Preregistration for TAO-RRT Project

October 30, 2024

1 Data collection

Have any data been collected for this study already?

No, no data have been collected.

2 Hypothesis

Our key hypothesis is that people better appreciate *within*-model uncertainty than *between*-model uncertainty.

We developed a theory according to which agents need to pick a model m in order to form a belief about a random variable y. Multiple models are possible, but due to a lack of knowledge of all possible models or cognitive constraints, agents only "think through" a finite number of possible models. The beliefs subjects hold average over the beliefs conditional on the models that they do think through.

Our theory makes several related predictions: 1) subjects assign too much probability to perceived-likely events and too little to perceived-unlikely events, are overprecise in that they 2) have a lower subjective variance than the objective variance, and 3) have an even higher mean squared error (MSE) relative to their subjective variance (these two would coincide with correct beliefs). This overprecision comes from the testable assumptions that 4) subjects don't fully appreciate between-model uncertainty (and if they only focus on one model, they fully neglect this uncertainty), but 5) largely do appreciate within-model uncertainty, which also then implies that 6) overprecision rises with between-model uncertainty. Some of our theoretical predictions arise from a "baseline assumption" that subjects all focus on one model, drawn independently from the correct distribution. Our baseline assumption adds the predictions that while subjects 7) will disagree given the same information, the wisdom-of-crowds means 8) the average beliefs will be well-calibrated. The theory also predicts that 9) this interpersonal disagreement rises with between-model uncertainty. Given the connection between overprecision and between-model uncertainty, this implies that 10) overprecision rises with interpersonal disagreement. Finally, if subjects have common information and focus on one model only, our theory predicts that variance overprecision rises 1-to-1 with between-model uncertainty and interpersonal disagreement, and MSE overprecision rises 2-to-1 with between-model uncertainty and interpersonal disagreement.

3 Experimental Design + Conditions

The experimental design presents participants with a series of problems, in each of which they make a prediction about the mean and variance of an outcome.

In each of 15 rounds, each participant sees a scatterplot image representing "sales" data over time. The task is to make a prediction about future sales, using two slider bars: one for the mean and the second for the variance around that mean.

The code that generates the images and selects the images to be shown to each subject appears on the project's website: https://osf.io/q86ev/ The 160 images vary with respect to: a) the trendline, b) noisiness of realizations, c) whether the true trendline is shown on the image, and d) the temporal distance to the prediction point from the sample data.

More specifically, the images participants see are constructed from a set of 40 "problems" each of which has four "conditions." A problem is a pattern of dots, and a baseline and alternative distance for the future prediction (which may be shorter or longer). The condition then selects one of the distances, and whether the trendline is given. We allow for 4 possible conditions: 1) a baseline problem (with no line), 2) the baseline with the alternative distance to the prediction point, 3) the baseline with the line depicted, 4) the baseline with both the line depicted and noise scaled such that the correct variance of the prediction is equal to that under condition 1.

For the main problems, we first randomize which 12 of the 40 problems they will see, and then randomly select one of the 4 conditions for each problem. We also randomize the order in which these images are shown.

Finally, for secondary analysis we select 3 of the 12 images to be shown again. One of these is an image with a line, one with no line and a relatively low distance, and one is with no line and relatively high distance. These "repeats" are randomly interspersed with the other images.

4 Variable Construction

Describe the key variable(s) specifying how they will be measured.

Our subjects will report subjective point estimates and variance estimates. In addition to this we construct a number of variables that are implied by the theory.

- Average subjective variance. This is our main dependent variable for the analysis: the average subjective variance for a particular image. We first take the average raw "slider bar" answer which corresponds to the square root of the standard deviation of the reported distribution, and then convert the average slider bar answer into a variance.

  The reason we aggregate in this way is that we believe that there is reporting error that is additive in the slider bar units (a "trembling hand"). Raising the slider variable to the power of four to compute variance would exacerbate the measurement error when the variance is high. Averaging before we compute the variance mitigates the problem.

- "True" variance of the prediction. Letting $\tilde{x}$ be the time period for which the subject is asked to make a prediction, the answer is:

$$\text{True variance} \equiv \hat{\sigma}^2 \left(1 + \frac{\tilde{x}^2}{\sum_1^n x_i^2}\right)$$

  where we use the unbiased estimator of the variance of the error:

$$\hat{\sigma}^2 \equiv \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-1}$$

NB: the constant in this regression that predicts $\hat{y}$ is fixed at 100, and so estimation of the constant does not contribute to the variance of the prediction.

