Visualizing a Pandemic: Lessons from COVID-19 About Data and Decisions

by

Madison J. Adkins

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Psychology)
in the University of Michigan
2023

Doctoral Committee:

Professor Priti Shah, Chair
Associate Professor Eytan Adar
Professor Aysecan Boduroglu, Koc University
Professor John Jonides
Professor Richard L. Lewis

Madison J. Adkins

mfansher@umich.edu

ORCID iD: 0000-0003-0009-4337

Dedication

to Tyler and my family

Acknowledgements

I would first like to thank Dr. Priti Shah, Dr. John Jonides, Dr. Rick Lewis, Dr. Aysecan Bodurolgu, and Dr. Eytan Adar for serving on my dissertation committee. I would like to thank my advisor, Dr. Priti Shah, for her guidance, support, and compassion over the last five years. I have found her unending curiosity and excitement about applications of psychology research to real world issues to be incredibly inspiring. I would also like to thank my secondary advisor, Dr. John Jonides for his invaluable guidance and genuine excitement and enthusiasm about the research we conducted together. I would like to thank Dr. Priti Shah, Dr. John Jonides, Dr. Rick Lewis, Dr. Aysecan Boduroglu, Dr. Han Zhang, Dr. Tyler Adkins, Dr. Pia Lalwani, Madelyn Quirk, and Madison Carlson who all contributed to the research I present in Chapters 2-4. Being able to work with such an exceptional group of people during a time of uncertainty, and finding inspiration in research together, was such a memorable experience. Thank you to the National Science Foundation for funding my time at Michigan as a Graduate Research Fellow and for funding the work I present in Chapters 2-4 (RAPID: Factors that affect understanding the risks of COVID-19). I would also like to thank the University of Michigan Psychology Department and members of the Basic and Applied Cognition Lab for their support over the years. Thank you to Dr. Sébastien Hélie and Dr. Thomas Redick who provided me with invaluable mentorship as an undergraduate and inspired me to apply to graduate school. Thank you to the CCN students who I have had so much fun with over the years – especially Natasha Vernooij, Kathy Xie, and Madelyn Quirk for their support and friendship. Thank you to my cohort buddy and friend – EunSeon Emery-Ahn for growing with me over the last five years and for the unending laughs and crafting sessions. Thank you to my husband, Tyler, for teaching me programming and analysis skills that were critical for completing this dissertation. Thank you for being my biggest supporter and best friend. Thank you to my family for their support over the years. I would especially like to thank my dad, Michael, and my mother, Kathryn.

Table of Contents

List of Tables

List of Figures

List of Appendices

Abstract

Researchers have long studied how data visualizations influence risk perception. It's possible that the spread of COVID throughout the population could have been reduced by using best practices from the data visualization and risk perception literatures to develop public communications about the severity of COVID that were effective and unbiased. In this dissertation I discuss three lines of research examining how people understood common data visualizations presented to the public during the pandemic and how understanding of COVID data was related to attitudes towards preventative measures. In Chapter 2, I show that people over- or under-estimated the exponential growth of COVID depending on the linearity of the data, that viewing tables of data improved forecasting accuracy compared to graphs, that viewing graphs was associated with false confidence in one's forecasts, and some evidence that attitudes towards social distancing was positively correlated with the magnitude of participants' forecasts. In Chapter 3, I show that people misunderstood the relationship between daily and cumulative case curves and that participating in a brief video intervention improved understanding of accumulation. The effects of the intervention were long lasting and transferred to contexts outside of COVID. Participating in the intervention was also associated with more favorable attitudes towards social distancing and social distancing policies. In Chapter 4, I show that viewing icon arrays illustrating the 1 in 1 million chance of experiencing the reported side effect from the Johnson & Johnson vaccine prevented significant increases in aversion towards the Johnson & Johnson vaccine as well as all COVID vaccines. Lastly, in Chapter 5, I provide a synthesis of the literature conducted during the pandemic on how people understood COVID visualizations and describe three main findings: (1) people misunderstood commonly used COVID visualizations, (2) data visualizations influenced risk perception, and (3) graphs were sometimes used to mislead the public during the pandemic. This research informs how data should be communicated with the public and provides guidelines for how data should be explained to the public with visualizations.

Chapter 1 Introduction

COVID-19 was introduced to the United States in early 2020 and has resulted in over 100 million cases and 1.1 million deaths among Americans at the time of writing this dissertation. The political polarization of COVID (Allcott et al., 2020; Calvillo et al., 2020; Christensen et al., 2020; Lammers et al., 2020), the rapid spread of misinformation through social media (Roozenbeek et al., 2020; van der Linden et al., 2020), the underestimation of COVID-related risks (Schlager & Whillans, 2022) and overestimation of risk associated with preventative measures like vaccination (Sallam, 2021) all contributed to the spread of COVID throughout the population. It's possible that many of these cases and deaths could have been prevented by using best practices from the data visualization and risk perception literatures to communicate the severity of COVID with the public. In this dissertation I discuss three lines of research examining how people understood common data visualizations presented to the public during the pandemic and how understanding of COVID data was related to attitudes towards preventative measures (Chapters 2-4). I conclude with a review of the body of literature on how people understood COVID visualizations (Chapter 5) followed by the theoretical contributions and broader application of my research (Chapter 6).

COVID Visualizations in the Media

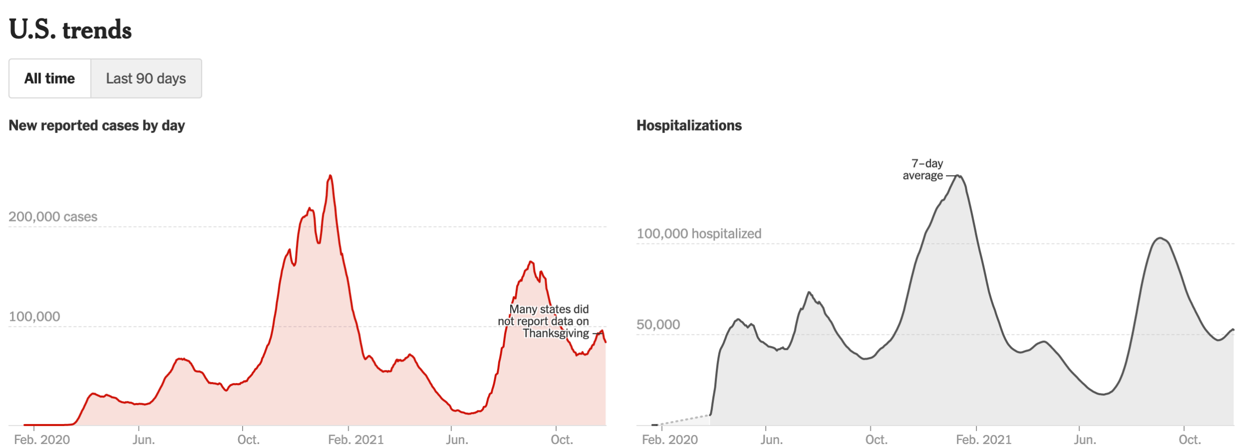
The COVID-19 pandemic brought the challenge of explaining biological, medical, and statistical concepts to a public audience with varying levels of literacy about these subjects. Understanding the growth patterns of COVID-19 has been a critical task for U.S. citizens since 2020, in that making well-informed, 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 illustrating vital pieces of information such as cumulative, daily, and active cases, as well as hospitalizations, number of vaccines administered, positivity rates, rate of testing, and number of deaths (Zacks & Franconeri, 2020a, 2020b). Kwon et al. (2021) found that reporting on COVID was twice as likely to include graphs compared to other news stories, with line and bar graphs being the most common types of visualizations. COVID graphs can be more generally categorized as showing temporal, geospatial, or multivariate data

