Memory & Cognition (2022) 50:1363–1380
https://doi.org/10.3758/s13421-022-01288-0

# How well do ordinary Americans forecast the growth of COVID-19?

Madison Fansher[1] · Tyler J. Adkins[1] · Richard L. Lewis[1,2,3] · Aysecan Boduroglu[4] · Poortata Lalwani[1] · Madelyn Quirk[1] · Priti Shah[1] · John Jonides[1]

Accepted: 5 February 2022 / Published online: 28 March 2022


**Abstract**
Across three experiments (N = 1565), we investigated how forecasts about the spread of COVID 19 are impacted by data trends, and whether patterns of misestimation predict adherence to social-distancing guidelines. We also investigated how mode of data presentation influences forecasting of future cases by showing participants data on the number of COVID-19 cases from a 5-week period in either graphical, tabular, or text-only form. We consistently found that people shown tables produced more accurate forecasts compared to people shown line-graphs of the same data; yet people shown line-graphs were more confident in their estimates. These findings suggest that graphs engender false-confidence in the accuracy of forecasts, that people's forecasts of future cases have important implications for their attitudes concerning social distancing, and that tables may be better than graphs for informing the public about the trajectory of COVID-19.

**Keywords** Forecasting · COVID-19 · Data visualization

## Introduction

Understanding the growth patterns of COVID-19 has been a critical task for US citizens in the last year, in that making well-informed, safe decisions depends on the rise and fall of cases. Much of the data presented on the spread of COVID-19 has been through various graphs presented by media outlets and health agencies (Zacks & Franconeri, 2020a, 2020b). Line graphs, bar graphs, tables, and heatmap data visualizations are popular methods for displaying the number of COVID-19 cases and deaths. Use of such visualizations allows for the communication of extensive information that would be difficult or impossible to illustrate with text alone. Due to the dynamic nature of the current COVID-19 outbreak, it has been argued that forecasting the future of the disease with accuracy is uniquely difficult (Makridakis et al., 2020) and research from the first week of the pandemic (11-16 March 2020) showed that individuals significantly underestimated their personal risk compared to that of the average American, average person in their state, and their neighborhood (Wise et al., 2020).

One influential variable may be a fundamental misunderstanding of the rate of growth of an exponential function as it relates to disease incidence. Consider this one problem: As of 1 March 2020, the Centers for Disease Control (CDC) reported that there had been 75 cases of COVID-19 and one death in the USA. On 18 March, there were 7,038 confirmed cases and 97 deaths due to the disease. How many cases and deaths would there be on 25 March? 1 April? 8 April? The purpose of this investigation was to examine whether people underestimated the growth of COVID-19 at the start of the pandemic, to test whether mode of data presentation (table vs. graph) influenced people's forecasts, and to test if forecasts of the virus' growth were related to reported adherence to social-distancing guidelines.

Madison Fansher and Tyler J. Adkins contributed equally to this work.

✉ Madison Fansher
mfansher@umich.edu

✉ Tyler J. Adkins
adkinsty@umich.edu

1 Department of Psychology, University of Michigan, Ann Arbor, MI, USA

2 Department of Linguistics, University of Michigan, Ann Arbor, MI, USA

3 Weinberg Institute for Cognitive Science, University of Michigan, Ann Arbor, MI, USA

4 Department of Psychology, Bogazici University, Istanbul, Turkey

## Understanding exponential growth

In judgmental forecasting tasks, participants are shown data, often as a function of time, and are asked to predict future values. Extrapolating trended data is a decision-making task susceptible to common heuristics and biases (Eggleton, 1982; Tversky & Kahneman, 1974). One well-documented bias in economic decision-making is *exponential growth bias*, in which people tend to perceive exponential functions as linear, thus underestimating the future growth of these trends (Levy & Tasoff, 2015). One proposed explanation for exponential growth bias is the "illusion of linearity"; the tendency to overgeneralize linear models and apply these models to situations where it is inappropriate (De Bock et al., 1998, 2002; Van Dooren et al., 2003). Another explanation may come from our understanding of *trend dampening*, describing the tendency to underestimate the growth of increasing trends and overestimate the growth of decreasing trends (Lawrence & Makridakis, 1989). Trend dampening is posited to result from the influence of ecological knowledge (Keren, 1983), and underestimation of exponential growth may result from such prior knowledge. For example, it may be a reasonable strategy to assume that exponential growth will decelerate considering that many real-life exponential growth trends are actually a part of a logistic growth trend that will eventually level off (Harvey & Reimers, 2013). Researchers have found that people underestimate the growth of exponential functions in judgmental forecasting tasks (Wagenaar & Sagaria, 1975; Wagenaar & Timmers, 1978, 1979), and that underestimation of nonlinearity increases with the size of the exponent (Wagenaar & Sagaria, 1975). When explicitly asked, people are aware of the tendency to underestimate exponential growth, but they continue to exhibit this bias nonetheless (Schonger & Sele, 2020).

The behavioral consequences of exponential growth bias have been examined in the context of economic decision making (Levy & Tasoff, 2016). Stango and Zinman (2009) found that people who exhibited exponential growth bias systematically underestimated interest rates for short-term loans and the benefits of long-term saving, and that more biased people borrowed more and saved less. Similar studies have shown that people mistakenly expect savings to accrue linearly rather than exponentially, leading them to underestimate the value of saving (Mckenzie & Liersch, 2011). Overall, these results suggest that people generally underestimate exponential growth and that this misestimation has real-life behavioral consequences. Thus, it is reasonable to wonder whether Americans underestimated the threat of COVID-19 due to exponential growth bias and whether this underestimation may have influenced real-life social-distancing behaviors.

## Tables versus graphs

Another factor that may influence understanding of the exponential spread of COVID-19 is the way in which data are displayed. Prior work has illustrated that data visualizations assist with the comprehension of quantitative information (see Hegarty, 2011), improve understanding of scientific concepts (van der Linden et al., 2014), and can enhance the communication of risk (Lipkus & Hollands, 1999). Much of public messaging surrounding COVID-19 is based on communicating risk and public health information via graphs, often displayed daily along with cumulative case counts.

Two common methods for displaying such time-series data are graphical (e.g., bar or line graphs) and tabular data presentations. There is mixed evidence on whether tables or graphs are most useful for presenting data (see DeSanctis, 1984; DeSanctis & Jarvenpaa, 1985; Goodwin & Wright, 1993, for reviews). DeSanctis (1984) reviewed the literature comparing graphs to tables on the following dimensions: interpretation speed and accuracy, decision-making/problem-solving quality and speed, information recall, preference, and decision-making confidence. Their review yielded inconsistent results. Out of the studies reviewed, 12 found tables to be better than graphs, seven found graphs to be better than tables, and ten found no difference between modes of presentation. However, this review was not limited to performance on judgmental forecasting tasks. Harvey and Bolger (1996) examined the influence of data presentation on judgmental forecasting and found that viewing data in tables was better for forecasting untrended data, while graphs were better for forecasting trended data. This finding was consistent regardless of data variability. Other researchers have found evidence that graphs are better for short-term forecasting while tables are better for long-term forecasting (Angus-Leppan & Fatseas, 1986; Lawrence et al., 1985). DeSanctis (1984) suggests that whether graphs or tables are more effective is highly dependent on the type of task, and Coll et al. (1991) found that the usefulness of tables or graphs depends on experience, with people working more efficiently with modes of presentation with which they were most familiar. Similarly, DeSanctis and Jarvenpaa (1985) found that while graphs may initially have no effect on decision making, graphs may aid decision making with repeated exposure.

## The current study

We examined whether Americans underestimated the exponential growth of COVID-19, and whether different modes of presenting COVID-19 data in news articles might influence forecasting judgments. Across three studies, participants viewed cumulative growth trends of COVID-19 cases as tables (Table group), as line graphs (Graph group), or as raw data embedded into the text of a fictional news article (the control or Text-only group). Participants were asked to predict the number of future cases for three future time points based on these trends, as well as their confidence in their responses. Given prior work on exponential growth bias, we hypothesized that participants would underestimate the growth of the virus. The impact of visualization on forecasting accuracy is less clear since there is mixed evidence on the effectiveness of tables versus graphs (see DeSanctis, 1984), and little work examining tables versus graphs in the context of extrapolating exponential functions (Wagenaar & Sagaria, 1975). We also examined how misestimation is related to real-life behavior given prior work showing that exponential growth bias influences real-life economic behaviors and decision-making (Levy & Tasoff, 2016). If underestimating the prevalence of COVID-19 leads to a lack of caution, then we expected to find a positive correlation between the number of forecasted cases and engagement in social distancing. Lastly, we examined whether forecasting could be improved with practice by having a subset of the participants complete the task multiple times during the pandemic (Keren, 1983; Wagenaar & Sagaria, 1975).

## Study 1

On 28 March 2020, participants were shown the cumulative COVID-19 case data from 29 February 2020 to 27 March 2020, and were asked to predict the number of cases on three future dates. Given work on exponential growth bias, we hypothesized that participants would underestimate the future trajectory of COVID-19 cases in the USA and that engaging in risk-reduction behaviors would be associated with greater estimates of the number of cases.

The main question of interest was whether participants would be more accurate if they viewed the graphs in tabular or graphical form. In addition, we included a control group for which participants viewed the raw data with no data visualization (text-only). Although one may assume that graphs would produce more accurate estimates given that participants would be able to visually view and extrapolate the trend, we did not pre-register specific hypotheses regarding the difference between tables and graphs as the evidence is mixed, and little work has compared the effectiveness of tables versus graphs in the context of extrapolating an exponential function. We did hypothesize that the text-only group would underestimate the growth of the virus more than the other two groups – given that they would have no data visualization on which to base their estimates. As such, we also hypothesized that those shown a data visualization (table and graph group) would be more confident in their estimates. Pre-registration for Study 1 may be viewed at https://aspredicted.org/blind.php?x=cd4a7h, although we note that the analysis reported significantly deviates from our pre-registration (see below).

