

# Quantum Sequential Sampler: Supplementary Materials

Jerome R. Busemeyer\* and Jiaqi Huang\*

Indiana University

Zo Ebel and Emmanuel M. Pothos

City, University of London

## Supplementary Material 1

In this section, we describe the two pilot experiments which were conducted to identify judgments and materials more likely to challenge classical constraints.

### *Pilot Experiments 1, 2*

*Participants.* For Pilot Experiment 1, 98 participants (70 male) from the USA were recruited on Amazon Mechanical Turk, to take part in a linked Qualtrics experiment. No other restrictions apart from location and the mandatory minimum age of 16 years were placed on participation. Each individual was paid \$1.75 to take part in the experiment, which was approximately 20 minutes long. To identify participants disengaged with the task, an attention check question was used (the same one as we used for the experimental investigation, described in the main text). Twenty-two participants failed to answer the attention check correctly and were excluded from any analyses. Another two participants were excluded, despite giving correct answers to the attention check question, since they failed to complete the task in a meaningful way by either rating all probabilities 100% or

0%. Hence the final sample size was reduced to 74 participants (52 male). 73 out of 74 participants were at least 25 years of age and therefore eligible to vote in the US.

For Pilot Experiment 2, 56 participants (25 male) from the USA were recruited in the same manner as above using Amazon Mechanical Turk. This was a shorter experiment, lasting approximately 10 minutes, for which participants were paid \$ 1.00 each. Again, participants failing a simple attention check question were excluded from subsequent analyses. Two participants were excluded in this way, resulting in a final sample of 54 (24 male). 49 out of those participants were at least 25 years of age and as such eligible to vote in the USA.

*Method.* For Pilot Experiment 1, participants were asked to provide 68 probability judgments each on the likelihood of one or both USA presidential candidates (Trump and Biden) winning the popular vote in certain states in the USA. Since the purpose of this experiment was to identify triplets of states that make conjunction fallacies more likely, four triplets of states were tested against each other:

- $T1$ : Ohio, Missouri, Michigan
- $T2$ : North Carolina, South Carolina, Pennsylvania
- $T3$ : Georgia, Montana, Nevada
- $T4$ : Ohio, North Carolina, Florida

Note, the reason why we examined triplets of states was because we intended the main manipulation to involve three events and their negations. The three basic events were Biden to win in state  $A$ ,  $B$  or  $C$ , with the negation of these three events corresponding to Trump winning in these states. Three events would allow better tests of Bayesian constraints,



for example, all two-way conjunctions in Bayesian theory are constrained by three-way conjunctions.

Participants were asked to judge the likelihood of possible combinations of events just for conjunctions, thus there were 12 conjunctions for each triplet, as well as 20 marginals, one for each state and candidate. Marginals and conjunctions were answered in two separate blocks and all questions were randomized within blocks. Examples for the questions used can be found in Figure S.1.1. All ratings were given by adjusting a slider between 0% and 100% in 1% increments, allowing for a total of 101 possible ratings along the scale. In the question text presented to participants, the candidate’s name was highlighted in the color usually associated with the republican (red) and democratic (blue) parties. This choice was made to enhance readability and allow participants to perceive the relevant information in the questions correctly, since the repetitive nature of the questions might cause attention to decline. Analogously, the ‘and’ in conjunctions and the ‘or’ in disjunctions were presented in boldface, so as to make them more noticeable to participants. All participants also answered three more questions corresponding to the Cognitive Reflection Test (CRT; Frederick, 2005), which were used to assess cognitive style. The version of the CRT used in this experiment included the following questions: (a) A bat and a ball cost \$1.10 in total. The bat costs \$1.00 more than the ball. How much does the ball cost? (b) If it takes five minutes for five machines to make five widgets, how long would it take for 100 machines to make 100 widgets? (c) In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

Pilot Experiment 2 was nearly identical to Pilot Experiment 1 and aimed to test an

additional triplet, triplet 5, which was:

- $T_5$ : Ohio, Wisconsin, Florida

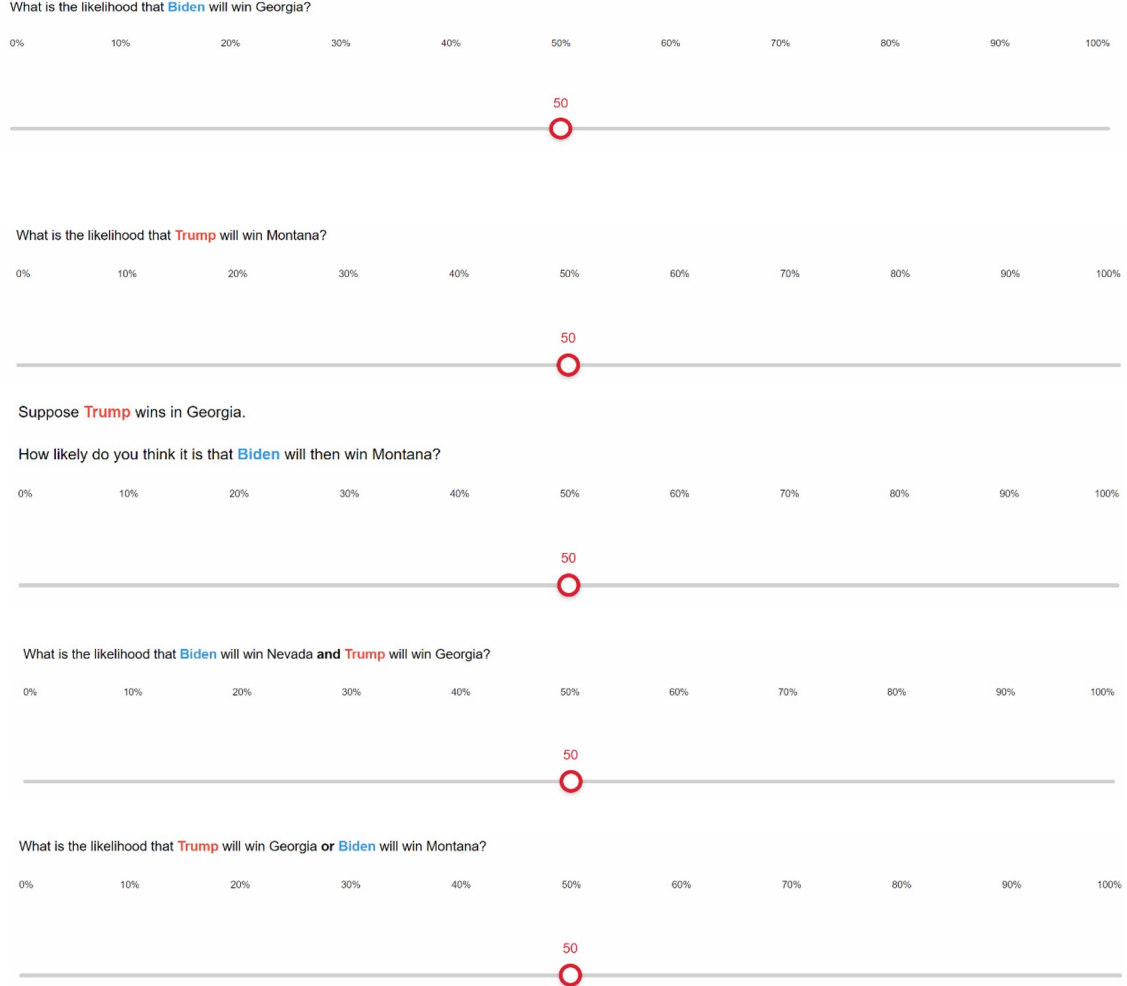


Figure S.1.1. Example questions for marginal, conditional, conjunction and disjunction probabilities.

## Result

The objective of the analyses was to establish which triplets of states were more likely to lead to conjunction fallacies. Straightforwardly, a conjunction fallacy is detected, if the conjunction of two events is rated more likely than one of the marginals. This means

that a conjunction fallacy is found if at least one marginal satisfies this condition. Clearly, since in these pilots we are not concerned with double conjunction fallacies, it suffices to compare the rating of a conjunction against the marginal with the lower rating. Accordingly, one-sided, paired Bayesian t-tests were conducted on all possible pairs of conjunctions and their corresponding lowest marginals. These t-tests are not correct inferential tests for the presence or absence of conjunction fallacies, because using the ‘lower than either marginal’ function can lead to a biased estimate of conjunction fallacy rates, if probabilities are noisy. However, here, we are only interested in whether there appear to be more conjunction fallacies for some candidate triplets vs. others, so the t-tests are useful (in main text, we rely directly on model fits).

We first consider the results of Pilot Experiment 1. Some conjunction fallacies were detected for all four triplets, as seen in Figure S.1.3. The results are mixed. For all triplets, we observed few significant results (i.e., conjunction fallacies), for questions that only consider one candidate at a time (these questions appear in the first block, for each triplet, in Figure S.1.3). A striking exception concerns the highly significant conjunction fallacy for the question “Trump wins Ohio & Missouri” (triplet 1,  $BF_{10} = 460$ ). By contrast, when both candidates were mentioned in a conjunction (see the second block of questions, for each triplet, in Figure S.1.3), there was a much higher rate of conjunction fallacies. The greatest number of conjunction fallacies was observed for triplet 4. However, for that triplet, the evidence for conjunction fallacies in cases when only one candidate was considered at a time was very weak (Bayes factors ranging between 1.50 and 6.02). Triplets 2 and 3 fall in between the extremes of triplet 1 and triplet 4, with three significant conjunction fallacies respectively and Bayes factors ranging between 2.66 and 135 for triplet 2 and between 1.49

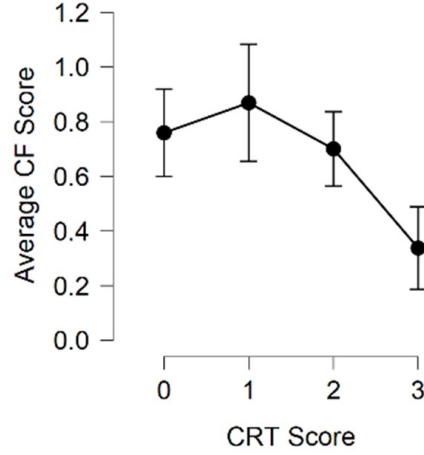
and 8.04 for triplet 3. Note that all the figures here also contain the results for triplet 5, for ease of comparison, which will be discussed in the next section.

It is possible that the preponderance of conjunction fallacies is impacted upon by style of thinking, more reflective vs. more reflective, and previous research has offered some corresponding evidence (Yearsley & Trueblood, 2018). To examine this possibility, we computed the CRT score from each participant, which was a number between 0 and 3, depending on the questions answered correctly by each participant. In our sample, 21 participants scored 0 points, 16 participants scored 1 point, 24 scored 2 points and the remaining 13 scored 3 points. We also computed an approximate measure of each participant’s tendency to produce conjunction fallacies, using the equation (as in Yearsley & Trueblood, 2018), noting again that this equation is problematic, if probability estimates are noisy (it is not used in the main part of our investigation; it is only used here, in the context of these pilots):

$$CF_{measure} = \frac{1}{\sigma} \sum_{i,j} \max(P(\text{state } i \cap \text{state } j) - \min(P(\text{state } i), P(\text{state } j)), 0) \quad (1)$$

This equation offers a heuristic measure of the tendency for a participant to produce conjunction fallacies, since it increases with every instance of a conjunction fallacy, while it stays constant (0 is contributed to the sum), when there is no conjunction fallacy. The measure is normalized by  $\sigma$ , the standard deviation of the probability judgments of each participant, to take into account participant variability in the utilization of the ratings scale.

A Bayesian regression analysis was conducted between the  $CF_{measure}$  and the CRT, yielding positive evidence for a correlation between the two variables ( $r = -.35, BF_{10} = 14.1$ ), such that higher CRT scores were associated with larger values in the  $CF_{measure}$ . Post-hoc Bayesian t-tests were conducted to unveil the exact relationship between conjunction



*Figure S.1.2.* This figure shows the  $CF_{measure}$  against different levels of the CRT score. It can be seen that participants with a CRT score of 3 were associated with a weaker  $CF_{measure}$ , compared to participants in other groups.

fallacies and CRT scores. We observed strong evidence that participants with a CRT score of 3 showed weaker conjunction fallacies compared to any other group ( $BF_{10} = 17.8$ ,  $BF_{10} = 34.7$  and  $BF_{10} = 10.9$ , respectively), see Figure S.1.2. This finding is consistent with expectation from Yearsley and Trueblood (2018), since a higher CRT score would, generally, be indicative of more reflective thinking.

Given these results, it is possible that a cleaner picture regarding the preponderance of conjunction fallacies for different triplets might be obtained if we excluded participants with a CRT score of 3, since these participants are more likely to avoid apparent classical fallacies. Repeating the above analysis with this exclusion restricted the sample to 61 participants. As shown in Figure S.1.4 (new results are indicated by shading), for all triplets there was an additional significant conjunction fallacy, with weak to strong evidence (Bayes factors ranging between 1.90 to 9.03). For triplet 3, we observed two further conjunction fallacies, but one of the previously noted ones dropped below the significance threshold.

The strength of different conjunction fallacies, which can be heuristically approximated

using the  $CF_{measure}$ , is of particular interest for testing formal probabilistic models (both Bayesian and quantum, since, even though conjunction fallacies are allowed in the latter case, they are bounded and so particularly strong conjunction fallacies might be outside quantum models, Yearsley & Trueblood, 2018).

The results from Pilot Experiment 1 offered a somewhat unclear picture regarding the triplets of states more likely to produce strong conjunction fallacies. Triplets 1 and 3 produced the highest (average)  $CF_{measure}$ , while triplets 2 and 4 offered the strongest results, in terms of the significance of Bayesian tests comparing probabilities for conjunctions against marginals.

For Pilot Experiment 2, as can be seen in Figures S.1.3, S.1.4, triplet 5 was associated with fewer significant instances of conjunction fallacies and weaker  $CF_{measure}$ . In this case, excluding participants with  $CRT = 3$  decreased the number of significant conjunction fallacies, contrary to what we observed for triplets 1-4.

Taking together all results, a few options present themselves for reasonable expectation regarding the emergence of results problematic for a classical perspective. Bearing in mind that the sampling for the main experiment would be much larger than for the pilot studies, (somewhat arbitrarily) we placed more faith on the higher  $CF_{measure}$  values observed for triplets 1 and 3 and chose these for the next steps:

- $T1$ : Ohio, Missouri, Michigan
- $T3$ : Georgia, Montana, Nevada

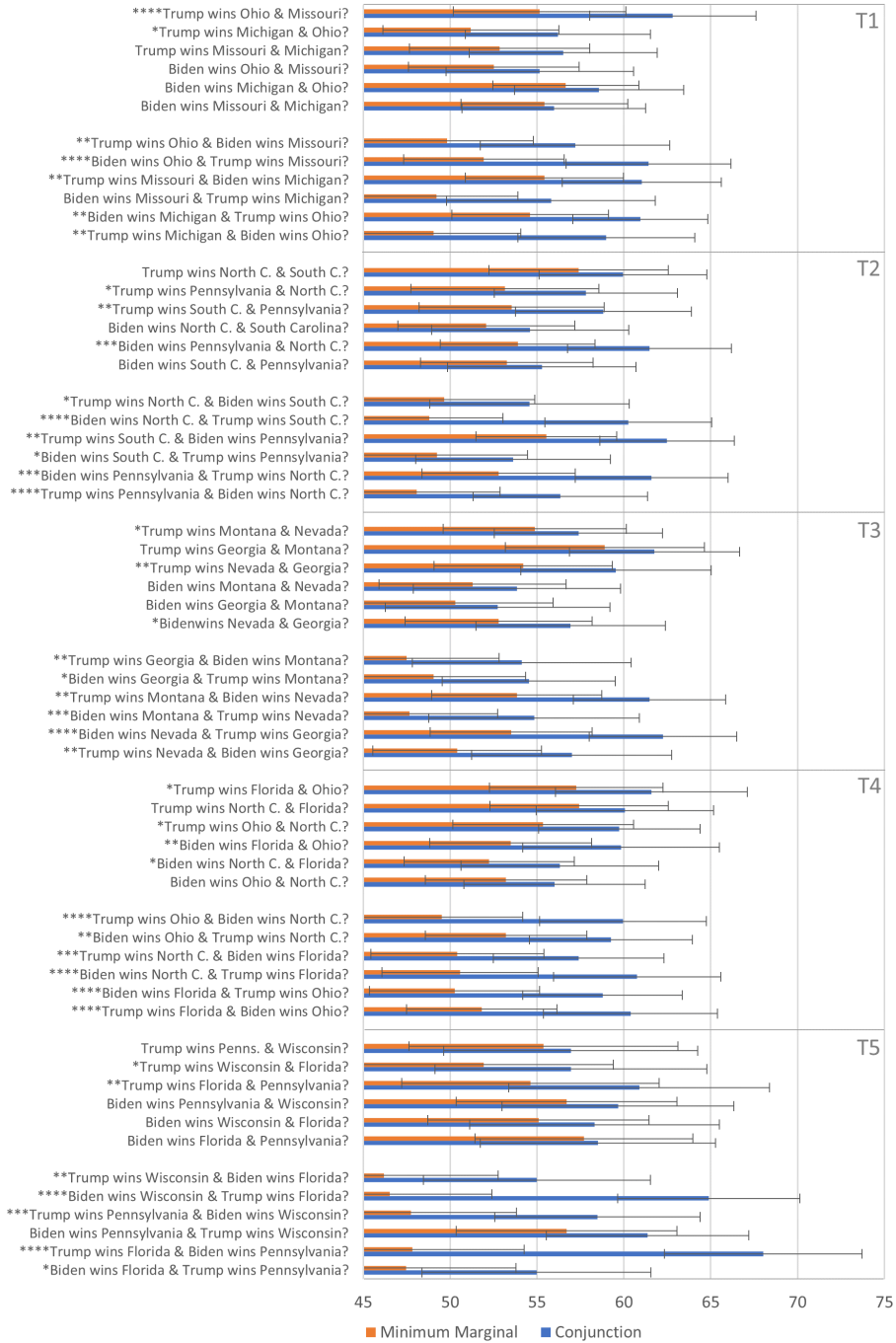


Figure S.1.3. Average conjunction probabilities (blue lines) compared to the lower corresponding marginal (red lines) for the four triplets in Pilot Experiment 1 (triplets 1-4) and Pilot Experiment 2 (triplet 5). We indicate with stars the  $BF_{10}$  evidence strength (\* = weak evidence, \*\* = positive evidence, \*\*\* = strong evidence, \*\*\*\* = very strong evidence).

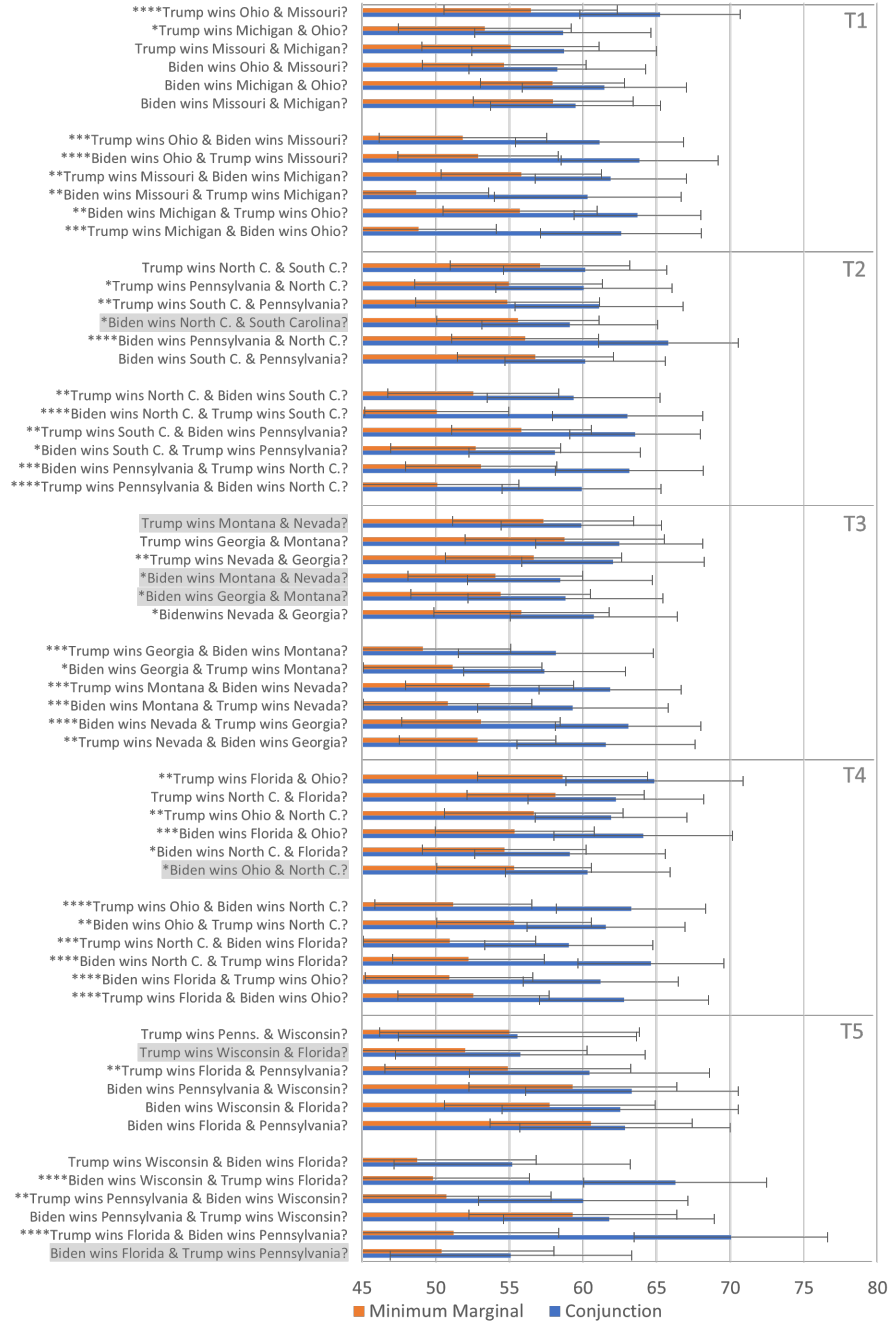


Figure S.1.4. Average conjunction probabilities (blue lines) compared to the lower corresponding marginal (red lines) for the four triplets in Pilot Experiment 1 (triplets 1-4) and Pilot Experiment 2 (triplet 5), when excluding participants with CRT = 3. We indicate with stars the BF10 evidence strength (\* = weak evidence, \*\* = positive evidence, \*\*\* = strong evidence, \*\*\*\* = very strong evidence). Shading indicates differences relative to Figure S.1.3.