(Zhang et al., 2021). Use of such visualizations allows for the communication of extensive information that would be difficult or impossible to illustrate with text alone. In an analysis of 668 COVID visualizations, Zhang et al. (2021) identified six key messages conveyed by COVID visualizations, including (1) informing of the severity; (2) forecasting trends and influences; (3) explaining the nature of the crisis; (4) guiding risk mitigation; (5) communicating risk, vulnerability, and equity; and (6) gauging the multifaceted impacts of the crisis.

The Importance of Graph Literacy

While graphical displays have been powerful tools used by the media throughout the pandemic to communicate the latest news, these visualizations aren't always properly understood. COVID-19 graphs are often complex, describing nonlinear functions with various logarithmic and smoothing functions applied, some with multiple y-axes, and others incorporating multiple plots into one. For example, **Figure 1** shows images from the New York Time's COVID-19 dashboard, where to understand the presented graphs, one must recognize that although the graphs are side-by-side, the y-axes are on different scales, and the right panel illustrates 7-day averages while the left panel does not.

Figure 1 *Graphs from the New York Times COVID-19 Dashboard in November 2021*



An estimated one third of Americans have low graph literacy (Galesic & Garcia-Retamero, 2011), with low graph literacy especially affecting some marginalized groups in the United States (Rodríguez et al., 2013). Graph interpretation is a complex task with many challenges (Franconeri et al., 2021; Glazer, 2011; Shah et al., 2005), thus solely relying on these visualizations to communicate the state of the pandemic without explanation may have been problematic. Fan et al. (2022) assessed how everyday people interpreted seven different types of common pandemic visualizations, such as heat maps, tables, and line graphs illustrating data like number of deaths over time and cases. They examined the depth of participant's understanding of these visualizations and found that participants commonly incorrectly understood the visualizations and were confident in their incorrect interpretations. The media rarely provided scaffolding to help the public understand COVID visualizations, for example, failing to explain the meaning of logarithmic case or death data (Hammes et al., 2021). Tabak and Dubovi (2021) examined participants' graph usage and skills during the pandemic and found that people self-reported using visualizations such as graphs, tables, and interactive visualizations more during COVID than prior to the pandemic regardless of graph literacy, however those with better graph interpretation skills reported a larger increase in the use of visualizations than participants with lower graph interpretation skills.

Visualizations and Risk Perception

Researchers have long studied how risk perception is affected by the presence and understanding of data visualizations (Ancker et al., 2006; Fagerlin et al., 2011; Hawley et al., 2008; Zipkin et al., 2014). Public health messaging tools like "Flatten the Curve" may have been less effective than intended because the general public simply did not understand the visualizations being presented to them. It's possible that misunderstanding these visualizations led to low-risk perception by some Americans and therefore more risky behaviors like continuing to socialize at a time when cases and deaths were growing exponentially and there were no vaccines readily available. For example, the most popular visualizations used by the media were graphs of cumulative cases, even though there is a substantial literature finding that people generally fail to understand the concept of accumulation regardless of motivation, graph literacy, cognitive capacity, education, and domain experience (Brunstein et al., 2010; Cronin et al., 2009; Sweeney & Sterman, 2000). It's possible that Americans misunderstood these data because they did not understand the concept of a cumulative function.

Research Questions

The introduction of COVID-19 in the United States at the beginning of 2020, along with the ensuing pandemic, brought a unique opportunity to study principles established in the visualization and risk perception literature in the "wild". In this scenario I could study how the presence and interpretation of data visualizations influenced real-life risky behaviors like failing to wear a mask or to socially distance. The use of large online subject recruitment platforms with thousands of participants allowed me to study these topics in samples that were representative of the U.S. population at a unique time point in history.

In this work I combine three lines of research in which I study how people understand COVID-19 data and visualizations, whether explanations of data through visualization improve understanding of COVID-related concepts, and lastly, whether understanding of data is related to attitudes towards preventative actions like engaging in social distancing, mandated social distancing by governments, and vaccination.

In Chapter 2 (Fansher et al., 2022b) I present three experiments in which participants were tasked with predicting the future number of cumulative COVID cases based on the most recent COVID data. These experiments were conducted at the beginning of the pandemic in March and April of 2020. I hypothesized that participants would underestimate the growth of the virus (i.e., predict more linear growth) due to the existing judgmental forecasting literature showing that people tend to underestimate exponential growth.

In Chapter 2, participants were tasked with extrapolating cumulative COVID case curves. I was surprised to have found that 27% of participants misunderstood the task at hand and predicted that the cumulative curve could *decrease*, suggesting that participants misunderstood the concept of accumulation as it related to COVID cases. This finding motivated Chapter 3 in which I study how people understand the relationship between daily and cumulative COVID cases. In this Chapter 3 I implemented an 8-minute narrated and animated video intervention to teach participants about the concept of accumulation and how it relates to the idea of "flattening" the cumulative curve.

Lastly, in April 2021 the CDC announced that there would be a pause in the administration of the Johnson & Johnson (J&J) vaccine due to reports of a rare blood clotting side effect. In Chapter 4 (Fansher et al., 2022c) I examine whether the probability language (i.e., 1 in 1 million, .0001%, 6 people) used to convey this message would impact attitudes towards vaccination and whether viewing icon arrays illustrating the 1 in 1 million probability of incurring the rare side effect would prevent increases in vaccine hesitancy.

Chapter 2 How Well Do Ordinary Americans Forecast the Growth of COVID-19?

2.1 Introduction

Consider this one problem: As of March 1st, 2020, the CDC reported that there had been 75 cases of COVID-19 in the United States. On March 18th, there were 7,038 confirmed cases. How many cases would there be on March 25th? April 1st? April 8th? Due to the dynamic nature of the spread of COVID-19 at the beginning of the pandemic, it was argued that forecasting the future of the disease with accuracy was a uniquely difficult challenge (Makridakis et al., 2020). Research from the first week of the pandemic (March 11-16, 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. 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.

