Does Interacting Help Users Better Understand the Structure of Probabilistic Models?

Supplemental Material

Evdoxia Taka, Sebastian Stein, and John H. Williamson

S.I Participants' Training

The training part of the user study comprised of four training videos. The training videos were same for both groups of participants; the static (SG) and interaction (IG) group, with some slight differences based on the visualization condition. Participants could ask questions once they had watched the videos.

The *first training video* (IG: https://youtu.be/6yrBrL6amiQ, and SG: https://youtu.be/zeonqIgHspk) was introducing the researcher and provided her contact details, explaining the freedom to withdraw and the purpose of the user study, and provided a description of the tasks and structure of the user study.

The *second training video* (IG: https://youtu.be/iPf8bwdxKy8, and SG: https://youtu.be/q0ZCM5KOxbI) was making an introduction to basic probabilistic concepts such as random variables, probability, (joint) probability mass function/-density, sampling, density/scatter/rug plots.

The *third training video* (IG: https://youtu.be/zQhy-LYJGQ4, and SG: https://youtu.be/ow86A6cvHjE) was presenting and demonstrating the use of the IPP tool.

The *fourth training video* (IG: https://youtu.be/9mfhepxCeRU, and SG: https://youtu.be/uDcqwLqQFDA) was presenting example tasks (one per RQ) and how the presented visualizations could be used to answer the questions. For example, participants in the SG were shown how they can interpret the shape of a pair plot in terms of relations between variables when they think of it on the basis of conditioning.

S.II Task Models

S.II.1 Model 1

The first model was designed to predict the mean November temperature (○C) in Scotland. The model consists of an observed random variable for the predicted $\mathrm{temperature}$ and a set of unidentified parameters a, b, and c. The mean value of the prior distribution for the mean value of $\mathrm{temperature}$'s distribution was set to 2 because out of prior experience of living in Scotland, the temperatures at this time of the year are usually nearly above 0. The standard deviations of the prior distributions were set to 10 to make them weakly informative.

$$\mathrm{a} \sim \mathrm{Uniform}(\mathrm{lower} = 80, \mathrm{upper} = 100)$$
$$\mathrm{b} \sim \mathrm{Normal}(\mu = 2, \sigma = 10)$$
$$\mathrm{c} \sim \mathrm{Half\text{-}Normal}(\sigma = 10)$$
$$\mathrm{temperature} \sim \mathrm{Normal}(\mu = \mathrm{b}, \sigma = \mathrm{c})$$

The PyMC3 code of the model can be found in https://github.com/evdoxiataka/ipme/tree/master/examples/user_study/min_temperature. The prior samples from the model used in the study could be found in the file `study_analysis\data\min_temperature.npz` in [*S1*]. The data used for the definition of the likelihood was the average minimum temperature in Scotland in month November for the years 1884-2020 (retrieved from https://www.metoffice.gov.uk/pub/data/weather/uk/climate/datasets/Tmin/date/Scotland.txt).

- *E. Taka is with the School of Computing Science, University of Glasgow, UK. E-mail: e.taka.1@research.gla.ac.uk*
- *S. Stein and J. H. Williamson are with the School of Computing Science, University of Glasgow, UK.*

Manuscript received February 22, 2022; revised August 23, 2022.

S.II.2 Model 2

The second model was designed to predict the output of an engine that generates random real numbers. The model consists of an observed random variable for the predicted $\mathrm{random\_number}$ and a set of unidentified parameters a, b, and c. For the parameterization of the uniform likelihood's bounds, we subtract a positive number c (sampled from a half-normal distribution) from a number a (sampled from a normal distribution centered around 0) to set the lower bound and we add it to a to set the upper bound.

$$\mathrm{a} \sim \mathrm{Normal}(\mu = 0, \sigma = 10)$$

$$\mathrm{b} \sim \mathrm{Half\text{-}Normal}(\sigma = 10)$$

$$\mathrm{c} \sim \mathrm{Half\text{-}Normal}(\sigma = 20)$$

$$\mathrm{random\_number} \sim \mathrm{Uniform}(\mathrm{lower} = \mathrm{a} - \mathrm{c}, \mathrm{upper} = \mathrm{a} + \mathrm{c})$$

The PyMC3 code of the model can be found in https://github.com/evdoxiataka/ipme/tree/master/examples/user_study/random_number_generator. The prior samples from the model used in the study could be found in the file `study_analysis\data\transformation.npz` in [*S1*]. The data used for the definition of the likelihood was synthetically created.

S.II.3 Model 3

The third model was designed to predict the reaction time (in msec) of lorry drivers under sleep deprivation conditions. The model consists of observed random variables for the predicted $\mathrm{reaction\_time}$ of each lorry driver ($i \in 1, 2, ..., 18$), a set of priors a, b, $\mathrm{sigma_i}$ and d, and a set of hyper-priors c, e, f, g and h. The day variable takes values in the $day \in 1, 2, ..., 10$. The visualizations of the tasks in the user study regarding this problem included only the parameters a, b, c, d, and the $\mathrm{reaction\_time}$ observed variable.

For setting the prior for parameter a, namely the intercept of the $\mathrm{reaction\_time}$'s mean value, we set a hyper-prior for the mean value of its prior distribution with mean value equal to 100 msec (0.1 sec) and standard deviation 150 msec (0.15 sec). Crudely, this would represent the mean value and standard deviation of the drivers' reaction time on day 0. We were expecting the drivers to have some small reaction time above 0 on day 0 of driving, because they were well-rested, and this reaction time to increase as the days pass by and the drivers become sleep-deprived. For setting the prior for parameter b, namely the slope of the $\mathrm{reaction\_time}$'s mean value that represents the amount of time in msec that the reaction time of the driver increases in each day, we set a hyper-prior for the mean value of its prior distribution with mean value equal to 10 msec (0.01 sec) and standard deviation 100 msec (0.1 sec). We were expecting that the drivers' reaction time would increase with day, but we had no previous knowledge of how much this increase could be.

We set the standard deviation of the a parameter to a higher value (150 msec) than the standard deviation of the b parameter (100 msec), as were expecting more variation to the drivers' reaction times at rest as this could reflect their individual traits, than to the effect of sleep deprivation on drivers (we thought that tiredness more or less affects drivers in the same ways). Finally, for the prior distribution for the standard deviation of the $\mathrm{reaction\_time}\}$'s likelihood a hyper-prior was set with standard deviation equal to 200 msec to account for bigger variations among drivers and days.

$$\mathrm{c} \sim \mathrm{Normal}(\mu = 100, \sigma = 150)$$

$$\mathrm{e} \sim \mathrm{Half\text{-}Normal}(\sigma = 150)$$

$$\mathrm{f} \sim \mathrm{Normal}(\mu = 10, \sigma = 100)$$

$$\mathrm{g} \sim \mathrm{Half\text{-}Normal}(\sigma = 100)$$

$$\mathrm{h} \sim \mathrm{Half\text{-}Normal}(\sigma = 200)$$

$$\mathrm{a_i} \sim \mathrm{Normal}(\mu = \mathrm{c}, \sigma = \mathrm{e})$$

$$\mathrm{b_i} \sim \mathrm{Normal}(\mu = \mathrm{f}, \sigma = \mathrm{g})$$

$$\mathrm{sigma_i} \sim \mathrm{Half\text{-}Normal}(\sigma = \mathrm{h})$$

$$\mathrm{d} \sim \mathrm{Normal}(\mu = 0, \sigma = 10)$$

$$\mathrm{reaction\_time_i} \sim \mathrm{Normal}(\mu = \mathrm{a_i} + day \cdot \mathrm{b_i}, \sigma = \mathrm{sigma_i})$$

The PyMC3 code of the model can be found in https://github.com/evdoxiataka/ipme/tree/master/examples/user_study/reaction_times. The prior samples from the model used in the study could be found in the file `study_analysis\data\reaction_times_hierarchical.npz` in [*S1*]. The data used for the definition of the likelihood was taken from the study presented in [*S2*].

S.III Analysis

The Bayesian models used for the analysis of the user study's collected data were designed and interpreted in PyMC3. The code for the models' specification is presented here and could be found as Python Jupyter Notebooks along with the collected data from the user study in [*S1*]. Please note that the variables' names are slightly different in the presented code below to be in alignment with Kruschke-style diagrams of the models presented in Fig. 5 of the paper.

S.III.1 Accuracy's Model

Two separate models were used for the analysis of accuracy; one for RQ1 tasks and the other for the RQ2-RQ3 tasks. The models were different in the likelihood used for the accuracy. A binomial likelihood was used for RQ1 tasks because multiple selections were allowed. A Bernoulli likelihood was used for the rest of tasks because only a single selection was allowed.

Beta priors with $\alpha = 1.0$ and $\beta = 1.0$ were set in both models for the probabilities of success ($\mathrm{thetaIG}$ and $\mathrm{thetaSG}$). These priors correspond to a uniform distribution with bounds between 0 and 1 and is a reasonable uninformative option in this case.

S.III.1.1 Model for RQ1

```
import pymc3 as pm
import numpy as np

coords = {"task": t_ids}
with pm.Model(coords=coords) as model:
    #priors
    thetaIG = pm.Beta("thetaIG", alpha = 1.0, beta = 1.0, dims = 'task')
    thetaSG = pm.Beta("thetaSG", alpha = 1.0, beta = 1.0, dims = 'task')

    #likelihood
    accuracyIG = pm.Binomial("accuracyIG", n = n_i,
                                          p = thetaIG[t_indices_i],
                                          observed = answers_i)
    accuracySG = pm.Binomial("accuracySG", n = n_s,
                                     p = thetaSG[t_indices_s],
                                     observed = answers_s)

    #comparisons
    diff_of_thetas = pm.Deterministic("difference of thetas",
                                      thetaIG - thetaSG,
                                      dims='task')

    #inference
    trace = pm.sample(2000)
```

S.III.1.2 Model for RQ2-RQ3

```
import pymc3 as pm
import numpy as np

coords = {"task": t_ids}
with pm.Model(coords=coords) as model:
    #priors
    thetaIG = pm.Beta("thetaIG", alpha = 1.0, beta = 1.0, dims = 'task')
    thetaSG = pm.Beta("thetaSG", alpha = 1.0, beta = 1.0, dims = 'task')

    #likelihood
    accuracyIG = pm.Bernoulli("accuracyIG", p = thetaIG[t_indices_i],
                                           observed = answers_i)
    accuracySG = pm.Bernoulli("accuracySG", p = thetaSG[t_indices_s],
                                      observed = answers_s)

    #comparisons
    diff_of_thetas = pm.Deterministic("difference of thetas",
                                      thetaIG - thetaSG,
                                      dims='task')

    #inference
    trace = pm.sample(2000)
```

S.III.2 Response Times' Model

The response times of the participants were continuous values and we assumed a normal likelihood to model them. A normal prior distribution was set for the μ and a half-normal prior distribution for the σ parameter of the response times' likelihood. The user study was designed so that each participant spends 2-3 min on average on each task. So, we set $\mu = 120$ sec for the priors and allowed for a variance of 60 sec to account for the fact that some tasks could be completed in less or more time depending on the complexity of the presented structure.

```
import pymc3 as pm
import numpy as np

coords = {"task": t_ids}
with pm.Model(coords=coords) as model:
    #priors
    groupIG_mean = pm.Normal("groupIG_mean", mu = 120, sd = 60, dims = 'task')
    groupIG_std = pm.HalfNormal("groupIG_std", sd = 90, dims = 'task')

    groupSG_mean = pm.Normal("groupSG_mean", mu = 120, sd = 60, dims = 'task')
    groupSG_std = pm.HalfNormal("groupSG_std", sd = 90, dims = 'task')

    #likelihood
    rtIG = pm.Normal("rtIG", mu = groupIG_mean[t_indices_i],
                                      sd = groupIG_std[t_indices_i],
                                      observed = times_i)# sec
    rtSG = pm.Normal("rtSG", mu = groupSG_mean[t_indices_s],
                                 sd = groupSG_std[t_indices_s],
                                 observed = times_s)# sec

    #comparisons
    diff_of_means = pm.Deterministic("difference of means",
                                     groupIG_mean - groupSG_mean,
                                     dims = 'task')
    effect_size = pm.Deterministic("effect size",
    diff_of_means / np.sqrt((groupIG_std ** 2 + groupSG_std ** 2) / 2),
                                    dims = 'task')

    #inference
    trace = pm.sample(2000)
```