## Supplementary Material 2

In this section, we offer some additional notes, figures, and tables, related to the standard statistical analyses of the behavioral results in main text. Note that in some cases it is less (or not at all) relevant to separate out results by triplet and so, where sensible, for brevity, we present combined results.

|   |   | Prior Odds | Posterior Odds | $BF_{10,U}$         | error%                  |
|---|---|------------|----------------|---------------------|-------------------------|
| 0 | 1 | 0.414      | 0.527          | 1.273               | $2.583 \times 10^{-5}$  |
| 1 | 2 | 0.414      | 652643         | $1.576 \times 10^6$ | $8.045 \times 10^{-13}$ |
| 2 | 3 | 0.414      | 1787           | 4314                | $1.384 \times 10^{-8}$  |

Table S2.1: Linear repeated contrasts and Bayesian post-hoc tests for the effect of CRT scores on the frequency of conjunction fallacies. Note: the posterior odds have been corrected for multiple testing by fixing to 0.5 the prior probability that the null hypothesis holds across all comparisons (Westfall, Johnson, & Utts, 1997). Individual comparisons are based on the default t-test with a Cauchy ( $0, r = \frac{1}{\sqrt{2}}$ ) prior. The “U” in the Bayes factor denotes that it is uncorrected.

|   |   | Prior Odds | Posterior Odds      | $BF_{10,U}$         | error%                  |
|---|---|------------|---------------------|---------------------|-------------------------|
| 0 | 1 | 0.414      | 0.043               | 0.104               | $3.68 \times 10^{-4}$   |
| 1 | 2 | 0.414      | 0.261               | 0.631               | $9.123 \times 10^{-6}$  |
| 2 | 3 | 0.414      | $2.391 \times 10^6$ | $5.772 \times 10^6$ | $8.616 \times 10^{-12}$ |

Table S2.2: Linear repeated contrasts and Bayesian post-hoc tests for the effect of CRT scores on the frequency of disjunction fallacies. Note: the posterior odds have been corrected for multiple testing by fixing to 0.5 the prior probability that the null hypothesis holds across all comparisons (Westfall, Johnson, & Utts, 1997). Individual comparisons are based on the default t-test with a Cauchy ( $0, r = \frac{1}{\sqrt{2}}$ ) prior. The “U” in the Bayes factor denotes that it is uncorrected.

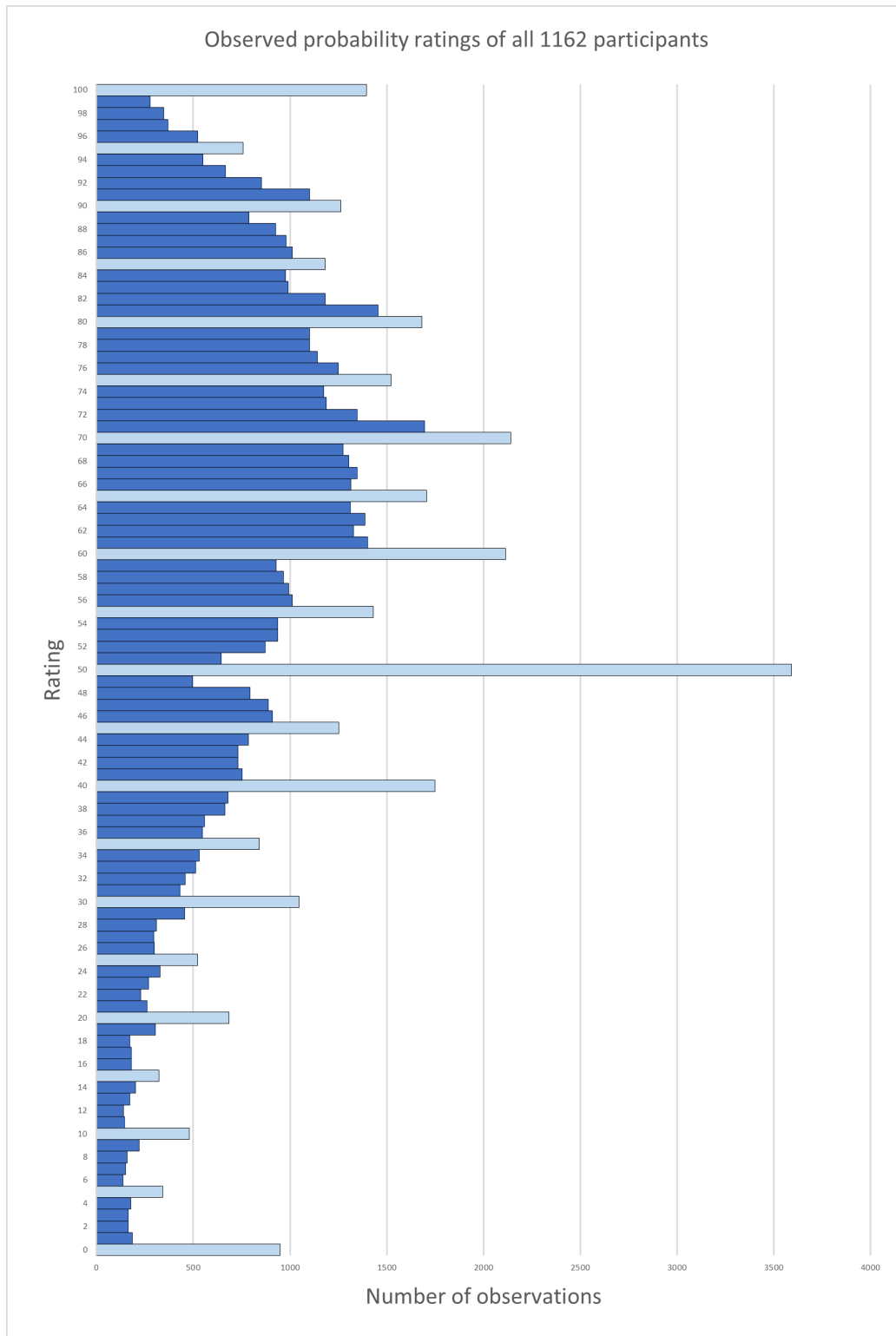


Figure S.2.1. A histogram of all observed probability ratings in the experiment. Ratings which are a multiple of 5 are highlighted.

| Order | Effect              | $BF_{10}$ | error%                 |
|-------|---------------------|-----------|------------------------|
| COE   | $A1 \ A2$           | 0.149     | $9.857 \times 10^{-6}$ |
| COE   | $A1 \ \neg A2$      | 0.104     | $1.403 \times 10^{-5}$ |
| COE   | $\neg A1 \ A2$      | 0.047     | $3.104 \times 10^{-5}$ |
| COE   | $\neg A1 \ \neg A2$ | 0.047     | $3.146 \times 10^{-5}$ |
| COE   | $A2 \ A3$           | 0.05      | $2.957 \times 10^{-5}$ |
| COE   | $A2 \ \neg A3$      | 0.047     | $3.096 \times 10^{-5}$ |
| COE   | $\neg A2 \ A3$      | 0.158     | $9.313 \times 10^{-6}$ |
| COE   | $\neg A2 \ \neg A3$ | 0.071     | $2.056 \times 10^{-5}$ |
| COE   | $A3 \ A1$           | 0.139     | $1.057 \times 10^{-5}$ |
| COE   | $A3 \ \neg A1$      | 0.103     | $1.417 \times 10^{-5}$ |
| COE   | $\neg A3 \ A1$      | 0.140     | $1.049 \times 10^{-5}$ |
| COE   | $\neg A3 \ \neg A1$ | 0.057     | $2.567 \times 10^{-5}$ |
| DOE   | $A1 \ A2$           | 0.211     | $6.929 \times 10^{-6}$ |
| DOE   | $A1 \ \neg A2$      | 0.120     | $1.220 \times 10^{-5}$ |
| DOE   | $\neg A1 \ A2$      | 0.047     | $3.112 \times 10^{-5}$ |
| DOE   | $\neg A1 \ \neg A2$ | 0.105     | $1.389 \times 10^{-5}$ |
| DOE   | $A2 \ A3$           | 0.161     | $9.084 \times 10^{-6}$ |
| DOE   | $A2 \ \neg A3$      | 0.069     | $2.125 \times 10^{-5}$ |
| DOE   | $\neg A2 \ A3$      | 0.047     | $3.127 \times 10^{-5}$ |
| DOE   | $\neg A2 \ \neg A3$ | 0.050     | $2.908 \times 10^{-5}$ |
| DOE   | $A3 \ A1$           | 0.100     | $1.470 \times 10^{-5}$ |
| DOE   | $A3 \ \neg A1$      | 0.059     | $2.499 \times 10^{-5}$ |
| DOE   | $\neg A3 \ A1$      | 0.057     | $2.590 \times 10^{-5}$ |
| DOE   | $\neg A3 \ \neg A1$ | 0.047     | $3.148 \times 10^{-5}$ |

Table S2.3: Bayesian one sample t-tests of all conjunction order effects (COE) and disjunction order effects (DOE) for Triplet 1, suggesting strong evidence against order effects for all questions pairs.

Note: for all tests, the alternative hypothesis specifies that the population mean differs from 0.

| Order Effect            | $BF_{10}$                             | error %                                   |
|-------------------------|---------------------------------------|-------------------------------------------|
| COE $A1 \ A2$           | 0.233                                 | $6.169 \times 10^{-6}$                    |
| COE $A1 \ \neg A2$      | 0.059                                 | $2.465 \times 10^{-5}$                    |
| COE $\neg A1 \ A2$      | 0.120                                 | $1.205 \times 10^{-5}$                    |
| COE $\neg A1 \ \neg A2$ | 0.047                                 | $3.080 \times 10^{-5}$                    |
| COE $A2 \ A3$           | 0.115                                 | $1.250 \times 10^{-5}$                    |
| COE $A2 \ \neg A3$      | 0.051                                 | $2.849 \times 10^{-5}$                    |
| COE $\neg A2 \ A3$      | 0.051                                 | $2.946 \times 10^{-6}$                    |
| COE $\neg A2 \ \neg A3$ | 0.047                                 | $3.077 \times 10^{-5}$                    |
| COE $A3 \ A1$           | 0.054                                 | $2.688 \times 10^{-5}$                    |
| COE $A3 \ \neg A1$      | 0.051                                 | $2.823 \times 10^{-5}$                    |
| COE $\neg A3 \ A1$      | 0.114                                 | $1.267 \times 10^{-5}$                    |
| COE $\neg A3 \ \neg A1$ | 0.063                                 | $2.304 \times 10^{-5}$                    |
| DOE $A1 \ A2$           | 0.065                                 | $2.234 \times 10^{-5}$                    |
| DOE $A1 \ \neg A2$      | 0.052                                 | $2.794 \times 10^{-5}$                    |
| DOE $\neg A1 \ A2$      | 0.060                                 | $2.423 \times 10^{-5}$                    |
| DOE $\neg A1 \ \neg A2$ | 0.051                                 | $2.830 \times 10^{-5}$                    |
| DOE $A2 \ A3$           | 0.143                                 | $1.010 \times 10^{-5}$                    |
| DOE $A2 \ \neg A3$      | 0.052                                 | $2.782 \times 10^{-5}$                    |
| DOE $\neg A2 \ A3$      | 0.047                                 | $3.076 \times 10^{-5}$                    |
| DOE $\neg A2 \ \neg A3$ | 0.052                                 | $2.785 \times 10^{-5}$                    |
| DOE $A3 \ A1$           | 0.050                                 | $2.910 \times 10^{-5}$                    |
| DOE $A3 \ \neg A1$      | <b><math>3.049 \times 10^7</math></b> | <b><math>4.366 \times 10^{-14}</math></b> |
| DOE $\neg A3 \ A1$      | 0.097                                 | $1.486 \times 10^{-5}$                    |
| DOE $\neg A3 \ \neg A1$ | 0.051                                 | $2.850 \times 10^{-5}$                    |

Table S2.4: Bayesian one sample t-tests of all conjunction order effects (COE) and disjunction order effects (DOE) for Triplet 2, suggesting strong evidence against order effects for all but one (printed in boldface) question pair. Note: for all tests, the alternative hypothesis specifies that the population mean differs from 0.

| Measure 1            | Measure 2            | $BF_{01}$ triplet 1     | error%                  | $BF_{01}$ triplet 2     | error%                  |
|----------------------|----------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| $P(A1 A2)$           | $P(A2 A1)$           | 0.016                   | $2.385 \times 10^{-8}$  | 0.010                   | $1.436 \times 10^{-8}$  |
| $P(A1 \neg A2)$      | $P(\neg A2 A1)$      | <b>10.431</b>           | $1.528 \times 10^{-5}$  | $1.085 \times 10^{-5}$  | $1.436 \times 10^{-14}$ |
| $P(\neg A1 A2)$      | $P(A2 \neg A1)$      | <b>15.018</b>           | $2.202 \times 10^{-5}$  | $1.851 \times 10^{-9}$  | $2.426 \times 10^{-15}$ |
| $P(\neg A1 \neg A2)$ | $P(\neg A2 \neg A1)$ | $4.028 \times 10^{-6}$  | $5.671 \times 10^{-12}$ | $1.434 \times 10^{-4}$  | $1.989 \times 10^{-10}$ |
| $P(A2 A3)$           | $P(A3 A2)$           | $1.487 \times 10^{-12}$ | $1.952 \times 10^{-18}$ | $9.055 \times 10^{-18}$ | $1.033 \times 10^{-23}$ |
| $P(A2 \neg A3)$      | $P(\neg A3 A2)$      | <b>3.658</b>            | $5.349 \times 10^{-6}$  | 0.871                   | $1.248 \times 10^{-6}$  |
| $P(\neg A2 A3)$      | $P(A3 \neg A2)$      | <b>1.884</b>            | $2.757 \times 10^{-6}$  | $6.104 \times 10^{-4}$  | $8.517 \times 10^{-10}$ |
| $P(\neg A2 \neg A3)$ | $P(\neg A3 \neg A2)$ | $2.888 \times 10^{-13}$ | $3.755 \times 10^{-19}$ | $1.658 \times 10^{-21}$ | $1.749 \times 10^{-27}$ |
| $P(A3 A1)$           | $P(A1 A3)$           | $4.763 \times 10^{-5}$  | $6.767 \times 10^{-11}$ | $1.477 \times 10^{-9}$  | $1.932 \times 10^{-15}$ |
| $P(A3 \neg A1)$      | $P(\neg A1 A3)$      | 0.054                   | $7.859 \times 10^{-8}$  | <b>12.678</b>           | $1.829 \times 10^{-5}$  |
| $P(\neg A3 A1)$      | $P(A1 \neg A3)$      | 0.003                   | $3.677 \times 10^{-9}$  | <b>7.378</b>            | $1.063 \times 10^{-5}$  |
| $P(\neg A3 \neg A1)$ | $P(\neg A1 \neg A3)$ | $2.581 \times 10^{-6}$  | $3.623 \times 10^{-12}$ | $6.349 \times 10^{-10}$ | $8.262 \times 10^{-16}$ |

Table S2.5: Bayesian paired sample t-tests comparing conditional probabilities to their reciprocal counterpart, for both triplets together. Results confirming the assumption of reciprocity are printed in boldface.

|                               | $BF_{01}$     | error%       |                                  | $BF_{01}$     | error%       |
|-------------------------------|---------------|--------------|----------------------------------|---------------|--------------|
| <b><math>Z_1A1, A2</math></b> | <b>1.199</b>  | <b>0.000</b> | <b><math>Z_9A1, A2</math></b>    | <b>27.306</b> | <b>0.011</b> |
| <b><math>Z_1A2, A1</math></b> | <b>13.706</b> | <b>0.006</b> | <b><math>Z_9A2, A3</math></b>    | <b>9.120</b>  | <b>0.004</b> |
| <b><math>Z_1A2, A3</math></b> | <b>11.802</b> | <b>0.005</b> | <b><math>Z_9A3, A1</math></b>    | <b>15.034</b> | <b>0.006</b> |
| <b><math>Z_1A3, A2</math></b> | <b>24.521</b> | <b>0.010</b> | $Z_{10}A1, A2$                   | 0.000         | 0.000        |
| <b><math>Z_1A3, A1</math></b> | <b>1.271</b>  | <b>0.000</b> | $Z_{10}A2, A3$                   | 0.000         | 0.000        |
| <b><math>Z_1A1, A3</math></b> | <b>1.637</b>  | <b>0.000</b> | $Z_{10}A3, A1$                   | 0.000         | 0.000        |
| <b><math>Z_2A1, A2</math></b> | <b>26.892</b> | <b>0.011</b> | $Z_{11}A1, A2$                   | 0.000         | 0.000        |
| <b><math>Z_2A2, A1</math></b> | <b>19.309</b> | <b>0.008</b> | $Z_{11}A2, A3$                   | 0.000         | 0.000        |
| <b><math>Z_2A2, A3</math></b> | <b>30.067</b> | <b>0.012</b> | $Z_{11}A3, A1$                   | 0.000         | 0.000        |
| <b><math>Z_2A3, A2</math></b> | <b>15.176</b> | <b>0.006</b> | $Z_{12}A1, A2$                   | 0.000         | 0.000        |
| <b><math>Z_2A3, A1</math></b> | <b>21.341</b> | <b>0.009</b> | $Z_{12}A2, A3$                   | 0.000         | 0.000        |
| <b><math>Z_2A1, A3</math></b> | <b>29.938</b> | <b>0.012</b> | $Z_{12}A3, A1$                   | 0.000         | 0.000        |
| $Z_3A1, A2$                   | 0.000         | 0.000        | $Z_{13}A1, A2$                   | 0.000         | 0.000        |
| $Z_3A2, A1$                   | 0.000         | 0.000        | $Z_{13}A2, A3$                   | 0.000         | 0.000        |
| $Z_3A2, A3$                   | 0.000         | 0.000        | $Z_{13}A3, A1$                   | 0.000         | 0.000        |
| $Z_3A3, A2$                   | 0.000         | 0.000        | <b><math>Z_{14}A1A2</math></b>   | <b>27.045</b> | <b>0.011</b> |
| $Z_3A3, A1$                   | 0.000         | 0.000        | <b><math>Z_{14}A2, A3</math></b> | <b>30.189</b> | <b>0.012</b> |
| $Z_3A1, A3$                   | 0.000         | 0.000        | <b><math>Z_{14}A3, A1</math></b> | <b>25.437</b> | <b>0.010</b> |
| $Z_4A1, A2$                   | 0.000         | 0.000        | $Z_{15}A1, A2$                   | 0.000         | 0.000        |
| $Z_4A2, A1$                   | 0.000         | 0.000        | $Z_{15}A2, A1$                   | 0.000         | 0.000        |
| $Z_4A2, A3$                   | 0.000         | 0.000        | $Z_{15}A2, A3$                   | 0.000         | 0.000        |
| $Z_4A3, A2$                   | 0.000         | 0.000        | $Z_{15}A3, A2$                   | 0.000         | 0.000        |
| $Z_4A3, A1$                   | 0.000         | 0.000        | $Z_{15}A3, A1$                   | 0.000         | 0.000        |
| $Z_4A1, A3$                   | 0.000         | 0.000        | $Z_{15}A1, A3$                   | 0.000         | 0.000        |

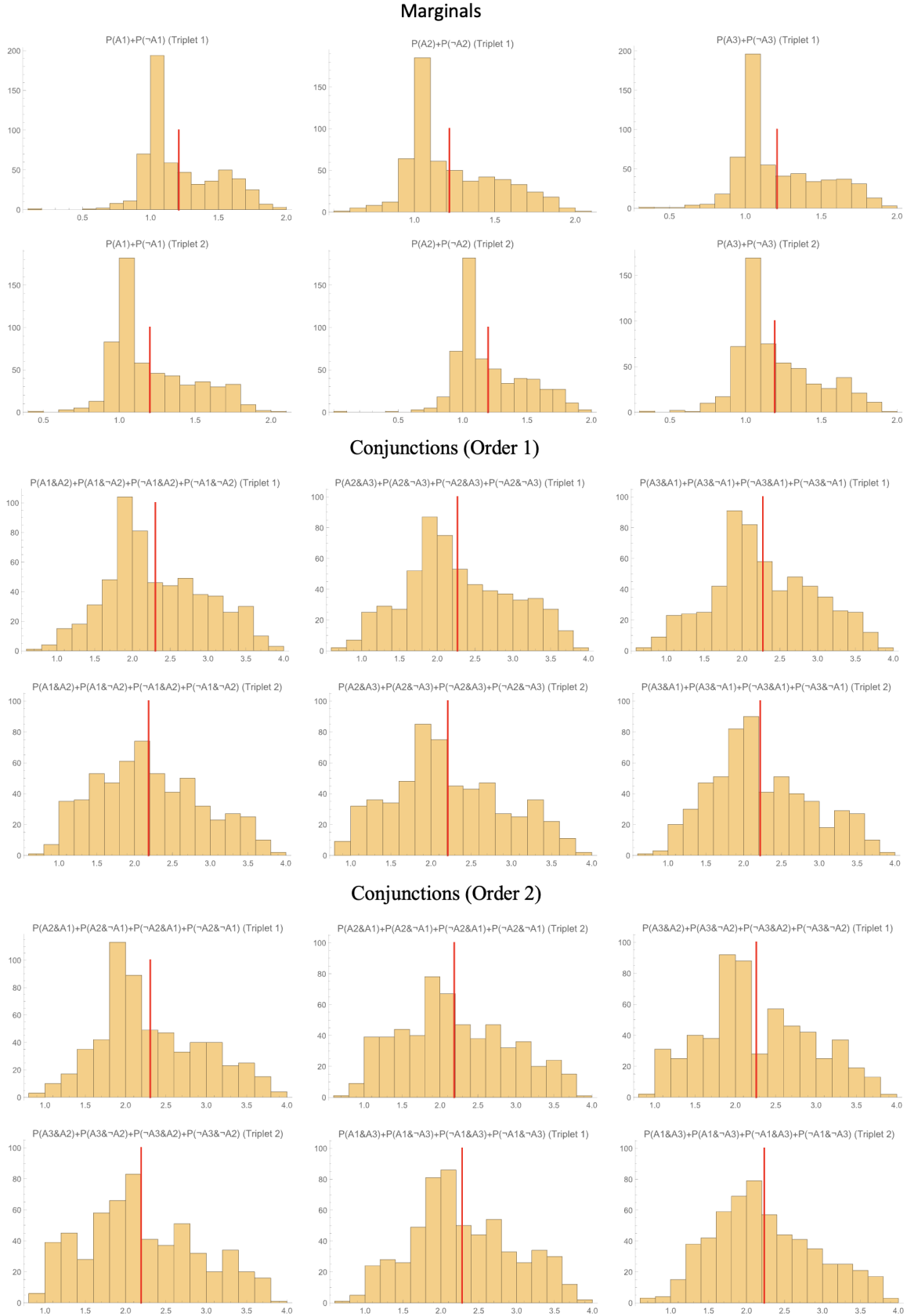
|             |       |       |                |       |       |
|-------------|-------|-------|----------------|-------|-------|
| $Z_6A1, A2$ | 0.000 | 0.000 | $Z_{17}A1, A2$ | 0.000 | 0.000 |
| $Z_6A2, A1$ | 0.000 | 0.000 | $Z_{17}A2, A1$ | 0.000 | 0.000 |
| $Z_6A2, A3$ | 0.000 | 0.000 | $Z_{17}A2, A3$ | 0.000 | 0.000 |
| $Z_6A3, A2$ | 0.000 | 0.000 | $Z_{17}A3, A2$ | 0.000 | 0.000 |
| $Z_6A3, A1$ | 0.000 | 0.000 | $Z_{17}A3, A1$ | 0.000 | 0.000 |
| $Z_6A1, A3$ | 0.000 | 0.000 | $Z_{17}A1, A3$ | 0.000 | 0.000 |
| $Z_7A1, A2$ | 0.000 | 0.000 | $Z_{18}A1, A2$ | 0.000 | 0.000 |
| $Z_7A2, A1$ | 0.000 | 0.000 | $Z_{18}A2, A1$ | 0.000 | 0.000 |
| $Z_7A2, A3$ | 0.000 | 0.000 | $Z_{18}A2, A3$ | 0.000 | 0.000 |
| $Z_7A3, A2$ | 0.000 | 0.000 | $Z_{18}A3, A2$ | 0.000 | 0.000 |
| $Z_7A3, A1$ | 0.000 | 0.000 | $Z_{18}A3, A1$ | 0.000 | 0.000 |
| $Z_7A1, A3$ | 0.000 | 0.000 | $Z_{18}A1, A3$ | 0.000 | 0.000 |
| $Z_8A1, A2$ | 0.000 | 0.000 |                |       |       |
| $Z_8A2, A1$ | 0.000 | 0.000 |                |       |       |
| $Z_8A2, A3$ | 0.000 | 0.000 |                |       |       |
| $Z_8A3, A2$ | 0.000 | 0.000 |                |       |       |
| $Z_8A3, A1$ | 0.000 | 0.000 |                |       |       |
| $Z_8A1, A3$ | 0.000 | 0.000 |                |       |       |

