Title

Extracting Statistical Relationships from Observational Data: Predicting with Full or Partial Information

Authors

Fréchette, Guillaume R

Vespa, Emanuel

Yuksel, Sevgi

Publication Date

2025-05-01

DOI

10.1257/pandp.20251108

Extracting Statistical Relationships from Observational Data: Predicting with Full or Partial Information*

Guillaume R. Fréchette Emanuel Vespa Sevgi Yuksel

Decision makers often rely on past observations to learn about the statistical relationships between different variables. For example, history informs voters' understanding of which type of policies and candidates generate desirable economic or social outcomes. Similarly, but in a very different context, hiring managers learn from their past experiences which observable characteristics of a job candidate are predictive of their productivity. An observation that motivates this paper is that to make predictions in settings like these, agents will have to adjust how they aggregate past information depending on the observables that are available at the moment the prediction is made.

To illustrate the problem, consider a hiring manager who is tasked with predicting whether a job candidate is a good fit for his firm. They have observed characteristics of previous candidates—for instance, experience and education level—and whether they have been successful on the job. As a first case, suppose that the manager observes both experience and education level for the potential candidate. Given the set of observations (which

*We would like to thank numerous seminar participants for helpful comments. We thank Jay Cizeski, Jimena Galindo, Yiyuan Hu, Jie Liao, and Dingyue Liu for outstanding research assistance. This research was supported by the National Science Foundation under Grants 2315663, 2315664, 2315665, 2437062. We are responsible for all errors. Fréchette: New York University, Department of Economics, gf35@nyu.edu. Vespa: UC San Diego, Department of Economics, evespa@ucsd.edu. Yuksel: New York University, Department of Economics, sevgiy@gmail.com.

we will refer to as data) the manager will find it useful to understand how experience and education *jointly* predict success. As an alternative case, assume that the manager only observes experience. Having partial information on observables at the prediction stage means that now the manager would find it useful to aggregate across education levels and use the *marginal* distribution for experience. Prior to the prediction stage, however, the agent may not know whether they will have full or partial information and when learning from data, both types of inference are potentially valuable.

However, it is not clear from an ex-ante perspective which type of inference might be behaviorally more difficult or to which people attend more readily. On the one hand, it seems easier to learn about how each variable separately predicts success relative to learning about how the two variables jointly predict success: the former simply requires identifying the unconditional correlation between the observable variable and the outcome of interest. On the other hand, if the data include information on both variables, it might be more straightforward for the manager to look for patterns in terms of how these two variables jointly predict the outcome of interest. Regardless of which approach the agent may take to learn from data, adjustments may be necessary at the prediction stage. For instance, suppose that the manager forms an understanding of the environment based on how observables operate jointly, but that only one of the observables is available at the prediction stage. Retrieving an unconditional correlation in this case requires aggregation. For example, to develop an understanding of how experience alone predicts success, the manager would need to aggregate over all education levels in the data and retrieve the marginal distribution for experience. In other words, given a specific way in which agents may have learned from the dataset, adjustments needed to accommodate the observables that are available at the prediction stage may introduce difficulties.

In this paper, we report results from a laboratory experiment that has the following general structure, which, despite its abstract framing, closely resembles the example of the above manager. In the first part of the experiment, participants are presented with data

sets. Each line of data consists of information on the realization of three binary variables. These variables are described as the status of lights and sound of a hypothetical machine; but, in the context of the previous example, can be interpreted as the level of experience, education, and success of each previous candidate encountered by the manager.

In the second part of the experiment, participants are presented with information from one line of data from the hypothetical machine (i.e. they know the status of some of the variables) and they have to predict a target variable (whether the machine makes a sound). For some predictions participants observe the status of both lights (full information on observables) and for other predictions they only see the status of one variable (partial information on observables). When participants face the data sets they are aware that they will later face a prediction task and they know that the data itself will not be available to them at that stage. However, while they inspect each data set, participants can take notes with the understanding that such notes will be available to them at the prediction stage.

Results presented in this paper document how predictions about the target variable change—for a given data set—when there is (at the prediction stage) partial or full information about the status of the other variables (the participants are shown all observable lights or only some). We focus on three measures that capture key aspects of behavior.