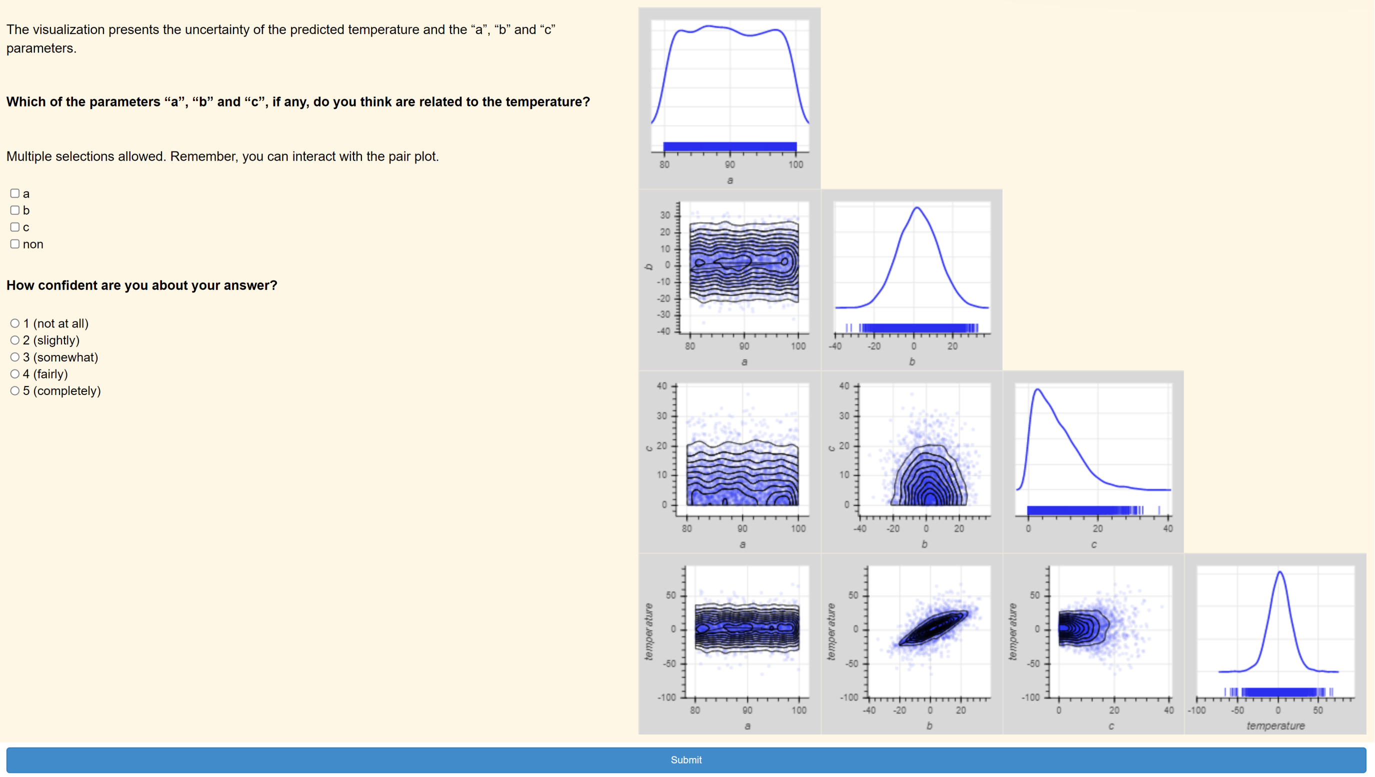
S.III.3 Confidence's Model

Although the structure of the model used for the analysis of confidence is the same as that used for the response times, the values of the parameters of priors were different. The recorded confidence levels of the participants were mapped to a $[-2, 2]$ scale. Thus, we centered the prior for the μ of the likelihood around 0, as we had no previous experience or knowledge about how high users' confidence would be, and allowed for a variance of 1 to create uninformative enough priors.

```
import pymc3 as pm
import numpy as np

coords = {"task": t_ids}
with pm.Model(coords=coords) as model:
    #priors
    groupIG_mean = pm.Normal("groupIG_mean", mu = 0, sd = 1, dims = 'task')
    groupIG_std = pm.HalfNormal("groupIG_std", sd = 1, dims = 'task')
    groupSG_mean = pm.Normal("groupSG_mean", mu = 0, sd = 1, dims = 'task')
    groupSG_std = pm.HalfNormal("groupSG_std", sd = 1, dims = 'task')

    #likelihood
    confIG = pm.Normal("confIG", mu = groupIG_mean[t_indices_i],
                                  sd = groupIG_std[t_indices_i],
                                  observed = conf_i)
    confSG = pm.Normal("confSG", mu = groupSG_mean[t_indices_s],
                            sd = groupSG_std[t_indices_s],
                            observed = conf_s)

    #comparisons
    diff_of_means = pm.Deterministic("difference of means",
                                 groupIG_mean - groupSG_mean,
                                 dims = 'task')
    effect_size = pm.Deterministic("effect size",
    diff_of_means / np.sqrt((groupIG_std ** 2 + groupSG_std ** 2) / 2),
                               dims = 'task')

    #inference
    trace = pm.sample(2000)
```

S.IV Tasks

Fig. S1-S19 present the study questions as they were presented to participants during the study in exactly the same order.



The visualization presents the uncertainty of the predicted temperature and the “a”, “b” and “c” parameters.

Which of the parameters “a”, “b” and “c”, if any, do you think are related to the temperature?

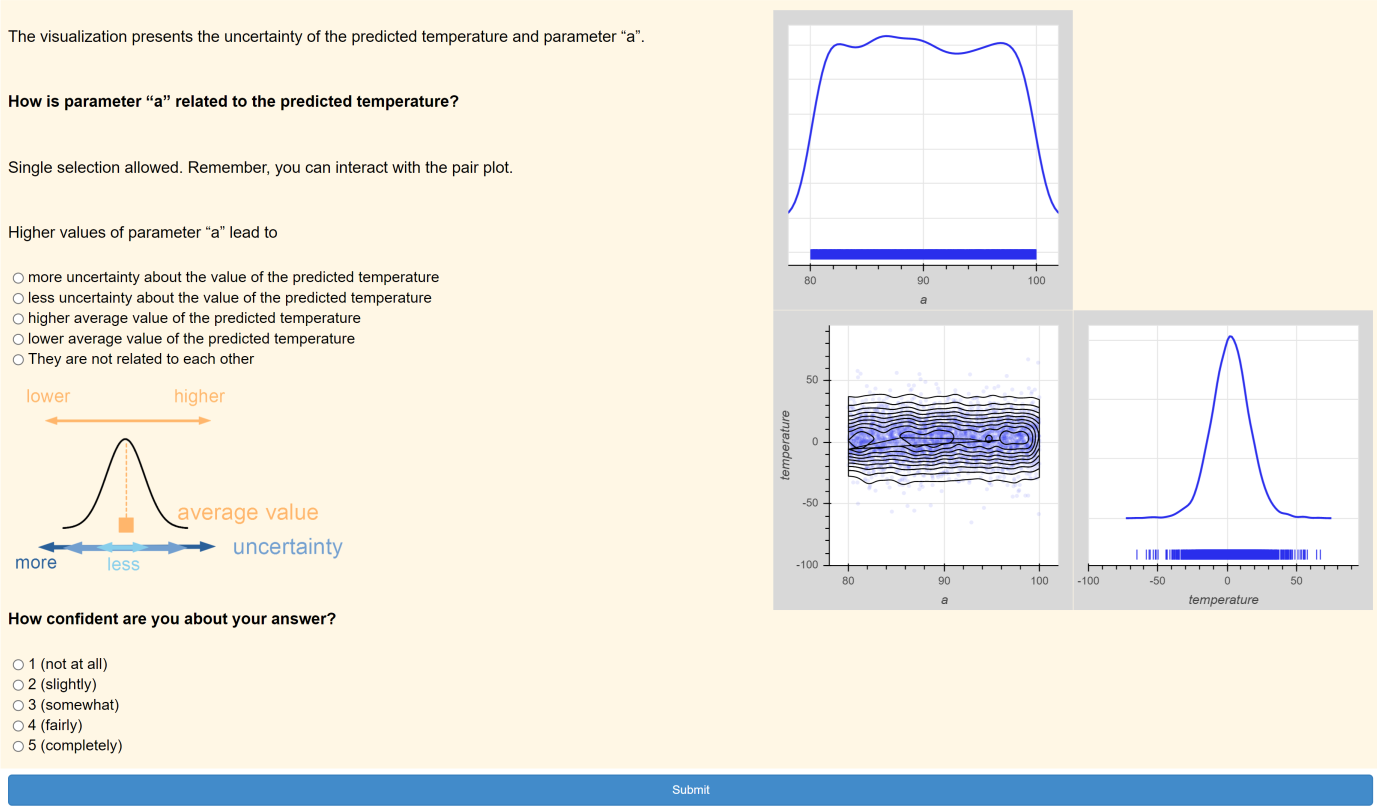
Multiple selections allowed. Remember, you can interact with the pair plot.

□ a
□ b
□ c
□ non

How confident are you about your answer?

○ 1 (not at all)
○ 2 (slightly)
○ 3 (somewhat)
○ 4 (fairly)
○ 5 (completely)

Fig. S1. **Model 1** - Task t1 (RQ1).



The visualization presents the uncertainty of the predicted temperature and parameter “a”.

How is parameter “a” related to the predicted temperature?

Single selection allowed. Remember, you can interact with the pair plot.

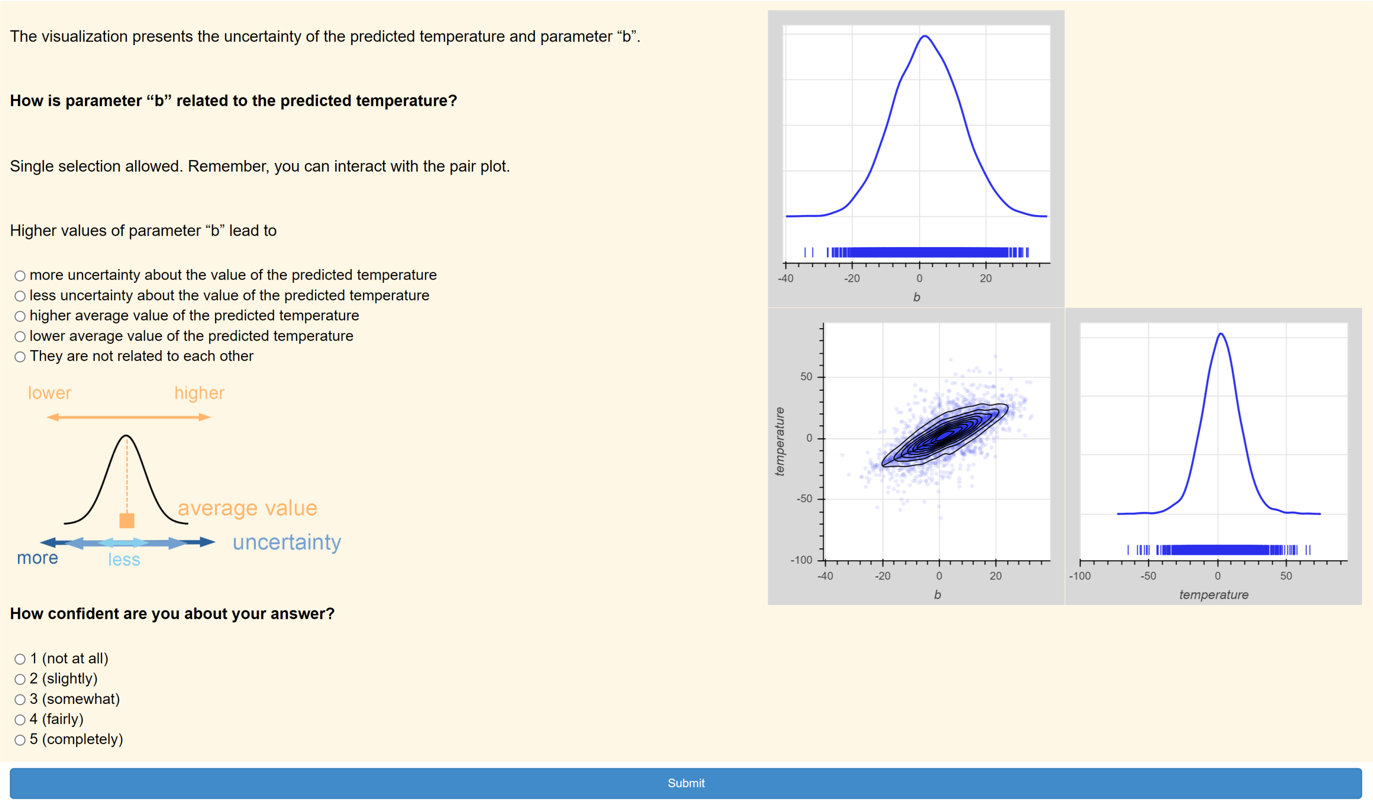
Higher values of parameter “a” lead to

○ more uncertainty about the value of the predicted temperature
○ less uncertainty about the value of the predicted temperature
○ higher average value of the predicted temperature
○ lower average value of the predicted temperature
○ They are not related to each other

How confident are you about your answer?