### Methods

#### Participants

We recruited a large convenience sample of 1,198 participants from Amazon Mechanical Turk to participate in an online experiment (*M* age: 37.8 years, *SD*: 12.3; 56.2% male, 43.8% female). 770 of these participants remained after applying our exclusion criteria (outlined below).

#### Design

Study 1 used a between-subjects design in which participants were randomly assigned to view news articles with COVID-19 data in either graphs (Fig. 1a), tables (Fig. 1b), or as raw data embedded in text (Fig. 1c). Those in the text-only group viewed a guide on proper hand-washing technique to serve as a control image (Fig. 1d).

#### Materials

Two articles in the format of an online news article were created for the purpose of this experiment. All stimuli used across experiments are available in the Online Supplemental Materials (OSM) (Stimuli S1–S8). Participants read a short vignette about COVID-19 in the USA. Participants viewed data on the total number of deaths and confirmed cases of COVID-19 in the USA in a graphical format (N = 409), a tabular format (N = 408), or a text-only format (N = 381). The five data points shown to participants were from the 5 weeks preceding the date of the study (27 March; Fig. 1a–c). Participants were then asked to estimate the number of confirmed cases, actual cases, and deaths, 3, 6, and 9 days after the article shown to them was written. They were also asked to report their confidence in each of these nine estimates on a scale of 0–100.

After providing their estimates, demographic and individual-difference data were collected (see OSM for measures). A subset of these questions is examined in the current work and concerns social-distancing behaviors. These questions include:

1. How successful have you been in engaging in social isolation? (Slider scale from 0 (Unsuccessful) to 100 (Very successful))
2. How successful will you be at engaging in social isolation in the next week? (Slider scale from 0 (Unsuccessful) to 100 (Very successful))

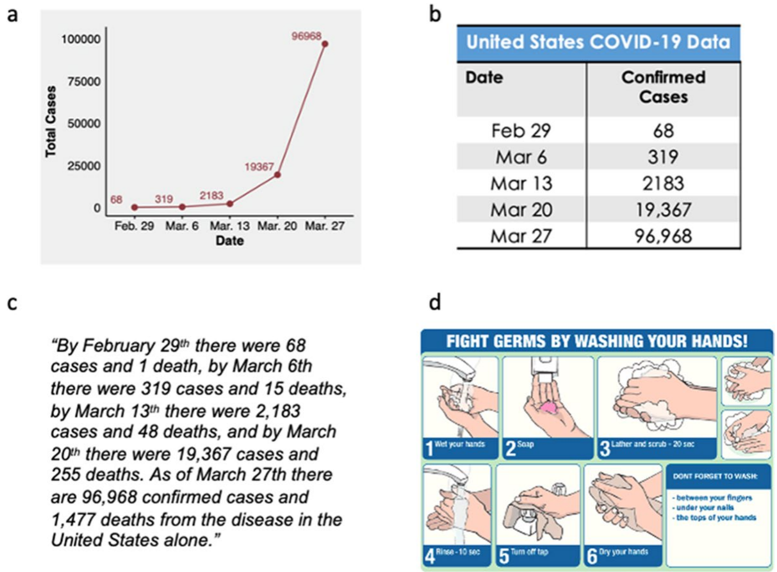



**Fig. 1** **a** Stimuli shown to those in the graph condition, **b** stimuli shown to those in the table condition, **c** raw data embedded in text shown to the text-only group, and **d** control image shown to the text-only group

3. Estimate how much time will pass before we can stop social distancing (1 week, 2 weeks, 3 weeks, 4 weeks, 2 months, 3 months, 4+ months)

We made social distancing the focus of our individual-difference analyses in order to examine the relationship between forecasting of exponential growth and an important real-life behavior (Greenstone & Nigam, 2020). In a separate investigation, not reported here, we use the same data set with structural equation modelling methods to examine the relationship between trait individual difference variables, social-distancing behaviors, and misestimation of the growth of COVID-19 (Quirk et al., in preparation; see Deviations from Pre-Registration for further detail). For the sake of transparency, we report all the individual difference measures that were collected at the time of the study even though they are not analyzed in the current report.

All materials and questionnaires were administered using Qualtrics survey software.

### Quality assurance

To ensure data quality, participants were asked to verify that they were not a robot with a CAPTCHA at the beginning of the survey. We also included two attention check items: an embedded question in the risk aversion scale that asked participants to "please select 6" for the question and a free-response item that asked participants to report the name of the president of the USA. In addition, we asked participants to self-report their perceived effort on the survey on a scale of 1–10. Participants were told that their rating would not affect their compensation for their participation.

### Procedure

Participants located in the USA were invited to take a survey via Amazon Mechanical Turk. They were told that they would read news articles and predict health-related data. After agreeing to participate, they were sent to a Qualtrics survey where they provided informed consent. They were next shown the news article associated with their randomly assigned condition and were immediately asked to report their estimates and confidence for the number of confirmed cases, actual cases, and deaths 3, 6, and 9 days later. Participants then completed the series of individual-difference and demographic questionnaires, rated their perceived effort on the task, and were debriefed. Participants were thanked and compensated US$1 after survey completion. This research was classified as exempt by the University of Michigan Institutional Review Board.

### Exclusion criteria

Our exclusion criteria are outlined below. All exclusion criteria were the same for Studies 1–3 and generally exclude participants who did not put effort into the task, failed to pay attention, or failed to follow instructions. Please see the OSM for further information about participants excluded in Studies 1–3. In Study 1 we excluded all the data from participants who:

1. Were younger than 18 years, N = 3
2. Did not provide a valid zip code (i.e., possibly not a US resident), N = 66
3. Reported an impossible forecast (i.e., misunderstood the task), N = 400
4. Failed the basic attention check trial (“Please select option 6”), N = 4
5. Failed to correctly identify the US President (free response), N = 11
6. Self-reported investing effort of less than 5 out of 10, N = 3
7. Took less than 30 s to complete the task (considered impossible based on the number of survey items), N = 0

And we excluded individual outlier forecasts:

8. Greater than 10x the last datum provided in the visualization or text.

We considered criteria 1–3 to be required for inclusion in the data analyses as they determine eligibility to participate in the study as well as a basic understanding of the task. Criterion 3 was necessary because participants were tasked with forecasting cumulative growth, so participants who forecasted a decrease were *not* forecasting cumulative growth. That the excluded participants were doing something categorically different from the rest is evident by their distinct distribution of forecasts, most of which were very low (< 1,000 cases) (Fig. S1B).

We adopted additional exclusion criteria measuring effort, attention, and task-understanding as we wanted our data to be of the highest quality possible given that the data were collected online. These exclusion criteria had a minimal impact on the sample size and key results. See the OSM for an analysis of the effects of our optional exclusion criteria (4–8) on the sample size and key results (table vs. graph) in Studies 1 and 2 (Tables S1 and S2).

Whereas most forecasts predicted under 1 million cases, a small number of outlier forecasts were as large as 40 million (N = 4 participants). Upon inspection of the forecast distributions (Fig. S1A), we found that using a cutoff of ten times the last datum provided to participants neatly eliminated outliers without affecting the distribution. The cutoff corresponded to forecasts of roughly 1 million cases for Study 1 and its replication and forecasts of roughly 4 million cases for Study 2 and its replication.

## Regression modeling

We modeled participants’ forecasts of future total confirmed COVID-19 cases using hierarchical regression models (see Fig. S1C for the distributions of responses modeled here). We also examined participants’ forecasts of *deaths* due to COVID-19 and “*actual*” COVID-19 cases and our results largely held for these other forecasts, although we omit these data from the main text for brevity (see the *Other Forecasts* section of the OSM). Our models of forecasts included fixed effects of forecast horizon (within-subject; 6–3 days, 9–6 days), data visualization group (between-subject; table–graph), and their interaction. The models allowed intercepts to vary randomly by state. We allowed intercepts to vary by state because at the time of the study the number and growth of COVID-19 cases varied dramatically among states. We implemented the model using the R-package {brms}, an open-source package for Bayesian multilevel modeling (Bürkner, 2017, 2018). This package translates input models into the probabilistic programming language stan, which supports approximate Bayesian inference over model parameters using Markov Chain Monte Carlo (MCMC) sampling (Carpenter et al., 2017).

When we modeled forecasts, we used a Gamma likelihood function rather than the default Gaussian because the distribution of forecasts was positive-only and had a very long right tail (Fig. S1C). To facilitate specification of priors and to obtain standardized effect size estimates, we rescaled our outcome variables by dividing by the standard deviation of all estimates (within the experiment). Our model of forecasts was specified as follows:

$$y \sim Gamma(\mu, \alpha)$$
$$\log(\mu) = \beta_0 + \beta X + \beta_0^{state}$$

The first expression above is the likelihood function and the second expression is the regression formula for the mean with a log link function. In the regression formula, $\beta_0$ is the population intercept, $\beta_0^{state}$ is a state’s ‘random’ intercept, $X$ denotes the predictors (delay, group, delay*group), and $\beta$ denotes the corresponding population-level regression coefficients. The auxiliary shape parameter of the gamma distribution is denoted by $\alpha$. We assigned the following weakly informative default priors to the model parameters (Gelman et al., 2008):

$$\beta_0 \sim Student\_t(3, 0, 2.5)$$
$$\beta \sim Student\_t(3, 0, 2.5)$$
$$\beta_0^{state} \sim Normal(0, \sigma_{state})$$
$$\sigma_{state} \sim HalfStudent\_t(3, 0, 2.5)$$
$$\alpha \sim Gamma(0.01, 0.01)$$

All MCMC chains passed visual inspection, all $\hat{R}$values were 1, and all effective sample sizes (ESS) were greater than 10,000, which has been recommended as the minimum ESS to obtain reliable MCMC estimates of 95% credible intervals (Kruschke, 2015). After fitting the models, we performed graphical posterior predictive checks using the R packages {bayesplot} (Gabry et al., 2019) and {loo} (Vehtari et al., 2017). To quantify uncertainty about the effects of interest, we report posterior standard deviations (SDs), 95% credible intervals (CIs) as well as probabilities of direction (PD). The PD is defined as the probability that an effect goes in the direction indicated by the median estimate (Makowski et al., 2019). For main effects of interest, we report the differences of means ($M_{diff}$, in native units) as well as standardized regression coefficients ($\beta_{effect}$, in sample SD units).

