Supplementary 2: Bayesian estimation of the posterior probability of game players' perception

Quan-Hoang Vuong^{1,2}, Manh-Toan Ho^{1,3}, Minh-Hoang Nguyen^{1,3}, Thanh-Hang Pham^{4,5}, Hoang-Anh Ho¹, Thu-Trang Vuong⁶, 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

In this Supplementary file, we present the materials, mathematical formulas, codes, results, and interpretation of the estimated posterior probability of game players' environmental perceptions. We also compare perceptions of players with differences in terms of geographical region, biological sex, and the highest educational level completed.

2. Materials

To estimate the probability of game players' perception, we employ three different types of demographic variables (geographical region, biological sex, and educational level) and the variable of perception "C2" (environmental perception) using two different datasets (dataset [1] with 584 responses and [2] with 503 responses, respectively).

Variable "A1" – region contains four levels: 'Asia', 'EU', 'US/Canada', and 'Other'.

Variable "A2" – sex contains two levels: 'Male' and 'Female'.

Variable "A3" – education contains four levels: 'Secondary school', 'High school', 'Undergraduate school', and 'Graduate school and higher'.

Variable "C2", one of 15 items in the revised New Environmental Paradigm (NEP) Scale (Dunlap, Van Liere, Mertig, & Jones, 2000), is used to measure the agreement level of the respondent with the following statement based on a 5-point Likert scale: *"Humans have the right to modify the natural environment to suit their needs"*. The 5-point Likert scale ranges from 'Strongly disagree' to 'Strongly agree'. However, in this analysis, the 'Unsure' level was excluded, and the other four levels were combined to obtain binary data with two levels: 'agree' and 'disagree'. The dataset [1] and [2] after discarding the 'Unsure' level consisted of 392 responses and 334 responses, respectively.

3. Model construction and results

3.1. Posterior probability of the agreement with perception C2

The formula for estimating the posterior probability of the agreement with perception C2 is as follows:

$$C2 \sim \text{bernoulli}(\theta_{C2})$$

This formula can be coded and simulated with the prior set as normal distribution at 0.3 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 and Bayesian statistics can be found in the following articles (Tran et al., 2020; Q.-H. Vuong et al., 2019; Q.-H. Vuong et al., 2018; Q.-H. Vuong, Ho, et al., 2020). The reporting of the results and limitations strictly follows the principle of Q. H. Vuong (2020).

```
library(bayesvl)

# Design the model
modelC2 <- bayesvl()
modelC2 <- bvl_addNode(modelC2, "C2_bin", "bern", "normal(0.3, 10)")

# Fit the model using appropriate technical parameters
modelC2 <- bvl_modelFit(modelC2, filtered, warmup = 2000, iter = 5000, chains = 4)

# Visualize the distribution of posterior probability
bvl_plotParams(modelC2, row=1, col=1, credMass = 0.95)
```

In order to ensure the robustness, tweaking the priors of each model and performing posterior estimation with two different datasets are employed. First of all, using dataset [1], we estimate the posterior probability under two different prior set-ups with mean at 0.3 and 0.6, which are equivalent to the 'disagree' and 'agree' with perception C2, respectively. Then, we replicate the posterior estimation process with dataset [2] for the second sensitivity check.

Table S2_1: Posterior probability of the agreement with perception C2 with the prior being Normal (0.3,10) and Normal (0.6,10)

Prior: $\theta_{C2} \sim \text{normal}(0.3, 10)$									
Dataset	mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
dataset [1]	0.47	0.03	0.42	0.45	0.47	0.48	0.52	4915	1
dataset [2]	0.48	0.03	0.43	0.46	0.48	0.50	0.53	4467	1
Prior: $\theta_{C2} \sim \text{normal}(0.6, 10)$									
Dataset	mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
dataset [1]	0.47	0.02	0.42	0.45	0.47	0.48	0.52	4973	1
dataset [2]	0.48	0.03	0.43	0.46	0.48	0.5	0.54	4791	1