○ 1 (not at all)
○ 2 (slightly)
○ 3 (somewhat)
○ 4 (fairly)
○ 5 (completely)

Fig. S2. **Model 1** - Task t2 (RQ2).



The visualization presents the uncertainty of the predicted temperature and parameter “b”.

How is parameter “b” related to the predicted temperature?

Single selection allowed. Remember, you can interact with the pair plot.

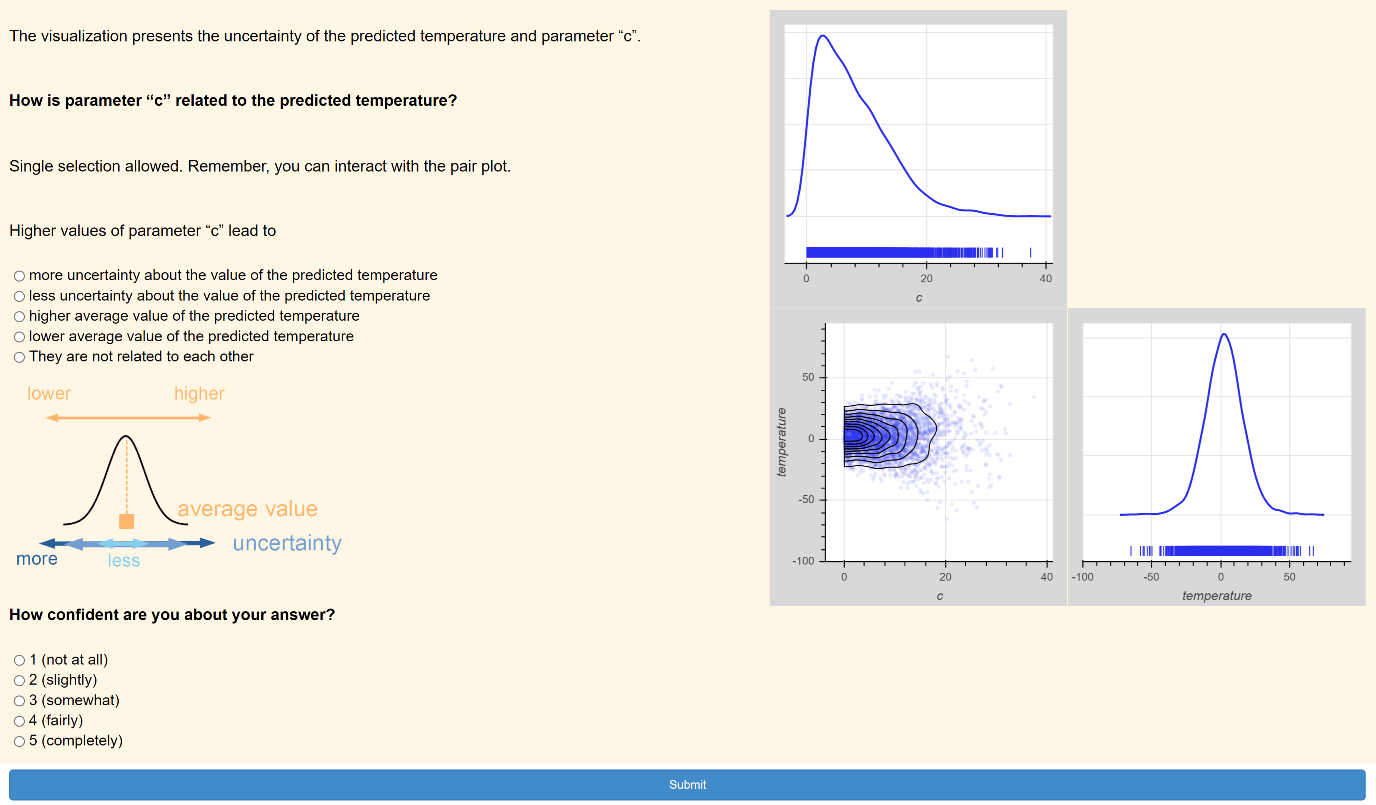
Higher values of parameter “b” lead to

○ more uncertainty about the value of the predicted temperature
○ less uncertainty about the value of the predicted temperature
○ higher average value of the predicted temperature
○ lower average value of the predicted temperature
○ They are not related to each other

How confident are you about your answer?

○ 1 (not at all)
○ 2 (slightly)
○ 3 (somewhat)
○ 4 (fairly)
○ 5 (completely)

Fig. S3. **Model 1** - Task t3 (RQ2).



The visualization presents the uncertainty of the predicted temperature and parameter “c”.

How is parameter “c” related to the predicted temperature?

Single selection allowed. Remember, you can interact with the pair plot.

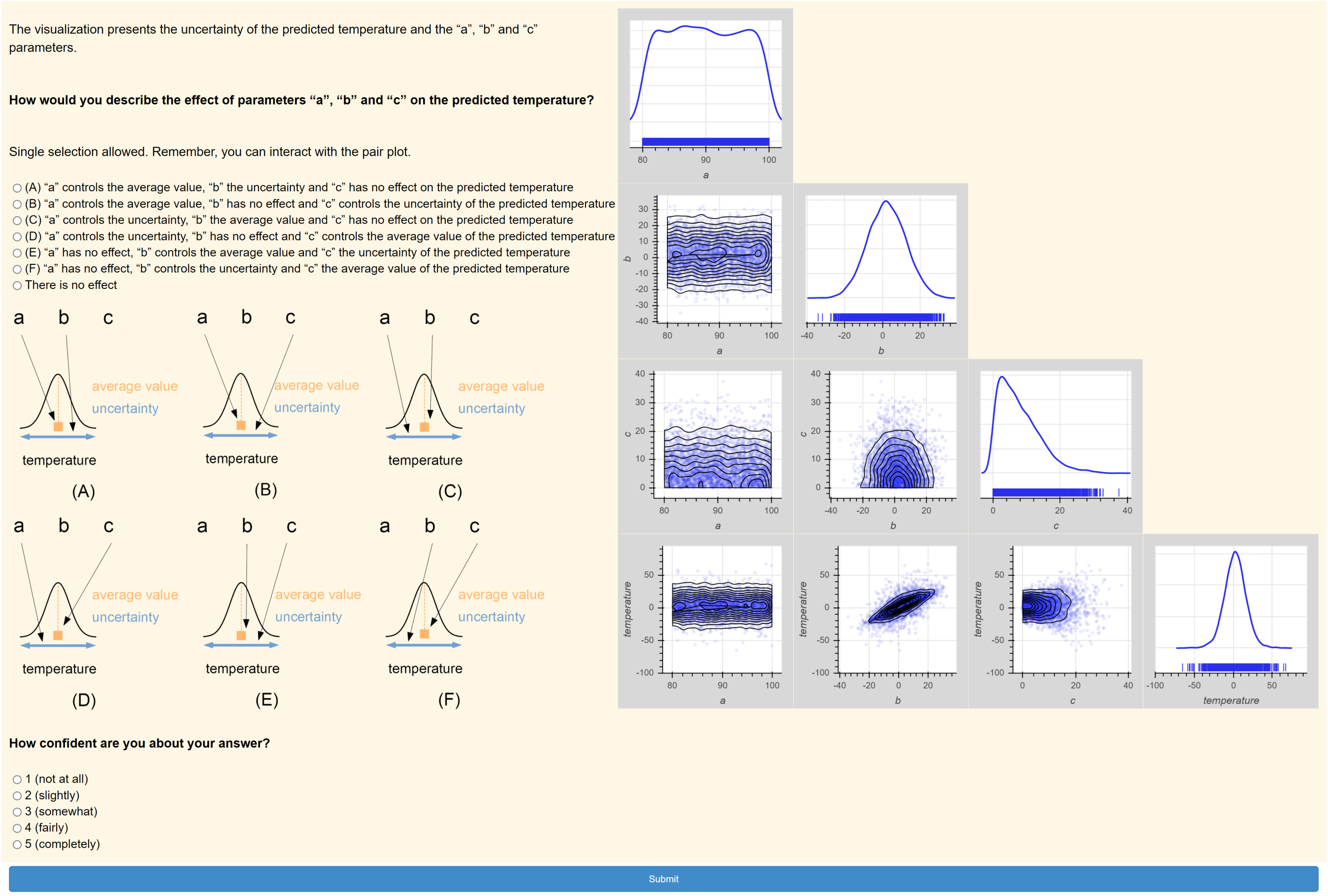
Higher values of parameter “c” lead to

○ more uncertainty about the value of the predicted temperature
○ less uncertainty about the value of the predicted temperature
○ higher average value of the predicted temperature
○ lower average value of the predicted temperature
○ They are not related to each other

How confident are you about your answer?

○ 1 (not at all)
○ 2 (slightly)
○ 3 (somewhat)
○ 4 (fairly)
○ 5 (completely)

Fig. S4. **Model 1** - Task t4 (RQ2).



The visualization presents the uncertainty of the predicted temperature and the “a”, “b” and “c” parameters.

How would you describe the effect of parameters “a”, “b” and “c” on the predicted temperature?