Understanding Exponential Growth

In judgmental forecasting tasks, participants are shown time-series data 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 & Timmers, 1978, 1979; Wagenaar & Sagaria, 1975), 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 vs. 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 displaying 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, 7 found graphs to be better than tables, and 10 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 vs graphs (see DeSanctis, 1984), and little work examining tables vs 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, 1983).

2.2 Study 1

On March 28, 2020, participants were shown the cumulative COVID-19 case data from Feb. 29, 2020 to March 27, 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 U.S. 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 vs. 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 in 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

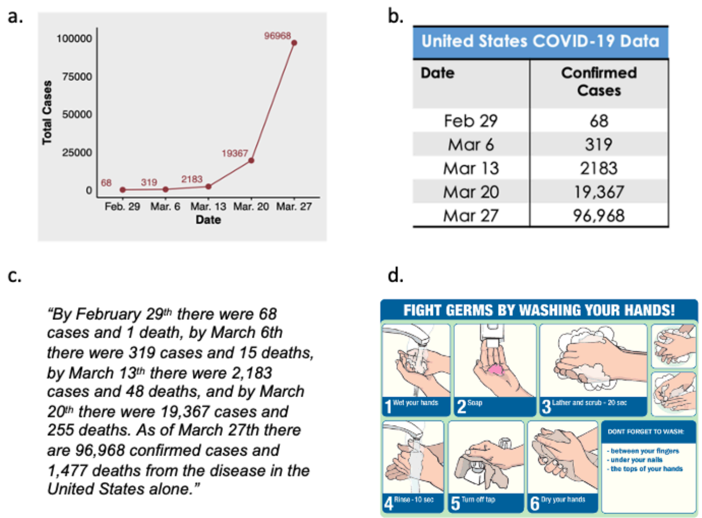
2.2.1 Methods

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, *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 (**Figure 2a**), tables (**Figure**

2b), or as raw data embedded in text (**Figure 2c**). Those in the text-only group viewed a guide on proper hand-washing technique to serve as a control image (**Figure 2d**).

Figure 2. *(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*



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 **Appendix A.** Participants read a short vignette about COVID-19 in the United States. Participants viewed data on the total number of deaths and confirmed cases of COVID-19 in the U.S. in either a graphical format (N = 409), tabular format (N = 408), or text-only format (N = 381). The five data points shown to participants were from the five weeks preceding the date of the study (March 27th; **Figure 2a-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 **Appendix B** for measures). A subset of these questions is examined in the current work and concern 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))

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 prep; 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 U.S. 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 United States 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 $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 **Appendix C** 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, N = 3
2. did not provide a valid zip code (i.e., possibly not a U.S. 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 U.S. President (free response), N = 11
6. self-reported investing effort of less than 5 out of 10, N = 3
7. took less than 30 seconds to complete the task (considered impossible based on the number of survey items), N = 0

And we exclude 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 (< 1000 cases).

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 **Appendix D** 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.

Whereas most forecasts predicted under one million cases, a small number of outlier forecasts were as large as 40 million (N = 4 participants). Upon inspection of the forecast distributions, we found that using a cutoff of 10 times the last datum provided to participants neatly eliminated outliers without affecting the distribution. The cutoff corresponded to forecasts of roughly one million cases for Study 1 and its replication and forecasts of roughly four 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 **Appendix E** 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 **Appendix F**). 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 (Mehrabian, 1996). 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. 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:

$$y \sim \mathrm{Gamma}(\mu, \alpha)$$

$$\log(\mu) = \beta_0 + \beta X + \beta_0^{\mathrm{state}}$$

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, β_0 is the population intercept, β_0^{state} is a state's 'random' intercept, X denotes the predictors (delay, group, delay*group) and β denotes the corresponding population-level regression coefficients. The auxiliary shape parameter of the gamma distribution is denoted by α. We assigned the following weakly informative default priors to the model parameters (Gelman et al., 2008):

$$\beta_0 \sim \mathrm{Student\_t}(3,0,2.5)$$

$$\beta \sim \mathrm{Student\_t}(3,0,2.5)$$

$$\beta_0^{\mathrm{state}} \sim \mathrm{Normal}(0, \sigma_{\mathrm{state}})$$

$$\sigma_{\mathrm{state}} \sim \mathrm{HalfStudent\_t}(3,0,2.5)$$

$$\alpha \sim \mathrm{Gamma}(0.01,0.01)$$

All MCMC chains passed visual inspection, all $\hat{R}$ values were 1, and all effective sample sizes (ESS) were greater than 10k, 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 (sd), 95% credible intervals (CI) 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 ($\mathrm{M_{diff}}$, in native units) as well as standardized regression coefficients (β_{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:

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

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:

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

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 ($\mathrm{P_{under}}$) and 95% credible intervals (CI) 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 large misestimation when comparing their forecasts to the predicted values of exponential models fit to the initial five data-points provided (**Appendix G**).

2.2.2 Results

On average, participants underestimated the number of cases on March 30th ($\mathrm{P_{under}} = 0.83, \mathrm{CI} = [0.79, 0.83], \mathrm{M_{est}} = 141\text{k cases},\ \mathrm{se_{est}} = 2.7\text{k},\ \text{Truth} = 166\text{k}$), April 2nd

($\mathrm{P_{under}} = 0.77, \mathrm{CI} = [0.73{,}0.81], \mathrm{M_{est}} = 207\mathrm{k}$ cases, $\mathrm{se_{est}} = 5.4\mathrm{k}$, $\mathrm{Truth} = 248\mathrm{k}$), and April 5th ($\mathrm{P_{under}} = 0.78, \mathrm{CI} = [0.74{,}0.82], \mathrm{M_{est}} = 270\mathrm{k}$ cases, $\mathrm{se_{est}} = 8.1\mathrm{k}, \mathrm{Truth} = 341\mathrm{k}$) (**Figure 3C**). Critically, the Table group produced more accurate estimates than the Graph group ($\mathrm{M_{diff}} = 14\mathrm{k}$, $\beta_{\mathrm{T-G}} = 0.05$, $\mathrm{sd} = 0.02$, $\mathrm{CI_{95\%}} = [0.01{,}0.10]$, $\mathrm{pd} = 0.99$) (**Figure 3C**). Further, the Table group forecasted greater growth in the number of cases from March 30th to April 2nd than the Graph group ($\beta_{\mathrm{A2-M30*T-G}} = 0.11$, $\mathrm{sd} = 0.06$, $\mathrm{CI_{95\%}} = [0.00{,}0.23]$, $\mathrm{pd} = 0.98$) (**Figure 3C**). However, the two groups forecasted similar increases in cases from April 2nd to April 5th ($\beta_{\mathrm{A5-M2*T-G}} = -0.01$, $\mathrm{sd} = 0.06$, $\mathrm{CI_{95\%}} = [-0.12{,}0.11]$, $\mathrm{pd} = 0.59$). We found that participants in a text-only control group produced virtually identical forecasts (on average) to those in the table group for March 30th (Text: ~145k cases, Table: ~141k, Graph: ~142k), April 2nd (Text: ~220k, Table: ~218k, Graph: ~195k), and April 5th (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 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 ($\beta_T = -0.01, \mathrm{CI_{95\%}} = [-0.06, 0.04], \mathrm{pd} = 0.66$).