We applied a similar Bayesian hierarchical regression model to participants' reported *confidence* (0–100) in their forecasts. This model used the same predictors (group and day) but used the default Gaussian likelihood function with an identity link function for the regression formula:

$$y \sim Normal(\mu, \sigma)$$
$$\mu = \beta_0 + \beta X + \beta_0^{state}$$

We also used a Bayesian hierarchical regression model to estimate the proportion of participants who underestimated the number of cases at a given time point. The model used a Bernoulli likelihood function with a logit link function:

$$y \sim Bernoulli(\mu)$$
$$logit(\mu) = \beta_0 + \beta_0^{state}$$

This model simply included one population intercept and varying intercepts by state, normally distributed around the population mean. We fit the model separately to forecasts at each forecast horizon (3, 6, and 9 days). In the *Results* section, we report the posterior mean ($P_{under}$) and 95% CIs for the probability of overestimation, after converting from log-odds to probability. While here we compare participants forecasts to actual case numbers (i.e., "true" total number of confirmed COVID-19 cases), participants still demonstrate a large misestimation when comparing their forecasts to the predicted values of exponential models fit to the initial five data-points provided (Table S4 and Fig. S2).

## Deviations from pre-registered analysis

Our reported analyses deviated from the analyses anticipated in our pre-registration forms. We anticipated using a Gaussian distribution in our regression models but discovered that a Gamma distribution was more appropriate, given that the forecasts were positive-only and had extreme skew. We also anticipated using participant-level random intercepts but realized that these estimates were too noisy given only three observations per subject, so we decided instead to use state-level random intercepts, as there were many more observations per state, and we expected that forecasts would likely vary by state given differences in social-distancing policies. Our analyses excluded "impossible" forecasts, but this exclusion criterion was not in our pre-registration form, simply because we did not anticipate that people would misunderstand the task in this way. In the OSM, we explore how the results vary (or not) under different exclusion criteria with a multiverse analysis. In addition, in our original pre-registration we anticipated including the text-only group as a control condition when modelling participants' forecasts; however, we realized that since we are using hierarchical regression modelling, we would be unable to compare table and graph groups as the text-only group would act as our reference condition. Thus, we only model the data from the table and graph groups, but we do report descriptive statistics for the text-only group as well as include their data in Fig. 2. We anticipated including many individual differences in our analyses but realized that these data should be thoroughly investigated in a separate paper using structural equation models (Quirk et al., in preparation). This investigation has many interesting results, but they don't concern mode of data presentation nor do these results affect the implications of the current investigation. Given the many deviations from our pre-registered analysis methods, the following results should be considered *exploratory* rather than confirmatory.

## Results

On average, participants underestimated the number of cases on 30 March ($P_{under}=0.83$, $CI=[0.79,0.83]$, $M_{est}=141k$ cases, $se_{est}=2.7k$, Truth = 166k), 2 April ($P_{under}=0.77$, $CI=[0.73,0.81]$, $M_{est}=207k$ cases, $se_{est}=5.4k$, Truth = 248k), and 5 April ($P_{under}=0.78$, $CI=[0.74,0.82]$, $M_{est}=270k$ cases, $se_{est}=8.1k$, Truth = 341k) (Fig. 2c). Critically, the Table group produced more accurate estimates than the Graph group ($M_{diff}=14k$, $\beta_{T-G}=0.05$, $SD=0.02$, $CI_{95\%}=[0.01,0.10]$, $PD=0.99$) (Fig. 2c). Further, the Table group forecasted greater growth in the number of cases from 30 March to 2 April than the Graph group ($\beta_{A2-M30*T-G}=0.11$, $SD=0.06$, $CI_{95\%}=[0.00,0.23]$, $PD=0.98$) (Fig. 2c). However, the two groups forecasted similar increases in cases from 2 April to 5 April ($\beta_{A5-M2*T-G}=-0.01$, $SD=0.06$, $CI_{95\%}=[-0.12,0.11]$, $PD=0.59$).We found that participants in a text-only control group produced virtually identical forecasts (on average) to those in the table group for 30 March (Text: ~145k cases, Table: ~141k, Graph: ~142k), 2 April (Text: ~220k, Table: ~218k, Graph: ~195k), and 5 April (Text: ~278k, Table: ~281k, Graph: ~263k). A regression model comparing forecasts of the Graph and Table groups to the Text group revealed that participants in the Graph group produced lower forecasts

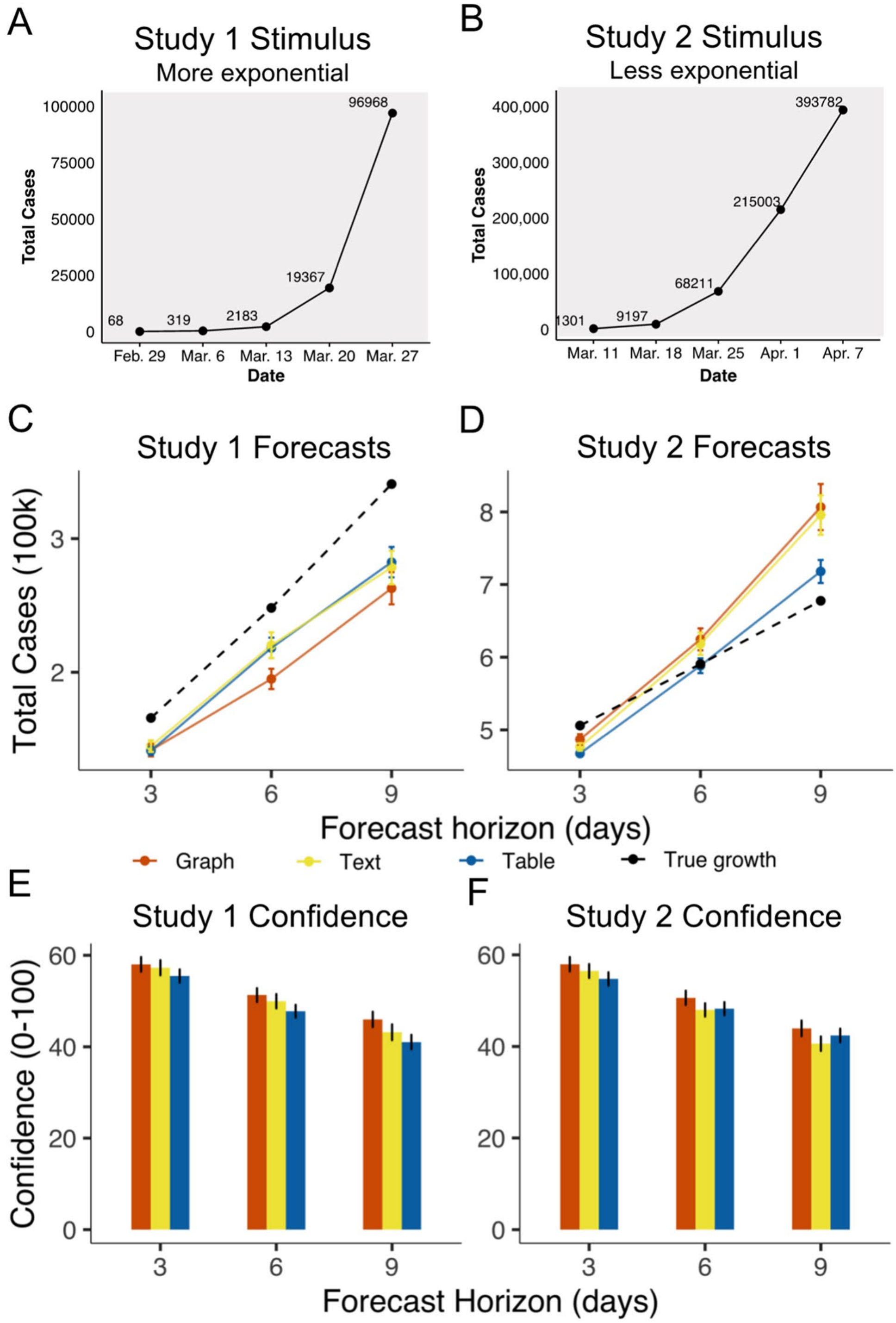



**Fig. 2** Results from Studies 1 and 2. People who viewed graphs of COVID-19 growth in the USA produced less accurate forecasts compared to people who viewed the same data in tables. Study 1 (N = 540) was conducted on 27 March 2020, and Study 2 (N = 543) was conducted on 7 April 2020. Participants in Study 1 were presented with the data from panel A, in graphical form (shown) or tabular form; Participants in Study 2 were presented with data from panel B. Participants' forecasts from Study 1 are shown in panel C; participants' forecasts from Study 2 are shown in panel D. The black data points reflect the "true" total number of confirmed COVID-19 cases in the USA according to worldometers.info, accessed on 27 April 2020. The colored lines show the mean forecasts for participants in the graph groups (red), text-only groups (yellow), and table groups (blue) and error bars represent ± 1 standard error of the mean. In Study 1 (panel B), participants tended to underestimate the future trend, whereas in Study 2 (panel D), participants tended to overestimate the future trend. In both studies, participants who were presented tabular data (blue) produced forecasts closer to the true values. Participants' reported confidence (mean +/- se) in their forecasts are shown in panels E and F. Participants were more confident in forecasts from graphs, despite those forecasts being less accurate, compared to forecasts from tables

compared to participants in the Text group ($\beta_G = -0.06, \mathrm{CI}_{95\%} = [-0.11, -0.01], \mathrm{PD} = 0.99$), while participants in the Table group produced forecasts of roughly equal magnitude to those in the Text group ($\beta_T = -0.01, \mathrm{CI}_{95\%} = [-0.06, 0.04], \mathrm{PD} = 0.66$).