First, we focus on the optimality of the predictions. Formally, we look at whether a prediction maximizes accuracy (i.e., the probability with which the guess is correct) given the available information on the status of other variables and given the true correlation structure (as reflected in the data made available to the participants). We document that predictions are less optimal (by six percentage points, p-value < 0.01) when there is partial information on the status of other variables relative to the full information case. That is, when learning from a data set consisting of three variables, we find people to be better at making optimal predictions about one variable when they are shown the status of the remaining two variables in contrast to being shown the status of only one variable. This suggests that the way in which participants summarize the properties of a data set makes it

relatively more challenging to make predictions the require using unconditional correlations. If we take as a reference that learning from the data set is done jointly for both variables, this suggests that aggregating across one variable (the unobserved one) to make predictions based on observing only the other is challenging.

Next, we study (in)consistency of behavior within and across subjects. Specifically, we compute disagreement rates: (i) the likelihood that a participant's prediction is consistent with their other predictions when given the exact same information;[1] and (ii) the likelihood that a participant's prediction is consistent with the predictions of others when given the exact same information. Within participants, we find that their predictions are not more inconsistent with partial relative to full information. However, across participants, the evidence indicates that disagreement increases with partial information.

This paper is related to a recent theoretical literature in economics that studies agents' possible misperceptions and how these impact what they learn.[2] In turn, theoretical work has motivated recent experimental studies.[3] Most of the recent experimental work focuses on participants who have access to data from the relevant environment and a narrative that provides an interpretation. The narrative may introduce a biased interpretation of the data. Some of the questions involve understanding how the narrative affects participants' reading of the data and choices that they subsequently make. In the main project behind this paper, Fréchette, Vespa & Yuksel (2024) (henceforth FVY), we experimentally study the types of mental models people form by learning from a set of observations; namely, we study the kinds of inferences individuals make regarding the statistical relationships among variables from examining data. FVY does not provide participants with narratives, but aims to understand the mental models that participants form when they only have access to a data set.[4] Complementing results from FVY, in this paper, we study whether the amount

[1] For each data set, subjects predict the sound multiple times for a given light configuration.

[2] For further detail, see e.g., Eliaz & Spiegler (2020) Schwartzstein & Sunderam (2021), and Aina (2023).

[3] For instance, see Charles & Kendall (2024), Kendall & Oprea (2024), Ambuehl & Thysen (2024), Barron & Fries (2024), and Aina & Schneider (2025).

[4] FVY also provides a more detailed description of its connection to other papers in the literature. There

of information that participants have at the prediction stage impacts people's ability to effectively use the statistical relationship between variables in the data sets.

Learning Task

The task and data come from one treatment of the experiment presented in FVY.[5] We refer the reader to FVY for details associated with the design. Here, we provide an outline.

Each session consists of two parts. In Part 1, participants are presented with 11 data sets, one at a time, and have the ability to type notes-to-self using the computer. In Part 2, participants do not have access to the data sets, and can only observe the notes-to-self they have taken earlier. For each data set they observe in Part 1, participants make predictions.

Specifically, each data set is framed as being generated by a different *machine*. Each line in the data set corresponds to a *trial*, an incidence of the machine's operation, and records three binary variables corresponding to three observable features of the machine: status of a blue and a red light (each are on or off) and whether or not the machine makes a sound. The relation between the visible lights and the sound is probabilistic and they see multiple trials with the same set of lights on or off. In Part 1 of the experiment, participants observe 27 trials corresponding to each machine. In Part 2 of the experiment, participants make 36 predictions about each machine; specifically, they are asked to predict whether the machine makes a sound after potentially observing some information about the status of the lights. At the end of the experiment, one prediction is chosen randomly for payment and the participant receives a bonus of $25 if their prediction is correct.

The data used in this paper come from the *Unspecified Prediction* treatment of FVY, where the participants were *not* told before Part 2 that they would make predictions about whether the machine would make sound.[6] The instructions simply informed participants

is a related literature in psychology that is further described there as well.

