Evidence for Optimal Human Estimates of Event Duration in Observed Real-World Contexts

Biocomplexity Institute; Metaculus; VDH

TBA

Abstract

Your abstract here.

Evidence for Optimal Human Estimates of Event Duration in Observed Real-World Contexts

Introduction

Human judgments of future events, what we will call forecasts, are ubiquitous and span the mundane (e.g., will the car in front of me stop suddenly) to the extraordinary–e.g., decisions made during military operation. The scientific study of human forecasting, fittingly, draws upon myriad domains and literature. The focus of the current report is what is called forecasting of event duration, a specific domain that is defined by forecasting how much longer an event will continue given the point in time for which the forecast is called. It is similar to being asked: "We've been at war for 2 months; how much longer will the war last?"

[DESCRIBE THE GOTTs formulation of the generic, uninformed prior to orient readers; and orient them to the Bayesian formulation of and solution to the generic problem; define t and t_{total}, etc.][1]

Over the past decade, we've accumulated evidence that (i) humans can estimate t_{total} in an optimal way given little evidence and (ii) prior knowledge is combined with new, observed data in an optimal way, (iii) non-normative, heuristic models do not account for the full range of experimental data (see Griffiths and Tenenbaum (2011))[2].

The literature, however, is incomplete in its account of how humans adapt in the face of changing information in real-world domains. In such cases, incremental updating of a prior via further information may not suffice because other issues may arise (e.g., error in situational assessment, rapid changes in the situation). We can imagine real-world situations with such characteristics–x, y, z. Our interest is in a difficult case:

[1] The Bayesian formulation of the general class of mechanisms of social causal attribution was raised in the 1970s and supported a host of experimental findings. This work was inspired by the somewhat rich history of Bayesian models of perception and basic information processing in cognitive psychology from the 1960s.[PUT REFS in here]

[2] The cognitive and neurobiological mechanisms, beyond the normative Bayes approach, are not necessarily well aligned with the normative account but see (ACT-R paper X).

dynamic informational inputs (to the decision making model) with little immediate environmental feedback in respect to the accuracy of the prediction.

In this article, we explore event duration prediction in the case of real-world epidemics. This kind of natural system, characterized by waves in series seems to have the potential to mask both t and the prior in respect to estimating event duration. Specifically, the setting of the current work are the waves of the COVID-19 delta and omicron variants, two waves that came in close succession. We observed four rounds of an online prediction tournament focused on characteristics of the COVID-19 epidemic in Virginia from late 2021 to early 2022. From these observations, we constructed a mathematical model of the decision process that, at its heart, represents the optimal Bayes prediction, given t and $prior$.

Because our model required two inputs, t and $prior$, we fixed t from epidemiological date and focused on estimating the prior. This decision has theoretical support in the literature [Cite]. Further, given the setting of the prediction tournament that we observed, it was clear that the focus of the forecasters was on omicron (we will show evidence of this in the results); for the early judgments in November 2021, the event predictions suggested that forecasters were estimating the beginning of omicron wave (with some uncertainty). Once it seemed clear to forecasters that the omicron wave was underway, we used a reasonable point in the epidemic curve (7-day average of daily cases) as t_0.

We address the following research questions:

I. Is there evidence that Ss make optimal estimates of t_{total} in real-world, dynamic environments?

II. Is there evidence for dynamic optimality (shifts in the prior)?

III. Is there evidence that external information drives dynamic optimality?

Methods

ADD ACCURACY ANALYSIS AND DATA STRUCTURE

Data Structure

These data were extracted from the first four rounds of an online prediction tournament run by Metaculus and using the Metaculus forecasters pool as the participants. The dates available for forecasting per rounds 1-4 were, respectively: 11-12-2021 to 12-03-2021; 12-03-2021 to 12-24-2021; 12-24-2021 to 01-14-2022; and 01-14-2022 to 02-04-2022. The prediction window for the forecasts per rounds 1-4 were, respectively:

Forecasts were at the day resolution.

The Metaculus forecasters pool....[describe here]

[**NOTE: Good place for a graphic of the task.**].

Forecasting Interface

Pre-processing of Forecasts

The design of the forecasting tournament necessitated some pre-processing of these data. First, we checked the degree to which forecasters were using the end-points of a round's prediction window, what we dub end-point forecasts; left end-point forecasts were at the earliest point of the prediction window; right end-point forecasts were at the latest point.