Table S2.6: Bayesian one sample t-tests for all the Z identities, for both triplets together (Table 2 in main text). Results that confirm Bayesian probability theory are printed in boldface. Note, we evaluated the identities separately for each order (for conjunctions, disjunctions). Note: for all tests, the null hypothesis specifies that the population mean equals 0. Figure 4 in main text illustrates the results differently, separating them out by triplet; it can be seen that there is hardly any difference between the two triplets, hence the aggregated presentation in this table.

|                                                                                                  | $BF_{10}$                | error %                  |
|--------------------------------------------------------------------------------------------------|--------------------------|--------------------------|
| $T1 \ P(A1) + P(\neg A1)$                                                                        | $1.216 \times 10^{56}$   | $8.807 \times 10^{-59}$  |
| $T1 \ P(A2) + P(\neg A2)$                                                                        | $1.020 \times 10^{56}$   | $9.258 \times 10^{-60}$  |
| $T1 \ P(A3) + P(\neg A3)$                                                                        | $6.273 \times 10^{53}$   | $5.756 \times 10^{-61}$  |
| $T1 \ P(A1 \wedge A2) + P(A1 \wedge \neg A2) + P(\neg A1 \wedge A2) + P(\neg A1 \wedge \neg A2)$ | $1.157 \times 10^{184}$  | $4.970 \times 10^{-188}$ |
| $T1 \ P(A2 \wedge A3) + P(A2 \wedge \neg A3) + P(\neg A2 \wedge A3) + P(\neg A2 \wedge \neg A3)$ | $1.216 \times 10^{201}$  | $1.546 \times 10^{-250}$ |
| $T1 \ P(A3 \wedge A1) + P(A3 \wedge \neg A1) + P(\neg A3 \wedge A1) + P(\neg A3 \wedge \neg A1)$ | $5.853 \times 10^{190}$  | $1.547 \times 10^{-195}$ |
| $T1 \ P(A2 \wedge A1) + P(A2 \wedge \neg A1) + P(\neg A2 \wedge A1) + P(\neg A2 \wedge \neg A1)$ | $4.127 \times 10^{204}$  | $4.930 \times 10^{-209}$ |
| $T1 \ P(A3 \wedge A2) + P(A3 \wedge \neg A2) + P(\neg A3 \wedge A2) + P(\neg A3 \wedge \neg A2)$ | $1.192 \times 10^{188}$  | $6.566 \times 10^{-193}$ |
| $T1 \ P(A1 \wedge A3) + P(A1 \wedge \neg A3) + P(\neg A1 \wedge A3) + P(\neg A1 \wedge \neg A3)$ | $2.036 \times 10^{189}$  | $4.822 \times 10^{-194}$ |
| $T2 \ P(A1) + P(\neg A1)$                                                                        | $1.186 \times 10^{53}$   | $8.359 \times 10^{-56}$  |
| $T2 \ P(A2) + P(\neg A2)$                                                                        | $3.444 \times 10^{50}$   | $1.032 \times 10^{-57}$  |
| $T2 \ P(A3) + P(\neg A3)$                                                                        | $5.499 \times 10^{50}$   | $6.395 \times 10^{-58}$  |
| $T2 \ P(A1 \wedge A2) + P(A1 \wedge \neg A2) + P(\neg A1 \wedge A2) + P(\neg A1 \wedge \neg A2)$ | $3.766 \times 10^{167}$  | $2.172 \times 10^{-170}$ |
| $T2 \ P(A1 \wedge A2) + P(A1 \wedge \neg A2) + P(\neg A1 \wedge A2) + P(\neg A1 \wedge \neg A2)$ | $3.766 \times 10^{167}$  | $2.172 \times 10^{-170}$ |
| $T2 \ P(A2 \wedge A3) + P(A2 \wedge \neg A3) + P(\neg A2 \wedge A3) + P(\neg A2 \wedge \neg A3)$ | $1.132 \times 10^{171}$  | $2.294 \times 10^{-174}$ |
| $T2 \ P(A3 \wedge A1) + P(A3 \wedge \neg A1) + P(\neg A3 \wedge A1) + P(\neg A3 \wedge \neg A1)$ | $8.408 \times 10^{182}$  | $1.025 \times 10^{-186}$ |
| $T2 \ P(A2 \wedge A1) + P(A2 \wedge \neg A1) + P(\neg A2 \wedge A1) + P(\neg A2 \wedge \neg A1)$ | $3.706 \times 10^{164}$  | $6.860 \times 10^{-168}$ |
| $T2 \ P(A3 \wedge A2) + P(A3 \wedge \neg A2) + P(\neg A3 \wedge A2) + P(\neg A3 \wedge \neg A2)$ | $9.390 \times 10^{165}$  | $4.566 \times 10^{-169}$ |
| $T2 \ P(A1 \wedge A3) + P(A1 \wedge \neg A3) + P(\neg A1 \wedge A3) + P(\neg A1 \wedge \neg A3)$ | $5.264 \times 10^{-178}$ | $9.444 \times 10^{-182}$ |

Table S2.7: One-sample Bayesian t-tests concerning consistency with the binary complementarity on marginals and four-way law of total probability. Note: For all tests, the null hypothesis is that the population mean equals 1. T1 and T2 refer to the two different triplets.

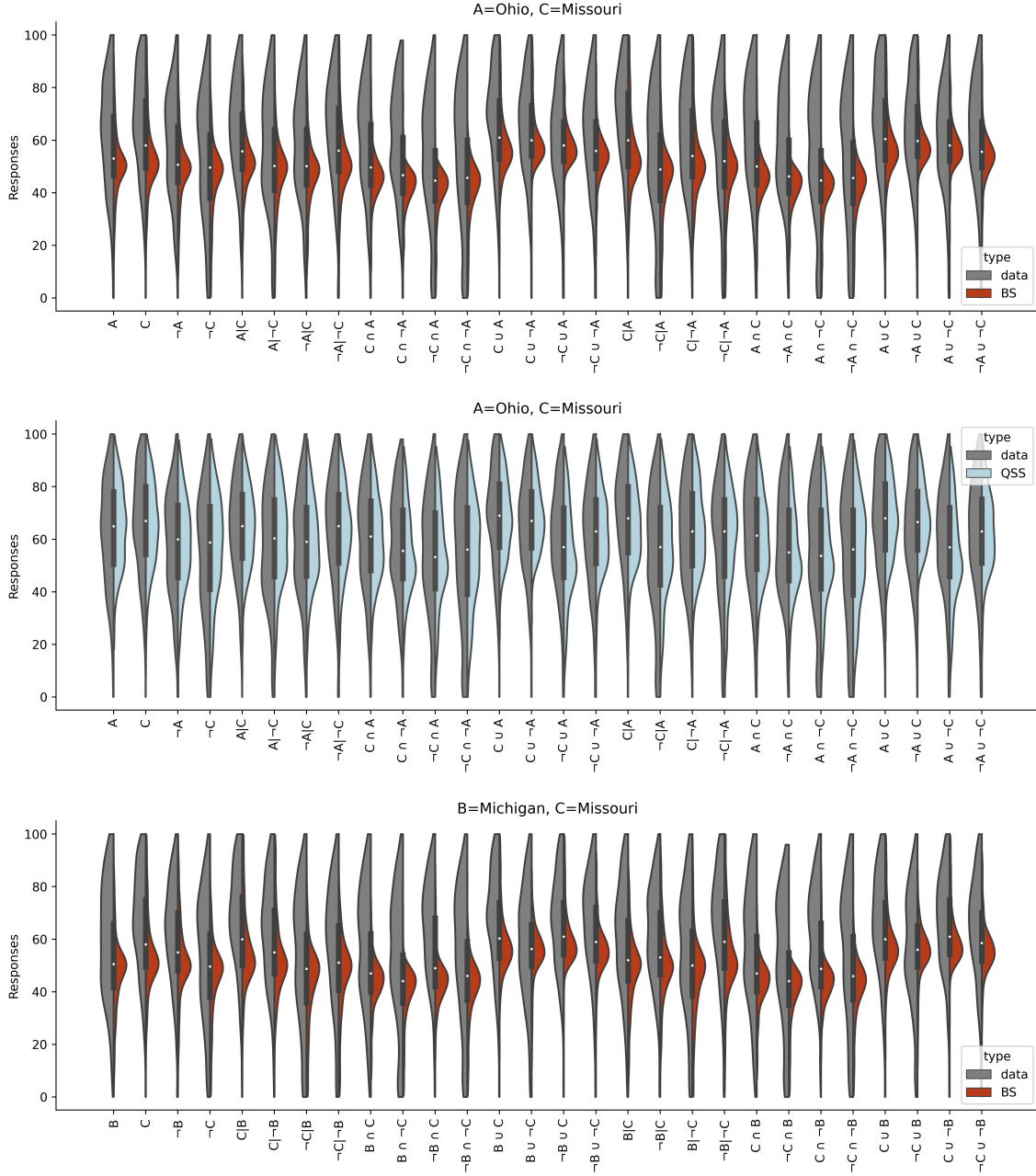


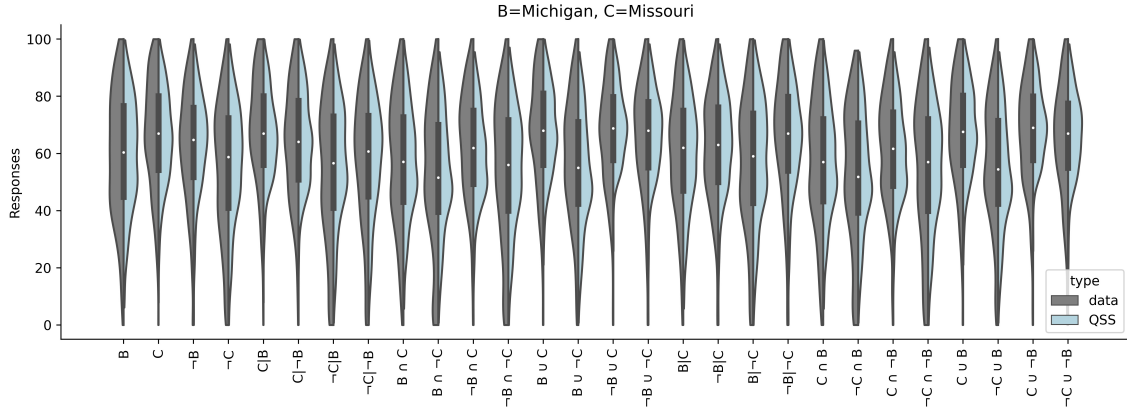


*Figure S.2.2.* Histograms of the quantities corresponding to binary complementarity on marginals and four-way the law of total probability, for different triplets and for different predicate orders for

## Supplementary Material 3

In this section, we offer some additional figures, illustrating the results of the comparison between the Quantum Sequential Sampler and the Bayesian Sampler models.





*Figure S.3.1.* Violin plots showing empirical data vs. the predictions of the Quantum Sequential Sampler, and empirical data (top panel) vs. the predictions of the Bayesian Sampler (bottom panel) for the first triplet. The data are for the  $\{A, C\}$  and  $\{B, C\}$  pairs of events of Triplet 1.

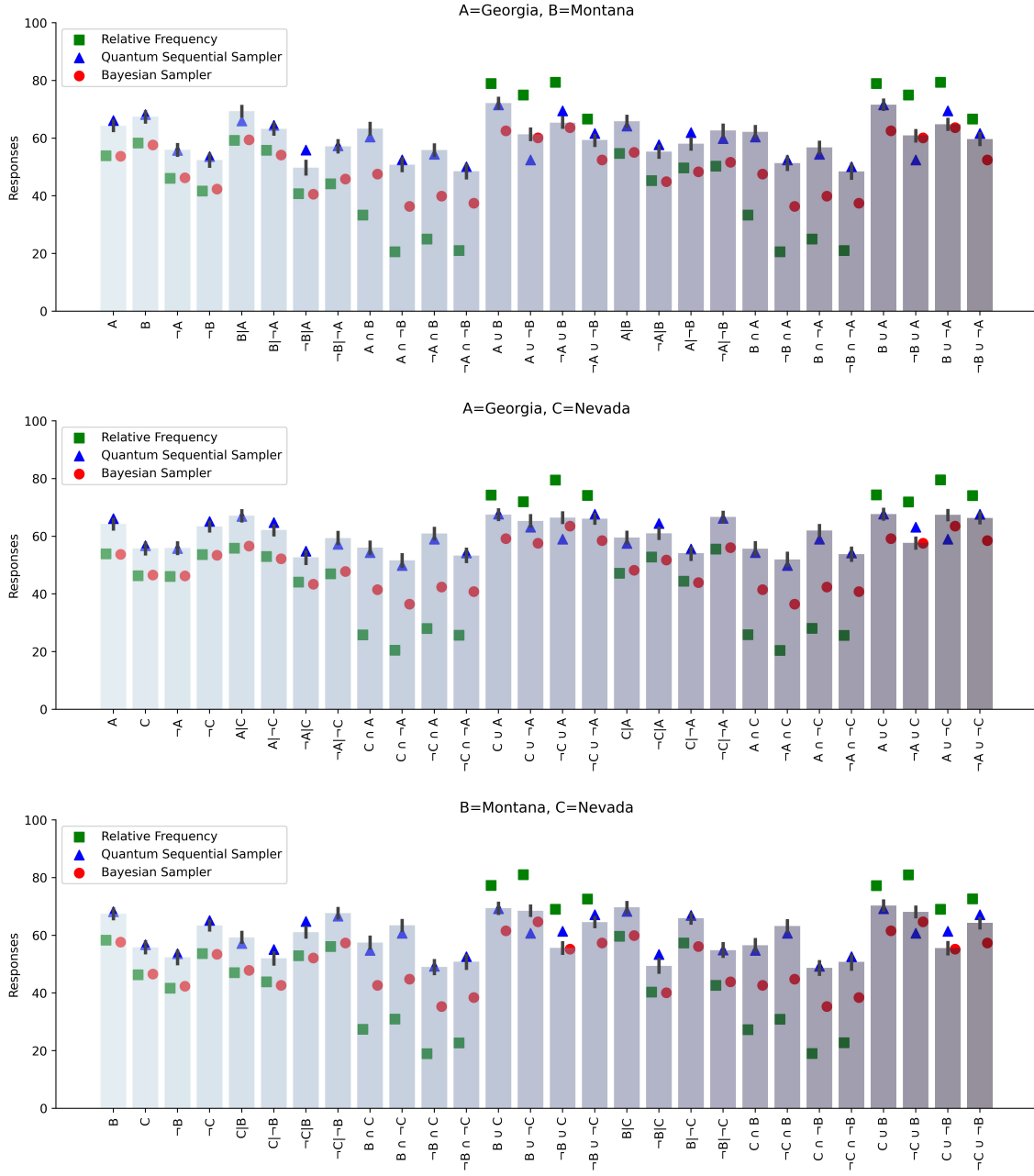
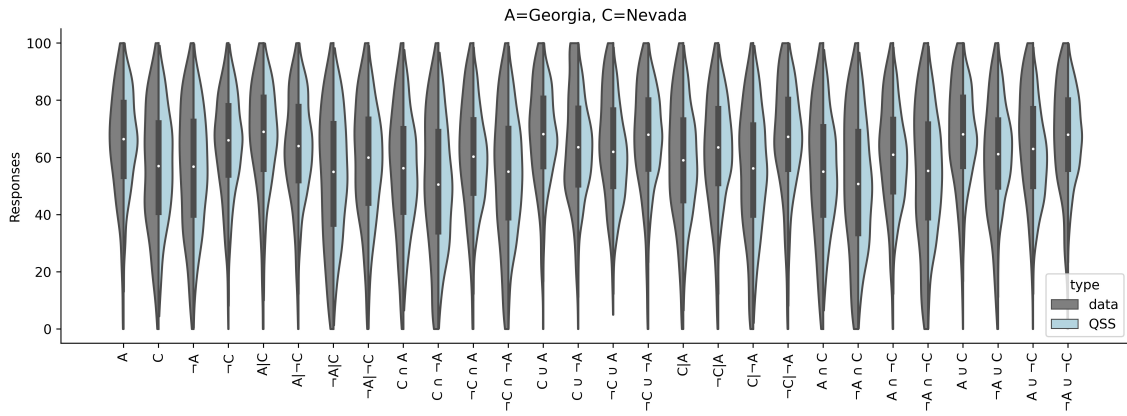
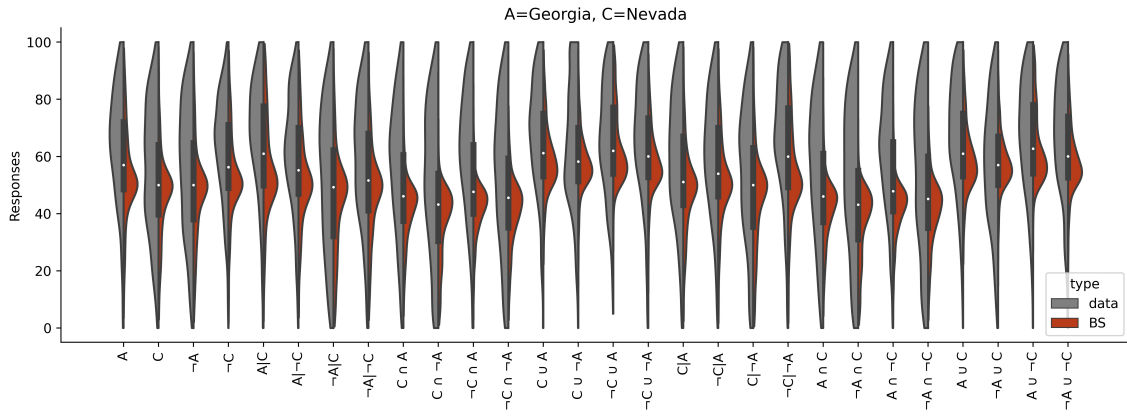
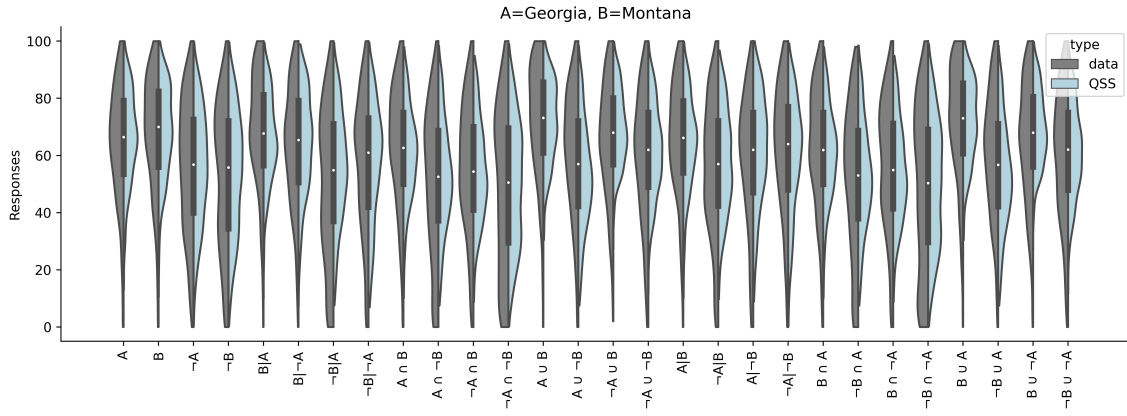
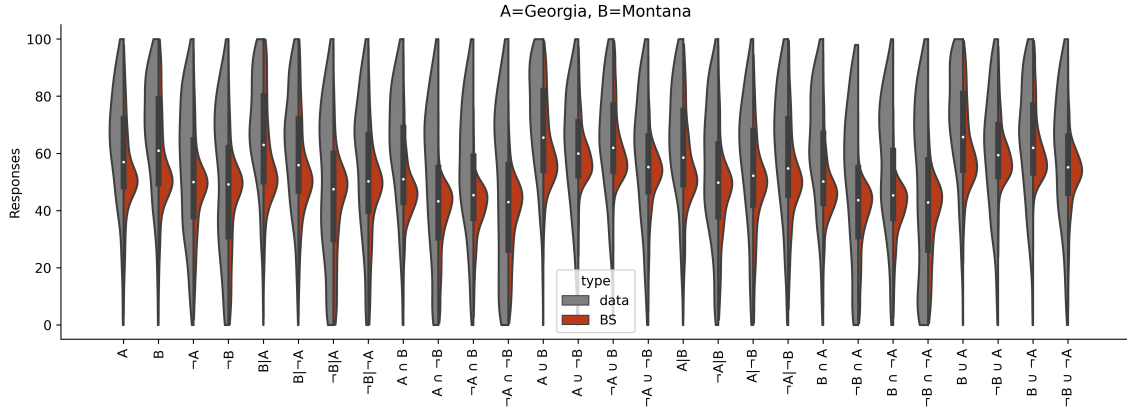


Figure S.3.2. Mean predictions of the Quantum Sequential Sampler, Bayesian Sampler, and relative frequency models, against empirical results. The data are for Triplet 2.



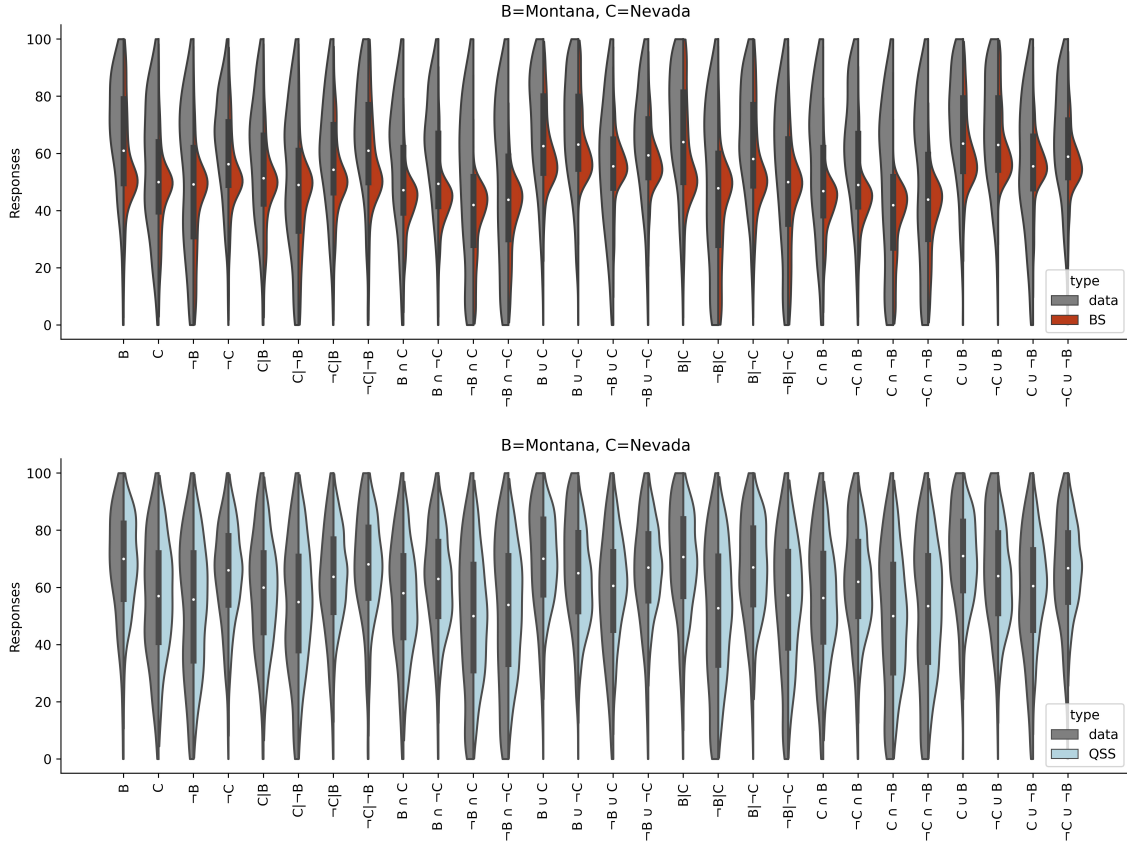


Figure S.3.3. Violin plots showing empirical data vs. the predictions of the Quantum Sequential Sampler, and empirical data (top panel) vs. the predictions of the Bayesian Sampler (bottom panel) for the second triplet. The data are for the  $\{A, B\}$ ,  $\{A, C\}$  and  $\{B, C\}$  pairs of events for Triplet 2.