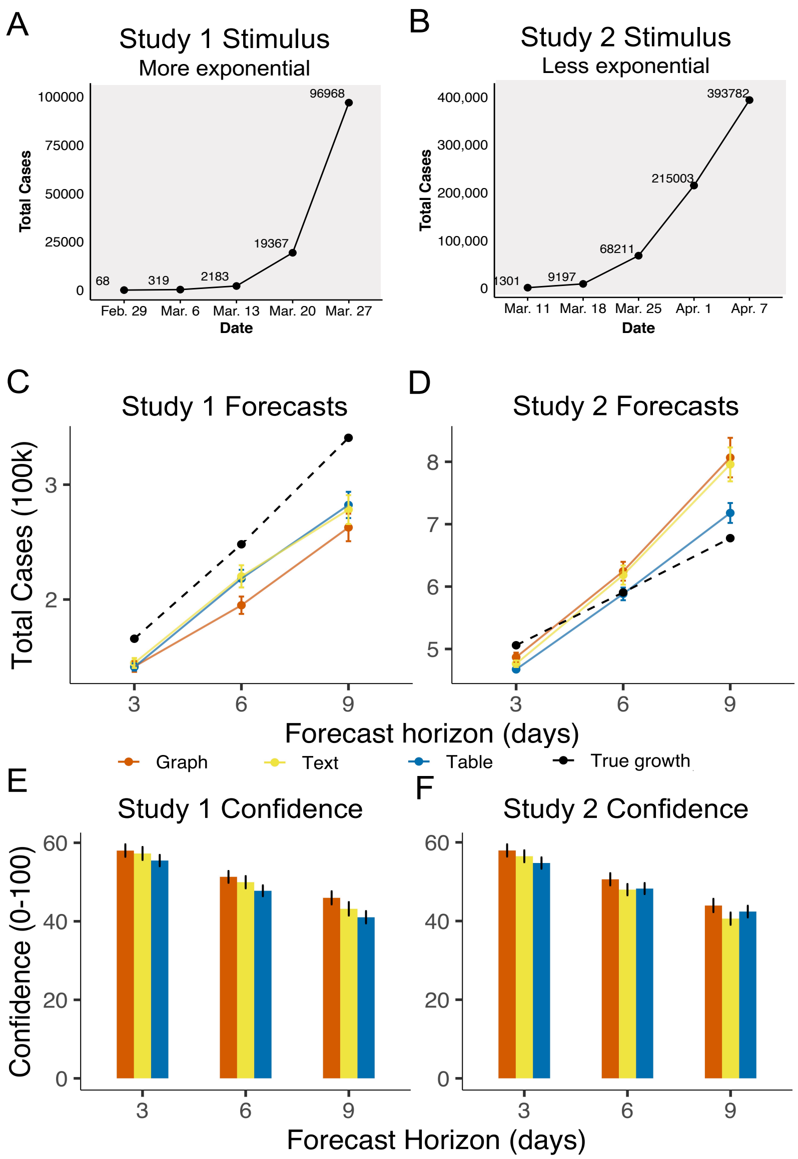
Although tables facilitated more accurate forecasting compared to graphs, people shown tables were less confident in their forecasts than people shown graphs ($\mathrm{M_{diff}} = -3.6$, $\beta_{\mathrm{T-G}} = -0.16, \mathrm{CI_{95\%}} = [-0.26, -0.07], \mathrm{pd} = 1.0$) (**Figure 3E**). Overall, confidence decreased over time as participants were less confident about their April 2nd forecast than their March 30th forecast ($\mathrm{M_{diff}} = -7.3, \beta_{\mathrm{A2-M30}} = -0.29, \mathrm{CI_{95\%}} = [-0.40, -0.18], \mathrm{pd} = 1.0$) and their April 5th forecast compared to their April 2nd forecast ($\mathrm{M_{diff}} = -5.9, \beta_{\mathrm{A5-A2}} = -0.23, \mathrm{CI_{95\%}} = [-0.35, -0.11], \mathrm{pd} = 1.0$). 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 March 30th (Graph: 58.0, Text: 57.3, Table: 55.5), April 2nd (Graph: 51.3, Text: 50.0, Table: 47.8), and April 5th (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 ($\beta_T = -0.07, \mathrm{CI}_{95\%} = [-0.17, 0.03], \mathrm{pd} = 0.90$).

2.2.3 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 vs. graphs in the context of extrapolating exponential growth; however, Wagenaar & 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.

Figure 3. *Results from Studies 1 and 2*



Note. 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 US according to worldometers.info, accessed on April 27th, 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.

2.3 Study 2

Study 2 aimed to replicate the findings from Study 1 in the context of newer data about the pandemic (trend up until April 8th) 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, 1978). Ten days after launching Study 1 (April 7th) we administered the online survey again to people in the U.S. with updated data to reflect the growth of COVID-19 in the U.S. from March 11th to April 7th (**Figure 4A-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 at https://aspredicted.org/blind.php?x=py8qp2

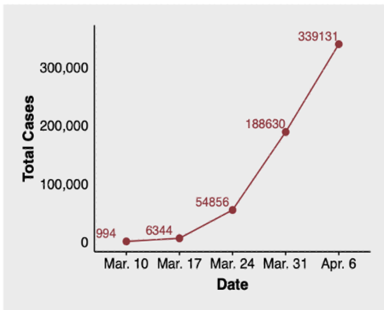
2.3.1 Methods

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

Figure 4. *(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 2d.*

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."

Materials. The materials were the same as in Study 1, except the five data points presented to participants were for the five weeks preceding April 7th instead of the weeks preceding March 27th (**Figure 4A-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.