Second, the punctuated point at 1/14...need a test of other questions for that point to convince reader that it is not just new roundness.

Third, the lack of end-point forecasts, is it driven by points trade-off? NOT SURE IF WE CAN ADDRESS THIS POTENTIAL ISSUE

Results

Overview of Key Finding

Figure 1 shows primary aggregated result. See Caption for description.

Figure 2 compares the deltas for the case-count curve and the 'human t_{total}'. The idea here is to explore the degree to which the case counts are driving the human judgments. See Caption for description.

Figure 3 shows a parallel result to Figure 1 but for the most active subjects.

Figure 4 provides a parallel result to Figure 2 but for individual subjects (the most active).

Discussion

NOTE: For Discussion The mechanism of punctuated prior switching (or t switching), may be driven as a meta event duration model or a hierarchical event duration model.

References

Griffiths, T. L., & Tenenbaum, J. B. (2011). Predicting the future as bayesian inference: People combine prior knowledge with observations when estimating duration and extent. *Journal of Experimental Psychology: General*, *140*(4), 725–743. doi: 10.1037/a0024899

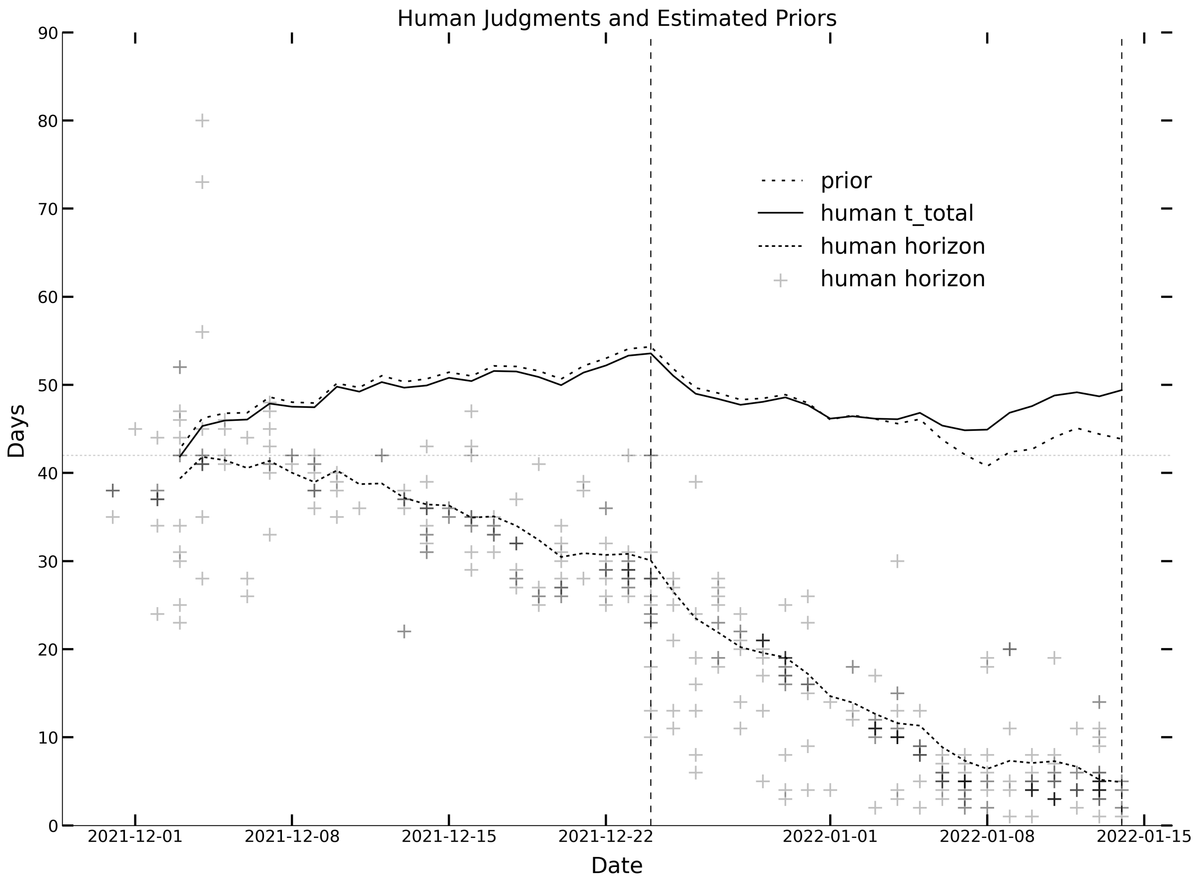

Figure 1. This figure depicts three components: 'Priors', 'human t_{total}' and 'human horizon'. Before defining these measures, we have the following preliminaries. The analysis represented depends on an assumed value of t_0 (the beginning of the epidemic wave); we assumed t_0 to be 11-30-21 (this was generated from the 7-day moving