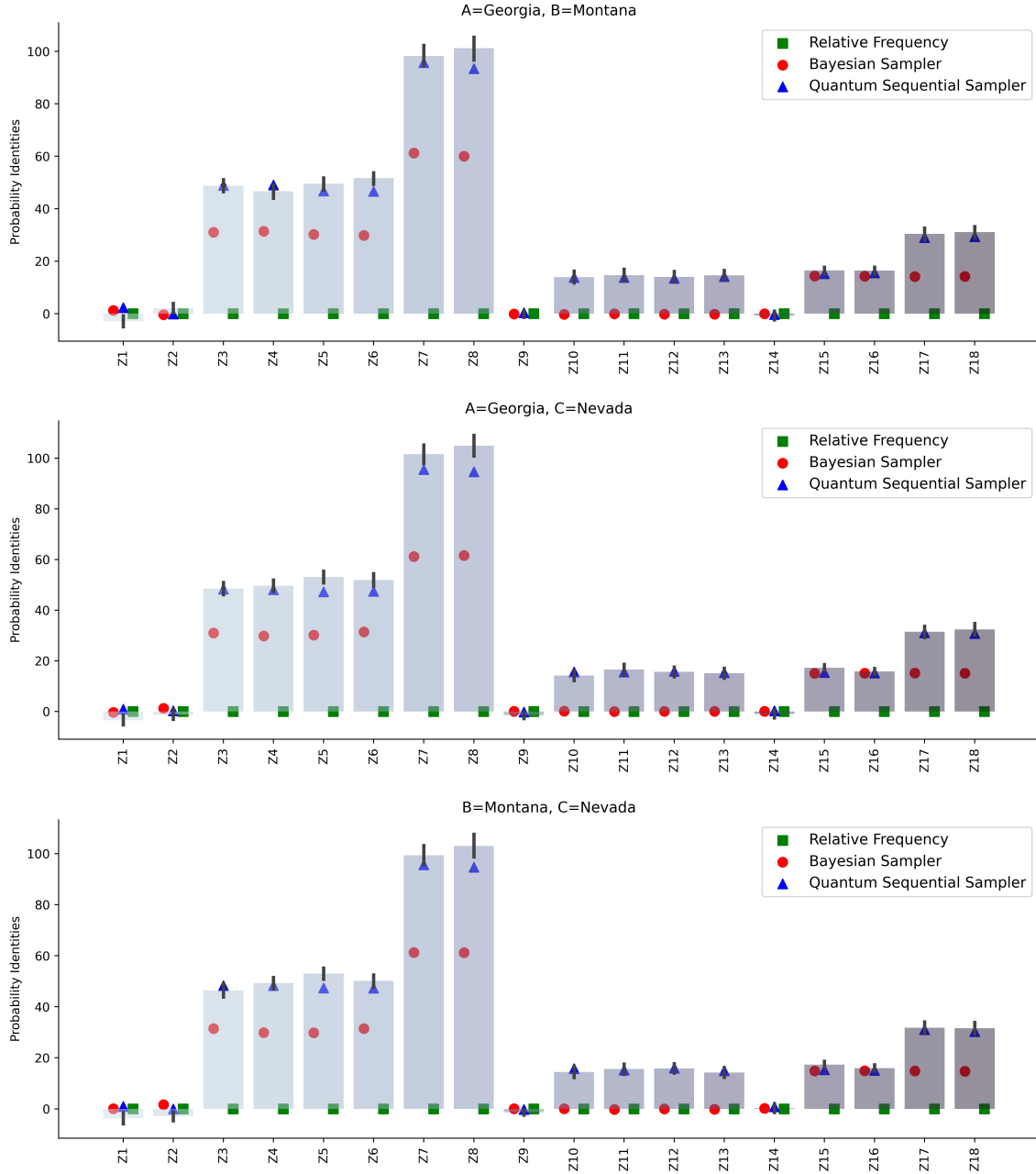


Figure S.3.4. The empirical value of probabilistic identities in Table 2, together with values computed from best-fit predictions, from the Quantum, Sequential Sampler, the Bayesian Sampler, and the relative frequency model. The data in this figure is from Triplet 2.

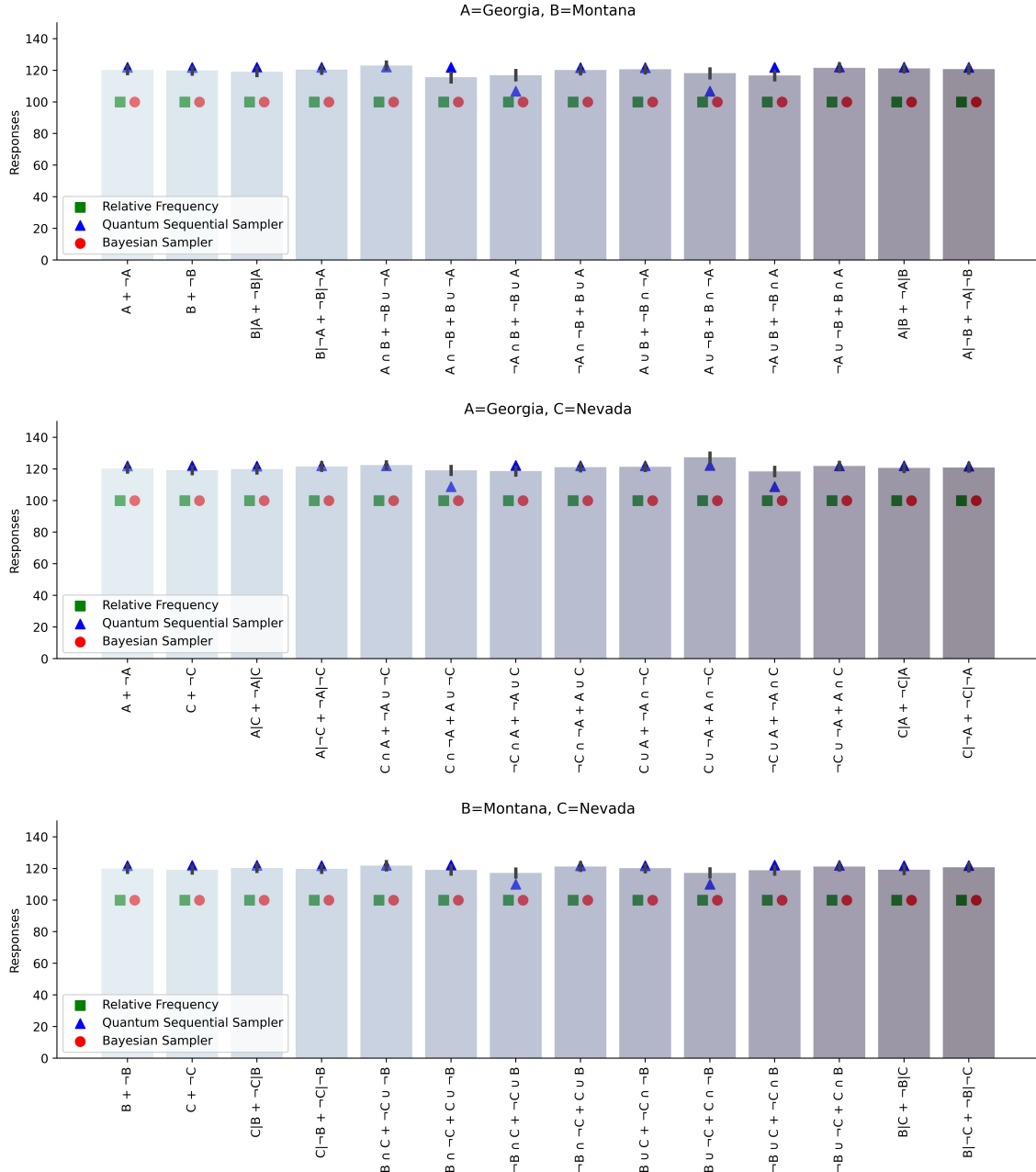


Figure S.3.5. Empirical values for binary complementarity, together with predictions from the Quantum Sequential Sampler, the Bayesian Sampler, and the relative frequency model, averaged across all participants. The results are for Triplet  $T2$ .



## Supplementary Material 4

In this section we offer some additional figures, illustrating the results of comparing the quantum and classical variant of Quantum Sequential Sampler, across all of the 1162 participants.



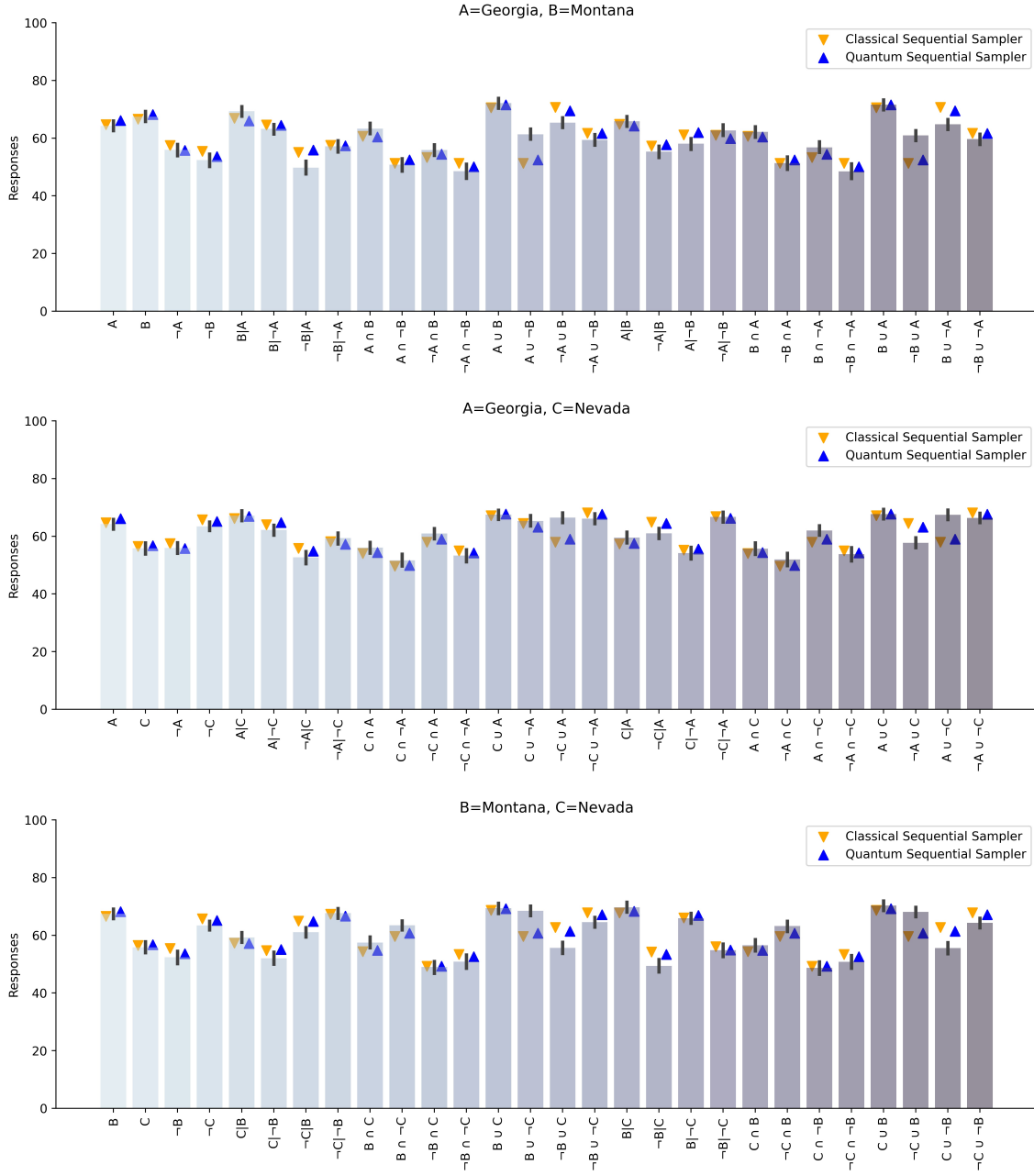
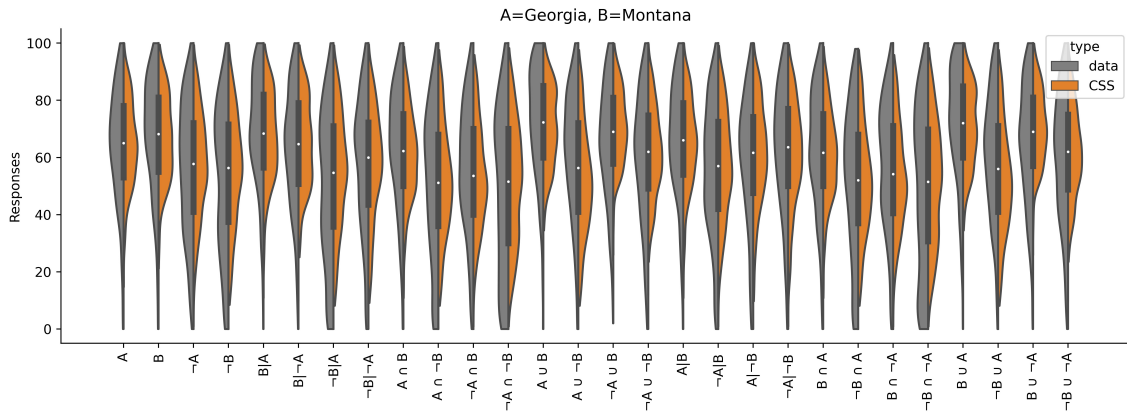
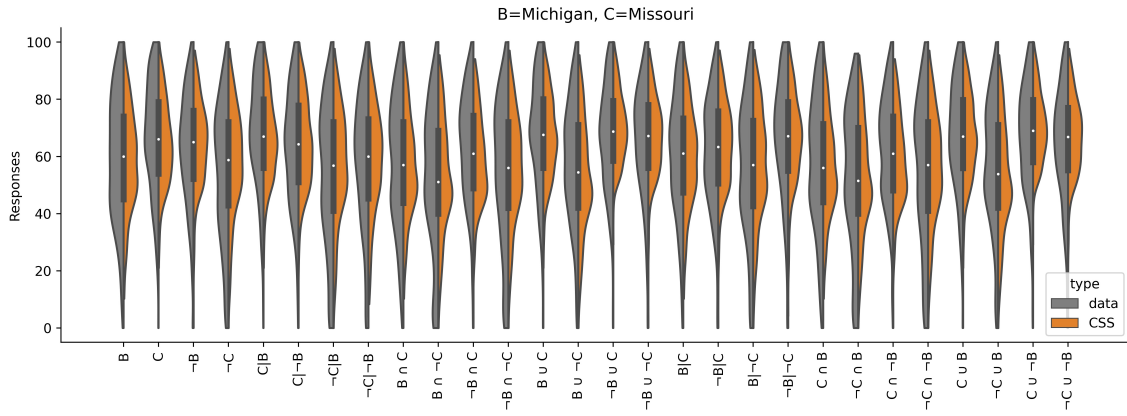
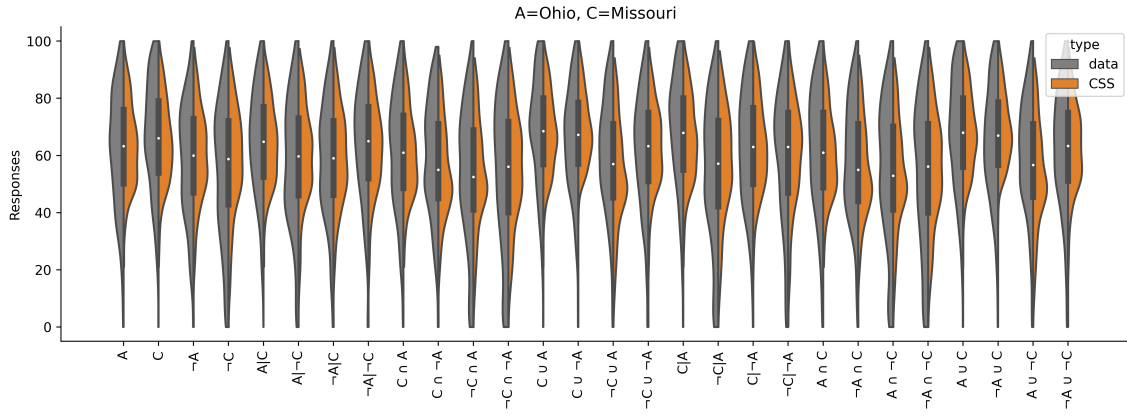
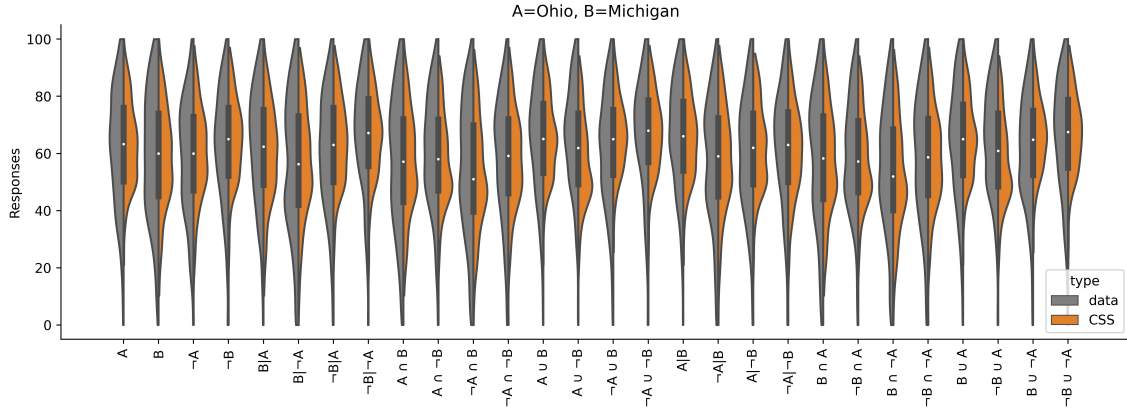


Figure S.4.1. Mean predictions of the quantum and classical variants of the Quantum Sequential Sampler against empirical results.



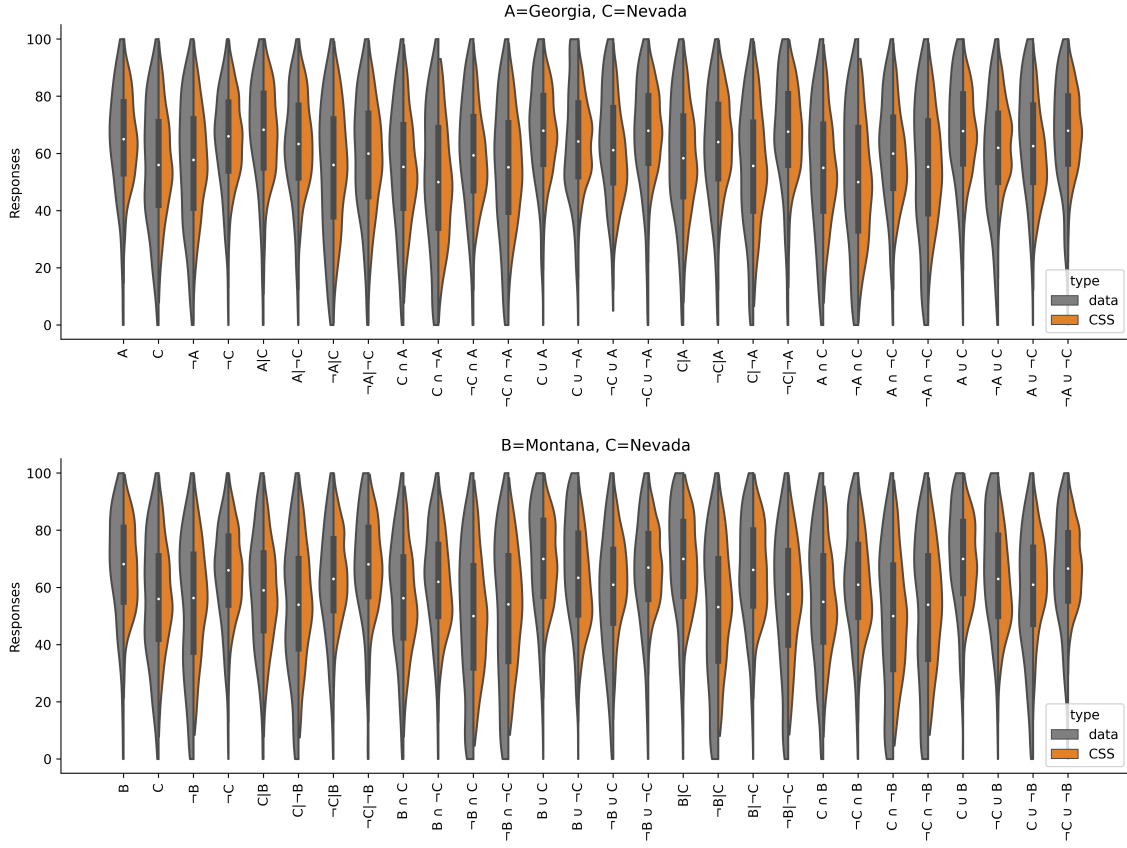
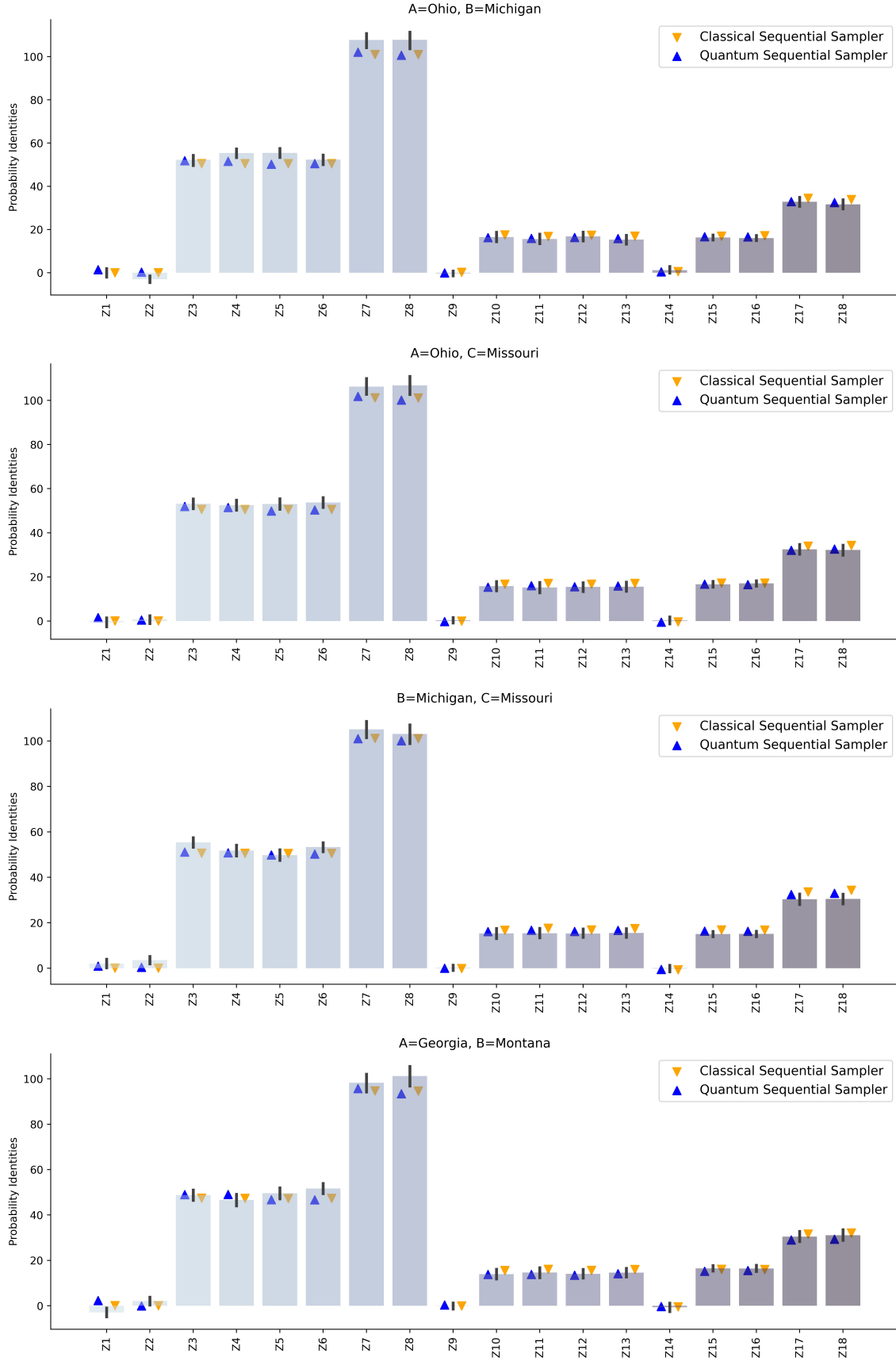


Figure S.4.2. Violin plots showing empirical data vs. the predictions of the classical variant of the Quantum Sequential Sampler model.