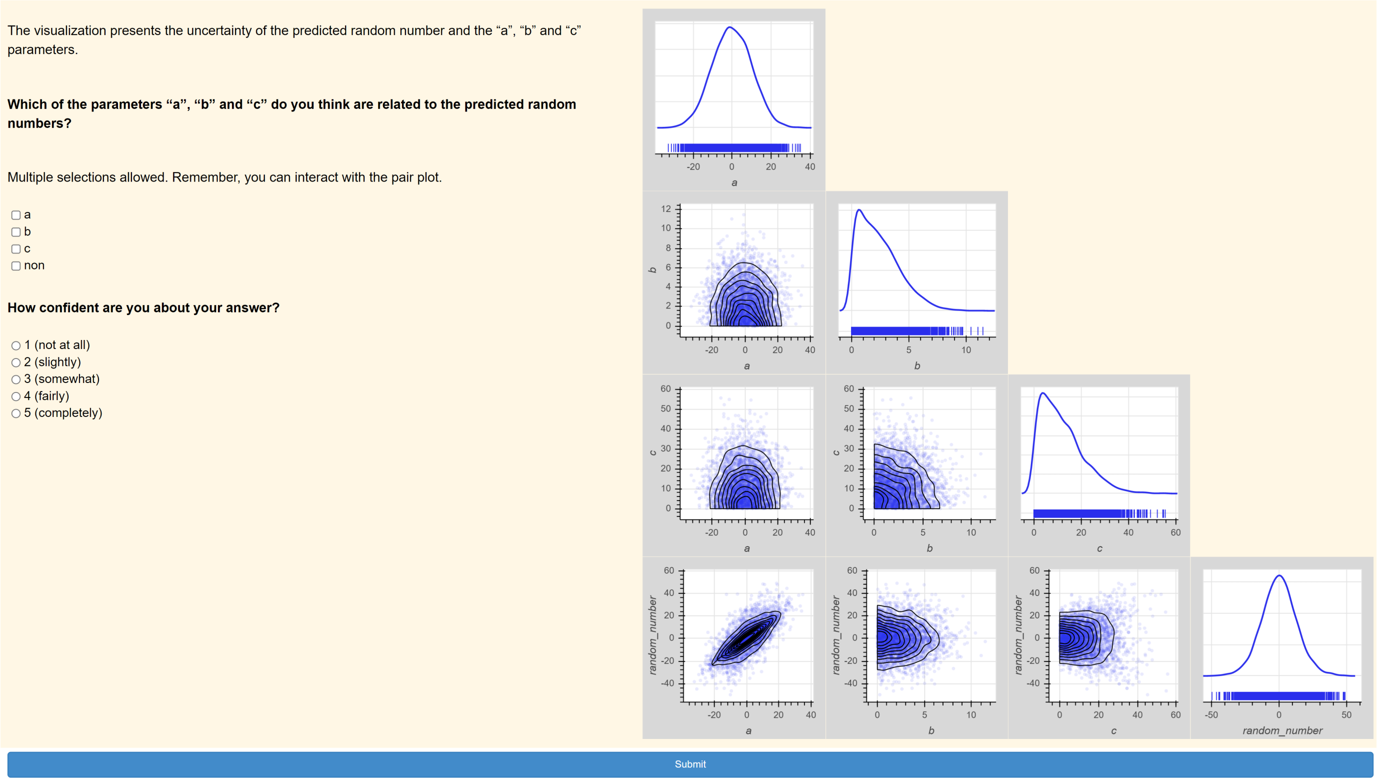
Single selection allowed. Remember, you can interact with the pair plot.

○ (A) “a” controls the average value, “b” the uncertainty and “c” has no effect on the predicted temperature
○ (B) “a” controls the average value, “b” has no effect and “c” controls the uncertainty of the predicted temperature
○ (C) “a” controls the uncertainty, “b” the average value and “c” has no effect on the predicted temperature
○ (D) “a” controls the uncertainty, “b” has no effect and “c” controls the average value of the predicted temperature
○ (E) “a” has no effect, “b” controls the average value and “c” the uncertainty of the predicted temperature
○ (F) “a” has no effect, “b” controls the uncertainty and “c” the average value of the predicted temperature
○ There is no effect

How confident are you about your answer?

○ 1 (not at all)
○ 2 (slightly)
○ 3 (somewhat)
○ 4 (fairly)
○ 5 (completely)

Fig. S5. **Model 1** - Task t5 (RQ3).



The visualization presents the uncertainty of the predicted random number and the “a”, “b” and “c” parameters.

Which of the parameters “a”, “b” and “c” do you think are related to the predicted random numbers?

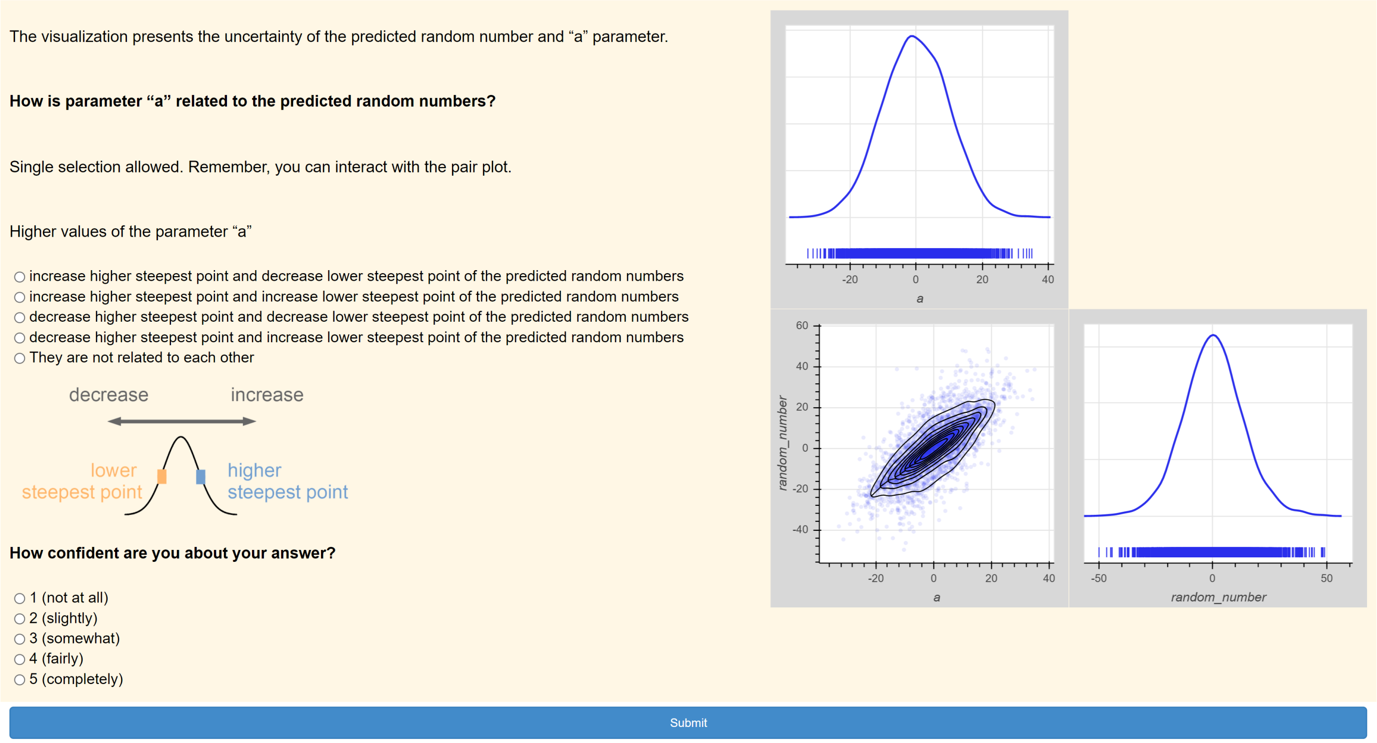
Multiple selections allowed. Remember, you can interact with the pair plot.

□ a
□ b
□ c
□ non

How confident are you about your answer?

○ 1 (not at all)
○ 2 (slightly)
○ 3 (somewhat)
○ 4 (fairly)
○ 5 (completely)

Fig. S6. **Model 2** - Task t6 (RQ1).



The visualization presents the uncertainty of the predicted random number and “a” parameter.

How is parameter “a” related to the predicted random numbers?

Single selection allowed. Remember, you can interact with the pair plot.

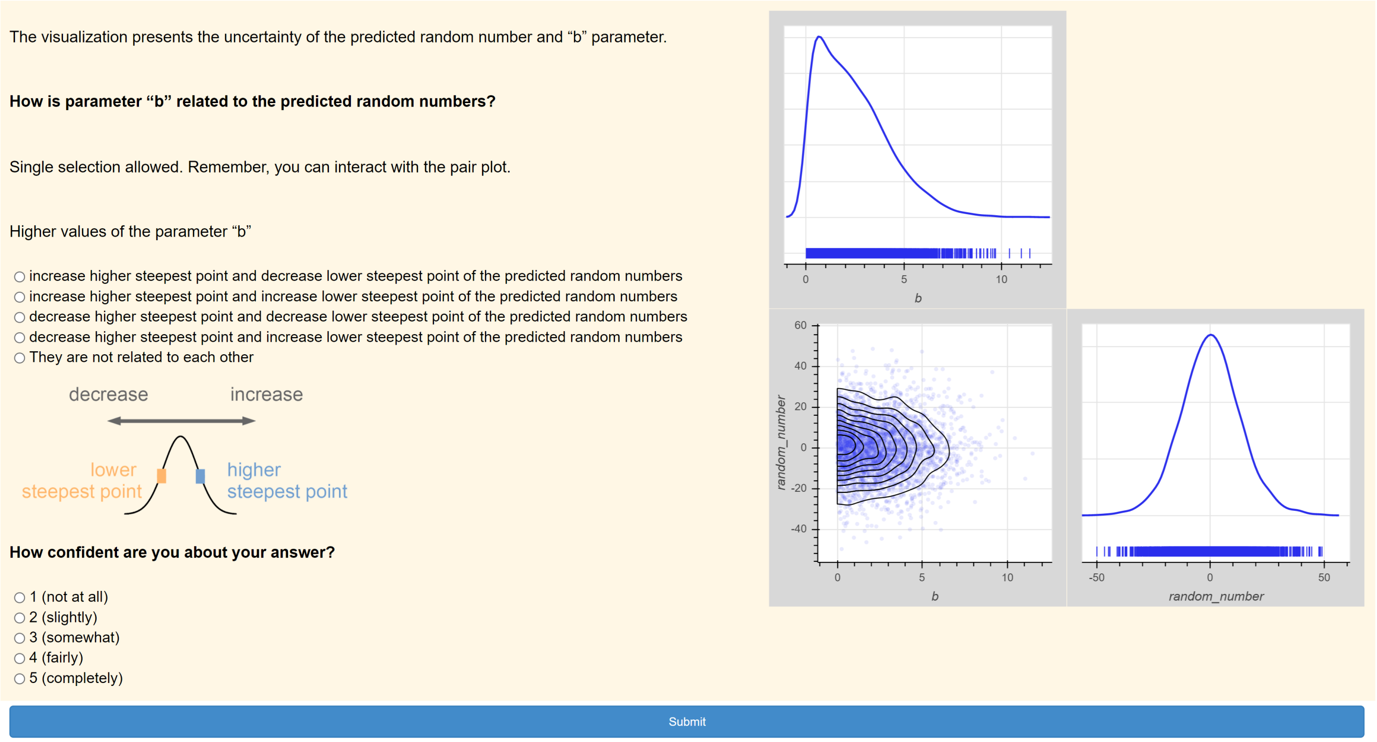
Higher values of the parameter “a”

○ increase higher steepest point and decrease lower steepest point of the predicted random numbers
○ increase higher steepest point and increase lower steepest point of the predicted random numbers
○ decrease higher steepest point and decrease lower steepest point of the predicted random numbers
○ decrease higher steepest point and increase lower steepest point of the predicted random numbers
○ They are not related to each other

How confident are you about your answer?

○ 1 (not at all)
○ 2 (slightly)
○ 3 (somewhat)
○ 4 (fairly)
○ 5 (completely)

Fig. S7. **Model 2** - Task t7 (RQ2).



The visualization presents the uncertainty of the predicted random number and “b” parameter.

How is parameter “b” related to the predicted random numbers?

Single selection allowed. Remember, you can interact with the pair plot.