average of daily case-counts in Virginia). We also define upfront 'judgment date' as the date on which a human judgment was made and 'prediction date' as actual predicted date of the event in question (e.g., 'I predict that the hottest day in May will be May 15th.' May 15th would be the prediction date). From these preliminaries, we can define our measures. The 'human horizon' reflects the observed human judgments ('human horizon', +) between 11-30-21 and 1-14-22 across all valid subjects (note, the darker + symbols reflect multiple observations conjunctive in date and value). This measure was defined as 'prediction date' minus 'judgment date'; it is called horizon to capture the notion of how far out is the prediction from the time of a judgment. The second component, 'human t_{total}' was defined as 'prediction date' minus t_0 and reflects the human's sense of the event's duration. The third component, 'prior', was constructed using a Bayesian model of event duration. This model expects t_0 and t_{total} as arguments and computes a posterior to reflect the Bayes optimal estimate of event duration (note: this is not a learning model, at least to date, so the posterior does not become integrated into the existing prior [no Bayes updating]). The measure 'prior' illustrates the best prior given the 'human t_{total}.' and thus captures a theoretical mental construct

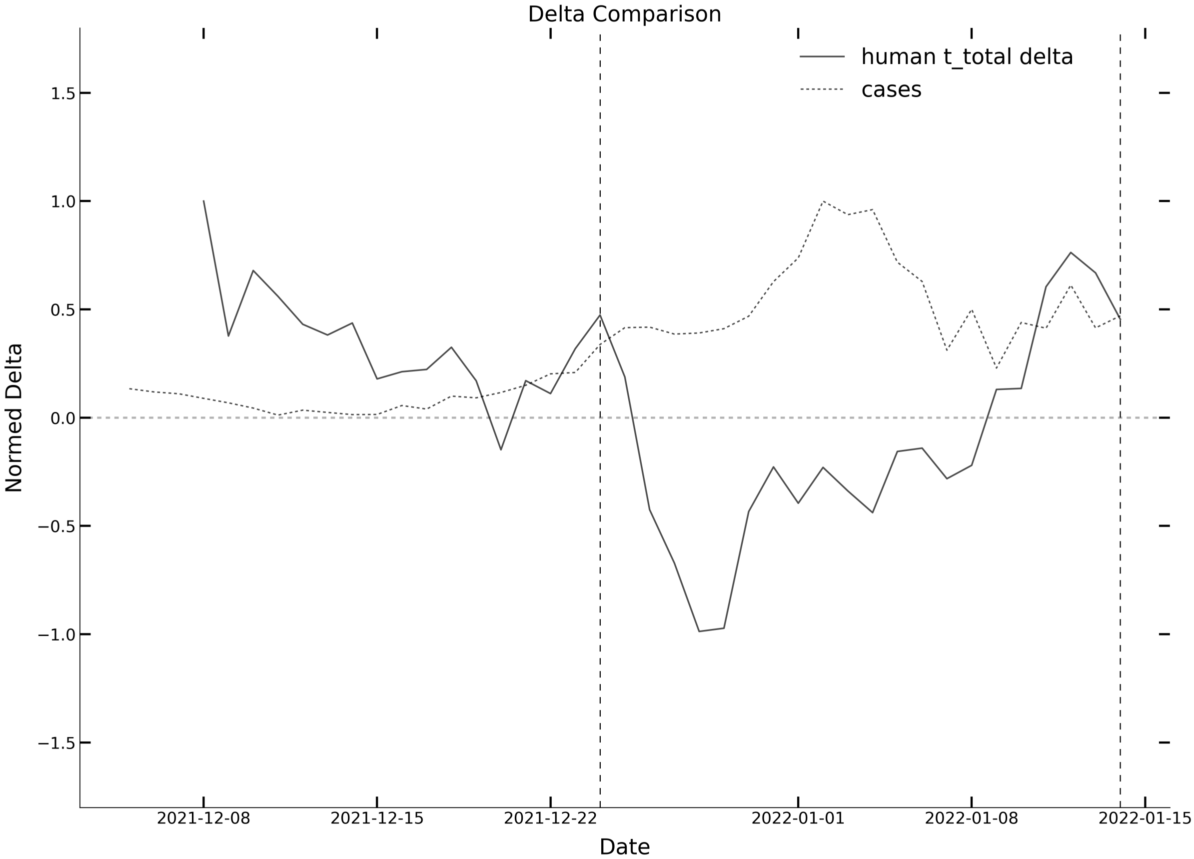

***Figure 2*.** Here we see the delta (defined over 5 day period) for the VA case counts and the 'human t_total.' See Figure caption 1 for definitions of measures.