Although tables facilitated more accurate forecasting compared to graphs, people shown tables were less confident in their forecasts than people shown graphs ($\mathrm{M}_{\mathrm{diff}} = -3.6, \beta_{\mathrm{T-G}} = -0.16, \mathrm{CI}_{95\%} = [-0.26, -0.07], \mathrm{PD} = 1.0$) (Fig. 2e). Overall, confidence decreased over time as participants were less confident about their 2 April forecast than their 30 March forecast ($\mathrm{M}_{\mathrm{diff}} = -7.3, \beta_{\mathrm{A2-M30}} = -0.29, \mathrm{CI}_{95\%} = [-0.40, -0.18], \mathrm{PD} = 1.0$) and their 5 April forecast compared to their 2 April forecast ($\mathrm{M}_{\mathrm{diff}} = -5.9, \beta_{\mathrm{A5-A2}} = -0.23, \mathrm{CI}_{95\%} = [-0.35, -0.11], \mathrm{PD} = 1.0$). This illustrates that even though participants misestimated the number of cases, their responses were still rational to an extent. We found that participants in the text-only control group reported intermediate confidence in their forecasts (on average) compared to participants in the other groups for 30 March (Graph: 58.0, Text: 57.3, Table: 55.5), 2 April (Graph: 51.3, Text: 50.0, Table: 47.8), and 5 April (Graph: 46.0, Text: 43.2, Table: 41.0). A regression model comparing confidence of the Graph and Table groups to the Text group revealed that participants in the Graph group was more confident than the Text group ($\beta_G = 0.09, \mathrm{CI}_{95\%} = [-0.01, 0.20], \mathrm{PD} = 0.96$), while participants in the Table group appeared less confident than the Text group ($\beta_T = -0.07, \mathrm{CI}_{95\%} = [-0.17, 0.03], \mathrm{PD} = 0.90$).

### Discussion

Study 1 provides evidence for our hypothesis that Americans generally underestimated the growth of COVID-19, exhibiting exponential growth bias. These forecasts were more accurate when participants were presented with data in tables or text rather than graphs, which comes as somewhat of a surprise given the documented benefits of graphical presentation (for review, see Hegarty, 2011). However, in the context of forecasting, some prior work has shown that graphs are more effective than tables for forecasting trended functions and short-term forecasts, both consistent with the task employed in the current study (Harvey & Bolger, 1996; Lawrence et al., 1985). As mentioned previously, little work has examined tables versus graphs in the context of extrapolating exponential growth; however, Wagenaar and Sagaria (1975) found that participants produced more accurate estimates of exponential growth when shown raw numbers (similar to a tabular format) in contrast to graphs of the same trends, aligning with the findings from Study 1. Surprisingly, participants who viewed raw data embedded in text exhibited behavior similar to those who were shown tables, even though they were not shown a visualization of the data. This may be because both groups of participants viewed the data without an exponential trend graphically imposed on the data (see *General discussion* for more). In Study 2 we aimed to replicate our findings using the most recent COVID-19 data (at the time) to see whether the benefits of tables over graphs would continue to be observed.

## Study 2

Study 2 aimed to replicate the findings from Study 1 in the context of newer data about the pandemic (trend up until 8 April) to test the robustness of the findings that participants underestimated exponential growth and that forecasting could be improved by viewing tables of data. We also recruited a subset of the sample from Study 1 to see if forecasting would improve with experience performing the task given mixed evidence on the effect of experience on forecasting (Keren, 1983; Wagenaar & Sagaria, 1975). Ten days after launching Study 1 (7 April) we administered the online survey again to people in the USA with uPDated data to reflect the growth of COVID-19 in the USA from 11 March to 7 April (Fig. 3a, b). We hypothesized that people would underestimate the number of confirmed cases and that this underestimation would be greatest for those in the graph condition. We also hypothesized that people in the graph condition would be more confident in their estimates. Pre-registration may be accessed athttps://aspredicted.org/blind.php?x=py8qp2

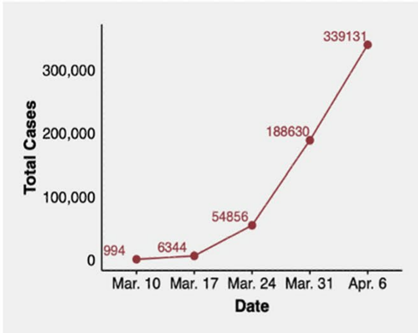
### Methods

#### Participants

We recruited a large convenience sample of 1,180 participants from Amazon Mechanical Turk to participate in an online experiment (*M* Age: 38.7 years, *SD*: 11.9; 53.6% male, 46.4% female). Half of the recruited participants had also participated in Study 1 so we could examine whether forecasting would improve with practice. 802 subjects remained after applying our exclusion criteria (outlined above).

#### Design

Participants were randomly assigned to data visualization groups (Graph $N = 379$, Table $N = 408$, Text = 393). Approximately half of the participants also participated in Study 1 ($N = 580$) and half of participants did not ($N = 600$). Returning participants were assigned to the same condition they had experienced previously. This allowed us to examine whether participants produced more accurate COVID-19 a




b

**United States COVID-19 Data**

| Date | Confirmed Cases |
|---|---|
| Mar 10 | 994 |
| Mar 17 | 6,344 |
| Mar 24 | 54,856 |
| Mar 31 | 188,630 |
| Apr 6 | 339,131 |

c

*"By March 10th there were 994 cases and 30 deaths, by March 17th there were 6,344 cases and 110 deaths, by March 24th there were 54,866 cases and 780 deaths, and by March 31st there were 188,630 cases and 4,053 deaths. As of April 6th there are 339,131 confirmed cases and 9,689 deaths from the disease in the United States alone."*

**Fig. 3** **a** Stimuli shown to those in the graph condition, **b** stimuli shown to those in the table condition, **c** stimuli shown to those in the text-only condition along with the diagram in Fig. 1d

forecasts with more exposure to exponential trends in the media, as prior work has shown that experience influences forecasting of exponential trends (Keren, 1983).

### Materials

The materials were the same as in Study 1, except the five data points presented to participants were for the 5 weeks preceding 7 April instead of the weeks preceding 27 March (Fig. 3a, b).

### Procedure

All procedures were the same as those used in Study 1.

### Modeling

In addition to the regression models of forecasts, confidence, and overestimation used in Study 1, in Study 2 we used a similar Bayesian hierarchical regression model to examine whether returning participants showed improved forecasting performance relative to participants who had not been tested previously. The outcome here was forecasting error, which we defined as the absolute error (*estimate – truth*) scaled by the truth (*absolute error/truth*) to normalize the error measure across studies and forecast horizons. The model used a gamma link function as in the forecast model described above. The model included main effects of group (table–graph), study (2–1), cohort (returning–new), and forecast horizon (6–3 days, 9–6 days), as well as the group by study, group by cohort, study by cohort, and group by cohort by study interactions. The priors were the same as those used in the previous models. The analysis again deviates from the pre-registered analysis methods in the same aspects discussed in Study 1, thus the following results should be considered *exploratory* rather than confirmatory.

## Results

Overall, participants *underestimated* the number of cases on 10 April ($P_{under} = 0.83, CI = [0.80, 0.87], M_{est} = 476k$ cases, $se_{est} = 3.9k, Truth = 506k$) but *overestimated* the number of cases on 13 April ($P_{under} = 0.54, CI = [0.49, 0.59], M_{est} = 602k$ cases, $se_{est} = 8.7k, Truth = 591k$) and 16 April ($P_{under} = 0.51, CI = [0.46, 0.56], M_{est} = 752k$ cases, $se_{est} = 16.3k, Truth = 678k$) (Fig. 2d). On average, the Table group was more accurate than the Graph group ($M_{diff} = 46.5k, \beta_{T-G} = -0.06, SD = 0.01, CI = [-0.08, -0.03], PD = 1.0$) (Fig. 2d) and there was a group by forecast horizon interaction, such that the Table group forecasted smaller increases in cases from 13 April to 16 April than the Graph group ($\beta_{A16-A13*T-G} = -0.05, SD = 0.04, CI = [-0.12, 0.02], PD = 0.93$) (Fig. 2d). The two groups forecasted similar increases in cases from 10 April to 13 April ($\beta_{A5-M2*T-G} = -0.01, SD = 0.04, CI = [-0.08, 0.06], PD = 0.64$). We found that participants in a text-only control group produced very similar forecasts (on average) to those in the graph group for 10 April (Text: 476k cases, Graph: 487k, Table: 467k), 13 April (Text: 618k, Graph: 625k, Table: 588k), and 16 April (Text: 796k, Graph: 807k, Table: 718k). A regression model comparing forecasts of the Graph and Table groups to the Text group revealed that participants in the Table group produced lower forecasts compared to participants in the Text group ($\beta_T = -0.05, CI_{95\%} = [-0.08, -0.02], PD = 1.0$), while participants in the Graph group produced forecasts of roughly equal magnitude to those in the Text group ($\beta_G = 0.01, CI_{95\%} = [-0.02, 0.05], PD = 0.82$).

The Table group was overall less confident (0–100) in their forecasts than the Graph group ($M_{diff} = -2.5, \beta_{T-G} = -0.11$, $SD = 0.05$, $CI = [-0.19, -0.01]$, $PD = 0.99$) (Fig. 2f). Participants were predictably less confident in their more distal forecasts as they were less confident about their 13 April forecast than their 10 April forecast ($M_{diff} = -6.7, \beta_{A13-A10} = -0.27, SD = 0.06, CI = [-0.38, -0.16], PD = 1.0$) and their 16 April forecast than their 13 April forecast ($M_{diff} = -6.1, \beta_{A16-A13} = -0.24, SD = 0.06, CI = [-0.35, -0.13], PD = 1.0$) (Fig. 2f). We found that participants in the text-only control group reported intermediate confidence (on average) compared to participants in the other groups for 10 April (Graph: 57.9, Text: 56.5, Table: 54.7) but lower confidence (on average) compared to the other groups for 13 April (Graph: 50.6, Table: 48.3, Text: 48.0) and 16 April (Graph: 43.9, Table: 42.4, Text: 40.6). A regression model comparing confidence of the Graph and Table groups to the Text group revealed that overall participants in the Graph group was more confident than the Text group ($\beta_G = 0.11, CI_{95\%} = [0.01, 0.21], PD = 0.99$), while participants in the Table group displayed approximately equal confidence compared to the Text group ($\beta_T = 0.02, CI_{95\%} = [-0.07, 0.12], PD = 0.69$).