- Between-model variance of the prediction. This is the portion of the variance that is attributable to uncertainty about the model:

$$\text{Between-model variance} \equiv \hat{\sigma}^2 \frac{\tilde{x}^2}{\sum_1^n x_i^2}$$

- Within-model variance of the prediction. This is the portion of the variance that is attributable to uncertainty over the error term.

$$\text{Within-model Variance} \equiv \hat{\sigma}^2$$

- Error in Variance. This is given by the difference between the subjective variance and the true variance of the prediction.

- Error in MSE. This is given by the difference between the subjective variance and the true mean-squared error.

- Disagreement. This is given by the variance of point predictions across subjects for a given image.

5 Analyses

Specify exactly which analyses you will conduct to examine the main question/hypothesis.

Our main analyses are regressions of the form

$$\overline{\text{Subj. Var.}}_j = \beta_0 + \beta_1 \cdot \text{w/in Var.}_j + \beta_2 \cdot \text{b/w Var.}_j + \varepsilon_j$$

Here j indexes an image, which is a problem-condition pair. The average on the left-hand side (regressand) is taken over all subjects who are exposed to image j.

For Bayesians, $\beta_1 \simeq \beta_2 \simeq 1$. By contrast, our theory predicts that subjects neglect between-model variance, so that $\beta_2 \in [0, 1)$ and, in the extremal case of our theory (where subjects only consider one model), $\beta_2 = 0$.

We will consider several versions of this regression, most of which differ by restricting the sample to subsets of the observations corresponding to the conditions from Section 3 above. In all of our main specifications we include problem fixed effects and cluster standard errors by problem.

For all regressions in this section our directional hypothesis is that $\beta_2 < \beta_1$, and our aggressive prediction is that $\beta_2 = 0$ and $\beta_1 = 1$.

- Our main analysis includes all four conditions. Both β_1 and β_2 are identified.

- Conditions 1 and 2. Restricting to this sample isolates variation due to changes in the distance to the prediction point. That distance manipulates across-model variance but not within model variance, and therefore β_2 is identified but β_1 is not.

- Conditions 1 and 3. Restricting to this sample isolates variation due to the presence of the line. The presence of the line removes across-model variance. Although it does not change the underlying within-model variance, it does change the estimate of within-model variance very slightly. Therefore β_2 is identified and β_1, though formally identified, is likely to be poorly estimated.

- Conditions 3 and 4. Restricting to this sample isolates variation due to changes in the variance of the error in the presence of the line. Because there is no between-model variance in the presence of the line, here we are able to identify β_1 but not β_2.

Our strategy of restricting the sample in order to manipulate the variation used for identification is chosen for expositional simplicity. There are other, more efficient approaches that would not throw away data (e.g., two-stage least squares using treatment indicators as an instrument). We will also report results for these specifications.

We will also report regression models for other outcomes that the model predicts will be affected differently by within- and between-model uncertainty:

$$\overline{\text{Outcome}}_j = \beta_0 + \beta_1 \cdot \text{w/in Var.}_{\cdot j} + \beta_2 \cdot \text{b/w Var.}_{\cdot j} + \varepsilon_j$$

All of these regressions will collapse to the problem-condition level, and we will pool all conditions and include problem fixed effects, with standard errors clustered at the problem level.

The outcomes we will report are:

1. Overprecision, as measured by the true variance for a condition minus the average reported variance (averaged in the slider bar and then converted to a variance). Note that this ends up being redundant with the regressions predicting subjective variance (in the sense that the coefficients on this regression plus the regression predicting subjective variance sum to 1. Still, our directional prediction for this regression is that $\beta_1 > \beta_2$. The aggressive prediction is that $\beta_1 = 0$ and $\beta_2 = 1$.

2. Bias in mean prediction, as measured by the average squared distance between the mean estimate and the true estimate. Our directional prediction for this regression is that $\beta_1 < \beta_2$. The aggressive prediction is that $\beta_1 = 0$ and $\beta_2 = 1$.

