WAIC, Bayesian stacking, Pseudo-BMA with Bayesian bootstrap, and Pseudo-BMA without Bayesian bootstrap (Vehtari, Gelman, & Gabry, 2017; Yao, Vehtari, Simpson, & Gelman, 2018). Both models obtain considerable weight, and no significant disparity is observed from the comparison. Model 1 has a higher weight in Bayesian stacking and Pseudo-BMA without Bayesian bootstrap, whereas Model 2 has a higher weight in other remaining categories. This result hints that both models are well-specified.

Table S4_3: Weight comparison between Model 1 and 2

Model	WAIC	Bayesian stacking	Pseudo-BMA with Bayesian bootstrap	Pseudo-BMA without Bayesian bootstrap
Model 1	0.417	0.574	0.497	0.510
Model 2	0.583	0.426	0.503	0.490
Model 1: C2 ~ BinomialLogit(1 b_Animal * Animal + b_Tree * Tree + a_Sex[Sex])				
Model 2: C2 ~ BinomialLogit(1 b_Animal * Animal + b_Tree * Tree + b_Animal_Tree * Animal * Tree + a_Sex[Sex])				

3. References

- Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics*, 7(4), 434-455.
- Canty, A., & Ripley, B. (2020). boot: Bootstrap R (S-Plus) Functions. Retrieved from <https://cran.r-project.org/web/packages/boot/>
- La, V.-P., & Vuong, Q.-H. (2019). bayesvl: Visually learning the graphical structure of Bayesian networks and performing MCMC with 'Stan'. Retrieved from <https://cran.r-project.org/web/packages/bayesvl/index.html>
- McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Boca Raton, FL, USA: CRC Press.
- Stan Development Team. (2020). RStan: the R interface to Stan. Retrieved from <http://mc-stan.org/>

- Vehtari, A., & Gabry, J. (2019). Bayesian Stacking and Pseudo-BMA weights using the loo package (Version loo 2.2.0). Retrieved from <https://mc-stan.org/loo/articles/loo2-weights.html>
- Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and computing*, 27(5), 1413-1432.
- Vuong, Q.-H., Bui, Q.-K., La, V.-P., Vuong, T.-T., Ho, M.-T., Nguyen, H.-K. T., . . . Ho, M.-T. (2019). Cultural evolution in Vietnam's early 20th century: A Bayesian networks analysis of Hanoi Franco-Chinese house designs. *Social Sciences & Humanities Open*, 1(1), 100001. doi:<https://doi.org/10.1016/j.ssaho.2019.100001>
- Vuong, Q.-H., Bui, Q.-K., La, V.-P., Vuong, T.-T., Nguyen, V.-H. T., Ho, M.-T., . . . Ho, M.-T. (2018). Cultural additivity: behavioural insights from the interaction of Confucianism, Buddhism and Taoism in folktales. *Palgrave Communications*, 4(1), 143. doi:10.1057/s41599-018-0189-2
- Vuong, Q.-H., Ho, M.-T., Nguyen, H.-K. T., Vuong, T.-T., Tran, T., Hoang, K.-L., . . . La, V.-P. (2020). On how religions could accidentally incite lies and violence: folktales as a cultural transmitter. *Palgrave Communications*, 6(1), 82. doi:10.1057/s41599-020-0442-3
- Vuong, Q.-H., La, V.-P., Nguyen, M.-H., Ho, M.-T., Ho, M.-T., & Mantello, P. (2020). Improving Bayesian statistics understanding in the age of Big Data with the bayesvl R package. *Software Impacts*, 4, 100016. doi:<https://doi.org/10.1016/j.simpa.2020.100016>
- Vuong, Q.-H., La, V.-P., Nguyen, M.-H., Ho, M.-T., Tran, T., & Ho, M.-T. (2020). Bayesian analysis for social data: A step-by-step protocol and interpretation. *MethodsX*, 100924.
- Vuong, Q. H. (2020). Reform retractions to make them more transparent. *Nature*, 582(7811), 149. doi:10.1038/d41586-020-01694-x
- Yao, Y., Vehtari, A., Simpson, D., & Gelman, A. (2018). Using stacking to average Bayesian predictive distributions (with discussion). *Bayesian Analysis*, 13(3), 917-1007.