2.3.2 Results

Overall, participants *underestimated* the number of cases on April 10th ($\mathrm{P_{under}} = 0.83, \mathrm{CI} = [0.80, 0.87], \mathrm{M_{est}} = 476\mathrm{k\ cases},\ \mathrm{se_{est}} = 3.9\mathrm{k},\ \mathrm{Truth} = 506\mathrm{k}$) but *overestimated* the number of cases on April 13th ($\mathrm{P_{under}} = 0.54, \mathrm{CI} = [0.49, 0.59], \mathrm{M_{est}} = 602\mathrm{k\ cases},\ \mathrm{se_{est}} = 8.7\mathrm{k},\ \mathrm{Truth} = 591\mathrm{k}$) and April 16th ($\mathrm{P_{under}} = 0.51, \mathrm{CI} = [0.46, 0.56], \mathrm{M_{est}} = 752\mathrm{k\ cases},\ \mathrm{se_{est}} = 16.3\mathrm{k},\ \mathrm{Truth} = 678\mathrm{k}$) (**Figure 3D**). On average the Table group was more accurate than the Graph group ($\mathrm{M_{diff}} = 46.5\mathrm{k},\ \beta_{\mathrm{T-G}} = -0.06,\ \mathrm{sd} = 0.01, \mathrm{CI} = [-0.08, -0.03],\ \mathrm{pd} = 1.0$) (**Figure 3D**) and there was a group by forecast horizon interaction, such that the Table group forecasted smaller increases in cases from April 13th to April 16th than the Graph group

($\beta_{\mathrm{A16-A13*T-G}} = -0.05$, $\mathrm{sd} = 0.04, \mathrm{CI} = [-0.12, 0.02]$, $\mathrm{pd} = 0.93$) (**Figure 3D**). The two groups forecasted similar increases in cases from April 10th to April 13th ($\beta_{\mathrm{A5-M2*T-G}} = -0.01$, $\mathrm{sd} = 0.04, \mathrm{CI} = [-0.08, 0.06]$, $\mathrm{pd} = 0.64$). We found that participants in a text-only control group produced very similar forecasts (on average) to those in the graph group for April 10th (Text: 476k cases, Graph: 487k, Table: 467k), April 13th (Text: 618k, Graph: 625k, Table: 588k), and April 16th (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 ($\beta_T = -0.05, \mathrm{CI}_{95\%} = [-0.08, -0.02], \mathrm{pd} = 1.0$), while participants in the Graph group produced forecasts of roughly equal magnitude to those in the Text group ($\beta_G = 0.01, \mathrm{CI}_{95\%} = [-0.02, 0.05], \mathrm{pd} = 0.82$).

The Table group was overall less confident (0-100) in their forecasts than the Graph group ($\mathrm{M_{diff}} = -2.5, \beta_{\mathrm{T-G}} = -0.11, \mathrm{sd} = 0.05, \mathrm{CI} = [-0.19, -0.01]$, $\mathrm{pd} = 0.99$) (**Figure 3F**). Participants were predictably less confident in their more distal forecasts as they were less confident about their April 13th forecast than their April 10th forecast ($\mathrm{M_{diff}} = -6.7, \beta_{\mathrm{A13-A10}} = -0.27, \mathrm{sd} = 0.06, \mathrm{CI} = [-0.38, -0.16], \mathrm{pd} = 1.0$) and their April 16th forecast than their April 13th forecast ($\mathrm{M_{diff}} = -6.1, \beta_{\mathrm{A16-A13}} = -0.24, \mathrm{sd} = 0.06, \mathrm{CI} = [-0.35, -0.13], \mathrm{pd} = 1.0$) (**Figure 3F**). We found that participants in the text-only control group reported intermediate confidence (on average) compared to participants in the other groups for April 10th (Graph: 57.9, Text: 56.5, Table: 54.7) but lower confidence (on average) compared to the other groups for April 13th (Graph: 50.6, Table: 48.3, Text: 48.0) and April 16th (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 ($\beta_G = 0.11, \mathrm{CI}_{95\%} = [0.01, 0.21], \mathrm{pd} = 0.99$), while participants in the Table group displayed approximately equal confidence compared to the Text group ($\beta_T = 0.02, \mathrm{CI}_{95\%} = [-0.07, 0.12], \mathrm{pd} = 0.69$).

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 ($\beta_{2-1} = -0.67, \mathrm{sd} = 0.03, \mathrm{CI} = [-0.73, -0.62], \mathrm{pd} = 1.0$) and that forecasting error was greater for the Graph group compared to the Table group ($\beta_{\mathrm{G-T}} = 0.13, \mathrm{sd} = 0.03, \mathrm{CI}_{95\%} = [0.07, 0.19], \mathrm{pd} = 1.0$) (**Figure 5A**). 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) ($\beta_{2-1*\mathrm{R-N}} = -0.10, \mathrm{sd} = 0.06, \mathrm{CI} = [-0.22, 0.01], \mathrm{pd} = 0.96$) (**Figure 5A**). 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) ($\beta_{2-1*\mathrm{R-N}*\mathrm{G-T}} = -0.10, \mathrm{sd} = 0.06, \mathrm{CI} = [-0.56, -0.10], \mathrm{pd} = 1.0$) (**Figure 5A**). 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.

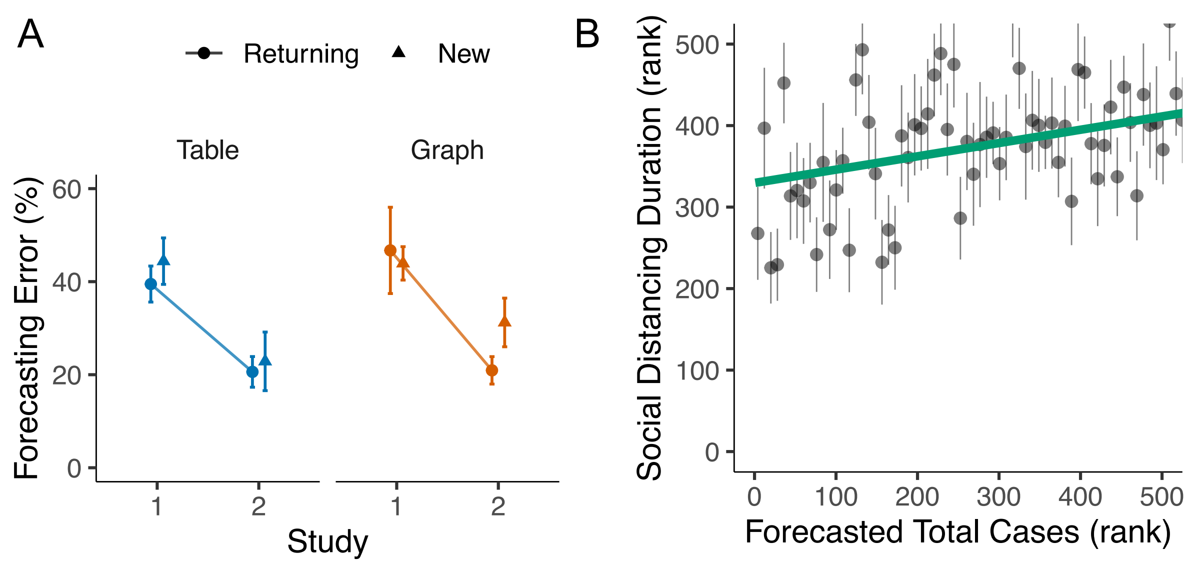
2.3.3 Discussion

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 April 7th, 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 themselves 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 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 under-estimation 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 under-estimation was due to the linearity/exponentiality of the data themselves, we would expect to replicate this pattern of over- and under-estimation.