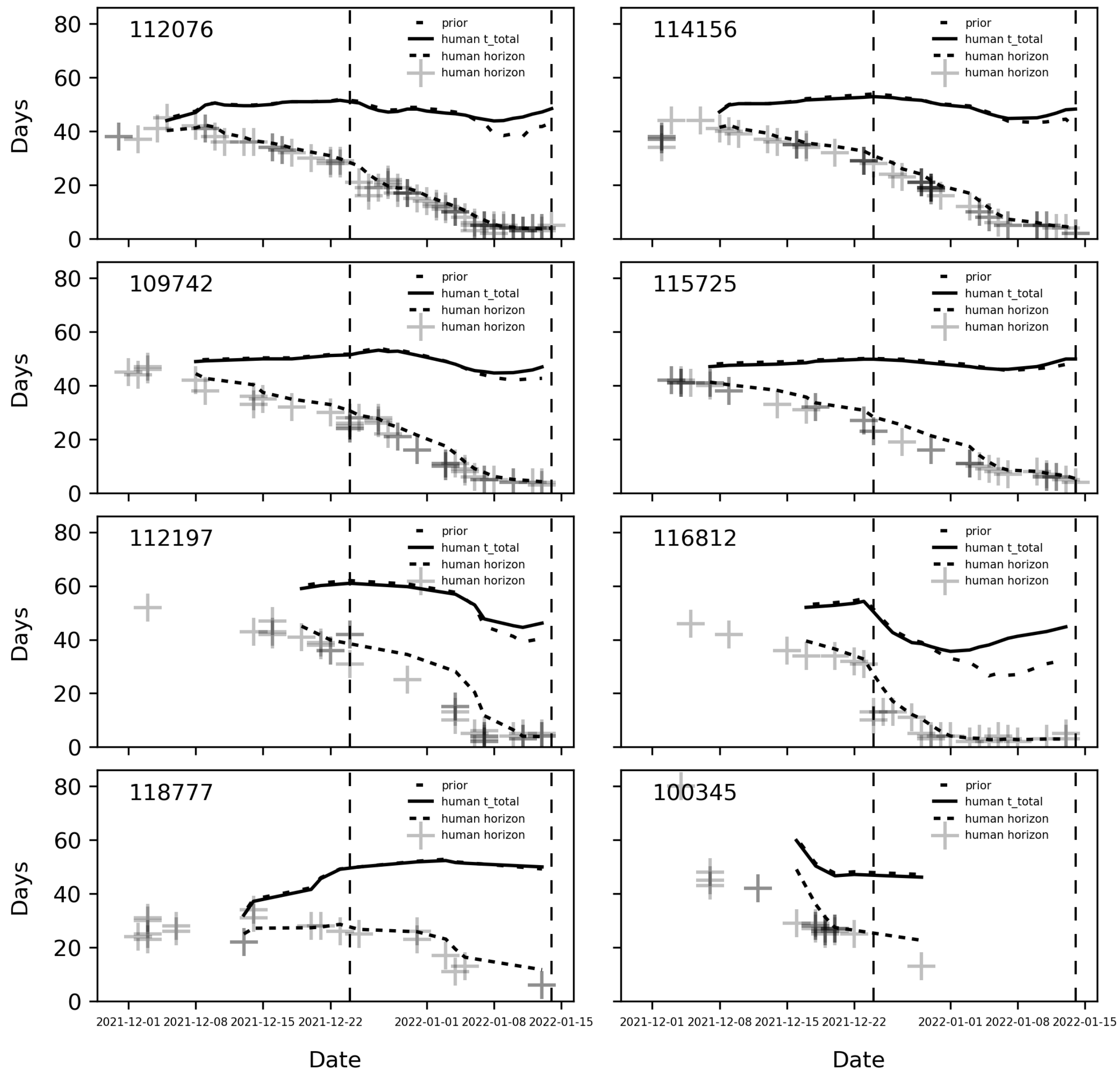

Figure 3. This figure depicts the same three components as in Figure 1 for the eight subjects with the highest response rate (>15 responses). Subject number is in the top-left corner of each panel.