Half of the participants in Study 2 had also participated in Study 1. Overall, forecasting error (|estimate – truth| / truth) was lower for these participants in Study 2 compared to Study 1 ($\beta_{2-1} = -0.67$, SD = 0.03, CI = [−0.73, −0.62], PD = 1.0) and that forecasting error was greater for the Graph group compared to the Table group ($\beta_{G-T} = 0.13$, SD = 0.03, $CI_{95\%}$ = [0.07,0.19], PD = 1.0) (Fig. 4a). Critically, the decrease in error from Study 1 to Study 2 was more pronounced for the *returning* participants when compared to participants new to the task (non-returning participants in Study 1 and newly recruited participants in Study 2) ($\beta_{2-1*R-N} = -0.10$, SD = 0.06, CI = [−0.22,0.01], PD = 0.96) (Fig. 4a). Practice effects were larger for the graph group compared to the table group, leading to a three-way interaction between study (2–1), cohort (returning–new), and group (table–graph) ($\beta_{2-1*R-N*G-T} = -0.10$, SD = 0.06, CI = [−0.56, −0.10], PD = 1.0) (Fig. 4a). This interaction suggests that practice with extrapolating exponential functions from graphs may lead to improved forecasting even though our results to this point have suggested that forecasting from tables is generally better than forecasting from graphs.

## Discussion

Despite the *prima facie* inconsistency of our two studies (graphs yielding lower estimates in Study 1 and higher estimates in Study 2), one critical pattern was resilient: tables facilitated more accurate forecasts than graphs, although graphs led to greater confidence in one's inaccurate forecasts.

Study 2 suggests that by 7 April, Americans began to overestimate the growth trajectory of COVID-19. One possible explanation for the inconsistencies between Study 1 and Study 2 is that there were critical differences in the structure of the data shown to participants, that is, the linearity/exponentiality of the functions. The data from Study 1 are fit better by an exponential model (Adjusted $R^2 = 0.99$) than the data of Study 2 (Adjusted $R^2 = 0.94$), whereas a linear model fits the data of Study 2 (Adjusted $R^2 = 0.84$) better than the data of Study 1 (Adjusted $R^2 = 0.54$). Prior work has shown that underestimation of exponential functions increases as the exponent increases, which could account for these differing results (Wagenaar & Sagaria, 1975) if participants were interpreting the function from Study 2 as more linear. Another possibility is that this inconsistency may have resulted from increased

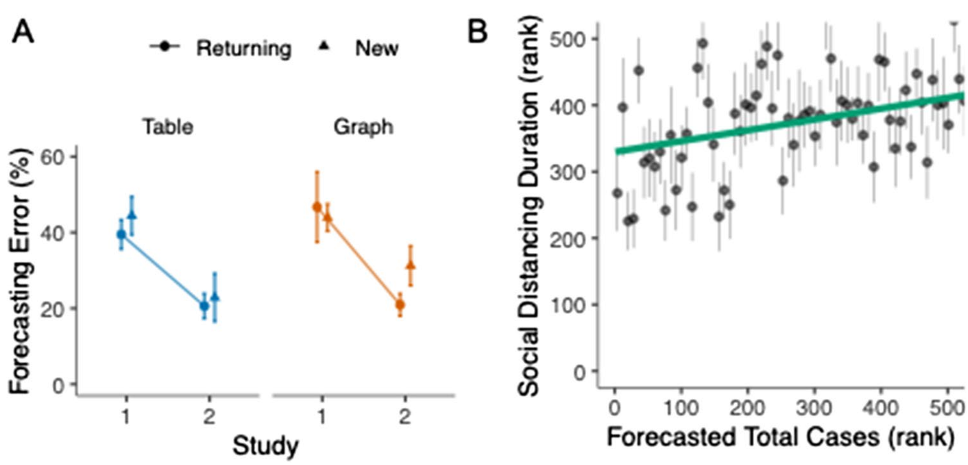



**Fig. 4** Forecast training and the link between forecasting and attitudes towards social distancing. Half of our participants in Study 2 also participated in Study 1. We found that returning participants in the graph group produced more accurate forecasts on their second attempt than new control participants. Panel ***A*** shows the mean percent error of participants forecasts compared to the truth (shown in Fig. 2a and b), separately for each Study, data visualization group, and cohort (new vs. returning). Error bars reflect standard errors of the means. Panel ***B*** shows the relationship between the average number of total cases forecasted and the forecasted time to desist all social distancing measures. The variables were ranked to place them on a common scale. For clarity, we show means and standard errors of y in 100 equally spaced bins of x. The green line represents a line of best fit for the raw data (not shown)

awareness of the spread of COVID-19 among the American public. Widespread news coverage of COVID-19 may have increased exposure to exponential functions, which led to overestimation of the future number of cases. Prior work has shown that over- and underestimation of exponential and linear growth may be influenced by prior experience engaging with such functions (Ebersbach et al., 2008; see *General discussion* for further detail).

To disentangle these possibilities, a third study was run in which participants were given the task from Study 1 or the task from Study 2. If it is the case that mere exposure to COVID-19 information and graphs or increased sensitivity to exponential growth led to greater estimates in Study 2, one may expect that participants would overestimate the number of cases regardless of the function shown to them. However, if the pattern of over- and underestimation was due to the linearity/exponentiality of the data themselves, we would expect to replicate this pattern of over- and underestimation.

## Study 3

Given that we wanted to show participants the exact stimuli from Studies 1 and 2, all mentions of the USA were removed from the original news article and replaced with references to a "hypothetical country." The purpose of this was twofold: (1) participants would be less tempted to look up the number of cases for the dates they were asked to forecast that had already occurred at this point, and (2) this would reduce the application of COVID-19 information specific to the USA to the scenario, such as lockdowns, mask mandates, and politicization of the virus, which would allow us to better understand how participants are interpreting the data themselves without context. If we were to replicate the pattern of over- and underestimation observed in Studies 1 and 2, this would suggest that it is something about the functions themselves that is leading to this pattern. If we consistently see overestimation regardless of whether participants view the data from Study 1 or Study 2, this would suggest that by the time of Study 2 people were generally more sensitive to the spread of the virus. We hypothesized that we would replicate the finding that tables would lead to more accurate estimates than graphs, given that this was consistent across Studies 1 and 2. Pre-registration may be viewed at https://aspredicted.org/blind.php?x=74SD4t

### Methods

#### Participants

We recruited a large convenience sample of 803 participants from Amazon Mechanical Turk to participate in an online experiment (*M* Age: 38.5 years, *SD*: 12.0; 57.1% male, 42.9% female). 442 participants remained after applying our exclusion criteria (outlined above).

#### Design

Study 3 used a 2 (timepoints: Study 1 data, Study 2 data) x 2 (data visualization: graph, table) factorial design. Participants were randomly assigned to view the data from Study 1 (Fig. 1; N = 411) or the data from Study 2 (Fig. 3; N = 392) and were also randomly assigned to view those data in either graphical (N = 203, Study 1 materials; N = 198, Study 2 materials) or tabular form (N = 208, Study 1 materials; N = 194, Study 2 materials).

#### Materials

All materials and questionnaires were the same as those in Studies 1 and 2, except for the mention of a hypothetical country.

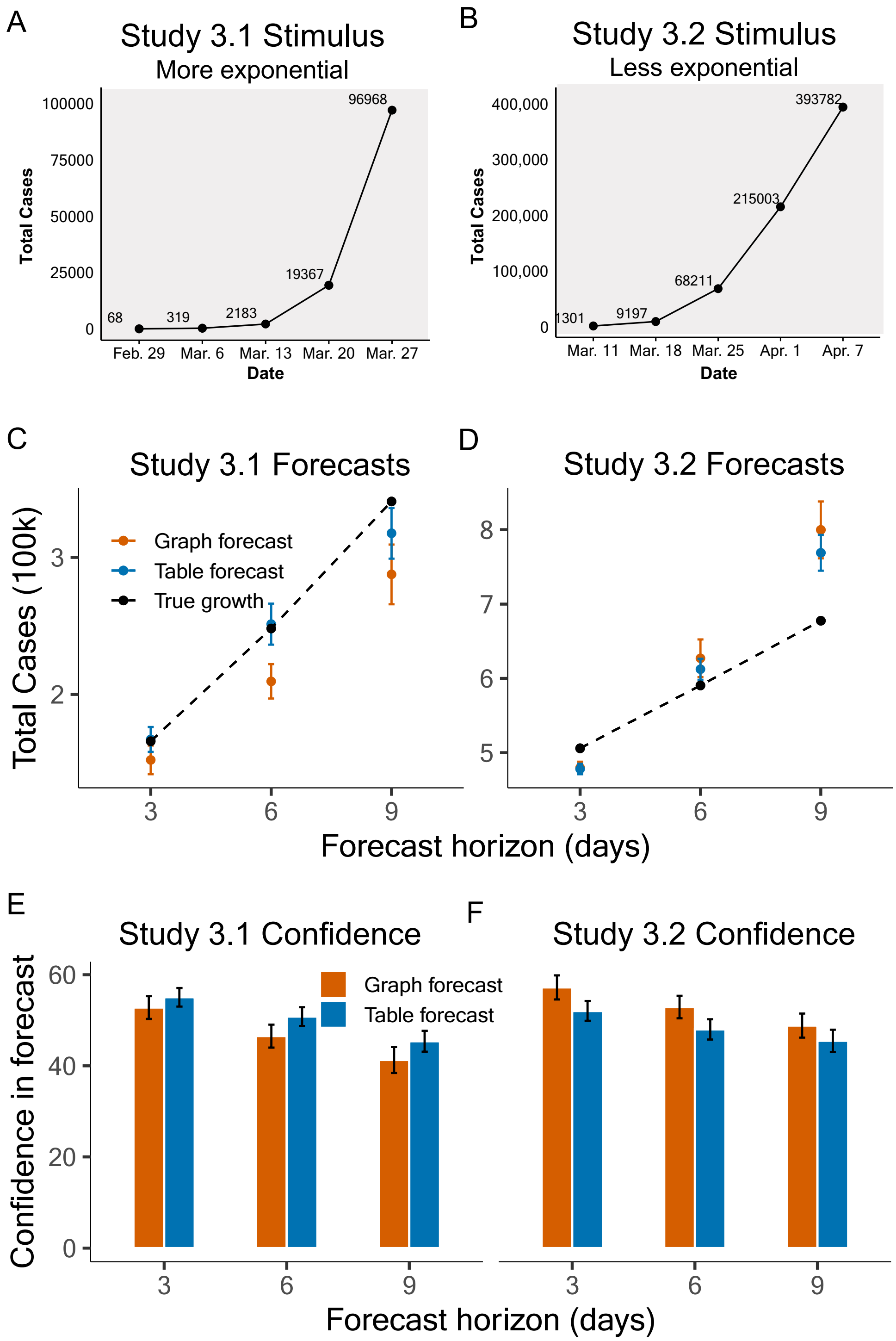
#### Procedure

All procedures were consistent with Studies 1 and 2 except participants were compensated US$0.75 instead of US$1 after survey completion.