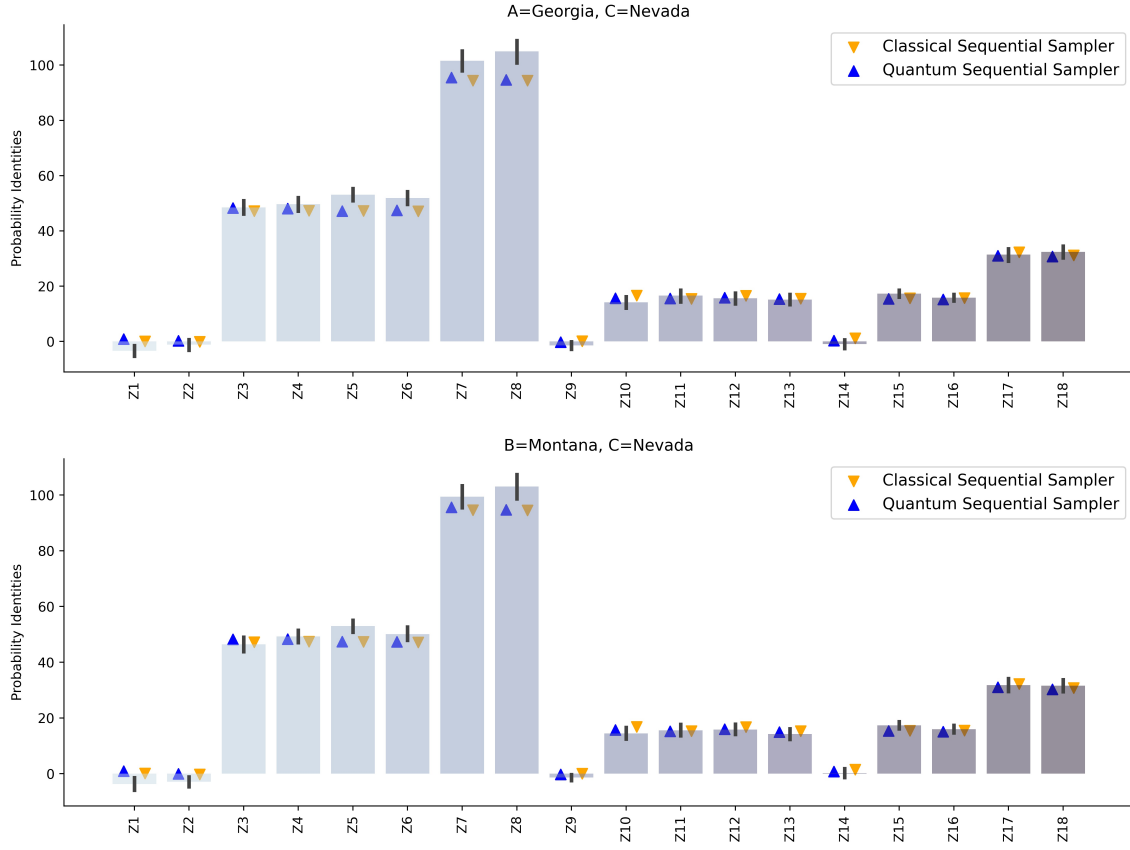
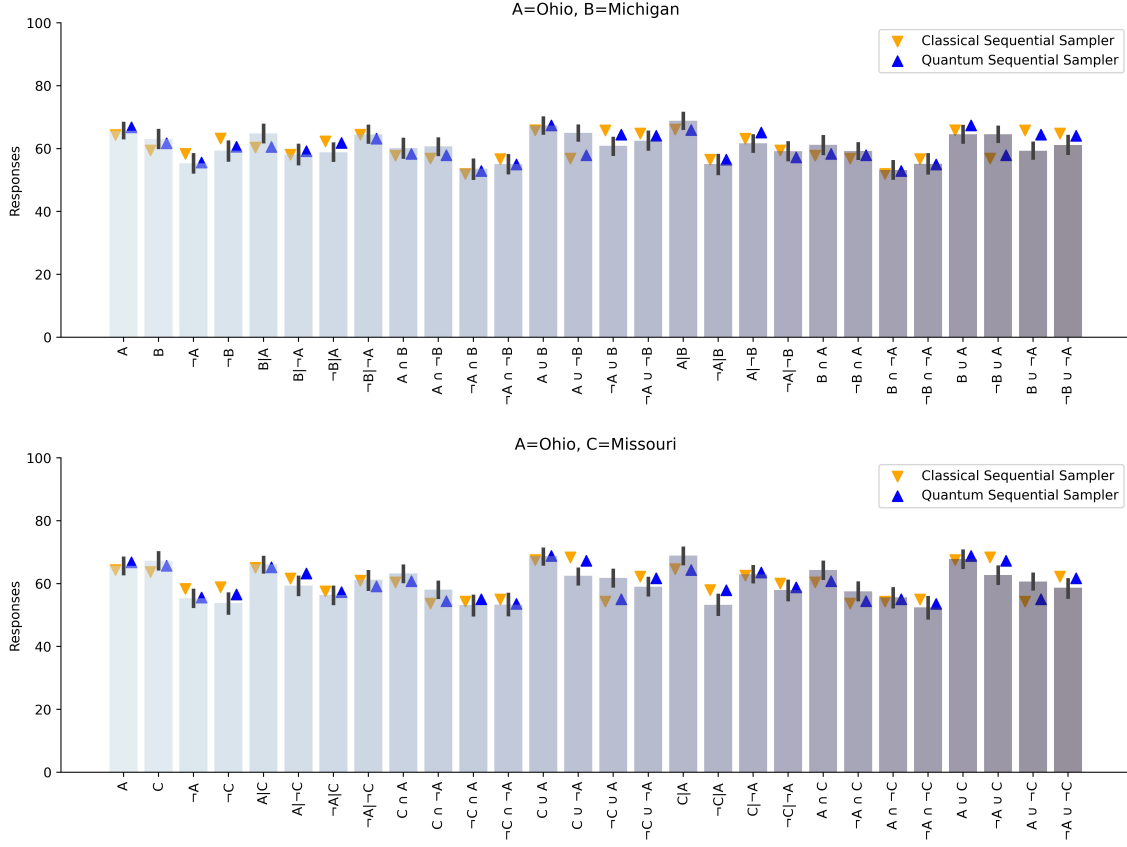
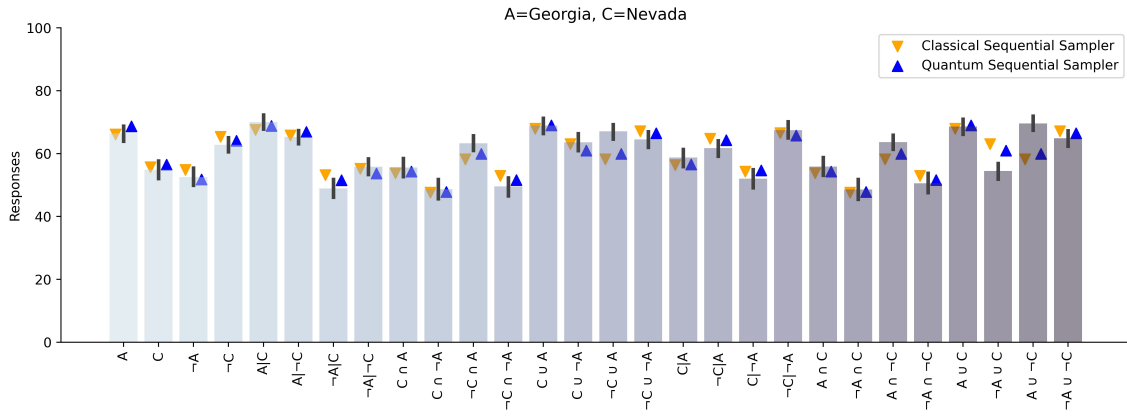
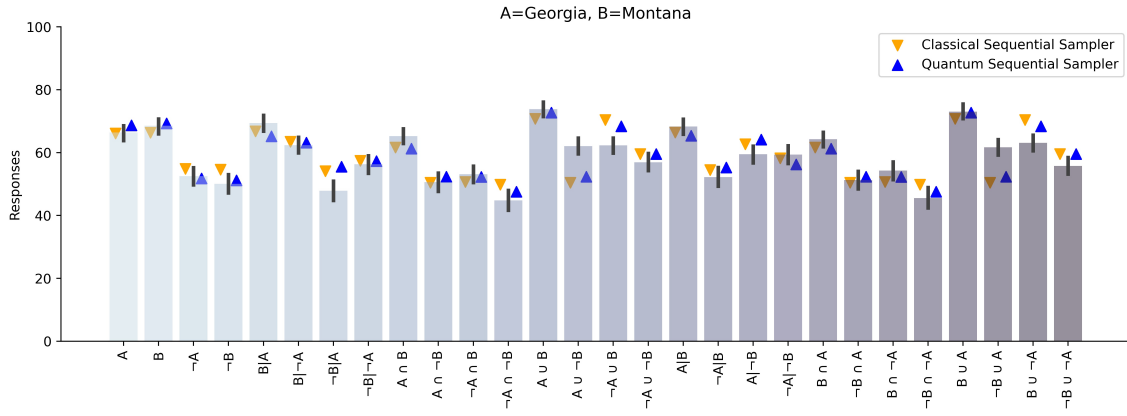
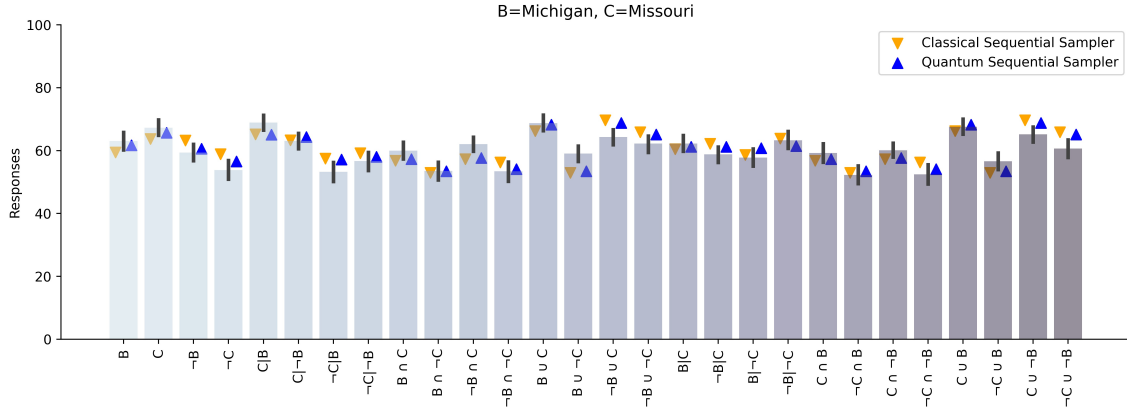


Figure S.4.3. The empirical value of probabilistic identities in Table 2, together with values computed from best-fit predictions, from the quantum and classical variants of the Quantum Sequential Sampler.

## Supplementary Material 5

In this section, we display additional figures comparing participants whose quantum interference parameters are significant by the two-sigma criterion ( $p < .05$ ) with those whose quantum interference parameters are significant by the four-sigma criterion ( $p < .00008$ ).







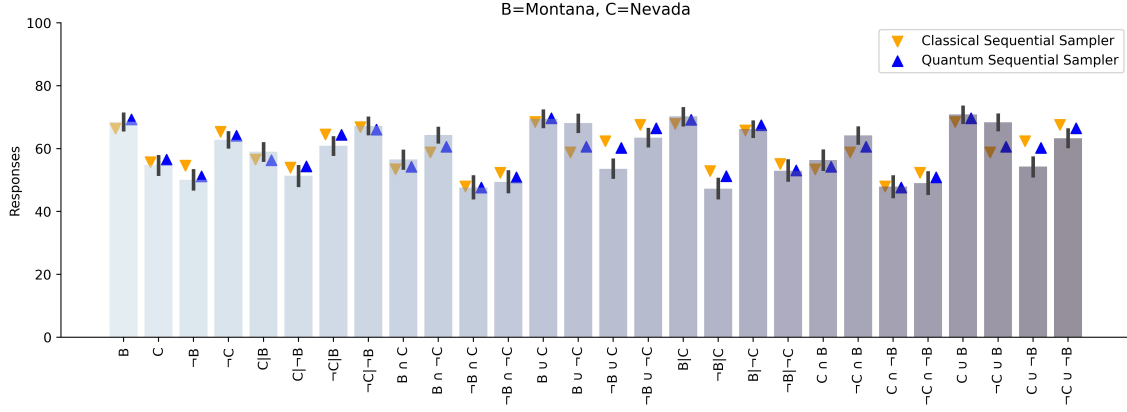
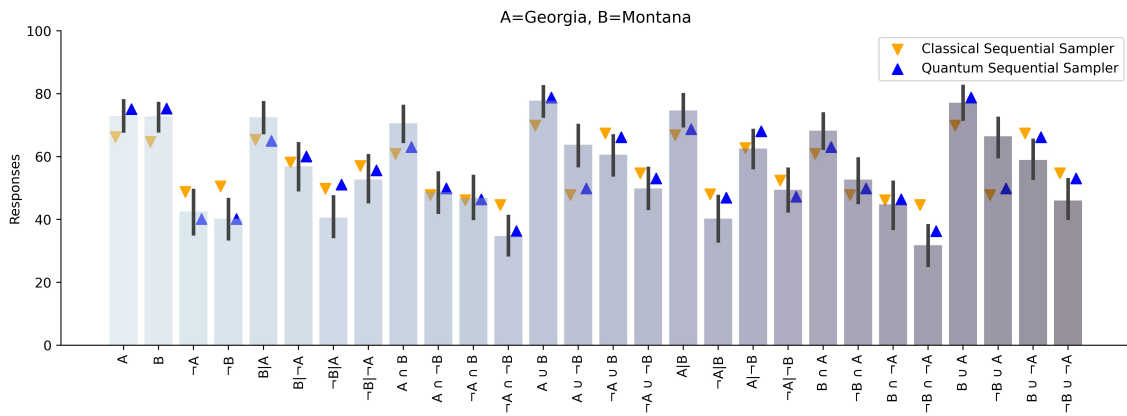
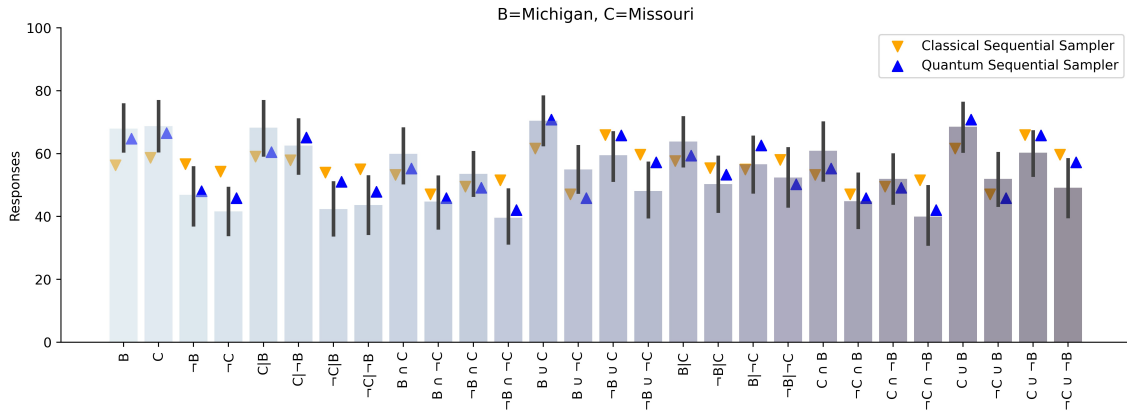
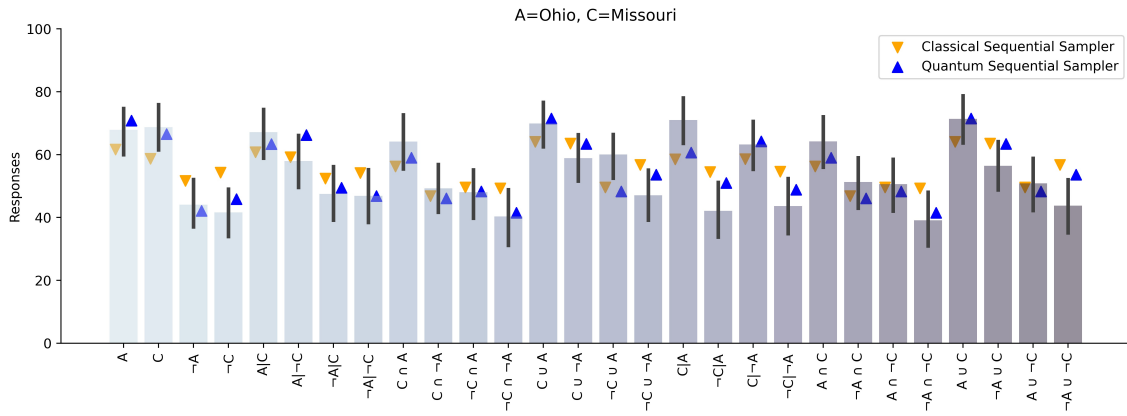
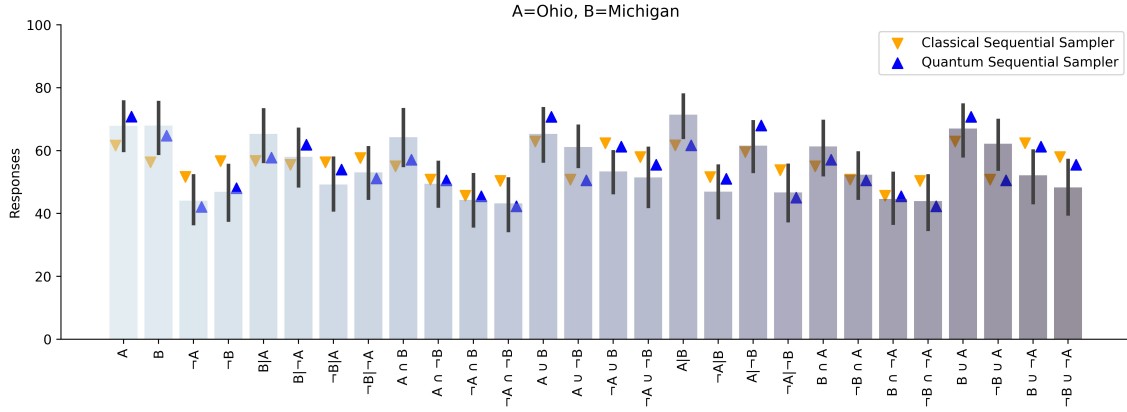


Figure S.5.1. Mean predictions of the quantum and classical variants of the Quantum Sequential Sampler against empirical results for participants whose quantum interference parameter is significant by the criterion ( $p < .05$ ).