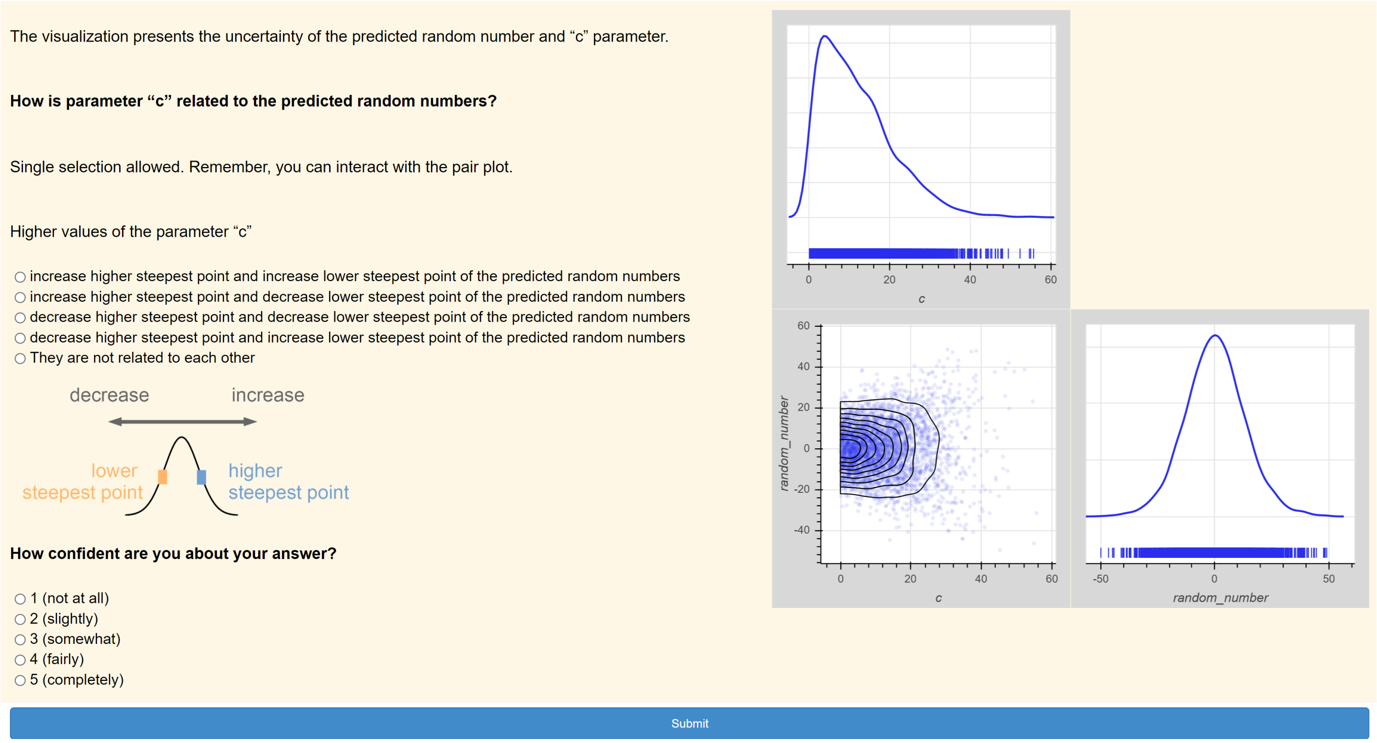
Higher values of the parameter “b”

○ increase higher steepest point and decrease lower steepest point of the predicted random numbers
○ increase higher steepest point and increase lower steepest point of the predicted random numbers
○ decrease higher steepest point and decrease lower steepest point of the predicted random numbers
○ decrease higher steepest point and increase lower steepest point of the predicted random numbers
○ They are not related to each other

How confident are you about your answer?

○ 1 (not at all)
○ 2 (slightly)
○ 3 (somewhat)
○ 4 (fairly)
○ 5 (completely)

Fig. S8. **Model 2** - Task t8 (RQ2).



The visualization presents the uncertainty of the predicted random number and “c” parameter.

How is parameter “c” related to the predicted random numbers?

Single selection allowed. Remember, you can interact with the pair plot.

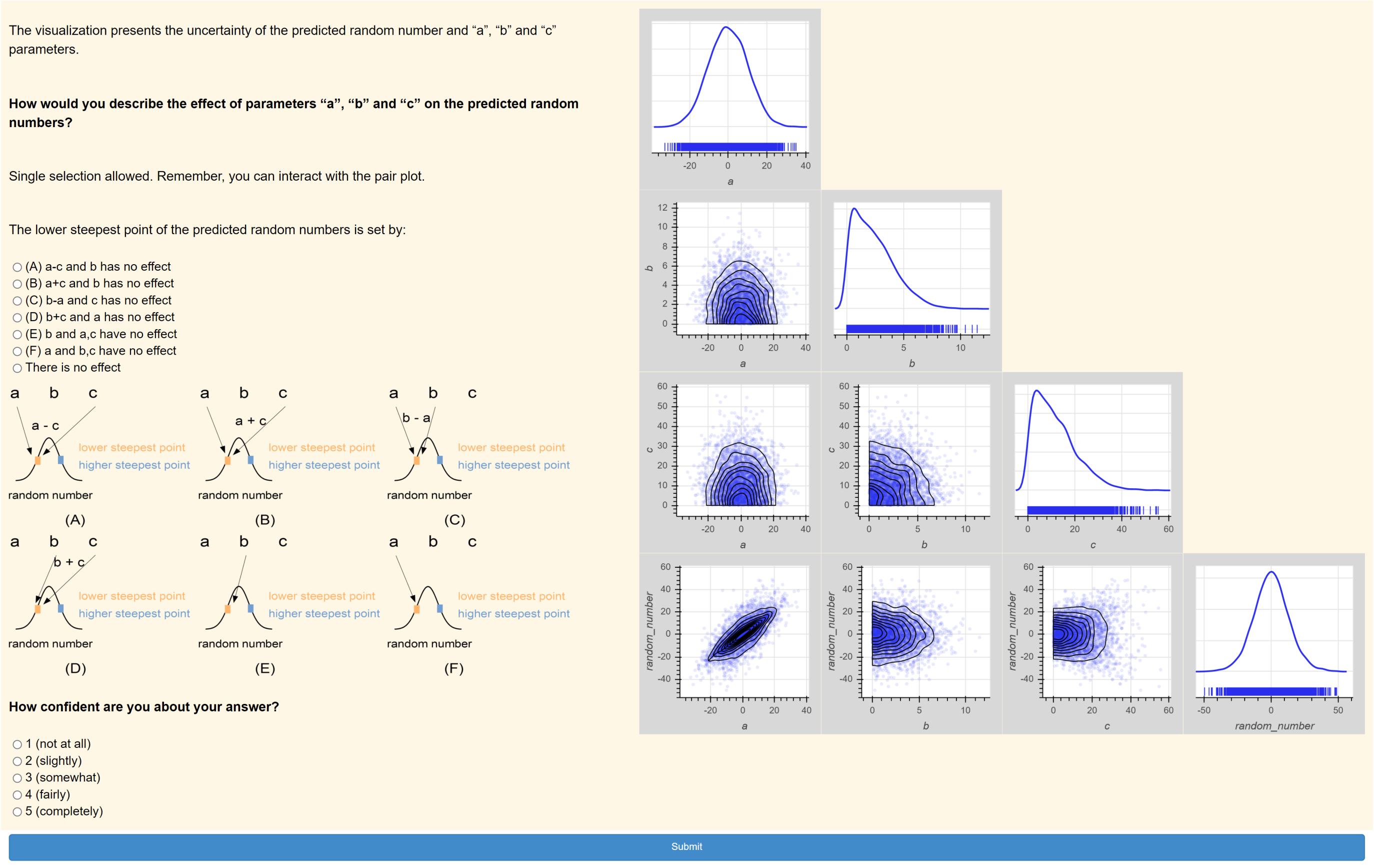
Higher values of the parameter “c”

○ increase higher steepest point and increase lower steepest point of the predicted random numbers
○ increase higher steepest point and decrease lower steepest point of the predicted random numbers
○ decrease higher steepest point and decrease lower steepest point of the predicted random numbers
○ decrease higher steepest point and increase lower steepest point of the predicted random numbers
○ They are not related to each other

How confident are you about your answer?

○ 1 (not at all)
○ 2 (slightly)
○ 3 (somewhat)
○ 4 (fairly)
○ 5 (completely)

Fig. S9. **Model 2** - Task t9 (RQ2).



The visualization presents the uncertainty of the predicted random number and “a”, “b” and “c” parameters.

How would you describe the effect of parameters “a”, “b” and “c” on the predicted random numbers?

Single selection allowed. Remember, you can interact with the pair plot.

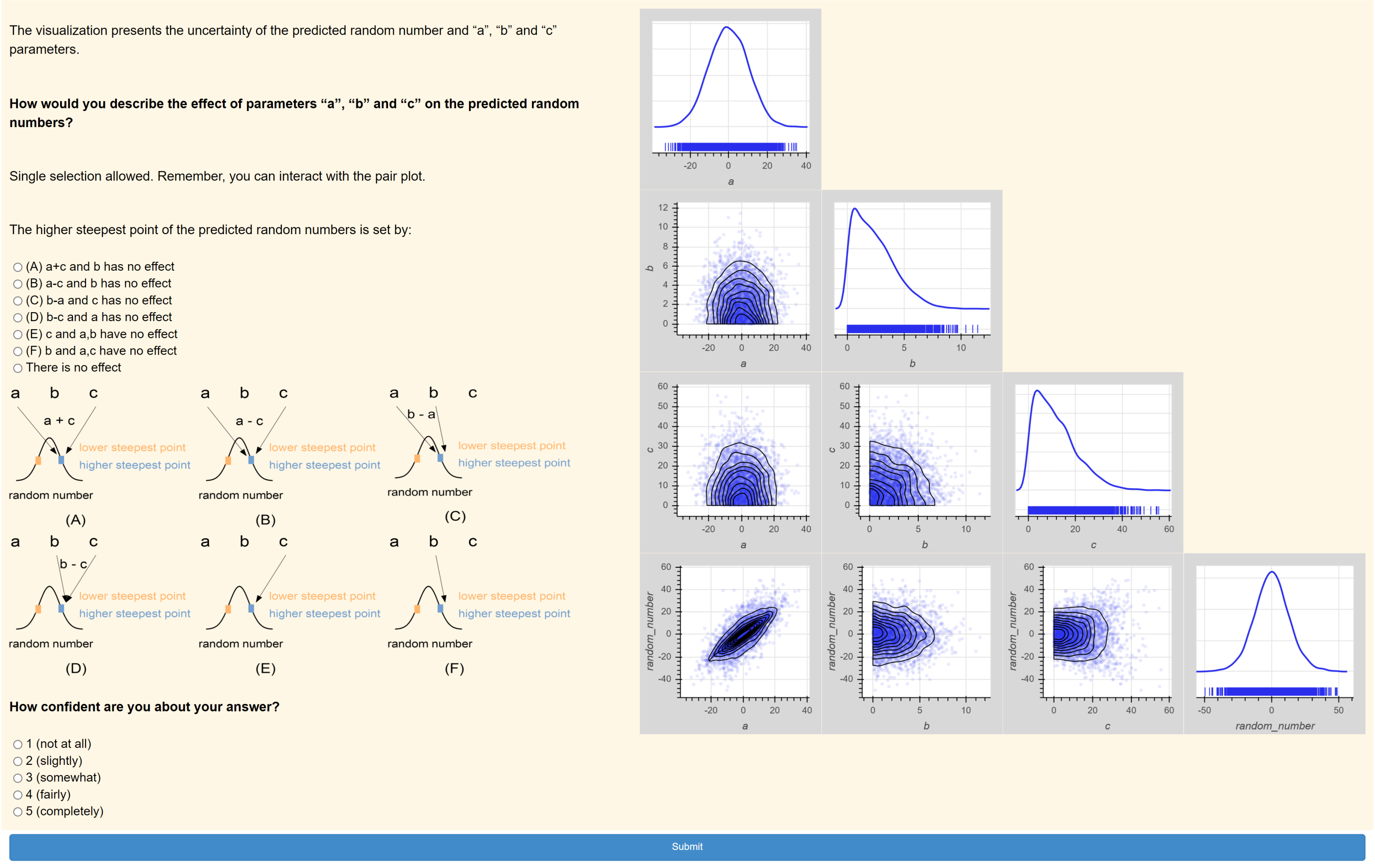
The lower steepest point of the predicted random numbers is set by:

○ (A) a-c and b has no effect
○ (B) a+c and b has no effect
○ (C) b-a and c has no effect
○ (D) b+c and a has no effect
○ (E) b and a,c have no effect
○ (F) a and b,c have no effect
○ There is no effect

How confident are you about your answer?