Table S2_1 shows the posterior estimation results of the posterior probability of the agreement with perception C2 under two different prior set-ups employing datasets [1] and [2]. In all circumstances, the values of sample size (n_eff) were greater than 1000, hinting at well-mixing chains (McElreath, 2020). Meanwhile, the Gelman Shrink Factor values of all parameters (Rhat) were equal to 1, indicating good signals of convergence (Brooks & Gelman, 1998). The probability

distribution of the agreement with perception C2 generated from two different datasets is illustrated in Figure S2_1.

Figure S2_1: The distribution of posterior probability of the agreement with perception C2 with the prior being Normal (0.3,10).

Figure S2_1a) and b) represent the distribution generated from the dataset [1] and [2], respectively. The posterior probabilities of both cases are normally distributed at around 0.46-0.48, suggesting the simulation's robustness given the change in sample size.

3.2. Posterior probability of the agreement with perception C2 by geographical regions

The formula for estimating the posterior probability of the agreement with perception C2 by the geographical regions is as follows:

$$C2 \sim \text{bernoulli}(\theta[\text{Region}]);$$

Where:

- $\theta[1] \sim \text{Asia}$
- $\theta[2] \sim \text{EU}$
- $\theta[3] \sim \text{Other}$
- $\theta[4] \sim \text{US/Canada}$

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

```
# Design the model
model_string <-
"data{
  int<lower=1> Nobs;
```

```

int<lower=1> NRegion;
int<lower=0,upper=1> C2_bin[Nobs];
int<lower=0,upper=NRegion> Region[Nobs];
}
parameters{
  real<lower=0,upper=1> theta[NRegion];
}
model{
  for(k in 1:NRegion) {
    theta[k] ~ beta(1, 1);
  }

  C2_bin ~ bernoulli(theta[Region]);
}
"

# Fit the model using appropriate technical parameters
datEnv <- with(filtered,
  list(Nobs      = length(A1),
       C2_bin    = as.numeric(C2_bin),
       NRegion   = length(unique(Region)),
       Region     = as.numeric(Region)))

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

# Visualize the distribution of posterior probability
exclude <- c("lp__")
par_names <- setdiff(dimnames(mstanRegion)[[3]], exclude)

postParams <- rstan::extract(mstanRegion, pars = par_names)
names(postParams) <- levels(filtered$Region)
ref <- reshape2::melt(postParams)
colnames(ref)[2:3] <- c("value", "Region")

ggplot2::ggplot(data=ref,aes_string(x="value",
color="Region"))+geom_density(size=1)+theme_minimal()

```

Table S2_2 presents the posterior probability of the agreement with perception C2 of respondents from four different regions: Asia, EU, US/Canada, and Others employing two different datasets. The n_{eff} and $Rhat$ indicate the good convergence of posteriors. The probability distribution of each region generated from the datasets [1] and [2] is shown in Figure S2_2.

Table S2_2: Posterior probability of the agreement with perception C2 by geographical regions the prior being Beta (1,1).

Dataset	Parameter	mean	sd	2.5%	25%	50%	75%	97.5%	n_{eff}	$Rhat$
Dataset [1]	theta [1]	0.47	0.05	0.38	0.44	0.47	0.50	0.56	14305	1
	theta [2]	0.28	0.06	0.17	0.24	0.28	0.32	0.40	13412	1
	theta [3]	0.31	0.12	0.10	0.22	0.30	0.39	0.58	14998	1
	theta [4]	0.53	0.03	0.46	0.50	0.53	0.55	0.59	14203	1
Dataset [2]	theta [1]	0.46	0.05	0.36	0.42	0.46	0.50	0.57	12108	1
	theta [2]	0.29	0.06	0.18	0.25	0.29	0.33	0.42	13108	1

	theta [3]	0.40	0.15	0.14	0.29	0.40	0.51	0.70	12215	1
	theta [4]	0.55	0.04	0.48	0.53	0.55	0.57	0.62	12249	1

a) Dataset [1]

b) Dataset [2]

Figure S2_2: The distribution of the posterior probability of the agreement with perception C2 by regions with the prior being Beta (1,1).