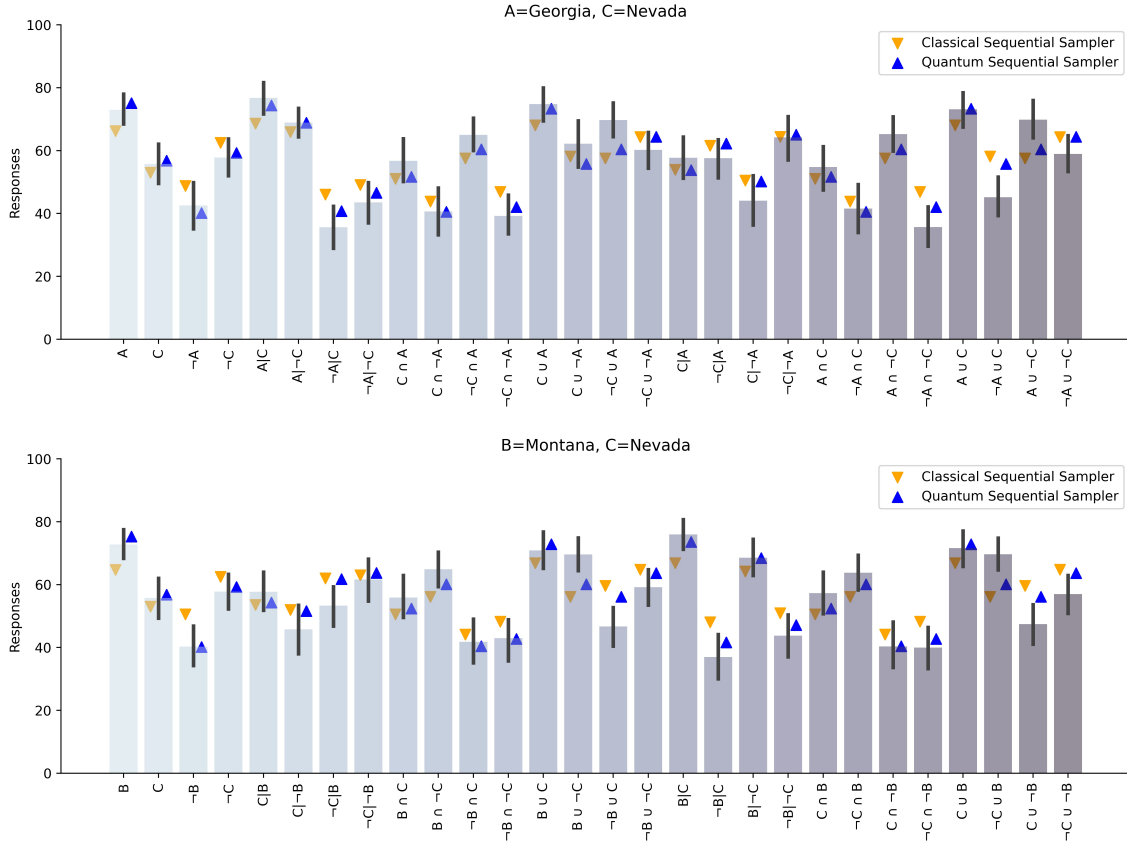
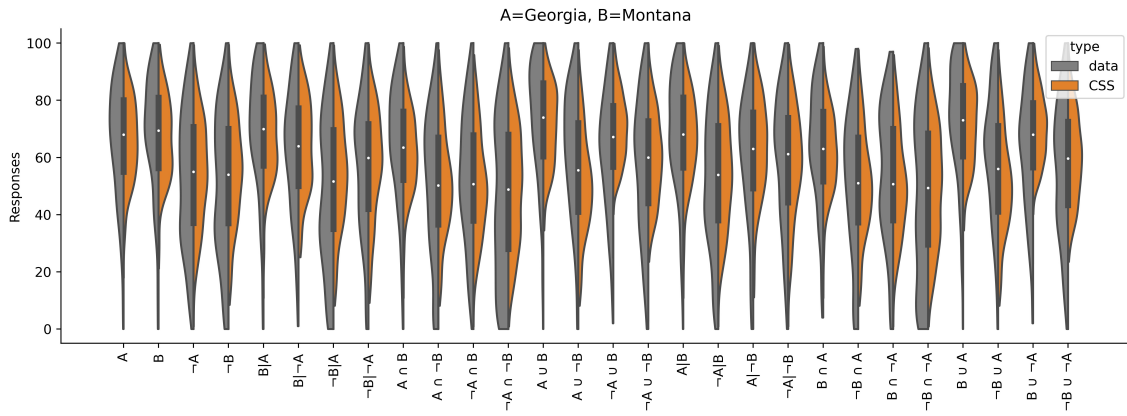
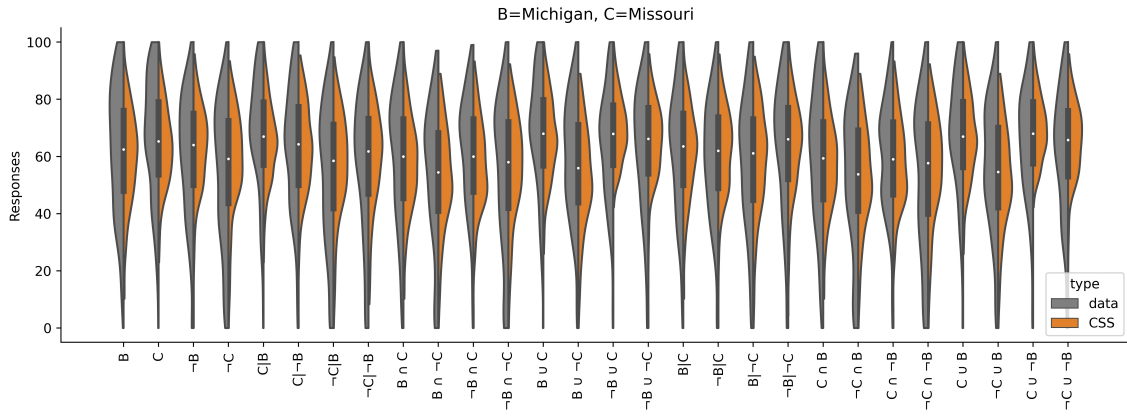
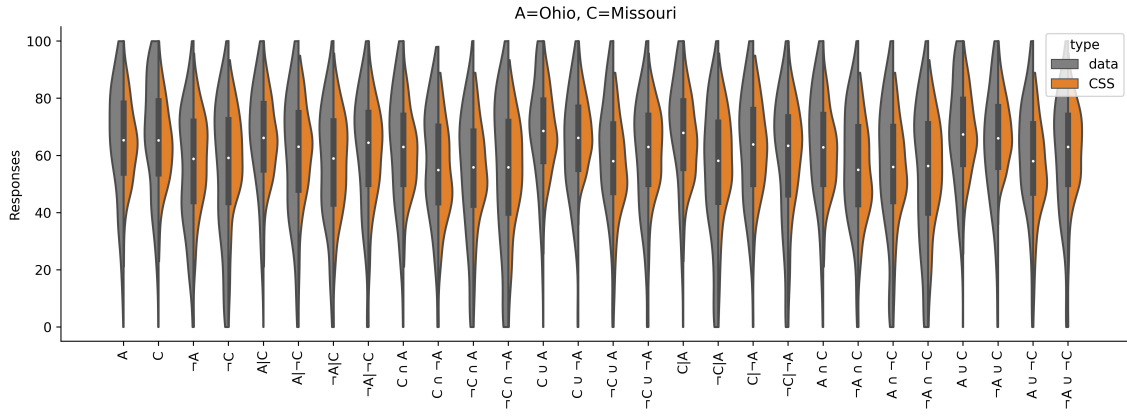
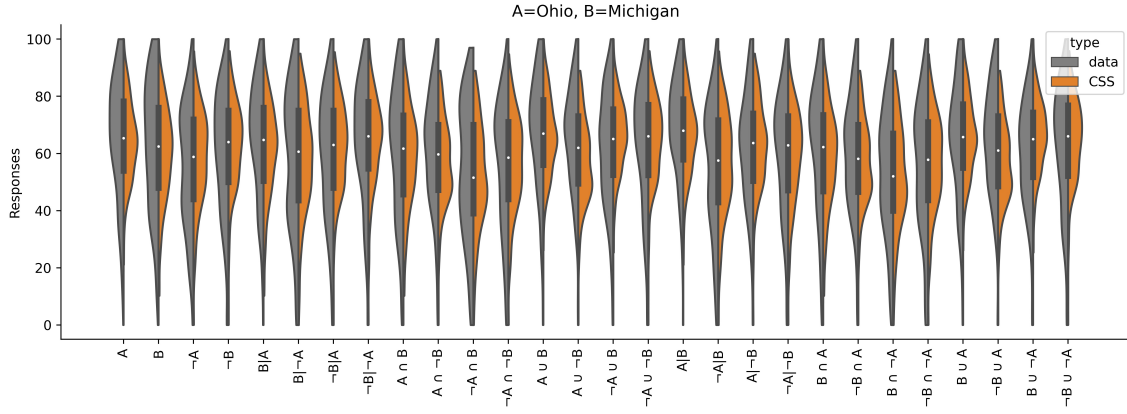
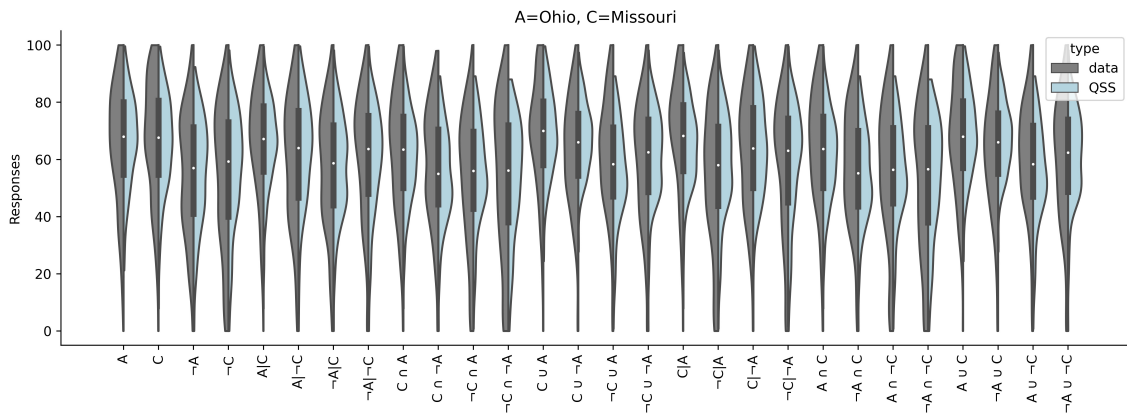
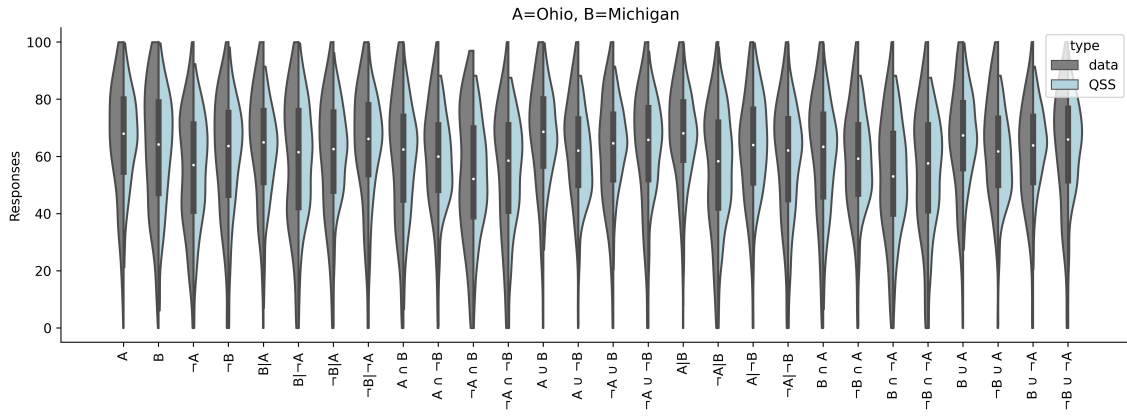
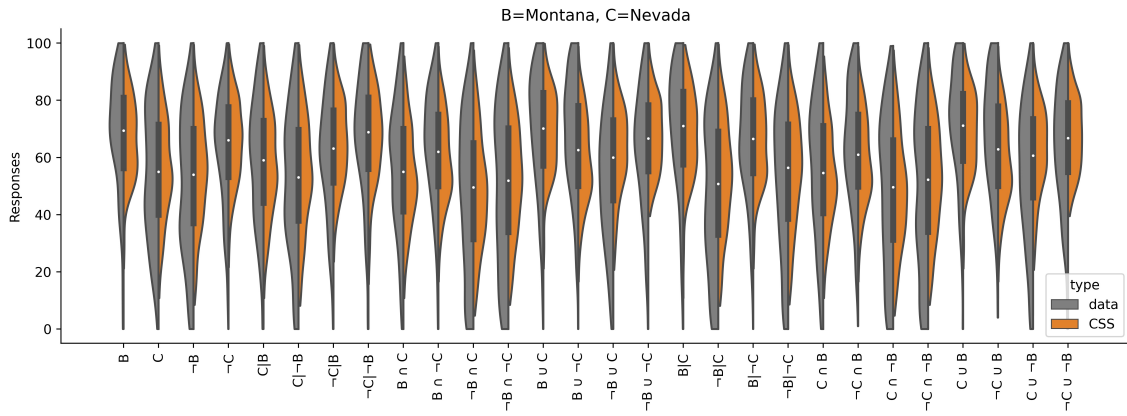
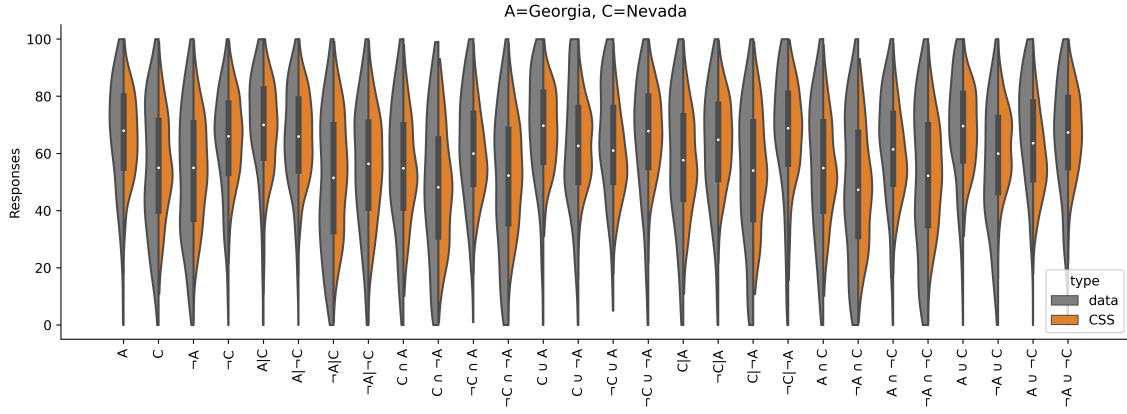
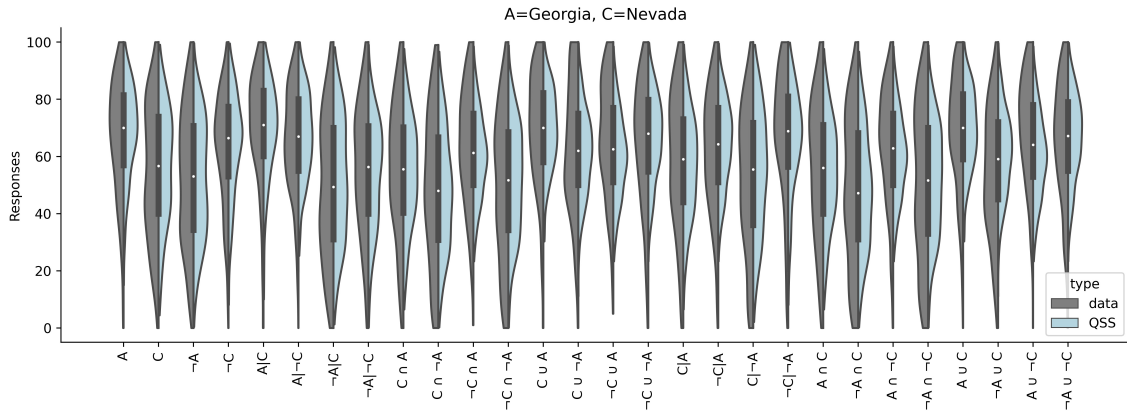
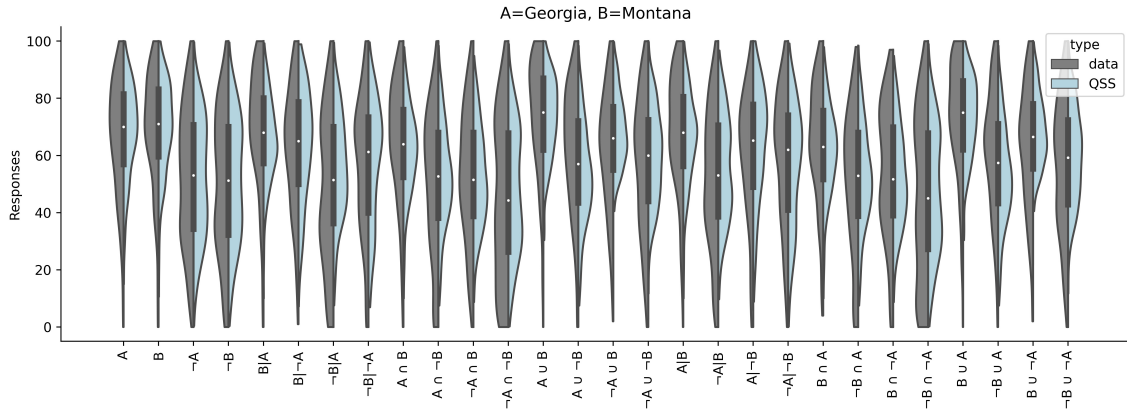
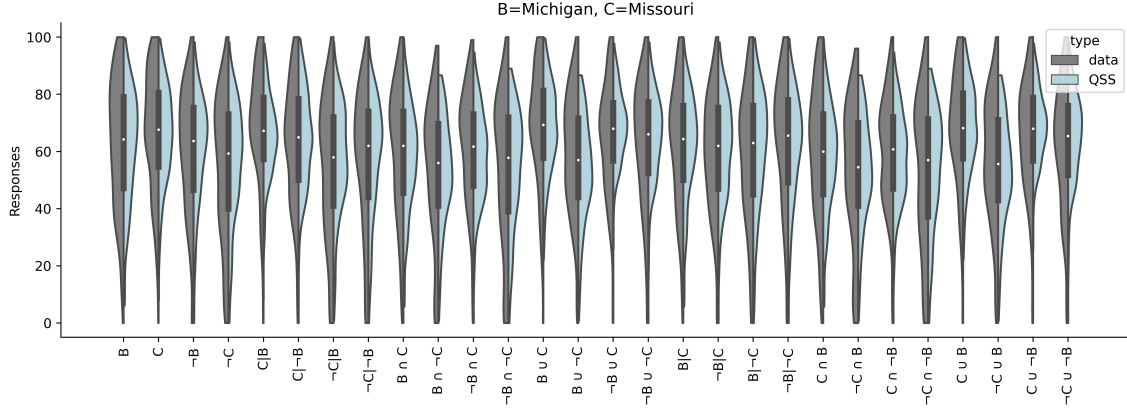


Figure S.5.2. Mean predictions of the quantum and classical variants of the Quantum Sequential Sampler against empirical results for participants whose quantum interference parameter is significant by the criterion ( $p < .00008$ ).