○ 1 (not at all)
○ 2 (slightly)
○ 3 (somewhat)
○ 4 (fairly)
○ 5 (completely)

Fig. S10. **Model 2** - Task t10 (RQ3).



The visualization presents the uncertainty of the predicted random number and “a”, “b” and “c” parameters.

How would you describe the effect of parameters “a”, “b” and “c” on the predicted random numbers?

Single selection allowed. Remember, you can interact with the pair plot.

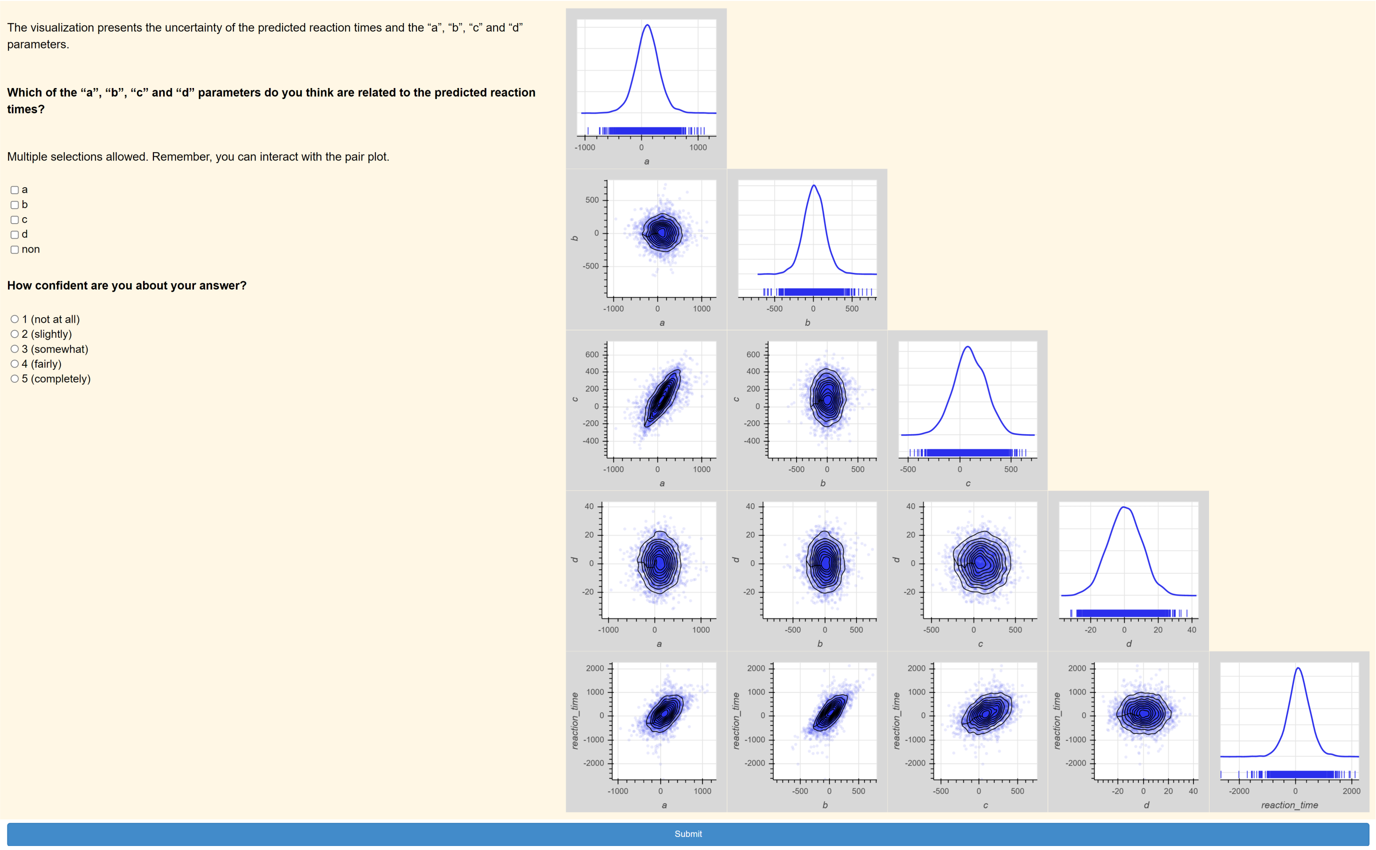
The higher steepest point of the predicted random numbers is set by:

○ (A) a+c and b has no effect
○ (B) a-c and b has no effect
○ (C) b-a and c has no effect
○ (D) b-c and a has no effect
○ (E) c and a,b have no effect
○ (F) b and a,c have no effect
○ There is no effect

How confident are you about your answer?

○ 1 (not at all)
○ 2 (slightly)
○ 3 (somewhat)
○ 4 (fairly)
○ 5 (completely)

Fig. S11. **Model 2** - Task t11 (RQ3).



The visualization presents the uncertainty of the predicted reaction times and the “a”, “b”, “c” and “d” parameters.

Which of the “a”, “b”, “c” and “d” parameters do you think are related to the predicted reaction times?

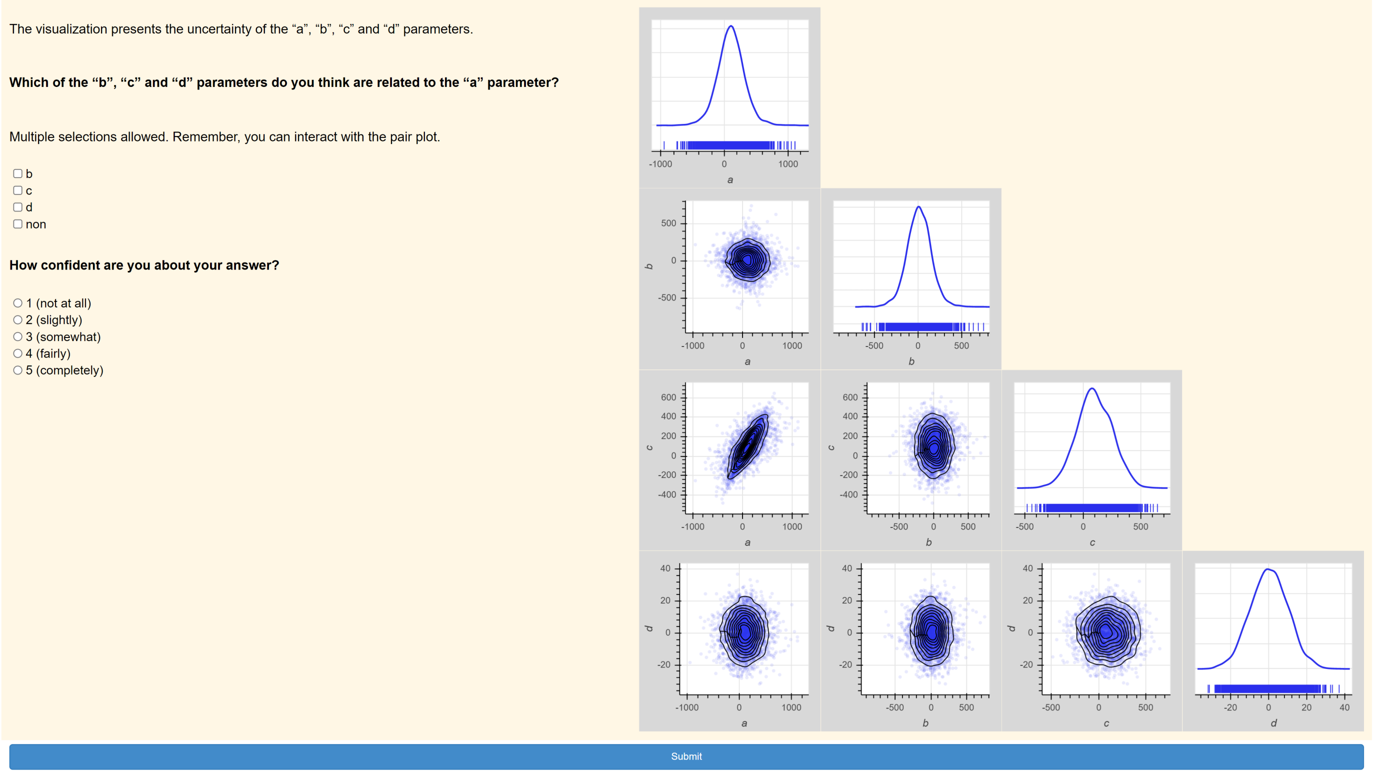
Multiple selections allowed. Remember, you can interact with the pair plot.

□ a
□ b
□ c
□ d
□ non

How confident are you about your answer?

○ 1 (not at all)
○ 2 (slightly)
○ 3 (somewhat)
○ 4 (fairly)
○ 5 (completely)

Fig. S12. **Model 3** - Task t12 (RQ1).



The visualization presents the uncertainty of the “a”, “b”, “c” and “d” parameters.

Which of the “b”, “c” and “d” parameters do you think are related to the “a” parameter?

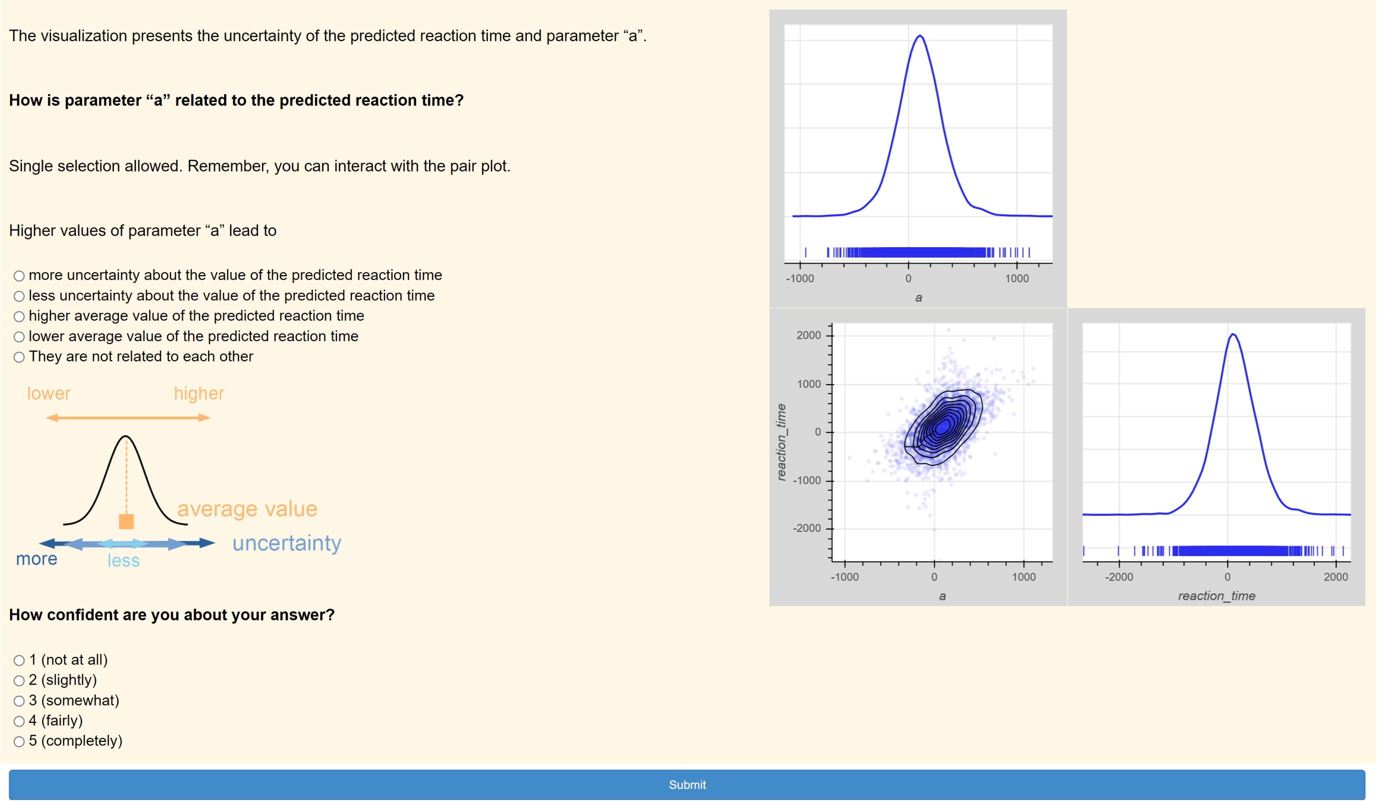
Multiple selections allowed. Remember, you can interact with the pair plot.