Figure 5. *Forecast training and the link between forecasting and attitudes towards social distancing.*



Note. 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 Figure 3A & 3B), 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).

2.4 Study 3

Given that we wanted to show participants the exact stimuli from Studies 1 and 2, all mentions of the U.S. 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 U.S. 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 under-estimation 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 over-estimation regardless of whether participants view the data from Study 1 or 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

2.4.1 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, *SD*: 12.0; 57.1% Male, 42.9% Female). Four hundred forty-two 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 (**Figure 2**; N = 411) or the data from Study 2 (**Figure 4**; 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 \$0.75 instead of \$1 after survey completion.

2.4.2 Results

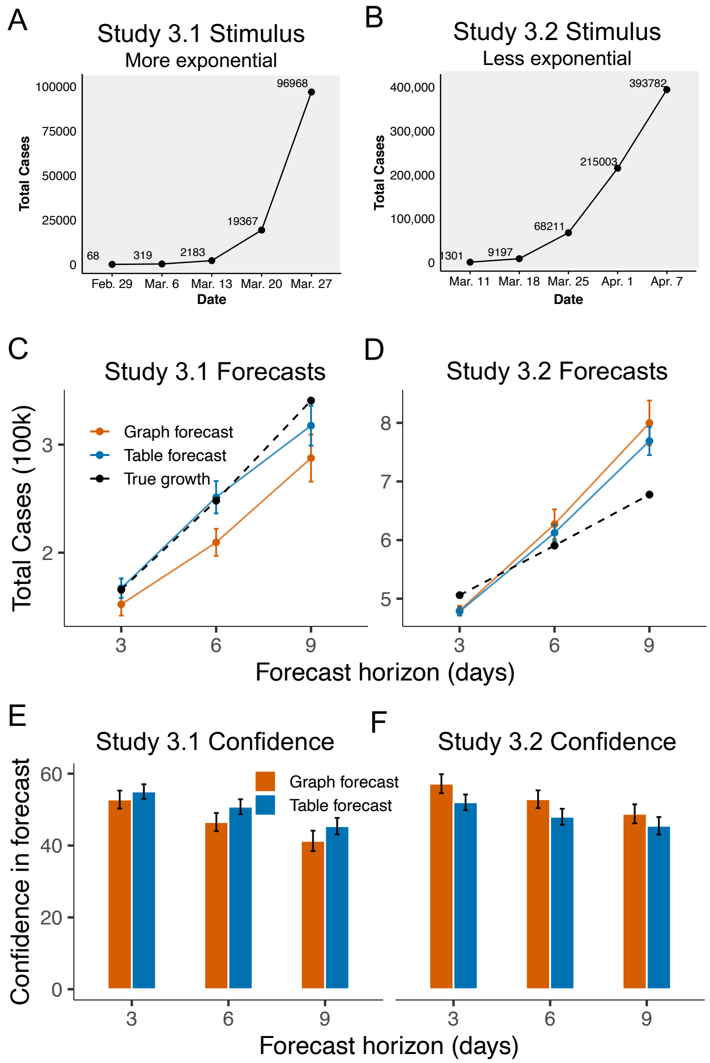
In the replication of Study 1, in which the presented data exhibited more exponential growth, participants underestimated the number of COVID-19 cases for March 30th ($\mathrm{P_{under}} = 0.79, \mathrm{CI} = [0.73, 0.85], \mathrm{M_{est}} = 161\mathrm{k}$ cases, $\mathrm{SE_{est}} = 6.85\mathrm{k}$, $\mathrm{Truth} = 166\mathrm{k}$), April 2nd ($\mathrm{P_{under}} = 0.69, \mathrm{CI} = [0.62, 0.75]$, $\mathrm{M_{est}} = 233\mathrm{k}$ cases, $\mathrm{SE_{est}} = 10.2\mathrm{k}$, $\mathrm{Truth} = 248\mathrm{k}$), and April 5th ($\mathrm{P_{under}} = 0.71, \mathrm{CI} = [0.64, 0.77]$) (**Figure 6C**). 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 U.S. for April 10th ($\mathrm{P_{under}} = 0.84, \mathrm{CI} = [0.79{,}0.90], \mathrm{M_{est}} = 479\mathrm{k}$ cases, $\mathrm{SE_{est}} = 5.29\mathrm{k},\ \mathrm{Truth} = 506\mathrm{k}$), slightly *overestimate* the number of cases for April 13th ($\mathrm{P_{under}} = 0.65, \mathrm{CI} = [0.47{,}0.63], \mathrm{M_{est}} = 619\mathrm{k}$ cases, $\mathrm{SE_{est}} = 13.7\mathrm{k},\ \mathrm{Truth} = 591\mathrm{k}$), and *overestimate* the number of cases for April 16th ($\mathrm{P_{under}} = 0.47, \mathrm{CI} = [0.49{,}0.55], \mathrm{M_{est}} = 783\mathrm{k}$ cases, $\mathrm{SE_{est}} = 21.6\mathrm{k},\ \mathrm{Truth} = 678\mathrm{k}$) (**Figure 6D**).

Overall, the Table groups were more accurate than the Graph groups as they underestimated less in the replication of Study 1 ($\mathrm{M_{diff}} = 28.2\mathrm{k},\ \beta_{\mathrm{T-G}} = 0.15,\ \mathrm{sd} = 0.04, \mathrm{CI} = [0.07{,}0.23],\ \mathrm{pd} = 1.0$) and overestimated less in the replication of Study 2 ($\mathrm{M_{diff}} = -19.7\mathrm{k},\ \beta_{\mathrm{T-G}} = -0.04,\ \mathrm{sd} = 0.02, \mathrm{CI} = [-0.08, -0.01], \mathrm{pd} = 0.95$) (**Figure 6C-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 ($\mathrm{M_{diff}} = 3.6, \beta_{\mathrm{T-G}} = 0.18, \mathrm{sd} = 0.08, \mathrm{CI}_{95\%} = [\,0.02,\ 0.33], \mathrm{pd} = 0.99$), however viewing graphs did lead to greater confidence in the replication of Study 2 ($\mathrm{M_{diff}} = -4.4, \beta_{\mathrm{T-G}} = 0.18, \mathrm{sd} = 0.08, \mathrm{CI} = [\,0.02,\ 0.33], \mathrm{pd} = 0.99$) (**Figure 6E-F**).