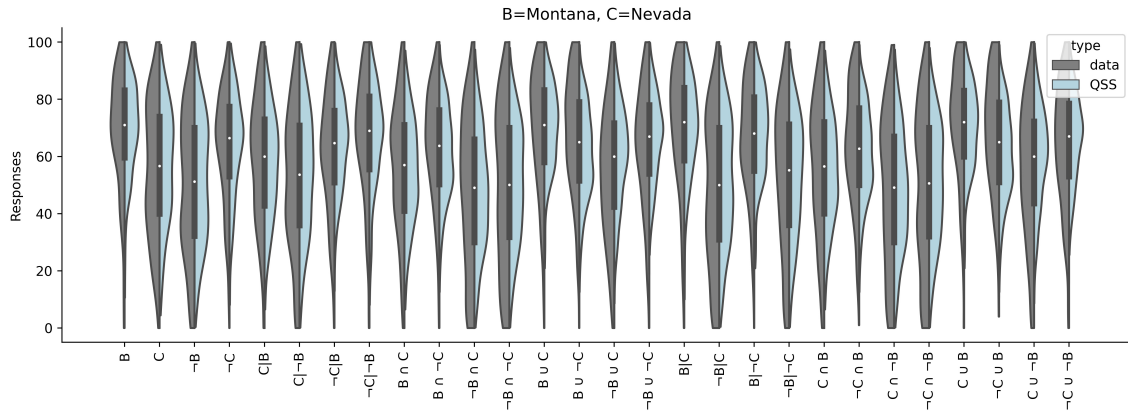
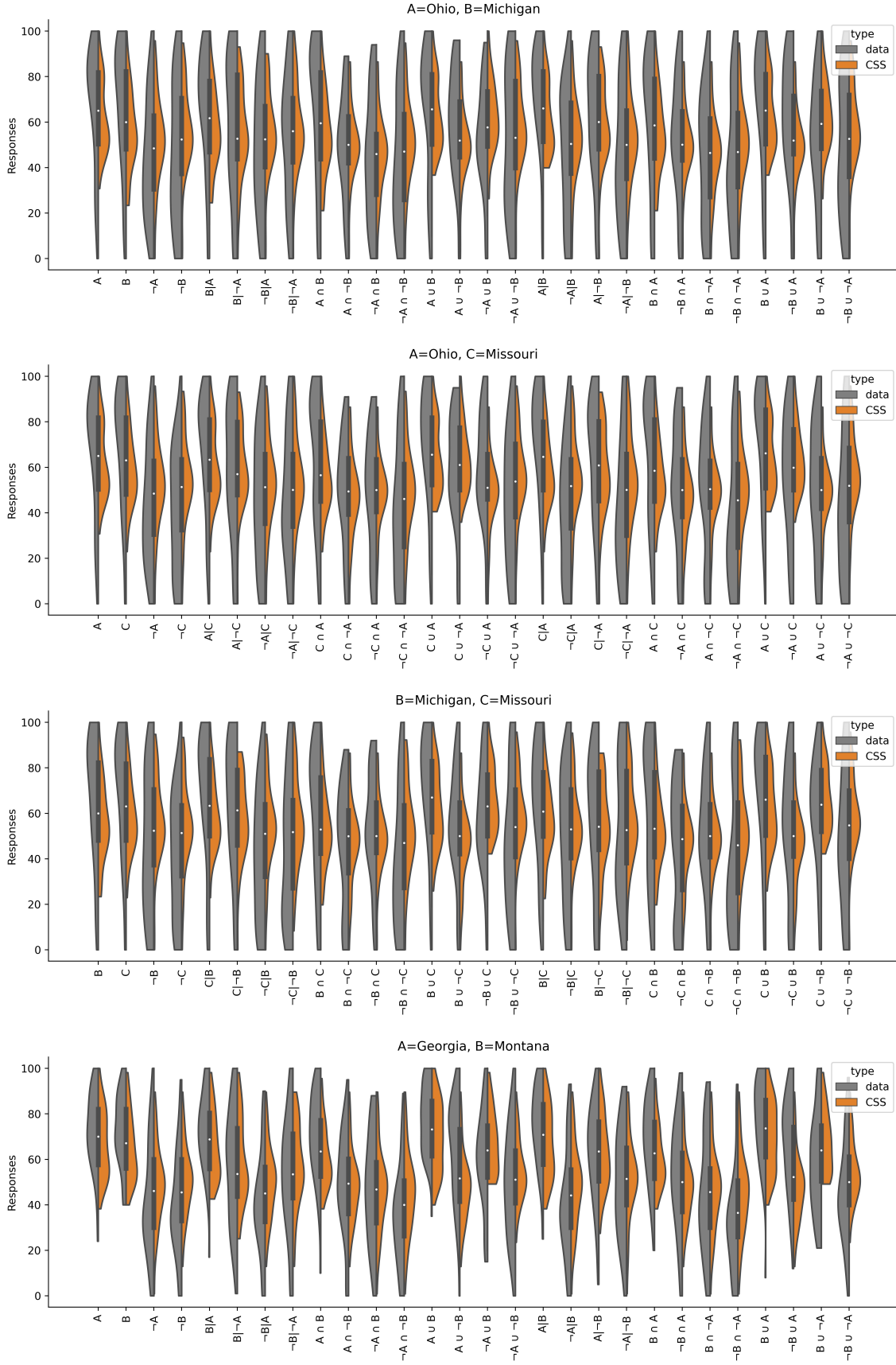
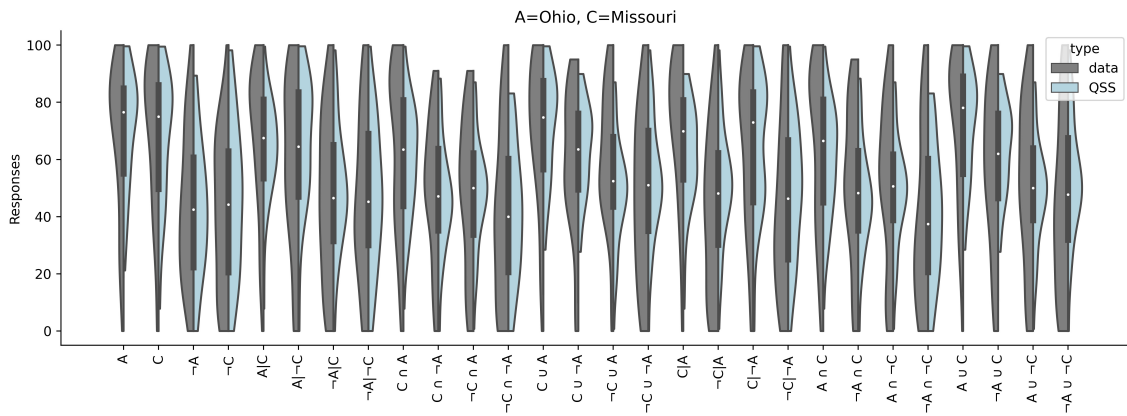
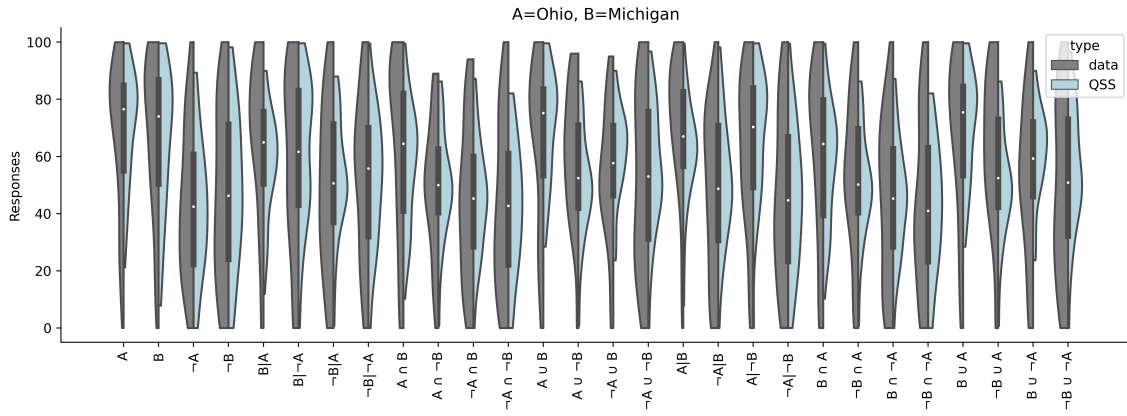
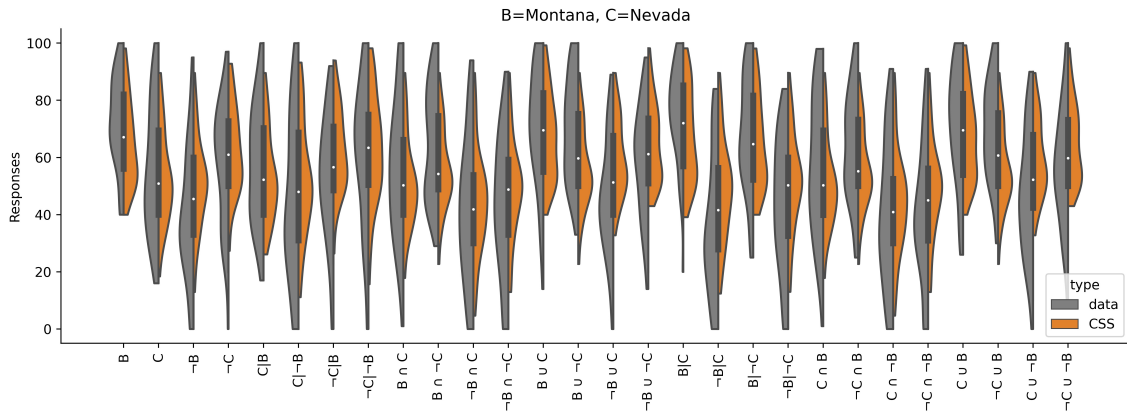
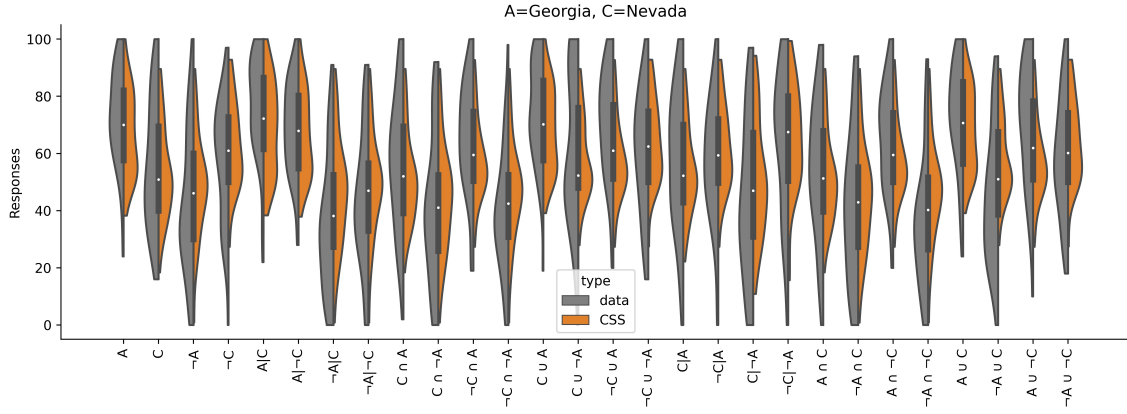
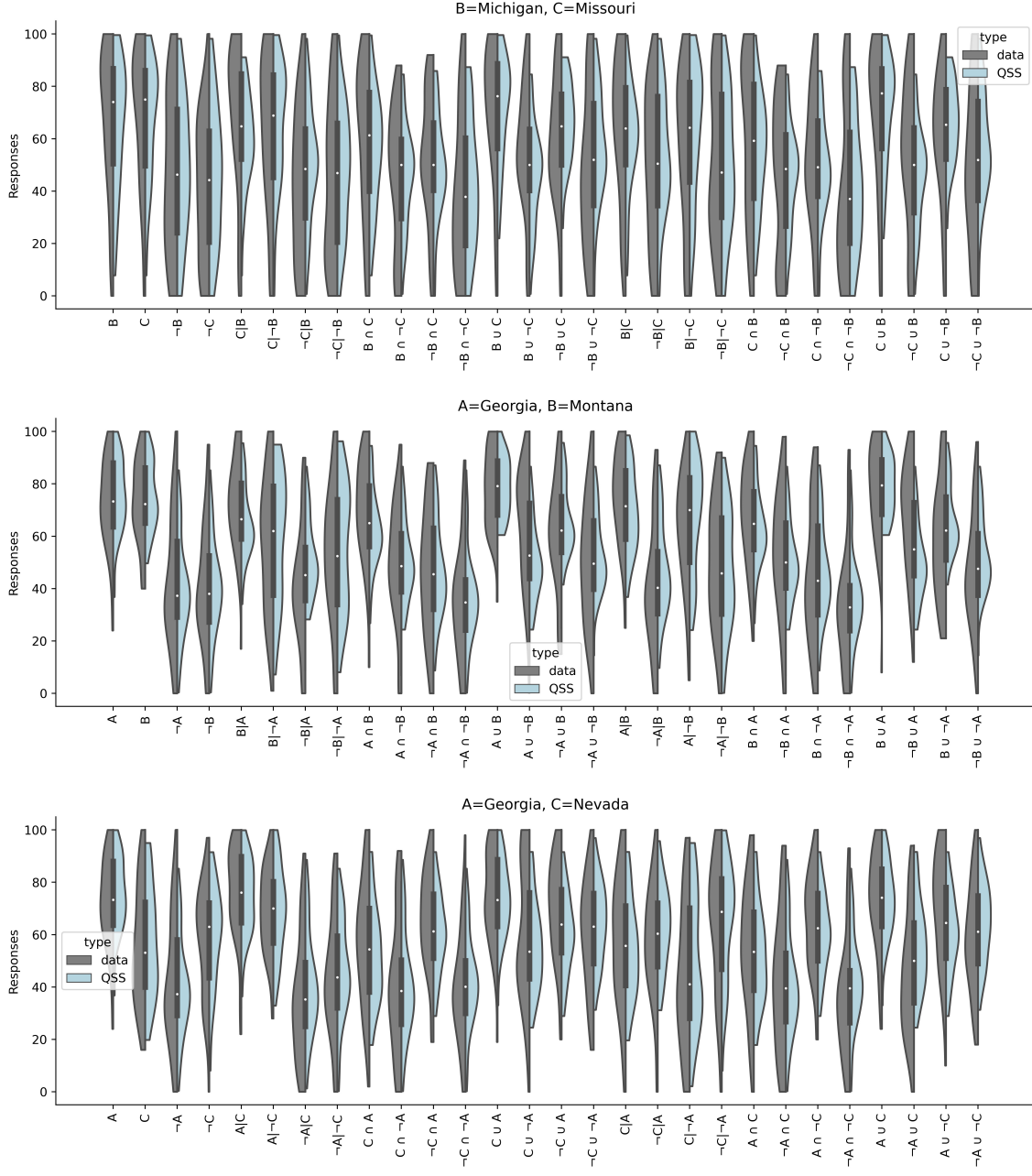


Figure S.5.3. Violin plots showing empirical data vs. the predictions of the classical and quantum variant of the Quantum Sequential Sampler model for participants whose quantum interference parameter is significant by the criterion ( $p < .05$ ).









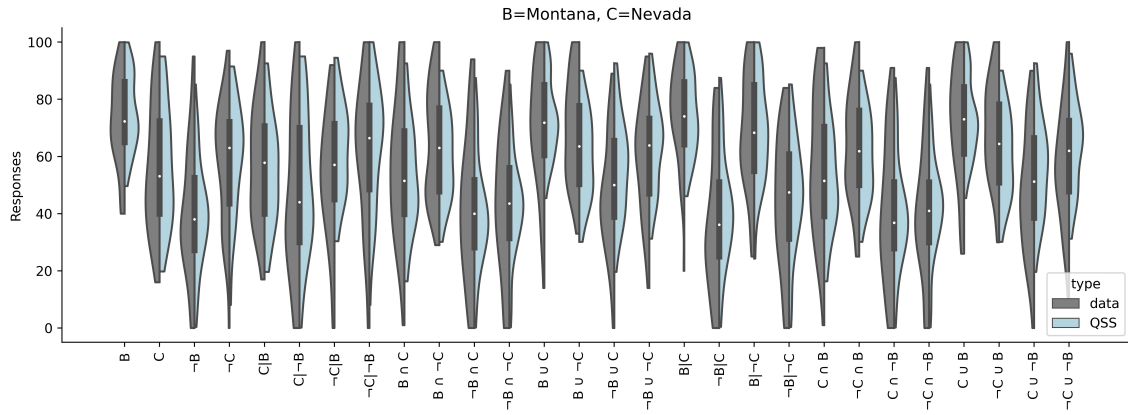
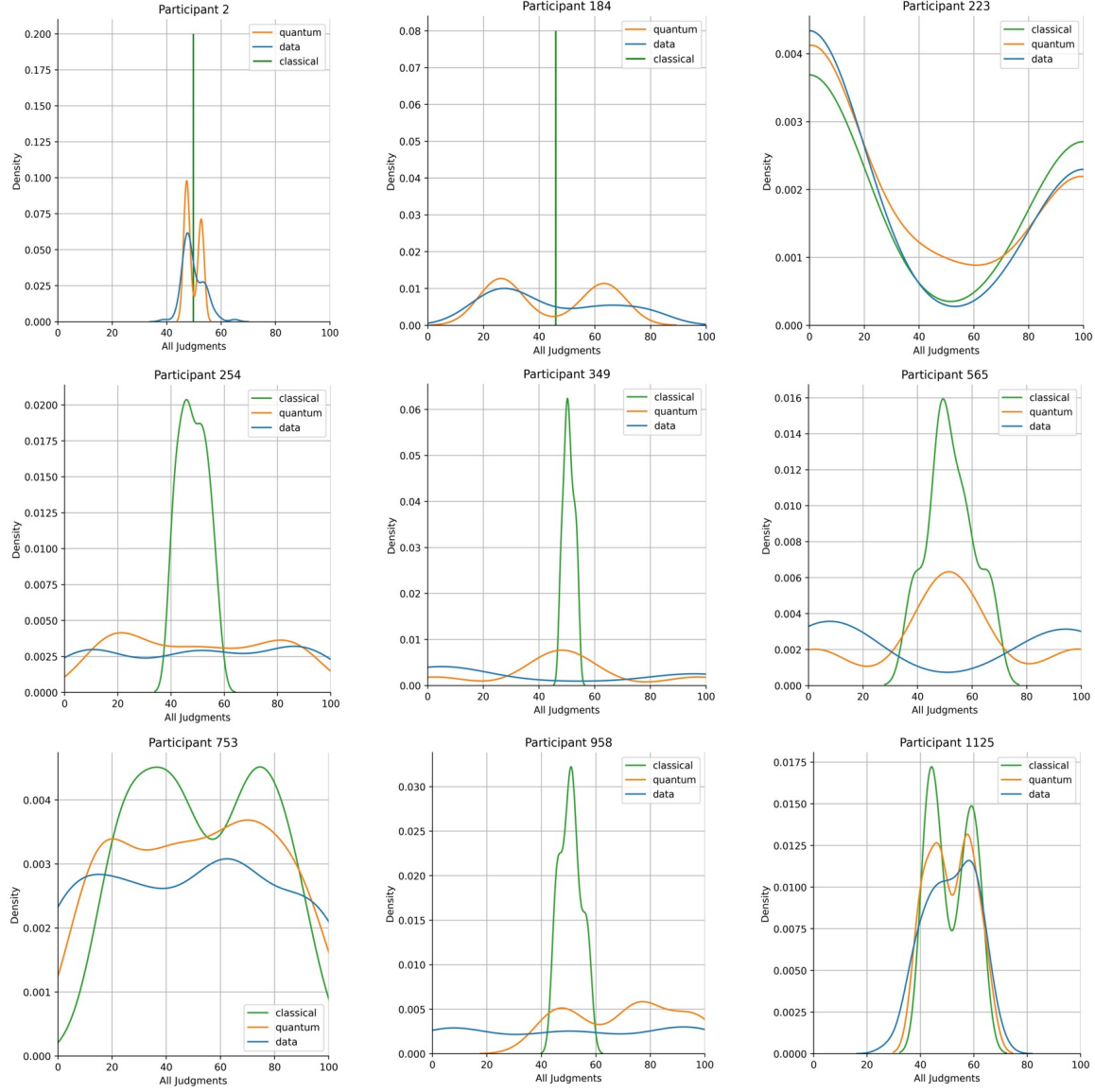


Figure S.5.4. Violin plots showing empirical data vs. the predictions of the classical and quantum variant of the Quantum Sequential Sampler model for participants whose quantum interference parameter is significant by the criterion ( $p < .00008$ ).

## Supplementary Material 6

In this section, we showcase the kernel density plots for more participants whose quantum interference parameter is significant by ( $p < .00008$ ).



*Figure S.6.1.* Example of kernel density plots of individual whose quantum interference parameters are significant by the criterion ( $p < .00008$ ). Classical refers to the classical variant of the quantum sequential sampler, and quantum refers to the quantum variant.

## Supplementary Material 7

In the following analysis, the values for the Z identities  $Z_1 - Z_{18}$  were compared between the probabilistic judgments of election events collected to test the Quantum Sequential Sampler model, as well as the judgments of weather events collected by Zhu et al. (2020) to test their Bayesian Sampler model.

First, in the case of the Bayesian Sampler, participants provided three ratings for each question. We computed the average of these three ratings, so that there was a single score for each probability judgment. In two out of the five weather conditions in their experiment (normal/typical, icy/frosty, windy/cloudy, cold/rainy and warm/snowy) 59 participants provided probability ratings, while in the other three 84 participants provided ratings. This results in a total of 370 data points (probability judgments) available for the computation of the Z identities.

For the Quantum Sequential Sampler, data collection was more extensive and involved multiple events and all possible orders thereof. For that reason, each Z identity could be computed three times for each participant (six times if the order of events is considered relevant). To keep the comparison between the two data sets consistent, the various iterations of the same Z identity were averaged for each of the 1162 participants, before comparing against the Z identities computed on the basis of the Bayesian Sampler data.

The Z identities were compared using classical and Bayesian two-sided, independent samples t-tests. According to the Bayesian t-tests, there was very strong evidence that the values of the identities computed from the present data were different from those computed from the Zhu et al. (2020) data for all identities, except from  $Z_1, Z_9, Z_{16}, Z_{17}$ , as seen in

Table S7.1.

The results using classical t-tests were identical (see Table S7.2). Please note that Levene's tests of homogeneity of variance indicated violations of the assumption of equal variances for most comparisons. As a consequence, Welch tests were also performed. There were no differences in the conclusions however and so, for simplicity, only the results using the classical t-tests are presented here.

The results are also shown in Figure S.7.1, highlighting the differences between the Z values computed from the two different data sets for most of the identities.

| Bayesian Independent Samples T-Test | $\mathbf{BF}_{10}$      | error %                  |
|-------------------------------------|-------------------------|--------------------------|
| Z1                                  | 0.099                   | 0.048                    |
| Z2                                  | $\infty$                | 0.000                    |
| Z3                                  | $3.586 \times 10^{232}$ | $2.565 \times 10^{-236}$ |
| Z4                                  | $1.776 \times 10^{25}$  | $1.468 \times 10^{-28}$  |
| Z5                                  | $7.227 \times 10^{257}$ | $8.398 \times 10^{-263}$ |
| Z6                                  | $4.175 \times 10^{236}$ | $1.802 \times 10^{-239}$ |
| Z7                                  | $7.268 \times 10^{145}$ | $2.204 \times 10^{-151}$ |
| Z8                                  | $3.492 \times 10^{119}$ | $3.744 \times 10^{-124}$ |
| Z9                                  | 0.100                   | 0.047                    |
| Z10                                 | $4.820 \times 10^7$     | $7.740 \times 10^{-11}$  |
| Z11                                 | $3.583 \times 10^{24}$  | $7.376 \times 10^{-28}$  |
| Z12                                 | $1.837 \times 10^7$     | $2.051 \times 10^{-10}$  |
| Z13                                 | $3.208 \times 10^{27}$  | $7.782 \times 10^{-31}$  |
| Z14                                 | $2.875 \times 10^8$     | $1.275 \times 10^{-11}$  |
| Z15                                 | 0.068                   | 0.069                    |
| Z16                                 | 0.087                   | 0.054                    |
| Z17                                 | $7.999 \times 10^6$     | $4.749 \times 10^{-10}$  |
| Z18                                 | $2.601 \times 10^{21}$  | $1.080 \times 10^{-24}$  |

Table S7.1: Two-sided, independent samples Bayesian t-tests comparing the Z identities computed from the data for the Quantum Sequential Sampler vs the Bayesian Sampler.



| Independent Samples T-Test | t       | df   | p                   |
|----------------------------|---------|------|---------------------|
| Z1                         | -0.887  | 1530 | 0.375               |
| Z2                         | -51.981 | 1530 | < .001 <sup>a</sup> |
| Z3                         | -39.630 | 1530 | < .001 <sup>a</sup> |
| Z4                         | -11.312 | 1530 | < .001 <sup>a</sup> |
| Z5                         | -42.618 | 1530 | < .001 <sup>a</sup> |
| Z6                         | -40.109 | 1530 | < .001 <sup>a</sup> |
| Z7                         | -29.277 | 1530 | < .001 <sup>a</sup> |
| Z8                         | -25.978 | 1530 | < .001 <sup>a</sup> |
| Z9                         | -0.906  | 1530 | 0.365               |
| Z10                        | -6.468  | 1530 | < .001 <sup>a</sup> |
| Z11                        | -11.156 | 1530 | < .001 <sup>a</sup> |
| Z12                        | -6.311  | 1530 | < .001 <sup>a</sup> |
| Z13                        | -11.804 | 1530 | < .001 <sup>a</sup> |
| Z14                        | 6.749   | 1530 | < .001 <sup>a</sup> |
| Z15                        | -0.179  | 1530 | 0.858               |
| Z16                        | -0.724  | 1530 | 0.469               |
| Z17                        | -6.173  | 1530 | < .001 <sup>a</sup> |
| Z18                        | -10.430 | 1530 | < .001 <sup>a</sup> |

Table S7.2: Two-sided, independent samples classical t-tests comparing the Z identities computed from the data for the Quantum Sequential Sampler vs the Bayesian Sampler.

<sup>a</sup> Levene's test is significant ( $p < .05$ ), suggesting a violation of the equal variances assumption

| Identity | Mean <sub>BS</sub> | Mean <sub>ass</sub> | SD <sub>BS</sub> | SD <sub>QSS</sub> |
|----------|--------------------|---------------------|------------------|-------------------|
| Z1       | -0.025             | -0.014              | 0.195            | 0.212             |
| Z2       | -0.371             | 0.000               | 0.215            | 0.066             |
| Z3       | -0.028             | 0.508               | 0.124            | 0.250             |
| Z4       | 0.344              | 0.508               | 0.216            | 0.251             |
| Z5       | -0.084             | 0.522               | 0.240            | 0.238             |
| Z6       | -0.002             | 0.523               | 0.154            | 0.236             |
| Z7       | 0.342              | 1.031               | 0.224            | 0.435             |
| Z8       | 0.367              | 1.045               | 0.314            | 0.470             |
| Z9       | -0.005             | -0.001              | 0.092            | 0.069             |
| Z10      | 0.067              | 0.153               | 0.250            | 0.214             |
| Z11      | 0.019              | 0.152               | 0.148            | 0.215             |
| Z12      | 0.072              | 0.154               | 0.236            | 0.212             |
| Z13      | 0.014              | 0.151               | 0.124            | 0.213             |
| Z14      | 0.054              | 0.002               | 0.231            | 0.067             |
| Z15      | 0.162              | 0.163               | 0.142            | 0.107             |
| Z16      | 0.157              | 0.162               | 0.142            | 0.116             |
| Z17      | 0.229              | 0.317               | 0.229            | 0.240             |
| Z18      | 0.176              | 0.315               | 0.172            | 0.237             |

Table S7.3: Descriptive data for the Z identities computed from the data for the Quantum Sequential Sampler vs the Bayesian Sampler.