### Results

While the analysis methods deviated from what was pre-registered for Study 3, the following results should be considered *confirmatory* considering that we aimed to replicate the main findings from Studies 1 and 2, thus it was necessary to use the same modelling methods that were implemented in Studies 1 and 2. The data for Study 3 were collected before the exact analysis methods for Studies 1 and 2 were established, which is why Study 3 was pre-registered with the same modelling methods pre-registered for Studies 1 and 2.

In the replication of Study 1, in which the presented data exhibited more exponential growth, participants underestimated the number of COVID-19 cases for 30 March ($P_{under}=0.79$, CI = [0.73,0.85], $M_{est}=161$k cases, $SE_{est}=6.85$k, Truth = 166k), 2 April ($P_{under}=0.69$, CI = [0.62,0.75], $M_{est}=233$k cases, $SE_{est}=10.2$k, Truth = 248k), and 5 April ($P_{under}=0.71$, CI = [0.64,0.77]) (Fig. 5c). In the replication of Study 2, in which the presented data were less exponential, participants tended to *underestimate* the number of COVID-19 cases in the USA for 10 April ($P_{under}=0.84$, CI = [0.79,0.90], $M_{est}=479$k cases, $SE_{est}=5.29$k, Truth = 506k), slightly *overestimate* the number of cases for 13 April ($P_{under}=0.65$, CI = [0.47,0.63], $M_{est}=619$k cases, $SE_{est}=13.7$k, Truth = 591k), and *overestimate* the number of cases for 16 April ($P_{under}=0.47$, CI = [0.49,0.55], $M_{est}=783$k cases, $SE_{est}=21.6$k, Truth = 678k) (Fig. 5d).




Overall, the Table groups were more accurate than the Graph groups as they underestimated less in the replication of Study 1 ($M_{diff} = 28.2k, \beta_{T-G} = 0.15, SD = 0.04, CI = [0.07, 0.23], PD = 1.0$) and overestimated less in the replication of Study 2 ($M_{diff} = -19.7k, \beta_{T-G} = -0.04, SD = 0.02, CI = [-0.08, -0.01], PD = 0.95$) (Fig. 5c, d). In the replication of Study 1 we did not replicate the finding that graphs led to false confidence as people shown tables were *more* confident in their forecasts than people shown graphs ($M_{diff} = 3.6, \beta_{T-G} = 0.18, SD = 0.08, CI_{95\%} = [0.02, 0.33], PD = 0.99$); however, viewing graphs did lead to greater confidence in the replication of Study 2 ($M_{diff} = -4.4, \beta_{T-G} = 0.18, SD = 0.08, CI = [0.02, 0.33], PD = 0.99$) (Fig. 5e, f).

◂**Fig. 5** Results from Study 3 (replications of Study 1 and Study 2). People who viewed graphs of COVID-19 growth produced less accurate forecasts compared to people who viewed the same data in tables. Participants in the replication of Study 1 (N = 215) were presented with the data from panel **A**, in graphical form (shown) or tabular form (not shown); Participants in replication of Study 2 (N = 227) were presented with data from panel **B**. Participants' forecasts from the replication of Study 1 are shown in panel **C**. Participants' forecasts from the replication of Study 2 are shown in panel **D**. The black points reflect the "true" total number of confirmed COVID-19 cases in the USA according to worldometers.info, accessed on 27 April 2020. The colored lines show the mean forecasts for participants in the graph groups (red) and table groups (blue) and error bars represent ± 1 standard error of the mean. In the replication of Study 1 (panel C), participants tended to underestimate the future trend, whereas in the replication of Study 2 (panel D), participants tended to overestimate the future trend. In both studies, participants who were presented tabular data (blue) produced more accurate forecasts. Participants' reported confidence (mean ± SE) in their forecasts is shown in panels **E** and **F**. Participants were more confident in forecasts from tables in the replication of Study 1 (where tables led to more accurate forecasts) and more confidence in forecasts from graphs in the replication of Study 2 (where graphs led to more accurate forecasts)

### Discussion

Study 3 replicated our previous results when all data were collected at the same time point and in the context of a hypothetical country. This suggests that it is the structure of the data themselves (e.g., linearity/exponentiality) that influences whether people over- or underestimate exponential trends. Viewing tables of COVID-19 data again led to more accurate forecasts than viewing graphs of COVID-19 data regardless of the data structure.

## Social-distancing analyses

To what extent do people's forecasts relate to their attitudes about social distancing? To provide some insight into this question, we conducted an exploratory set of rank-correlational analyses with data from Studies 1 and 2 and found across studies that the greater people's forecasts, the longer they expected social-distancing orders to remain in place ($\text{Study 1}: \tau = 0.15, p < 0.001; \text{Study 2}: \tau = 0.10, p < 0.001$) (Fig. 4b). Forecasted total number of cases was also positively correlated with prior ($\tau = 0.09$, $p < 0.001$) and future ($\tau = 0.06, p < 0.01$) adherence to social-distancing measures in Study 1,though there was no evidence for these relationships in Study 2 ($\text{prior}: \tau = -0.01, p = 0.80; \text{future}: \tau = 0.01, p = 0.72$).

Overall, these results suggest that forecasts about the cumulative spread of COVID-19 were related to people's attitudes about social distancing in Study 1, and there was a marginal relationship between forecasted cumulative cases and attitudes about social distancing in Study 2. Why the discrepancy between these two studies? To address this question, we examined data from the participants who participated in both Study 1 and Study 2 (N = 399). What we found is that the differential results between Study 1 and 2 shown above also held *within subject*. Forecasts were positively correlated with all three social-distancing measures in Study 1 (all $p < .01$), but only with the time to stop distancing measure in Study 2 ($p < .001$, other $p > .3$). It is therefore possible that increased COVID-19 knowledge among the general public attenuated the relationship between forecasts and social distancing behaviors and that the differences between Studies 1 and 2 could have resulted from an overall increase in social distancing by the time data were collected for Study 2. Although the difference in time between 27 March and 7 April may seem negligible, it is important to note that during this time many states were beginning to impose "stay at home" orders on their populations. Thus, it is possible that people were social distancing more by Study 2 than they were in Study 1 depending on their state or county's guidelines.

In Study 3, the data were shown to participants in the context of a hypothetical country. Thus, it is reasonable to assume that participants will reason about the state of the pandemic in a different country differently than they would their own. However, given the discrepancies between the Study 1 and Study 2 results, we decided to repeat the analyses using data collected from Study 3 for the hypothetical-country replication of Studies 1 and 2. We found no significant relationships between forecasts and future or past isolation for either study ($p > .2$). However, in the replication of Study 2 there was a positive correlation between forecasts and time-to-stop distancing ($\tau = 0.12$, $p = 0.017$); this relationship was not significant in the replication of Study 1 ($p = 0.441$).

## General discussion

This research adds to an existing body of showing that people are erroneous when engaging in judgmental forecasting by demonstrating that misestimation is impacted both by data structure and mode of presentation. While our participants were typically more accurate when they were forecasting based on data presented in tabular format, graphical formats led to a disproportionate confidence in estimates. In addition to mode of presentation, the nature of trends also impacted whether the trends were over- or underestimated. Lastly, we found slight evidence that judgmental forecasting accuracy was related to social-distancing behaviors.

## Misestimation

Why were the day-9 forecasts predominantly underestimations in Study 1, but overestimations in Study 2? Note that in Study 1, participants were presented with data that followed a more exponential trend, whereas in Study 2 participants were presented with a less exponential trend. In light of our successful replication of these results (Study 3), we reason that the behaviors observed in Studies 1 and 2 were not due to increased COVID-19 knowledge as the pandemic progressed, but instead resulted from the structure of the presented data (linearity/exponentiality). Consistent with this reasoning, prior work has shown that underestimation of exponential growth trends increases with an increasing exponent (Wagenaar & Sagaria, 1975; Wagenaar & Timmers, 1979). With a larger exponent, participants underestimated growth trends, and with a smaller exponent, participants actually overestimated growth trends. Thus, the most likely explanation for the deviation in our findings is that it was the difference in exponentiality/linearity of the functions shown to participants that led to this inconsistency.

Another factor that influences over- or underestimation of exponential functions is prior experience. Ebersbach et al. (2008) had children complete an exponential forecasting and a linear forecasting task and varied task order. They found that children's understanding was fragile in that forecasts were highly influenced by order effects. Those who first extrapolated an exponential curve overestimated the growth of a linear function and those who first extrapolated a linear curve underestimated the growth of an exponential function. Given these findings, it could be the case that in our Study 1, in which the function was more exponential, participants were used to extrapolating linear trends (i.e., the "illusion of linearity"), thus producing underestimates when shown an exponential function. By the time of Study 2, participants were more familiar with exponential functions given repeated exposure in the media; thus, they overestimated the growth of the more linear function shown to them in Study 2. However, these hypothesized order effects cannot account for why we *were* able to replicate the pattern of under- and overestimation from Studies 1 and 2 in Study 3.