2.4.3 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 under-estimate 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.

Figure 6. *Results from Study 3 (replications of Studies 1 and 2)*



Note. People who viewed graphs of COVID-19 growth produced less accurate forecasts compared to people who viewed the same data in tables. Participants in Study 3.1 (N = 215) were presented with the data from panel A, in graphical form (shown) or tabular form (not shown); Participants in Study 3.2 (N = 227) were presented with data from panel B. Participants' forecasts from Study 3.1 are shown in panel C; Participants' forecasts from Study 3.2 are shown in panel D. The black points reflect the "true" total number of confirmed COVID-19 cases in the US according to worldometers.info, accessed on April 27th, 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 Study 3.1 (panel C), participants tended to underestimate the future trend, whereas in Study 3.2 (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 are shown in panels E and F. Participants were more confident in forecasts from tables in Study 3.1 (where tables led to more accurate forecasts) and more confidence in forecasts from graphs in study 3.2 (where graphs led to more accurate forecasts).

2.5 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 a 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 (Study 1: $\tau = 0.15, p < 0.001$; Study 2: $\tau = 0.10, p < 0.001$) (**Figure 5B**). 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 (prior: $\tau = -0.01, p = 0.80$; 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 Studies 1 and 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 March 27th and April 7th 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 2 results we decided to repeat the analyses using data collected from the hypothetical-country replication studies 3.1 and 3.2. We found no significant relationships between forecasts and future or past isolation for either study ($p > .2$). However, in Study 3.2 there was a positive correlation between forecasts and time-to-stop distancing ($\tau = 0.12, p = 0.017$); this relationship was not significant in Study 3.1 ($p = 0.441$).

2.6 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 under-estimated. Lastly, we found slight evidence that judgmental forecasting accuracy was related to social distancing behaviors.

Misestimation

Why were the day-nine 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 the degree of underestimation of exponential growth trends increases with an increasing exponent (Wagenaar & Timmers, 1979; Wagenaar & Sagaria, 1975). 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 under-estimation 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 over-estimated 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 over-estimation from Studies 1 and 2 in Study 3.

Tables vs. 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 vs 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., 2020). 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. It's also possible that participants use entirely different strategies when forecasting with tables and graphs. Perhaps participants in the graph group attempt to mentally visualize extrapolating the curve and their performance could be improved by allowing them to physically extrapolate the curve with a drawing tool.

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., 2020). 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 Studies 1 and 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.

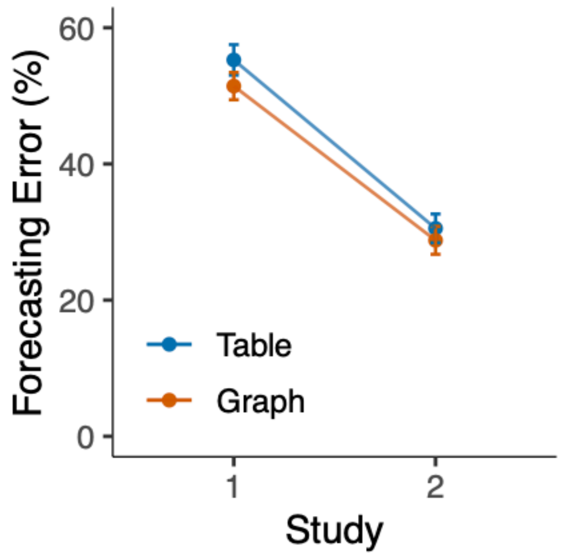
Broader Applications

It's 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 vs. financial data (Chapman & Johnson, 1995; Chapman, 1996). Exponential growth bias has been well documented across multiple domains, including economics and financial-decision-making (Levy & Tassoff, 2015) as well as reasoning about pandemic-related data (Banerjee et al., 2020; 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 vs. 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 an ~52% forecasting error in Study 1 and an ~30% forecasting error in Study 2 (see **Figure 7** and **Appendix G** for more details).

Figure *7. Forecasting Error*



Note. 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.

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 (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.

Lastly, it's possible that performance was influenced by motivation. If participants were incentivized to forecast with accuracy, then they may have shown better performance.

2.7 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.

Chapter 3 Flatten What Curve? Helping People Make Sense of Pandemic Incidence When Public Health Messaging Fails

3.1 Introduction

One concept that graphs were used to illustrate early in the pandemic was the importance of "flattening the curve". The language of "flattening the curve" was first used to explain how flattening the peak of the active-case curve could decrease pressure on the healthcare system, at a time where hospitals were overflowing with patients. In addition to referring to flattening the peak of the active-case curve, the phrase has also been used to describe a cumulative curve with a slope approaching zero. Since cumulative case curves illustrate the sum of the daily cases, a cumulative curve with a slope of zero indicates that zero new cases are being reported each day. To understand how to "flatten" either of these curves, one must know which curve needs to be flattened and what the different types of COVID-19 graphs represent. Conceptually, one must understand that in order to "flatten" the cumulative case curve, the number of daily cases must *decrease*, resulting in a cumulative curve with a slope approaching zero (thus appearing "flat"). Misunderstanding these curves may result in misguided decision-making – for example, if one falsely believes that the goal is to flatten the daily case curve, people may fail to engage in preventative behaviors at a time when the cumulative number of COVID-19 cases are continually rising.

In the current study, we investigated whether people understood the relationship between daily and cumulative case curves, whether this understanding could be improved by viewing a short, narrated video intervention, whether understanding of cumulative and daily cases was linked to social distancing support, and lastly, how individual differences predicted one's understanding of the relationship between cumulative and daily cases.

Misunderstanding Accumulation and the Correlation Heuristic

Early in the pandemic, Fansher et al. (2022b) had participants forecast the number of cumulative cases for three different timepoints when shown a graph of the cumulative number of cases thus far. They found that approximately 27% of participants falsely reported that the number of cumulative cases would *decrease* in the future, suggesting that participants did not understand the concept of accumulation. This misunderstanding comes as no surprise given the literature on how people understand accumulation. The literature on stock-flow reasoning highlights the difficulties people have with linking accumulation and rate of change functions, coined "stock-flow failure" (e.g., Cronin et al., 2009; Cronin & Gonzalez, 2007; Sweeney & Sterman, 2000). The concept of accumulation is relevant to many different domains, for example, the accumulation of greenhouse gases or visitors to a store. All accumulation problems have a "stock", that is continuously changing with an inflow and outflow. In the context of cumulative COVID-19 cases, the "stock" is the total or cumulative number of people who have been infected with COVID-19, while the "flow" is the inflow of daily new cases. In the case of the active-case curve, one would consider both the number of new infections and new recoveries.