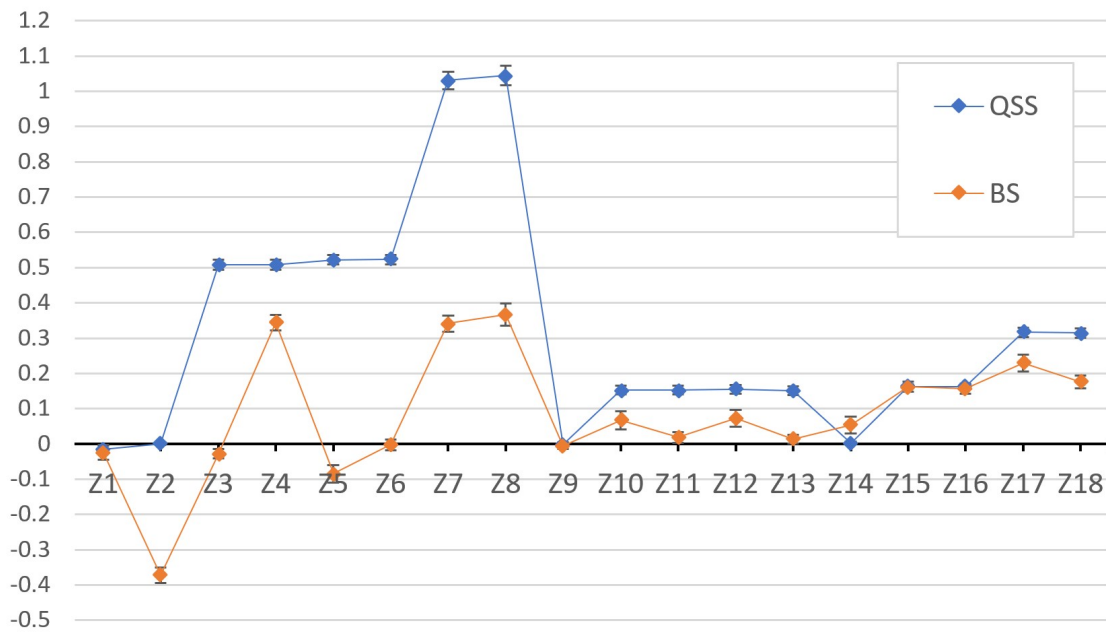
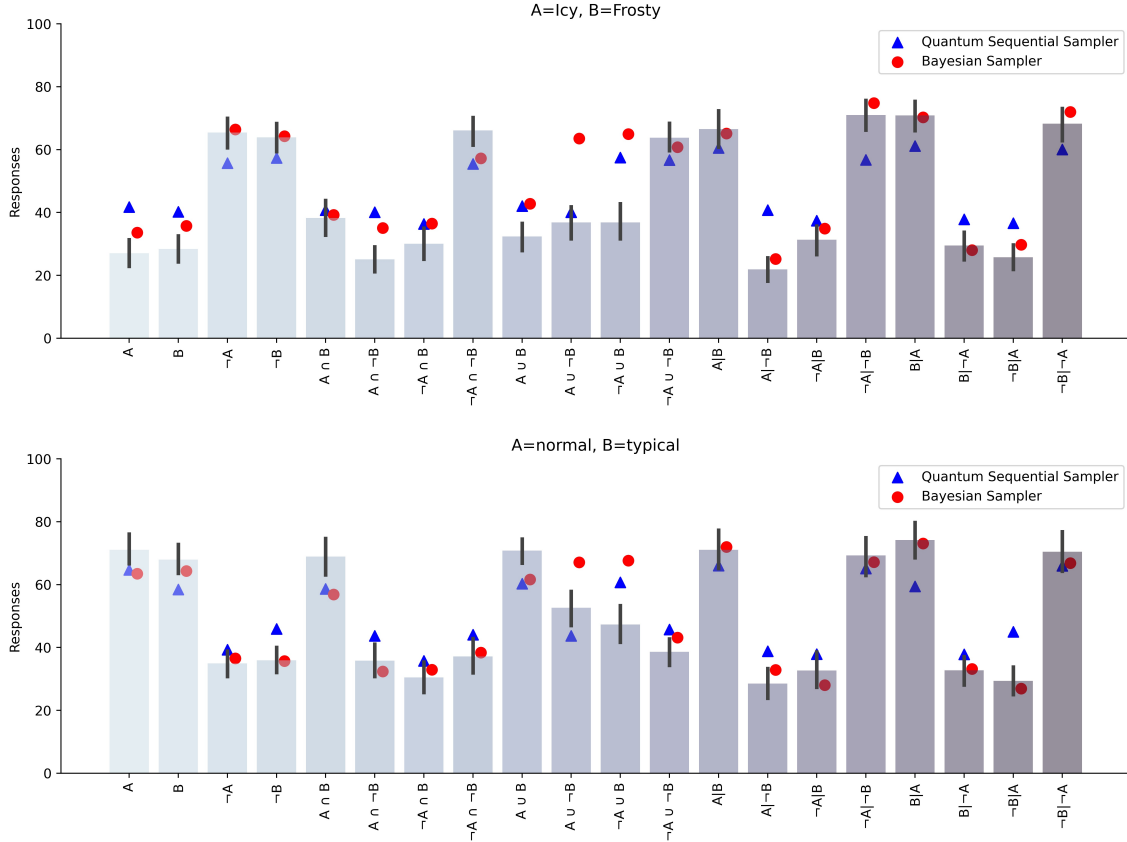


Figure S.7.1. Z identities computed from the data for the Quantum Sequential Sampler vs the Bayesian Sampler.

## Supplementary Material 8

In this section, we present all analysis plots that compare the Quantum Sequential Sampler (quantum variant) with the Bayesian Sampler model, using the dataset from Zhu et al. (2020) which includes repeated probability judgment measurements. It should be noted, as detailed in Appendix 2, that we have applied a rounding mechanism to the nearest 5 or 10 during our model fitting process to account for the significant presence of this rounding bias in Zhu et al.'s dataset. Consequently, the Bayesian Sampler's predictions in our analysis may slightly deviate from those reported in Zhu et al. (2020). Nonetheless, upon examination, it is evident that these differences are minor to be seen in plots and do not alter any analytical trend.



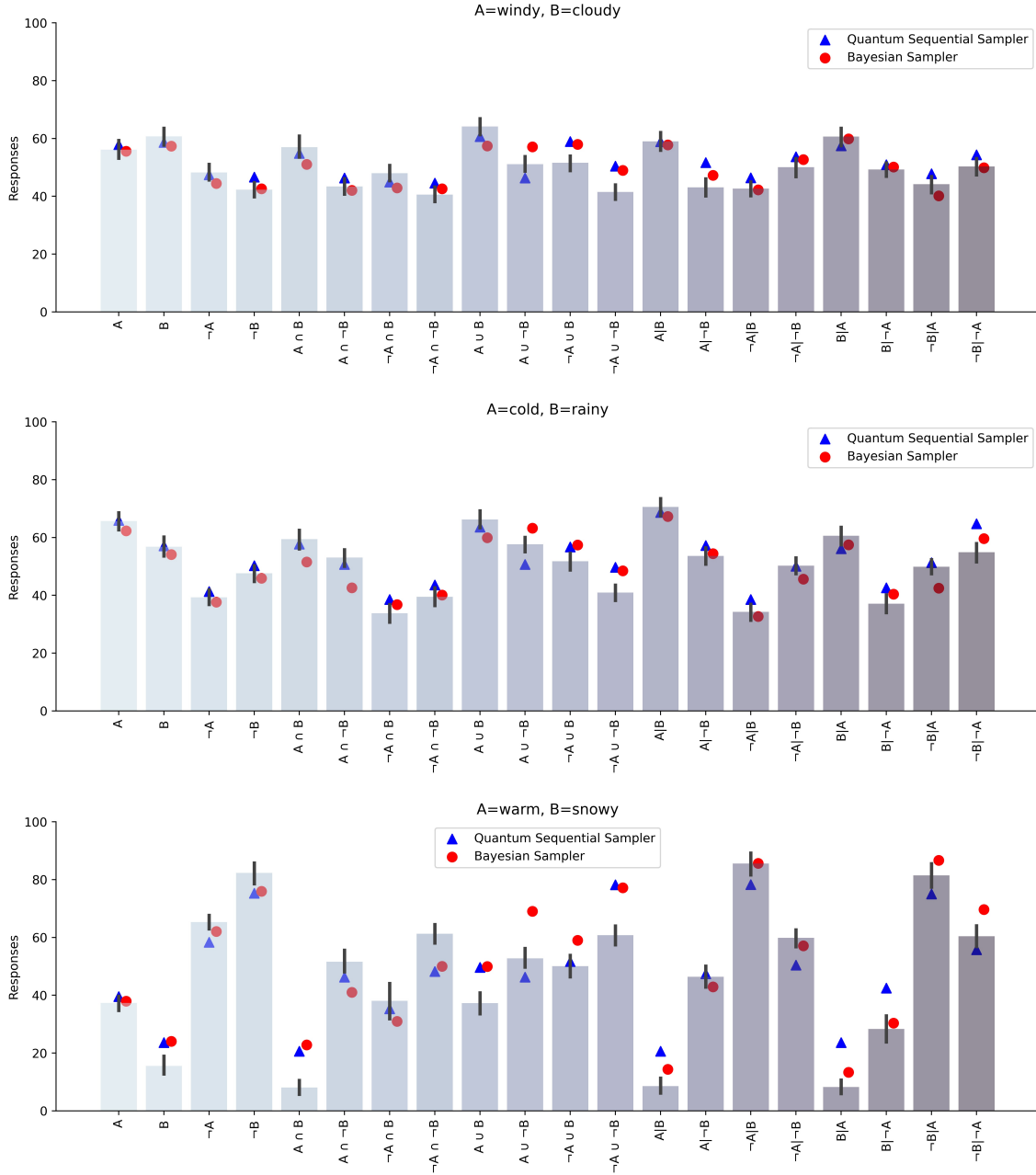
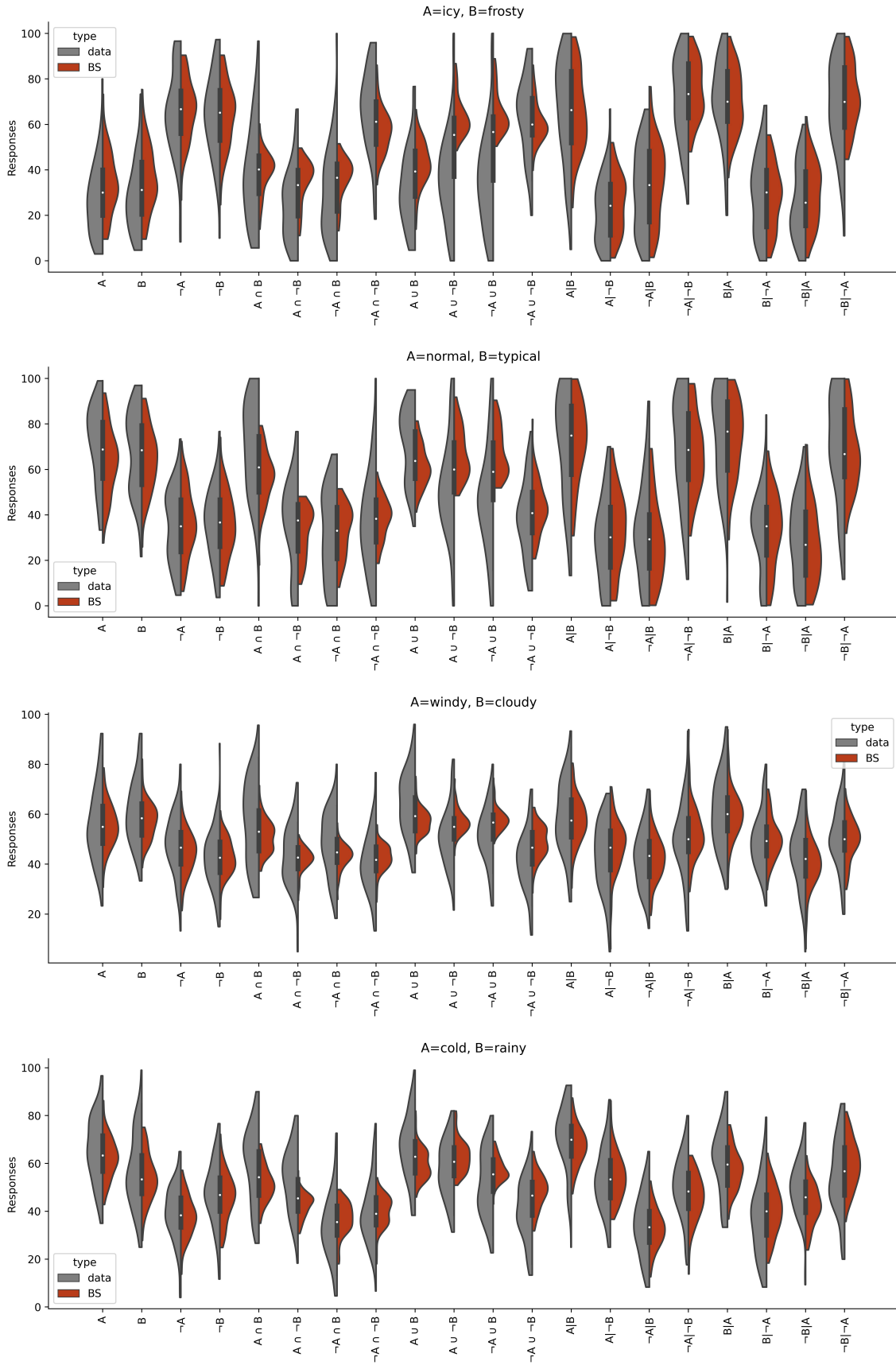


Figure S.8.1. Mean predictions of the Bayesian Sampler and the Quantum Sequential Sampler against empirical results in Zhu et al. (2020).



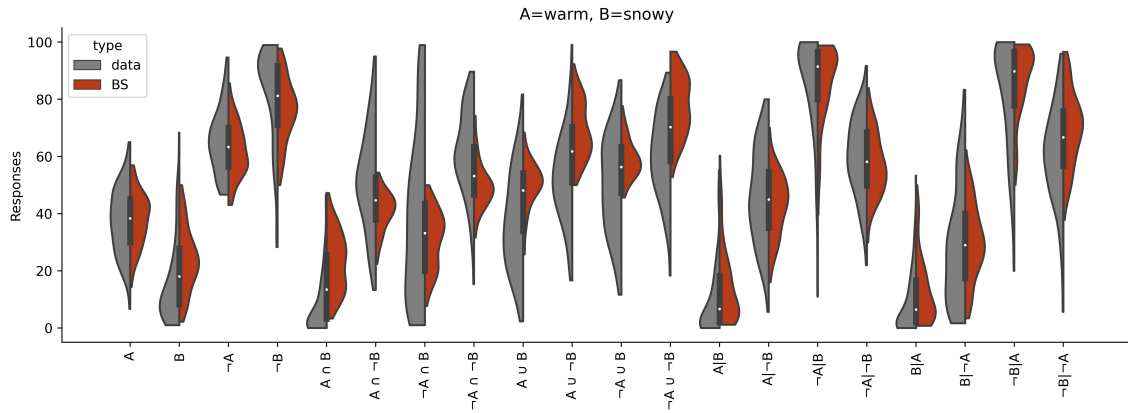
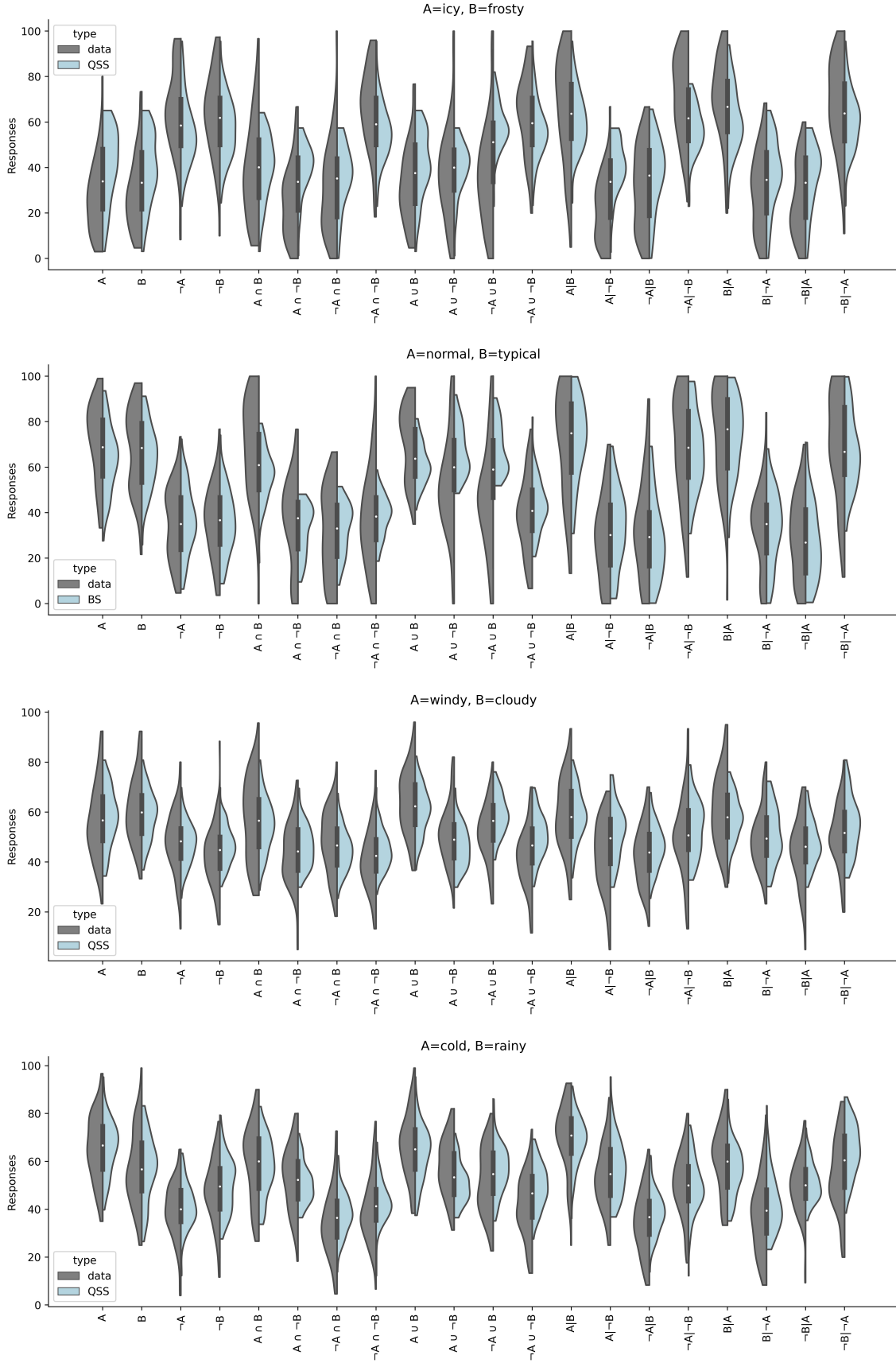


Figure S.8.2. Violin plots showing empirical data of Zhu et al. (2020) vs. the predictions of the Bayesian Sampler.





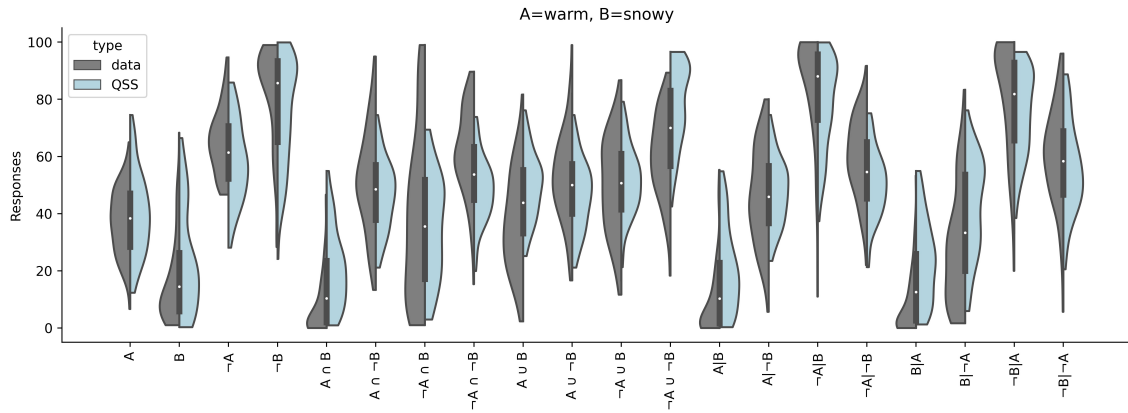
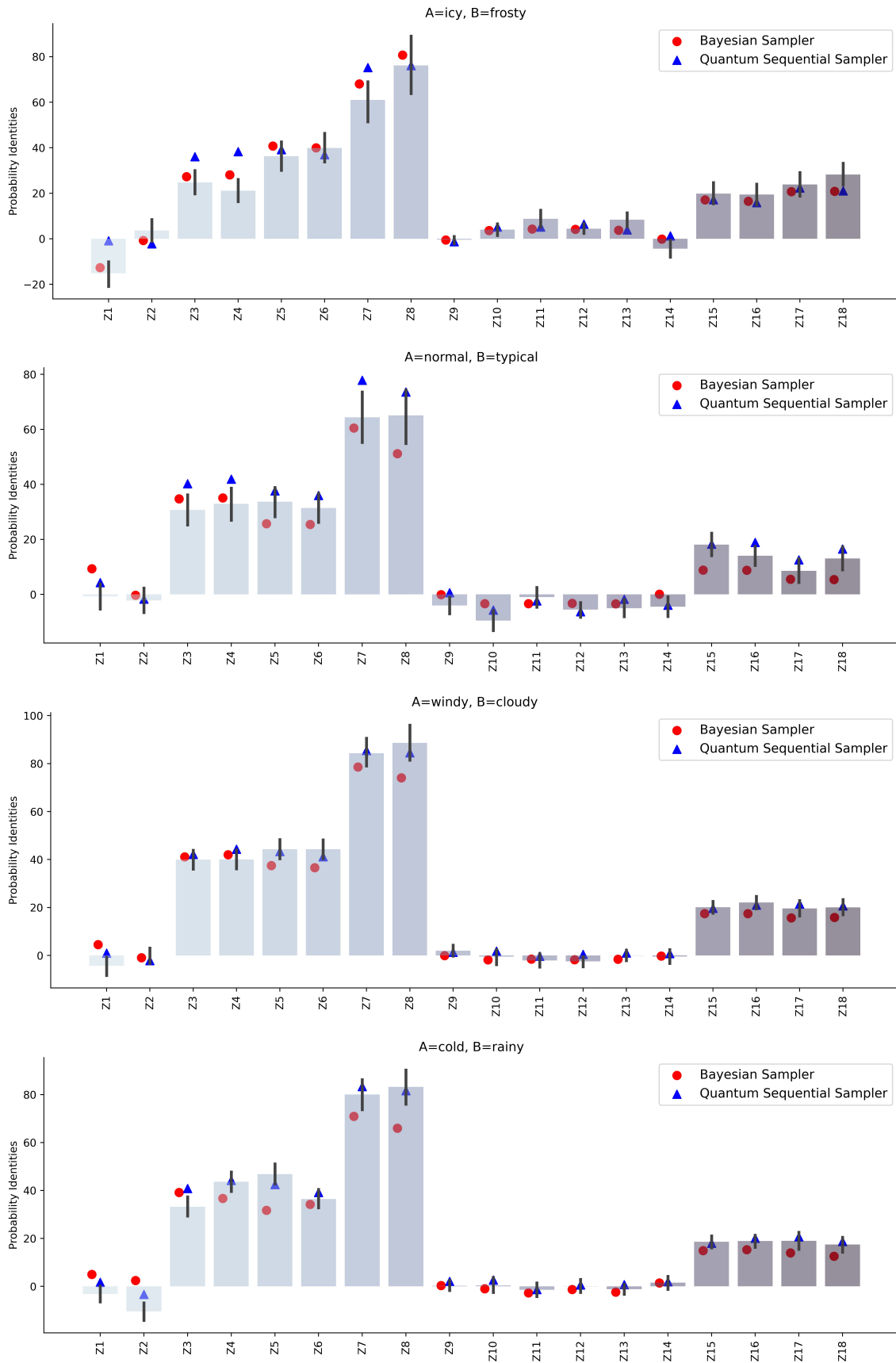


Figure S.8.2. Violin plots showing empirical data of Zhu et al. (2020) vs. the predictions of the Quantum Sequential Sampler.



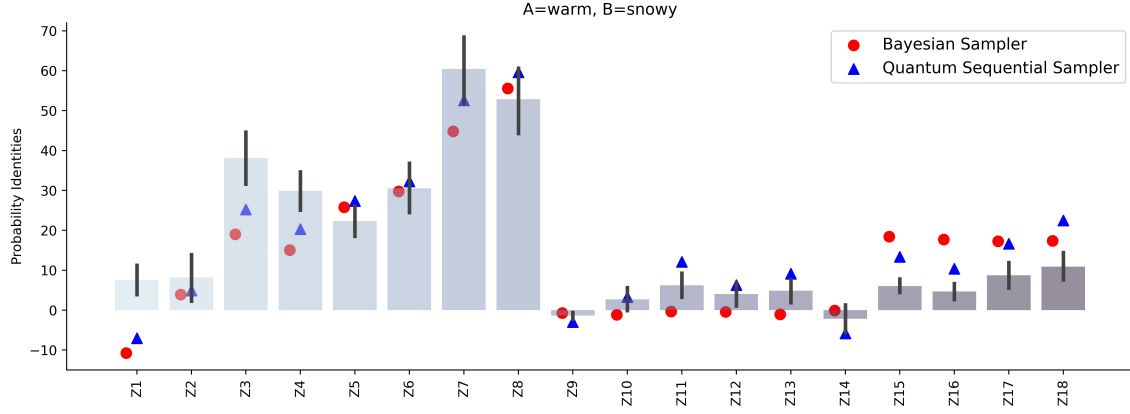


Figure S.8.3. The empirical value of probabilistic identities in Zhu et al. (2020), together with values computed from best-fit predictions, from the Bayesian Sampler and the Quantum Sequential Sampler.