## Tables versus graphs

Although better accuracy among those shown tables and false confidence in those shown graphs are the most robust findings in this investigation, the underlying causes of the differential effects of tables and graphs on forecasting are less clear. The advantage of tables over graphs for forecasting was somewhat surprising, given the rich literature that may suggest otherwise (Carey & White, 1991; Harvey & Bolger, 1996). For example, modern media tend to visualize data as graphs, and prior work has shown that people work better with visualizations with which they are familiar (Coll et al., 1991), and that graphs are more effective with repeated practice (DeSanctis & Jarvenpaa, 1985). Data were shown to participants as trended functions, and they were asked to produce short-term forecasts. Prior work has shown that graphs are more effective than tables for forecasting trended functions and short-term forecasts, both consistent with the task employed in the current study (Harvey & Bolger, 1996; Lawrence et al., 1985). Thus, it is somewhat surprising that tables consistently led to more accurate forecasts. One possibility is that the advantage of graphs – extracting trends from noisy data – was lost in the context of forecasting based on five data points; however, prior work suggests that more data do not necessarily mean more accurate forecasting (Wagenaar & Timmers, 1978). It is also crucial to remember that much of the prior research has studied graphs versus tables in the context of forecasting linear growth. In the context of forecasting exponential growth, consistent with our findings, prior work has shown that people tend to underestimate exponential trends more when shown graphs compared to tables (Wagenaar & Sagaria, 1975).

In alignment with these findings, since the inception of this work, other researchers have found that showing participants raw COVID-19 case counts (not in tabular form) for weeks 1, 2, and 3 led to increased forecasting accuracy for weeks 4 and 5 compared to viewing graphs of the same data (Banerjee et al., 2021). Future research should further explore the mechanisms by which tables improve the forecasting of exponential functions. One possibility is that participants used an advantageous heuristic when interpreting tables. Padilla et al. (2018) suggest that interpretations of data visualizations are susceptible to visual spatial biases that are driven by bottom-up attention, occurring early in the decision-making process. It could be that the perceptual features of tables make them better for forecasting exponential growth. For example, participants may be better able to see that at each time point ~1 digit is added to the number of cases, thus they may adopt the heuristic of adding a digit for each forecast, which is equivalent to forecasting exponential growth.

## Misestimation and social-distancing behaviors

Overall, our data suggest that forecasts about the cumulative spread of COVID-19 were related to people's attitudes about social distancing in Study 1 and there was a marginal relationship between forecasted cumulative cases and attitudes about social distancing in Study 2. We found some evidence that increased COVID-19 knowledge among the public attenuated the relationship between forecasts and social-distancing behaviors. It is also possible that increased politicization of the virus was driving behaviors in a way that makes it difficult to observe the effect of misestimation on social-distancing behaviors. It is difficult to interpret the Study 3 social distancing results given that the data were presented in the context of a hypothetical country. Our results are mixed, however; since the inception of this work, researchers have found that exponential prediction biases are associated with important COVID-related behaviors such as compliance with safety measures and perceived appropriateness of violating safety measures (Banerjee et al., 2021). In a short intervention, Lammers et al. (2020) showed that increasing understanding of exponential growth led to increased support for social distancing. Thus, our results from Studies 1 and 2 add to the mounting evidence that forecasting virus spread is related to preventative behaviors.

It is possible that the relationship observed between social-distancing and forecasting behaviors is due to a general personality trait, in that more cautious people will overestimate the growth of the pandemic and engage in preventative behaviors. Another interpretation is that understanding the magnitude of exponential growth leads to preventative behaviors as those who are aware of the exponential trajectory are more likely to understand the importance of slowing the spread of the virus. Our results provide evidence that the relationship isn't driven by a general personality trait given that for participants who were in both Study 1 and Study 2, there was a relationship between forecasts and social-distancing behavior in Study 1, but not Study 2. Consequently, social-distancing behaviors were generally not related to forecasts in Study 3, in which participants reasoned about data in a hypothetical context that would not affect their personal decision to socially distance.

## Broader applications

It is important to consider the broader applications of this work given that decision-making is often domain specific (Chapman, 1996). For example, people tend to engage in different cognitive processes when reasoning about health versus financial data (Chapman, 1996; Chapman & Johnson, 1995). Exponential growth bias has been well documented across multiple domains, including economics and financial decision-making (Levy & Tasoff, 2015) as well as reasoning about pandemic-related data (Banerjee et al., 2021; Lammers et al., 2020). Whether tables are better than graphs for forecasting other types of data is less clear. While few studies have examined tables versus graphs in the context of judgmental forecasting, Wagenaar and Sagaria (1975) did find that viewing tables of data led to more accurate forecasts than graphs when participants viewed data on indices of pollution. Thus, we are optimistic that our findings apply to contexts beyond reasoning about health data, though future work should further explore this possibility.

## Limitations

In this work we focus on forecasting the actual number of cases, which is a multivariate problem, and not whether participants are able to extrapolate the trend shown to them. It is possible that participants were inaccurate in their forecasts because they extrapolated the trended series shown to them, without accounting for unique variables associated with COVID-19 such as lockdown and testing. However, we found that participants showed high forecasting error even when their estimates are compared to an extrapolation of the trend shown to them, with a ~52% forecasting error in Study 1 and a ~30% forecasting error in Study 2 (see Fig. 6 and OSM for more details).

Another possibility is that participants failed to notice the difference between the time intervals of the presented data (7-day) and the forecasting task (3-day). This issue could affect the graph group more than the table group as prior work has shown that people often fail to pay close attention to graph axes (Lammers et al., 2020). However, if this were the case, then our participants (especially the graph group) should have consistently overestimated the number of future cases – but they did not. Future work could alter elements of the graph to try to improve forecasting, such as changing the specification of the axes and adding white space to allow participants to visually extrapolate the curve. Future work could also examine the use of interactive graphical interfaces

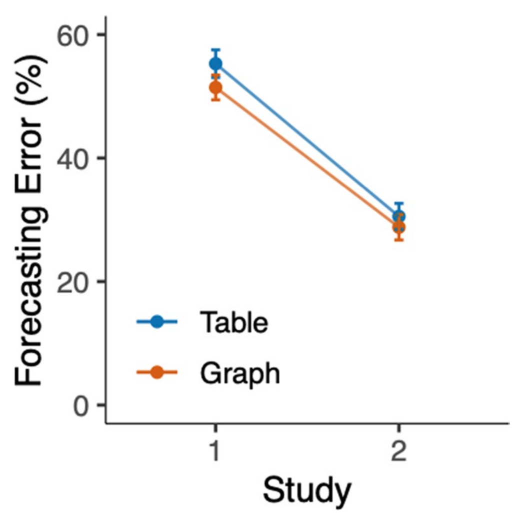



**Fig. 6** Forecasting error. Here we show the average % error of participants forecasts with respect to the predicted values from exponential models fit to the data presented in the stimuli. These results show that whether error is measured with respect to actual or predicted values, forecasting error is quite high

(Edmundson, 1990). For example, Schonger and Sele (2020) found that framing the spread of the disease in terms of doubling times rather than growth rates decreased exponential bias, and that reducing this bias was associated with better understanding of the benefits of non-pharmaceutical interventions such as social distancing and mask-wearing.

Another potential limitation of this research is that the same dates were used in Study 3 as in Study 1 and 2. It is possible that the differing dates may have influenced responses in addition to the different shaped growth curves. However, if participants were retrospectively considering the growth of the virus in weeks past, it is unlikely that participants would have continued to underestimate the growth of the virus especially given their new knowledge on the severity of the pandemic and knowledge that COVID-19 was growing exponentially at the beginning of the pandemic.

## Conclusions

In this investigation we contribute to the literature on data presentation in COVID-19 times as well as the more general forecasting literature. Our consistent finding that participants produced more accurate forecasts when presented with tables rather than graphs adds to the sparse literature on data presentation and extrapolation of exponential functions, and the finding that viewing graphs led to greater confidence in one's inaccurate forecasts is, to our knowledge, a novel contribution of this research that raises interesting questions in settings outside of COVID-19. For example, does showing people graphs of saving accumulation lead to false confidence in one's understanding of how savings accumulate? Our research also suggests that forecasting may be improved with repeated exposure, as participants who participated in Study 1 performed better in Study 2 when compared to participants without prior experience. We also add to the existing literature suggesting that exponential growth functions are underestimated depending on the size of the exponent, with our consistent finding that participants overestimated more linear and underestimated more exponential functions. Lastly, we add to the existing evidence that understanding exponential growth of COVID-19 is related to social-distancing behaviors.

**Supplementary Information** The online version contains supplementary material available athttps://doi.org/10.3758/s13421-022-01288-0.

**Code availability** Code is available on Github at https://github.com/adkinsty/forecasting.

**Authors' contributions** M.F. contributed to the experimental design, data collection, and manuscript preparation. T.J.A. contributed to experimental design, data analysis, and manuscript preparation. M.F. and T.J.A. are both to be considered as first author with equal contribution. J.J., R.L., A.B., M.Q., P.L., and P.S. contributed to experimental design and manuscript preparation.

**Funding** This work was funded by a grant from the National Science Foundation (NSF #2027822, awarded to JJ.

**Data availability** All data and materials are available at https://github.com/adkinsty/forecasting.

## Declarations

**Conflicts of interest** Not applicable.

**Ethics approval** This research was determined to be exempt by the University of Michigan Institutional Review Board.

**Consent to participate** All participants were given a digital consent form prior to participation.

**Consent for publication** Not applicable.

## References

Angus-Leppan, P., & Fatseas, V. (1986). The Forecasting Accuracy of Trainee Accountants Using Judgemental and Statistical Techniques. *Accounting and Business Research*, *16*(63), 179–188. https://doi.org/10.1080/00014788.1986.9729316

Banerjee, R., Bhattacharya, J., & Majumdar, P. (2021). Exponential-growth prediction bias and compliance with safety measures related to COVID-19. *Social Science & Medicine*, *268*, 113473.