□ b
□ c
□ d
□ non

How confident are you about your answer?

○ 1 (not at all)
○ 2 (slightly)
○ 3 (somewhat)
○ 4 (fairly)
○ 5 (completely)

Fig. S13. **Model 3** - Task t13 (RQ1).



The visualization presents the uncertainty of the predicted reaction time and parameter “a”.

How is parameter “a” related to the predicted reaction time?

Single selection allowed. Remember, you can interact with the pair plot.

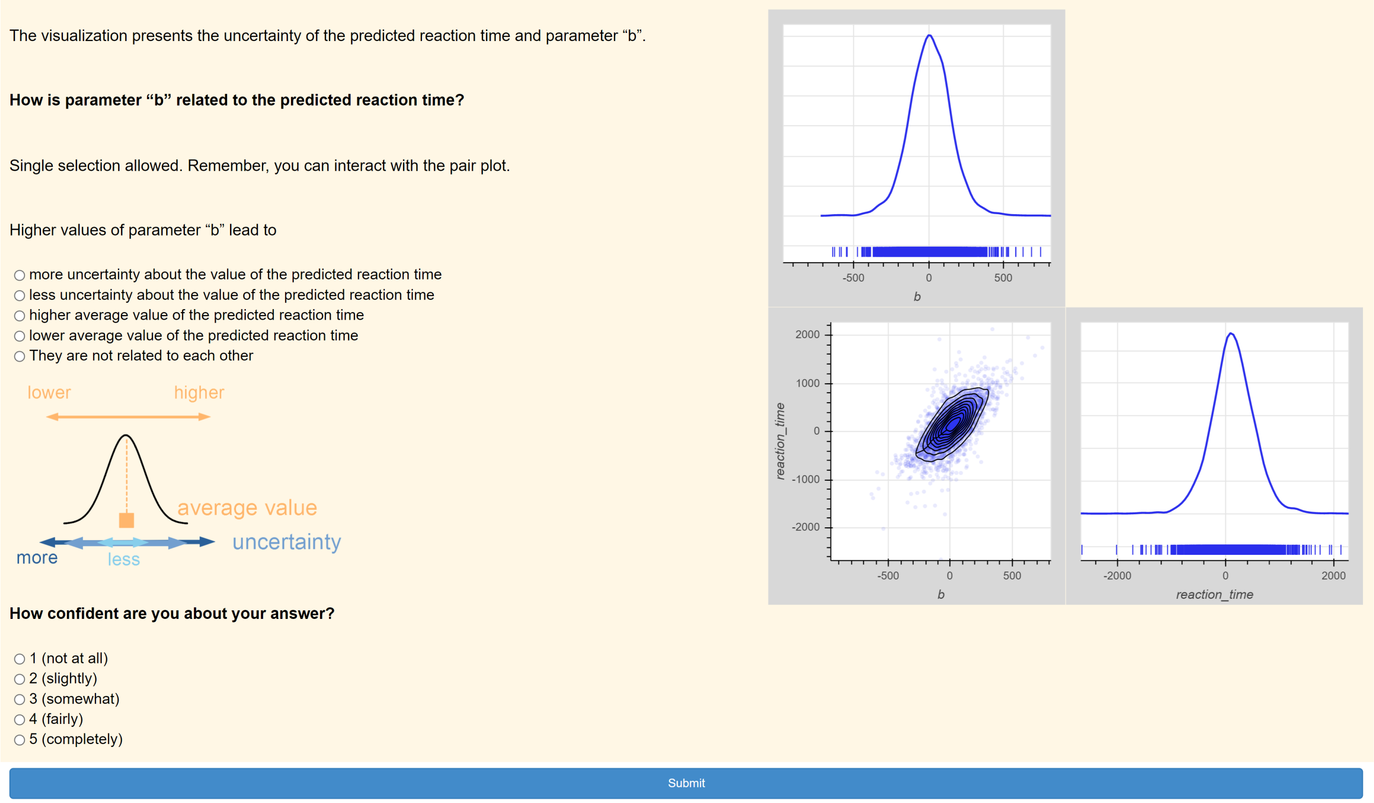
Higher values of parameter “a” lead to

○ more uncertainty about the value of the predicted reaction time
○ less uncertainty about the value of the predicted reaction time
○ higher average value of the predicted reaction time
○ lower average value of the predicted reaction time
○ They are not related to each other

How confident are you about your answer?

○ 1 (not at all)
○ 2 (slightly)
○ 3 (somewhat)
○ 4 (fairly)
○ 5 (completely)

Fig. S14. **Model 3** - Task t14 (RQ2).



The visualization presents the uncertainty of the predicted reaction time and parameter “b”.

How is parameter “b” related to the predicted reaction time?

Single selection allowed. Remember, you can interact with the pair plot.

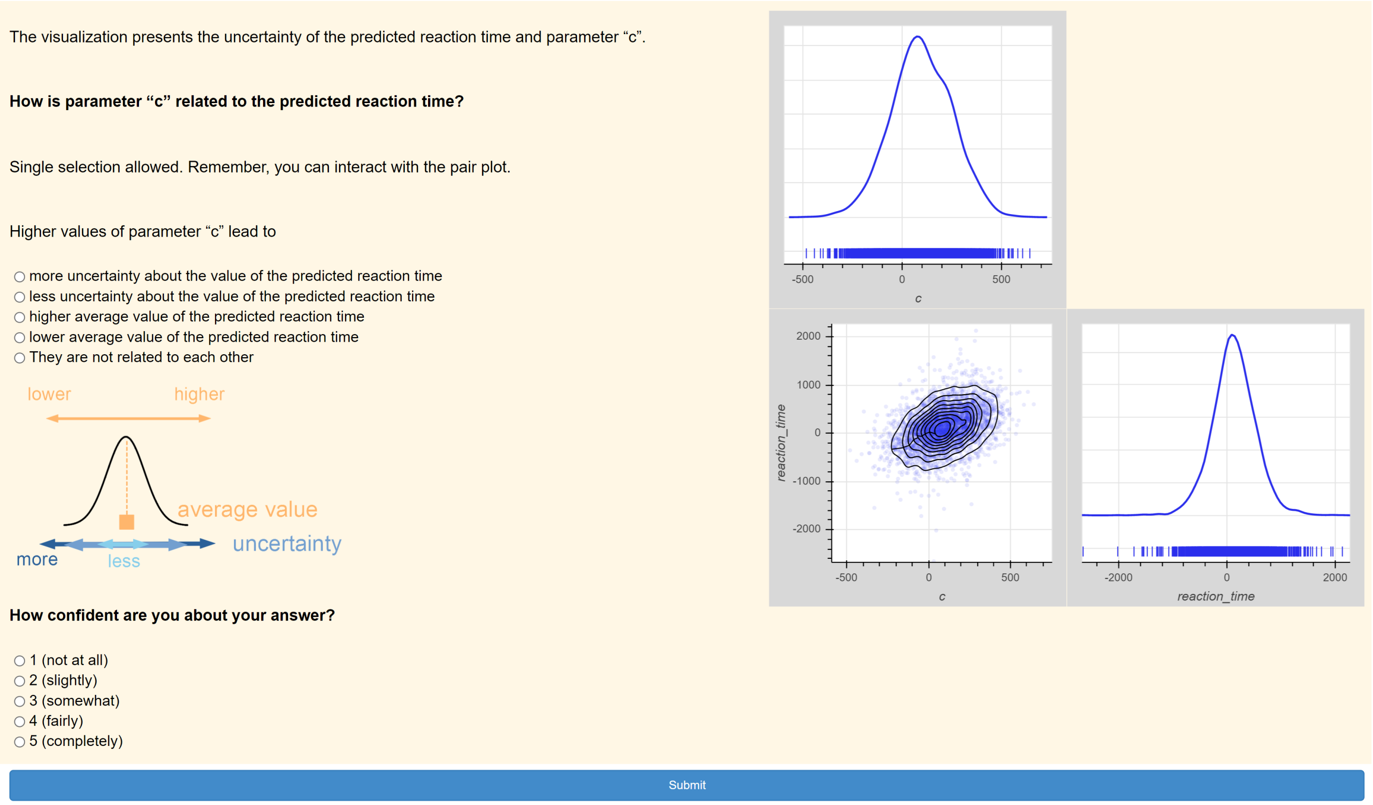
Higher values of parameter “b” lead to

○ more uncertainty about the value of the predicted reaction time
○ less uncertainty about the value of the predicted reaction time
○ higher average value of the predicted reaction time
○ lower average value of the predicted reaction time
○ They are not related to each other

How confident are you about your answer?

○ 1 (not at all)
○ 2 (slightly)
○ 3 (somewhat)
○ 4 (fairly)
○ 5 (completely)

Fig. S15. **Model 3** - Task t15 (RQ2).



The visualization presents the uncertainty of the predicted reaction time and parameter “c”.

How is parameter “c” related to the predicted reaction time?

Single selection allowed. Remember, you can interact with the pair plot.

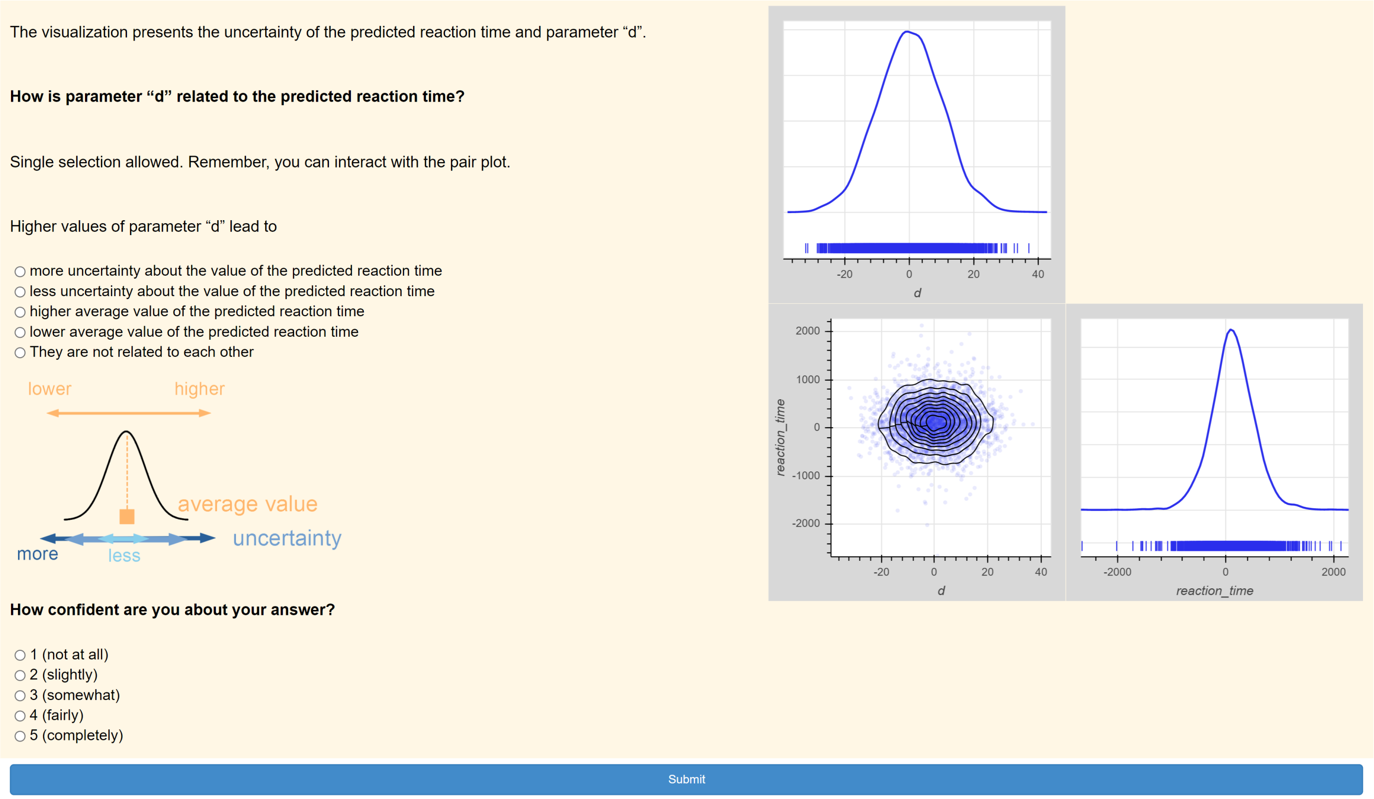
Higher values of parameter “c” lead to

○ more uncertainty about the value of the predicted reaction time
○ less uncertainty about the value of the predicted reaction time
○ higher average value of the predicted reaction time
○ lower average value of the predicted reaction time
○ They are not related to each other

How confident are you about your answer?