[5]The data associated with FVY will be available at https://doi.org/XXX/XXX.

[6]Four sessions were conducted for this treatment at the EconLab in UC San Diego. The data includes 70

that they would be making predictions about some feature of the machine after observing full, partial, or no information about the status of the other features of the machine. This treatment was designed specially with the goal of comparing predictions on the same variable (the sound) with full, partial, and no information about other variables (status of the lights).

Results

Our analysis focuses on the 10 machines (each machine refers to a data set) where the lights are indeed predictive of the sound; namely, the optimal prediction about the sound is a function of the status of the lights. We use the term *full information* to refer to prediction questions in which subjects make guesses about the sound knowing the status of both lights (whether the blue and red lights are on). By contrast, we use the term *partial information* to refer to prediction questions in which subjects make guesses about the sound knowing the status of only one light, i.e. they learn about whether the red light is on or the blue light is on and they are not informed about the status of the other light.

Our first result examines the optimality of predictions. Formally, a prediction is classified as *optimal* if it maximizes accuracy based on the available information about the status of other variables, given the true correlation structure (as reflected in the data provided to participants). To be more concrete, participants are making predictions about whether each machine makes a sound after observing some information about the status of the lights. Consider the full-information case, where the status of both lights is revealed, and both are on. Assume the joint distribution of these three variables (the two lights and the sound) is such that the frequency of the sound being on, conditional on both lights being on, is 80 percent. The optimal prediction is to guess that the sound will be on, as this maximizes the probability of an accurate guess, and consequently, the participant's likelihood of winning the bonus. Note that a 100 percent optimality rate is achievable with either full or partial

subjects. The experiment was conducting using otree (Chen et al. 2016).

Table 1: Average Behavior With Full and Partial Information

	Frequency of Optimal Prediction	Disagreement Rate (within-subject)	Disagreement Rate (across-subjects)
Full Information	81.5	15.0	29.3
Partial Information	75.5	11.0	36.6

Notes: See notes for Figure 1 for more information. All differences between full and partial information are significant (p-value < 0.01) using OLS regression with controls for each data set and subject-level clustering.

information about the lights. This contrasts with accuracy rates, which inherently depend on the level of information provided—that is, what has been revealed about the lights.

The first column of Table 1 reports the average frequency of optimal prediction with full and partial information on the status of the lights. There is a 6-percentage-point decline in optimality with partial information relative to full information. The left panel of Figure 1 plots the cumulative distribution of optimality rates (computed at the subject level) contrasting full with partial information. Note that the distributions are not symmetric around 0.5 suggesting that most participants are making use of the lights (in a somewhat optimal way) to guide their predictions, i.e., they are doing better than random. However, the distributions are not concentrated around 100 percent. There are substantial deviations from optimal behavior and more so with partial information than full information. The distribution with full information first-order stochastically dominates that with partial information.[7] Overall, this is clear evidence that participants are much better at making use of the information on the lights to guide their predictions on the sound when they are presented with the status of both lights relative to only one light.

Next, we investigate how consistent individual participants are in their own predictions. The second column of Table 1 presents the average rate of disagreement at the individual level, comparing the cases of full and partial information. Specifically, we calculate the rate at which a participant's predictions differ from their other predictions when given identical

[7]It is worth emphasizing that deviations from optimality are costly. On average, guessing accuracy is at least 10 percentage points lower than the best achievable with full or partial information.

Figure 1: Optimality and Disagreement with Full and Partial Information

Notes: In the left panel, each observation corresponds to the frequency with which a subject's guesses were optimal given the available information on lights. In the middle panel, each observation corresponds to the rate at which a subject's own guesses disagreed after observing the same configuration of lights. In the right panel, each observation corresponds to the rate at which a subject's own guesses disagreed with those of others after observing the same configuration of lights.

information (i.e., the same status of the lights for the same machine).[8] The disagreement rate decreases by four percentage points when moving from full to partial information. The cumulative distribution plotted in the middle panel of Figure 1 provides further information about how the consistency of predictions at the individual level changes with the level of information provided to them about the status of the lights. In general, the evidence indicates that participants' predictions, on average, are more consistent with partial information.