Bürkner, P.-C. (2017). **brms**: An *R* Package for Bayesian Multilevel Models Using *Stan*. *Journal of Statistical Software*, *80*(1). https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, *10*(1), 395. https://doi.org/10.32614/RJ-2018-017

Carey, J. M., & White, E. M. (1991). The Effects of Graphical Versus Numerical Response on the Accuracy of Graph-Based Forecasts. *Journal of Management*, *17*(1), 77–96. https://doi.org/10.1177/014920639101700106

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). *Stan*: A Probabilistic Programming Language. *Journal of Statistical Software*, *76*(1). https://doi.org/10.18637/jss.v076.i01

Chapman, G. B. (1996). Temporal discounting and utility for health and money. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(3), 771.

Chapman, G. B., & Johnson, E. J. (1995). Preference reversals in monetary and life expectancy evaluations. *Organizational Behavior and Human Decision Processes, 62,* 300-317.

Coll, R., Thyagarajan, A., & Chopra, S. (1991). An experimental study comparing the effectiveness of computer graphics data versus computer tabular data. *IEEE Transactions on Systems, Man, and Cybernetics*, *21*(4), 897–900. https://doi.org/10.1109/21.108306

De Bock, D., Verschaffel, L., & Janssens, D. (1998). The predominance of the linear model in secondary school students' solutions of word problems involving length and area of similar plane figures. *Educational Studies in Mathematics*, *35*(1), 65–83. https://doi.org/10.1023/A:1003151011999

De Bock, D., Van Dooren, W., Janssens, D., & Verschaffel, L. (2002). Improper use of linear reasoning: An in-depth study of the nature and the irresistibility of secondary school students' errors. *Educational Studies in Mathematics*, *50*(3), 311–334. https://doi.org/10.1023/A:1021205413749

DeSanctis, G. (1984). Computer graphics as decision aids: Directions for research. *Decision Sciences*, *15*(4), 463–487. https://doi.org/10.1111/j.1540-5915.1984.tb01236.x

DeSanctis, G. & Jarvenpaa, S. L. (1985). An Investigation of the "Tables Versus Graphs" controversy in a learning environment. ICIS 1985 Proceedings. 23.http://aisel.aisnet.org/icis1985/23

Ebersbach, M., Lehner, M., Resing, W. C. M., & Wilkening, F. (2008). Forecasting exponential growth and exponential decline: Similarities and differences. *Acta Psychologica*, *127*(2), 247–257. https://doi.org/10.1016/j.actpsy.2007.05.005

Edmundson, R. H. (1990). Decomposition; a strategy for judgemental forecasting. *Journal of Forecasting*, *9*(4), 305–314. https://doi.org/10.1002/for.3980090403

Eggleton, I. R. C. (1982). Intuitive Time-Series Extrapolation. *Journal of Accounting Research*, *20*(1), 68. https://doi.org/10.2307/2490763

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization in Bayesian workflow. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *182*(2), 389–402. https://doi.org/10.1111/rssa.12378

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, *2*(4), 1360–1383. https://doi.org/10.1214/08-AOAS191

Goodwin, P., & Wright, G. (1993). Improving judgmental time series forecasting: A review of the guidance provided by research. *International Journal of Forecasting*, *9*(2), 147–161. https://doi.org/10.1016/0169-2070(93)90001-4

Greenstone, M., & Nigam, V. (2020). Does Social Distancing Matter? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3561244

Harvey, N., & Bolger, F. (1996). Graphs versus tables: Effects of data presentation format on judgemental forecasting. *International Journal of Forecasting*, *12*(1), 119–137. https://doi.org/10.1016/0169-2070(95)00634-6

Harvey, N., & Reimers, S. (2013). Trend damping: Under-adjustment, experimental artifact, or adaptation to features of the natural environment? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(2), 589–607. https://doi.org/10.1037/a0029179

Hegarty, M. (2011). The Cognitive Science of Visual-Spatial Displays: Implications for Design. *Topics in Cognitive Science*, *3*(3), 446–474. https://doi.org/10.1111/j.1756-8765.2011.01150.x

Keren, G. (1983). Cultural differences in the misperception of exponential growth. *Perception & Psychophysics*, *34*(3), 289–293. https://doi.org/10.3758/BF03202958

Kruschke, J. K. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd Ed.). Academic Press.

Lammers, J., Crusius, J., & Gast, A. (2020). Correcting misperceptions of exponential coronavirus growth increases support for social distancing. *Proceedings of the National Academy of Sciences*, *117*(28), 16264–16266. https://doi.org/10.1073/pnas.2006048117

Lawrence, M., & Makridakis, S. (1989). Factors affecting judgmental forecasts and confidence intervals. *Organizational Behavior and Human Decision Processes*, *43*(2), 172–187. https://doi.org/10.1016/0749-5978(89)90049-6

Lawrence, M. J., Edmundson, R. H., & O'Connor, M. J. (1985). An examination of the accuracy of judgmental extrapolation of time series. *International Journal of Forecasting*, *1*(1), 25–35. https://doi.org/10.1016/S0169-2070(85)80068-6

Levy, M. R., & Tasoff, J. (2015). Exponential-growth bias and overconfidence. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2574069

Levy, M., & Tasoff, J. (2016). Exponential-growth bias and lifecycle consumption: Exponential-growth bias and lifecycle consumption. *Journal of the European Economic Association*, *14*(3), 545–583. https://doi.org/10.1111/jeea.12149

Lipkus, I. M., & Hollands, J. G. (1999). The visual communication of risk. *JNCI Monographs*, *1999*(25), 149–163. https://doi.org/10.1093/oxfordjournals.jncimonographs.a024191

Makowski, D., Ben-Shachar, M. S., Chen, S. H. A., & Lüdecke, D. (2019). Indices of effect existence and significance in the Bayesian framework. *Frontiers in Psychology*, *10*, 2767. https://doi.org/10.3389/fpsyg.2019.02767

Makridakis, S., Hyndman, R. J., & Petropoulos, F. (2020). Forecasting in social settings: The state of the art. *International Journal of Forecasting*, *36*(1), 15–28. https://doi.org/10.1016/j.ijforecast.2019.05.011

Mckenzie, C. R. M., & Liersch, M. J. (2011). Misunderstanding Savings Growth: Implications for Retirement Savings Behavior. *Journal of Marketing Research*, *48*(SPL), S1–S13. https://doi.org/10.1509/jmkr.48.SPL.S1

Okan, Y., Galesic, M., & Garcia-Retamero, R. (2016). How people with low and high graph literacy process health graphs: Evidence from eye-tracking: Graph literacy and health graph processing. *Journal of Behavioral Decision Making*, *29*(2–3), 271–294. https://doi.org/10.1002/bdm.1891

Padilla, L. M., Creem-Regehr, S. H., Hegarty, M., & Stefanucci, J. K. (2018). Decision making with visualizations: A cognitive framework across disciplines. *Cognitive Research: Principles and Implications*, *3*(1), 29. https://doi.org/10.1186/s41235-018-0120-9

Quirk, M., Fansher, M., Adkins, T.J., Lalwani, P., Boduroglu, A., Lewis, R.L., Shah, P. & Jonides, J. (2021). Conservatism, caution, and the COVID-19 pandemic. [Manuscript in preparation].

Schonger, M., & Sele, D. (2020). How to better communicate the exponential growth of infectious diseases. *PLoS One, 15*(12), e0242839.

Stango, V., & Zinman, J. (2009). Exponential Growth Bias and Household Finance. *The Journal of Finance*, *64*(6), 2807–2849. https://doi.org/10.1111/j.1540-6261.2009.01518.x

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

van der Linden, S. L., Leiserowitz, A. A., Feinberg, G. D., & Maibach, E. W. (2014). How to communicate the scientific consensus on climate change: Plain facts, pie charts or metaphors? *Climatic Change*, *126*(1–2), 255–262. https://doi.org/10.1007/s10584-014-1190-4

Van Dooren, W., De Bock, D., Depaepe, F., Janssens, D., & Verschaffel, L. (2003). The illusion of linearity: Expanding the evidence towards probabilistic reasoning. *Educational Studies in Mathematics*, *53*(2), 113–138. https://doi.org/10.1023/A:1025516816886

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Wagenaar, William A., & Sagaria, S. D. (1975). Misperception of exponential growth. *Perception & Psychophysics*, *18*(6), 416–422. https://doi.org/10.3758/BF03204114

Wagenaar, W. A., & Timmers, H. (1978). Extrapolation of exponential time series is not enhanced by having more data points. *Perception & Psychophysics*, *24*(2), 182–184. https://doi.org/10.3758/BF03199548

Wagenaar, Willem A., & Timmers, H. (1979). The pond-and-duckweed problem; Three experiments on the misperception of exponential growth. *Acta Psychologica*, *43*(3), 239–251. https://doi.org/10.1016/0001-6918(79)90028-3

Wise, T., Zbozinek, T. D., Michelini, G., Hagan, C. C., & Mobbs, D. (2020). *Changes in risk perception and protective behavior during the first week of the COVID-19 pandemic in the United States* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/dz428

Zacks, J. M., & Franconeri, S. L. (2020a). Designing Graphs for Decision-Makers. *Policy Insights from the Behavioral and Brain Sciences*, *7*(1), 52–63. https://doi.org/10.1177/2372732219893712

Zacks, J. M., & Franconeri, S. L. (2020b). Reading the pandemic data. *Editor's Blog*. https://blogs.sciencemag.org/editors-blog/2020/05/12/reading-the-pandemic-data/. Accessed Oct 2021

**Open practices statement** Pre-registration forms were created for all three studies:

Study 1 - https://aspredicted.org/blind.php?x=cd4a7h
Study 2 - https://aspredicted.org/blind.php?x=py8qp2
Study 3 - https://aspredicted.org/blind.php?x=74SD4t