3. "Excess MSE", which is the sum of overprecision and bias as different above.[1] Our directional prediction for this regression is that $\beta_1 < \beta_2$. The aggressive prediction is that $\beta_1 = 0$ and $\beta_2 = 2$.

4. Disagreement, as measured by the variance in the mean estimates among all subjects in the condition. Our directional prediction for this regression is that $\beta_1 < \beta_2$. The aggressive prediction is that $\beta_1 = 0$ and $\beta_2 = 1$.

Finally, we estimate two regression models where disagreement (as defined above) is the independent variable:

$$\overline{\text{Outcome}}_j = \gamma_0 + \gamma_1 \cdot \text{Disagreement}_j + \eta_j$$

When the Outcome is overprecision, the directional prediction is that $\gamma_1 > 0$, and the aggressive prediction is that $\gamma_1 = 1$. When we use Excess MSE as the outcome, the directional prediction is

[1] As derived in the theory, this is the difference between the average subjective variance of subjects and their average mean squared error for predictions, two quantities which would coincide for a Bayesian.

that $\gamma_1 > 0$ when we use Error in Variance as the regressand, and $\gamma_1 = 2$ when we use Error in MSE as the regressand.

Also as in the above, we will report estimates for different sample restrictions. Variation in disagreement is, in our model, primarily tied to between-model uncertainty, and so γ_1 will be most cleanly identified from the contrast between, e.g., conditions 1 and 2 or 1 and 3.

6 Outliers and Exclusions

We plan to exclude:

- Duplicate responses from the same IP addresses or worker ID.
- Respondents who never (across all rounds) move the variance slider off its origin at zero variance.
- Respondents who used a mobile device (identified by having a horizontal pixel count below 1000).
- Inattentive survey respondents who report outlier answers for the perceived mean on problems for which the mean is fully determined by the problem (our "with line" treatment). Outliers are defined by 3 times the interquartile range for deviations from the mean across all problems.
- Respondents who failed to complete all required questions.
- Respondents who answered less than 4 of our 6 comprehension check items correctly.
- Respondents who complete the survey in less than 5 minutes.

We hope that these criteria will exclude less than 20% of participants. We will redo all analyses with the full sample and report it in an appendix or supplement.

As a robustness check we will also rerun the analysis on the following subsample:

- Excluding images where the true 95% confidence interval lies outside of $[60, 140]$.

7 Sample Size

We will collect 600 participants from Prolific and then exclude those who run afoul of the above exclusion criteria. (This will leave us with less than 600.)

8 Other

Anything else you would like to preregister? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

We will maintain a statistical significance threshold of p=.005 (Benjamin et al., 2018). Between .005 and .05, p-values will qualify as marginally significant. We will register the survey and the stimuli (the 160 images) at the same time as this preanalysis plan (https://osf.io/q86ev/). To see the survey from the perspective of a participant, follow this link: https://berkeley.qualtrics.com/jfe/form/SV_8

9 Name

Give a title for this preregistration
TAO-RRT

10 Additional planned analyses?

- We will test the degree to which disagreement between individuals captures between-model uncertainty by regressing the error in the variance estimate on disagreement between individuals' best-guess forecasts; our theory predicts a regression weight of 1, in contrast to the normative baseline of 0.

- We will test the degree to which disagreement between individuals captures between-model uncertainty by regressing the error in the mean squared error on disagreement between individuals' best-guess forecasts; our theory predicts a regression weight of 2, in contrast to the normative baseline of 0.

- We will replicate our key hypothesis test in a purely between-subjects analysis using only data from the first trial, in order to assess the degree to which our effects depend on individuals' ability to see the contrasts between different experimental treatments.

- We will examine changes with practice: do participants become better-calibrated (reporting uncertainties closer to normative accuracy) over the course of the 15 trials?

- We will examine heterogeneous treatment effects with respect to overprecision. What is the distribution (across people) in weightings on the true between-model variation? Are there some subjects who more consistently neglect between-model variation than others?

- We will examine the differences subjects report in the repeated problem to see how much of this is driven by "trembling hand" response error, and how much may be driven by "drawing different models" in no-line problems. Our theory makes no clear prediction about whether the same individual should draw the same model upon observing the same problem later, which is why we place this under exploratory analysis.