Finally, we study the consistency of the predictions across participants. Formally, we compute the likelihood that two participants disagree in their predictions when they are given identical information (i.e., presented with the same status of the lights for the same machine).[9] The last column of Table 1 reports the average rate of disagreement contrasting full and partial information. Similarly, the right panel of Figure 1 plots the cumulative distribution of disagreement rates between subjects separately for the full and partial information cases. The evidence clearly shows that participants are more likely to disagree on what constitutes optimal behavior when they are shown partial information about the lights relative to full information. This is not just reflected in average values (Table 1), but also seen from the fact that the distribution associated with partial information first-order stochastic dominates that associated with full information (Figure 1).

[8]To demonstrate how the measure is constructed, consider a simpler example in which a participant makes eight predictions about a machine with full information: five after learning both lights to be on and three after learning both lights to be off. Assume that the participant predicts the sound to be on three out of five times in the first case and none of the three times in the later case. The disagreement rate of this participant for this machine (data set) would be computed as $\frac{5}{8}\left(\frac{3}{5}\times\frac{2}{5}+\frac{2}{5}\times\frac{3}{5}\right)+\frac{3}{8}(0)=\frac{3}{20}$.

[9]Consider two participants: for a given machine, one predicts the sound to be on in all decisions where both lights are revealed to be on, the other predicts the sound to be on for only half of these decisions. Their disagreement rate for this machine and light configuration is 50 percent. Their disagreement rate for this machine is computed by averaging over all light configurations (weighted by the frequency with which each configuration is observed).

Conclusion and Discussion

Our results can be summarized and interpreted jointly as follows.

First, we find that participants are better at correctly identifying statistical patterns when they have full, rather than partial, information. Specifically, their predictions are more consistent with the underlying data generating process when they are shown both lights. In contrast, they struggle to grasp or identify the unconditional relationship between a single variable and the dependent variable. This is consistent with subjects having difficulties aggregating across different realizations of the uncontrolled variable.

However, difficulties in learning unconditional distributions from data do not decrease the consistency of behavior at the individual level. As reported above, participants are as likely (even more likely) to make self-consistent predictions in these cases. This suggests that they are similarly relying on information they have extracted from the data sets given to them (on how different variables statistically related to each other) in guiding their predictions with full or partial information. What appears to change is the likelihood with which these patterns are optimal, that is, match the true correlation structure in the data.

Finally, as a consequence of the first two findings, participants exhibit greater disagreement in their predictions with partial information. Specifically, even when all participants are given identical data sets, they are more likely to draw conflicting conclusions about how the observable variables predict the outcome of interest. Understanding how disagreement arises from varying interpretations of the same information is important for explaining phenomena such as ideological polarization or excessive trading in financial markets. The results of this study suggest that the level of aggregation required in learning from data may plays a significant role in shaping such disagreement.

References

Aina, C. (2023), 'Tailored stories', *Working Paper* .

Aina, C. & Schneider, F. H. (2025), 'Weighting competing models', *Working Paper* .

Ambuehl, S. & Thysen, H. C. (2024), 'Choosing between causal interpretations: An experimental study', *Working Paper* .

Barron, K. & Fries, T. (2024), 'Narrative persuasion', *Working Paper* .

Charles, C. & Kendall, C. (2024), 'Causal narratives', *Working Paper* .

Chen, D. L., Schonger, M. & Wickens, C. (2016), 'otree an open-source platform for laboratory, online, and field experiments', *Journal of Behavioral and Experimental Finance* **9**, 88–97.

Eliaz, K. & Spiegler, R. (2020), 'A model of competing narratives', *American Economic Review* **110**(12), 3786–3816.

Fréchette, G. R., Vespa, E. & Yuksel, S. (2024), 'Extracting models from data sets: An experiment', *Working Paper* .

Kendall, C. W. & Oprea, R. (2024), 'On the complexity of forming mental models', *Quantitative Economics* **15**, 175–211.

Schwartzstein, J. & Sunderam, A. (2021), 'Using models to persuade', *American Economic Review* **111**(1), 276–323.