○ 1 (not at all)
○ 2 (slightly)
○ 3 (somewhat)
○ 4 (fairly)
○ 5 (completely)

Fig. S16. **Model 3** - Task t16 (RQ2).



The visualization presents the uncertainty of the predicted reaction time and parameter “d”.

How is parameter “d” related to the predicted reaction time?

Single selection allowed. Remember, you can interact with the pair plot.

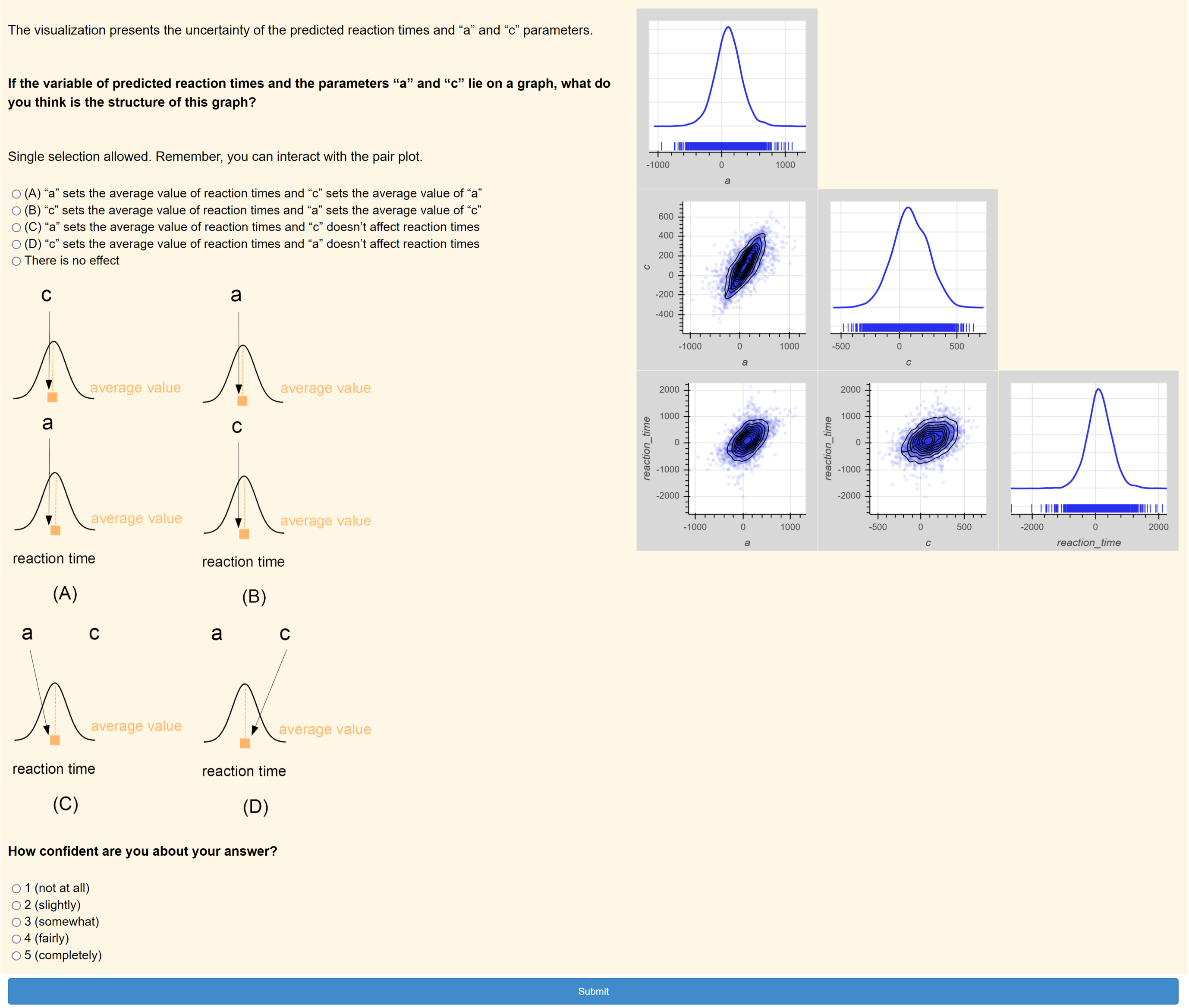
Higher values of parameter “d” lead to

○ more uncertainty about the value of the predicted reaction time
○ less uncertainty about the value of the predicted reaction time
○ higher average value of the predicted reaction time
○ lower average value of the predicted reaction time
○ They are not related to each other

How confident are you about your answer?

○ 1 (not at all)
○ 2 (slightly)
○ 3 (somewhat)
○ 4 (fairly)
○ 5 (completely)

Fig. S17. **Model 3** - Task t17 (RQ2).



The visualization presents the uncertainty of the predicted reaction times and “a” and “c” parameters.

If the variable of predicted reaction times and the parameters “a” and “c” lie on a graph, what do you think is the structure of this graph?

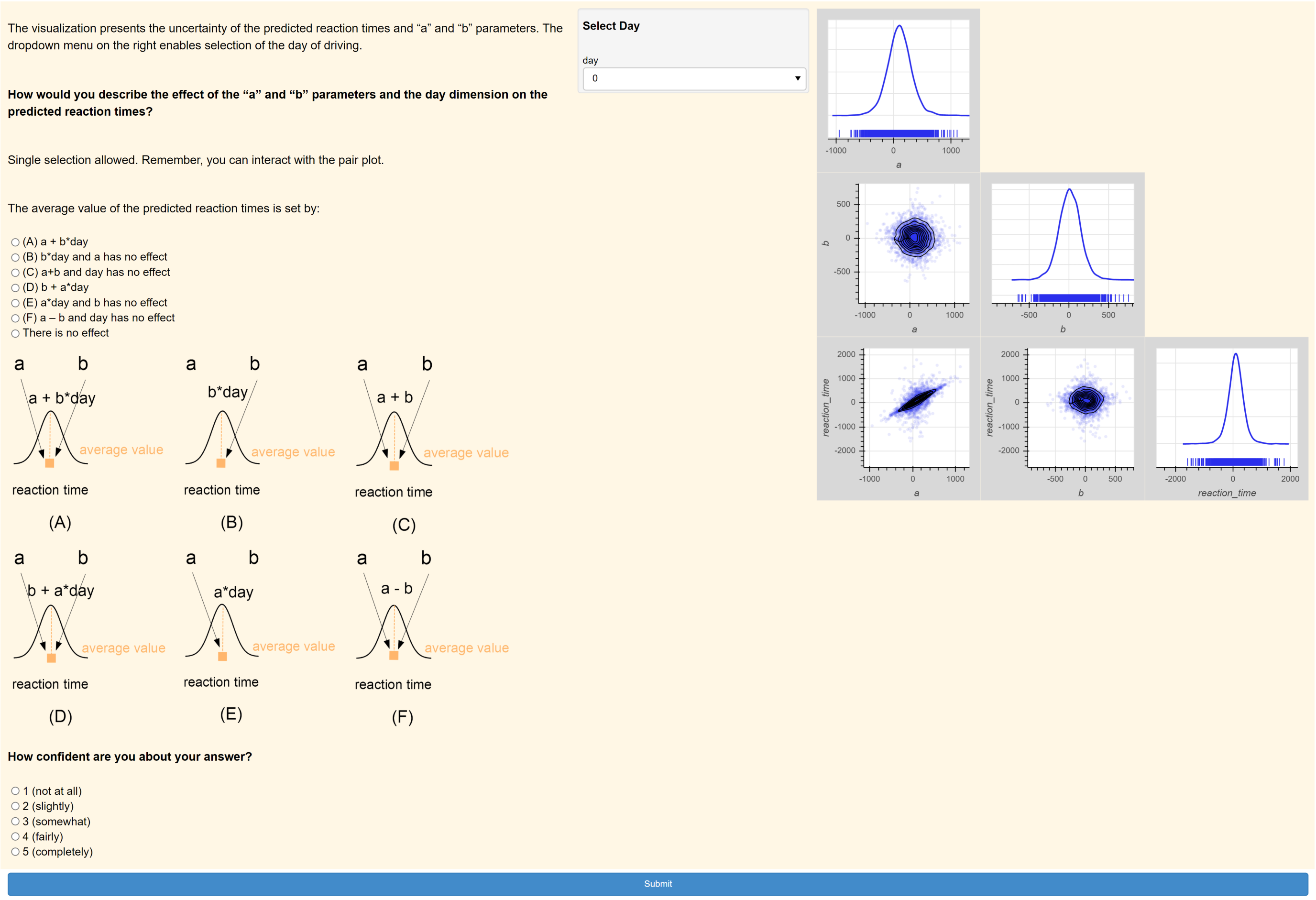
Single selection allowed. Remember, you can interact with the pair plot.

○ (A) “a” sets the average value of reaction times and “c” sets the average value of “a”
○ (B) “c” sets the average value of reaction times and “a” sets the average value of “c”
○ (C) “a” sets the average value of reaction times and “c” doesn’t affect reaction times
○ (D) “c” sets the average value of reaction times and “a” doesn’t affect reaction times
○ There is no effect

How confident are you about your answer?

○ 1 (not at all)
○ 2 (slightly)
○ 3 (somewhat)
○ 4 (fairly)
○ 5 (completely)

Fig. S18. **Model 3** - Task t18 (RQ3).



The visualization presents the uncertainty of the predicted reaction times and “a” and “b” parameters. The dropdown menu on the right enables selection of the day of driving.

How would you describe the effect of the “a” and “b” parameters and the day dimension on the predicted reaction times?

Single selection allowed. Remember, you can interact with the pair plot.

The average value of the predicted reaction times is set by:

○ (A) a + b*day
○ (B) b*day and a has no effect
○ (C) a+b and day has no effect
○ (D) b + a*day
○ (E) a*day and b has no effect
○ (F) a – b and day has no effect
○ There is no effect

How confident are you about your answer?

○ 1 (not at all)
○ 2 (slightly)
○ 3 (somewhat)
○ 4 (fairly)
○ 5 (completely)

Select Day

day: 0 ▾

Fig. S19. **Model 3** - Task t19 (RQ3).

References

[S1] E. Taka, S. Stein, and J. H. Williamson. Does interacting help users better understand the structure of probabilistic models? February 17, 2022. Distributed by University of Glasgow Enlighten Repository. http://dx.doi.org/10.5525/gla.researchdata.1248.

[S2] G. Belenky, N. J. Wesensten, D. R. Thorne, M. L. Thomas, H. C. Sing, D. P. Redmond, M. B. Russo, and T. J. Balkin, “Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose-response study,” *Journal of Sleep Research*, vol. 12, no. 1, pp. 1–12, 2003. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2869.2003.00337.x