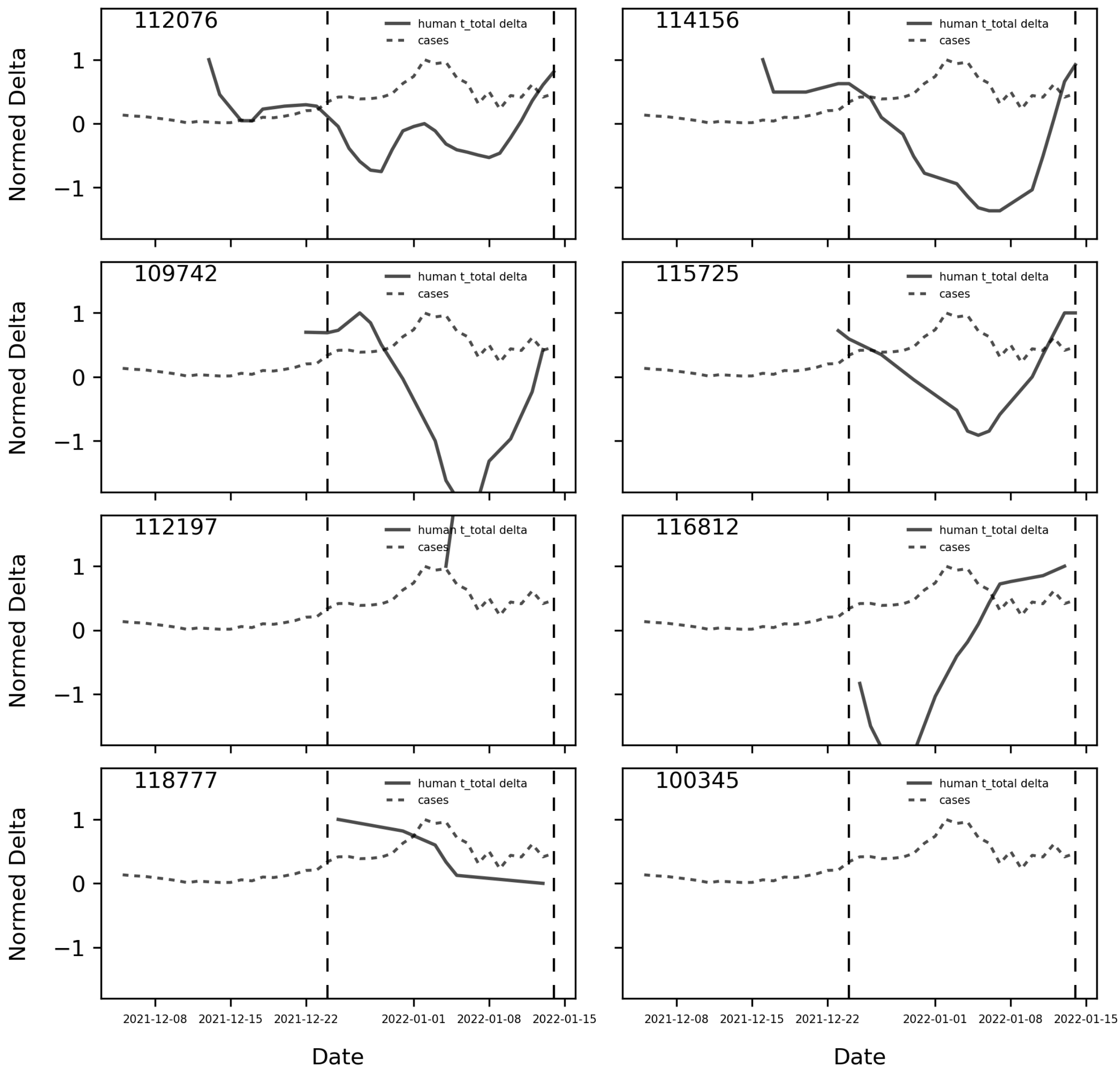

Figure 4. This figure depicts the delta analysis as shown in Figure 2 for the eight subjects with the highest response rate (>15 responses). Subject number is in the top-left corner of each panel.

Appendix

APPENIX STUFF

Hi I'm your Appendix