Figure S2_2a) and b) represent the distribution generated from the dataset [1] and [2], respectively. The probability distribution is relatively different among different geographical regions. Respondents from the US/Canada have the highest probability to agree that “humans have the right to modify the natural environment to suit their needs”, whereas respondents from the EU are least likely to agree with that statement.

3.3. Posterior probability of the agreement with perception C2 by biological sex

The formula for estimating the posterior probability of the agreement with perception C2 by the biological sex is as follows:

$$C2 \sim \text{bernoulli}(\theta[\text{Sex}]);$$

Where:

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

This formula can be coded and estimated by the following commands:

```
# Design the model
model_string <-
"data{
  int<lower=1> Nobs;
  int<lower=1> NSex;
  int<lower=0,upper=1> C2_bin[Nobs];
  int<lower=0,upper=NSex> A2[Nobs];
}
parameters{
  real<lower=0,upper=1> theta[NSex];
}
model{
  for(k in 1:NSex) {
    theta[k] ~ beta(1, 1);
  }

  C2_bin ~ bernoulli(theta[A2]);
}
"
```

```
# Fit the model using appropriate technical parameters
datEnv <- with(filtered,
  list(Nobs      = length(A1),
       C2_bin    = as.numeric(C2_bin),
       NSex      = length(unique(A2)),
       A2        = as.numeric(A2)))

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

```
# Visualize the distribution of posterior probability
exclude <- c("lp__")
par_names <- setdiff(dimnames(mstanSex)[[3]], exclude)

postParams <- rstan::extract(mstanSex, pars = par_names)
```

```
names(postParams) <- levels(filtered$A2)
ref <- reshape2::melt(postParams)
colnames(ref)[2:3] <- c("value", "Gender")

ggplot2::ggplot(data=ref, aes_string(x="value",
color="Gender"))+geom_density(size=1)+theme_minimal()
```

Table S2_3: Posterior probability of the agreement with perception C2 by gender with the prior being Beta (1,1).

Dataset	Parameter	mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
Dataset [1]	theta [1]	0.67	0.04	0.59	0.64	0.67	0.70	0.75	11985	1
	theta [2]	0.36	0.03	0.30	0.34	0.36	0.38	0.42	10222	1
Dataset [2]	theta [1]	0.68	0.04	0.60	0.66	0.68	0.71	0.76	10387	1
	theta [2]	0.36	0.03	0.30	0.34	0.36	0.38	0.42	10140	1

Table S2_3 presents the posterior probability of the agreement with perception C2 of males and females employing two different datasets. The n_{eff} and R_{hat} indicate the good convergence of posteriors. The probability distribution of each sex generated from the datasets [1] and [2] is shown in Figure S2_3.

Figure S3_3: The distribution of the posterior probability of the agreement with perception C2 by sex with the prior being Beta (1,1).

Figure S3_3a) and b) represent the distribution generated from the dataset [1] and [2], respectively. The probability distribution is relatively different between males and females. Male respondents have a much higher likelihood to agree “humans have the right to modify the natural environment to suit their needs” than female counterparts.

3.4. Posterior probability of the agreement with perception C2 by educational level

The formula for estimating the posterior probability of the agreement with perception C2 by the educational level is as follows:

$$C2 \sim \text{bernoulli}(\theta_{\text{Edu}});$$

Where:

- θ_1 ~ Graduate school and higher
- θ_2 ~ High school
- θ_3 ~ Secondary school
- θ_4 ~ Undergraduate school

This formula can be coded and estimated by the following commands:

```
# Design the model
model_string <-
"data{
  int<lower=1> Nobs;
  int<lower=1> NA3;
  int<lower=0,upper=1> C2_bin[Nobs];
  int<lower=0,upper=NA3> A3[Nobs];
}
parameters{
```



```

    real<lower=0,upper=1> theta[NA3];
  }
  model{
    for(k in 1:NA3) {
      theta[k] ~ beta(1, 1);
    }

    C2_bin ~ bernoulli(theta[A3]);
  }
  "

# Fit the model using appropriate technical parameters
datEnv <- with(filtered,
  list(Nobs      = length(A1),
       C2_bin    = as.numeric(C2_bin),
       NA3 = length(unique(A3)),
       A3        = as.numeric(A3)))

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

# Visualize the distribution of posterior probability
exclude <- c("lp__")
par_names <- setdiff(dimnames(mstanEdu)[[3]], exclude)

postParams <- rstan::extract(mstanEdu, pars = par_names)
names(postParams) <- levels(filtered$A3)
ref <- reshape2::melt(postParams)
colnames(ref)[2:3] <- c("value", "Education")

ggplot2::ggplot(data=ref,aes_string(x="value",
  color="Education"))+geom_density(size=1)+theme_minimal()+
  theme(legend.title = element_blank(), legend.position="bottom")

```

Table S2_4: Posterior probability of the agreement with perception C2 by educational level with the prior being Beta (1,1).

Dataset	Parameter	mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
Dataset [1]	theta [1]	0.45	0.05	0.35	0.42	0.45	0.49	0.55	13882	1
	theta [2]	0.44	0.06	0.33	0.41	0.44	0.48	0.55	13264	1
	theta [3]	0.33	0.13	0.11	0.24	0.32	0.42	0.61	13477	1
	theta [4]	0.49	0.03	0.42	0.47	0.49	0.51	0.56	14003	1
Dataset [2]	theta [1]	0.47	0.05	0.37	0.43	0.47	0.51	0.57	12436	1
	theta [2]	0.47	0.06	0.36	0.43	0.47	0.51	0.59	12507	1
	theta [3]	0.33	0.13	0.11	0.24	0.33	0.42	0.61	13853	1
	theta [4]	0.50	0.04	0.43	0.48	0.50	0.53	0.58	12953	1

Table S2_4 presents the posterior probability of the agreement with perception C2 of respondents with different educational levels employing two different datasets. The n_eff and Rhat indicate the good convergence of posteriors. The probability distribution of each educational level generated from the datasets [1] and [2] is shown in Figure S2_3.

Figure S2_4: The distribution of the posterior probability of the agreement with perception C2 by educational level with the prior being Beta (1,1).

Figure S2_4a) and b) represent the distribution generated from the dataset [1] and [2], respectively. The probability distribution is seemingly different among respondents with different educational levels. Respondents with undergraduate school education levels have the highest likelihood to agree “humans have the right to modify the natural environment to suit their needs”,

whereas those with secondary school education level have the lowest and greatest variance, hinting at the uncertain perception of respondents from this group.

4. 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.
- Dunlap, R. E., Van Liere, K. D., Mertig, A. G., & Jones, R. E. (2000). New trends in measuring environmental attitudes: measuring endorsement of the new ecological paradigm: a revised NEP scale. *Journal of social issues*, 56(3), 425-442.
- La, V.-P., & Vuong, Q.-H. (2019). bayesvl: Visually learning the graphical structure of Bayesian networks and performing MCMC with 'Stan': The Comprehensive R Archive Network (CRAN). 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/>
- Tran, T., Ho, M.-T., Pham, T.-H., Nguyen, M.-H., Nguyen, K.-L. P., Vuong, T.-T., . . . Vuong, Q.-H. (2020). How Digital Natives Learn and Thrive in the Digital Age: Evidence from an Emerging Economy. *Sustainability*, 12(9), 3819.
- 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