Stock-flow failure is thought to be a fundamental reasoning error affecting people regardless of their motivation, graph literacy, cognitive capacity, education, and domain experience (Brunstein et al., 2010; Cronin et al., 2009; Sweeney & Sterman, 2000). One possible reasoning error responsible for stock-flow failure is use of the correlation heuristic (Cronin et al., 2009). The correlation heuristic describes the tendency for people to believe that corresponding graphs of stocks and flows are perceptually similar to one another (Korzilius et al., 2014). For example, a person may believe that a flat daily case curve corresponds to a flat cumulative curve, because they exhibit similar trends, when in reality a flat daily curve is associated with a linearly increasing cumulative curve.

Implications of Stock-Flow Failure on Decision-Making

Historically, risk perception impacts decisions about personal health behaviors including vaccinations and medication adherence, as well as others (Brewer et al., 2007). For example, the risk of getting in an accident while driving a car is quite low, however as the number of times one engages in this activity increases, the risk increases as well. In this context, without the knowledge of cumulative risk, driving without wearing a seatbelt may seem like a small risk. When considering the risk over time, however, seatbelts become more attractive. The idea that continually participating in behaviors with relatively minor risks increases cumulative risk over time is a concept poorly understood by the public, leading to both overestimation and underestimation of risk, and in turn influencing behavior (De La Maza et al., 2019; Doyle, 1997; Slovic et al., 1978).

In the context of COVID-19, misunderstanding the relationship between daily and cumulative cases may lead to a lack of caution. For example, Amidon et al. (2021) argue that the concept of "flattening the curve" acts as a high-level abstraction of how social distancing will lead to decreased spread, and that understanding visual depictions of such risks leads to informed decision making. Failing to understand that a slowly increasing daily case curve is associated with exponential growth of a cumulative curve, or that a flat daily curve still leads to an increasing cumulative curve, may make people underestimate the prevalence of COVID-19, influencing their willingness to engage in preventative behaviors.

Improving Understanding with a Video Intervention

One possible method to teach people about the concept of accumulation is with the use of short, animated, narrated video clips. Prior work has shown that teaching with a video format can be more effective than presenting the same information with static images and can hold great benefits for education under certain conditions (Castro-Alonso et al., 2019; Fiorella & Mayer, 2018; Fyfield et al., 2019). For example, while watching such videos, learning increases when pictures are presented with narration, and learning further increases if this narration comes from a human voice compared to an automated one (Ginns, 2005). Cumulative and daily case graphs are multiple representations of the same underlying data (Ainsworth, 2008). With a video format, different graphs and tables may be animated simultaneously to illustrate this point and to help people better understand the one-to-one mapping between daily and cumulative cases. Animation may be used to deliberately guide the viewer's attention from point to point with the use of arrows or highlighting.

The Current Study

First, we aimed to show that people exhibited stock-flow failure in the context of COVID-19 cases (i.e., would fail to correctly associate cumulative and daily cases). We also investigated whether people would incorrectly apply the correlation heuristic when making these judgments. Second, we investigated whether viewing a video intervention would improve understanding of accumulation functions in both COVID and non-COVID-related contexts, in comparison to an active control condition. We implemented a longitudinal design to see if the intervention would improve understanding immediately after viewing the intervention, 1-2 weeks later, and then 6-7 weeks later (see **Figure 8**). Third, we examined whether understanding of accumulation functions was related to support for social distancing policies and engagement in preventative behaviors. Lastly, we ran an exploratory analysis investigating how understanding of accumulation functions and social distancing behavior were related to various individual difference traits including conservatism, graph literacy, subjective numeracy, education, and working memory capacity.

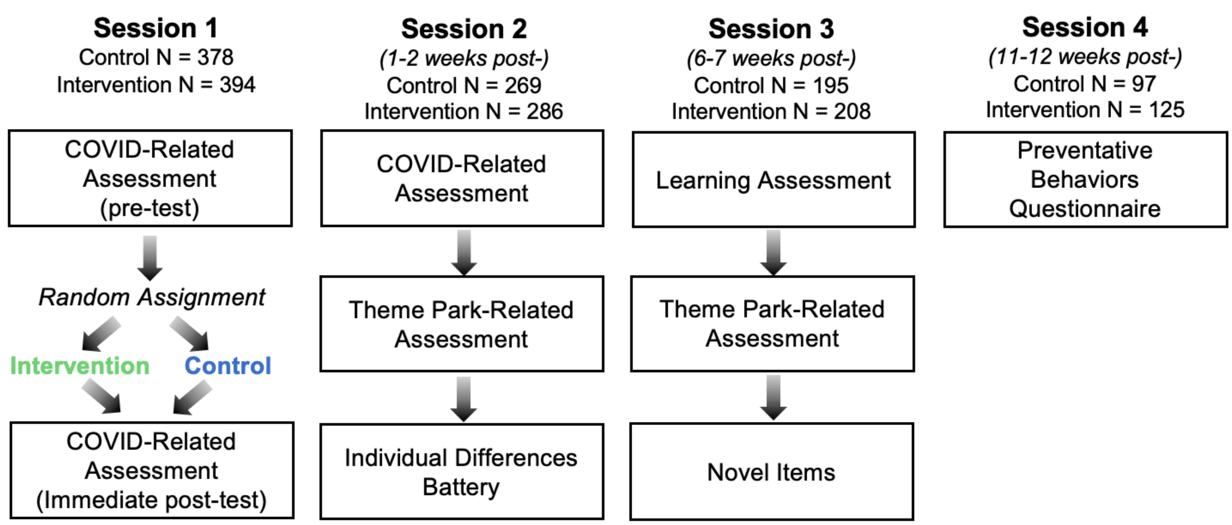
All materials, including video clips and assessments, data, and R scripts are available at https://osf.io/ewy9a/

3.2 Methods

Participants

A sample of 999 participants (Age M(SD) = 39.13(11.48); Gender = 45.15% Female, 54.60% Male, .2% Other) were recruited from Amazon Mechanical Turk in December 2020 to participate in the experiment. Participants were located in the United States and had at least a 95% approval rating for their prior participation in experiments. All procedures were approved by the University of Michigan IRB. For details on compensation, please see the Procedure section.

Figure 8. *Experimental design and session order*



Note. *N's indicate sample sizes after applying our exclusion criteria (see Results).

Design

Participants were randomly assigned to either a control or intervention condition and the experiment was separated into four sessions dispersed throughout twelve weeks (see **Figure 8**). In Session 1, we pre-tested participants' knowledge of the relationship between daily and cumulative COVID-19 graphs, gave them the intervention or control learning modules, and ended the session by giving the same assessment from the pre-test. Participants were invited back for Session 2 one week later, and they were given up to one week to complete the session. During Session 2 we assessed their knowledge of accumulation in both COVID and non-COVID contexts and we gave them a battery of individual difference measures. Six weeks after Session 1, participants were invited back for Session 3, and participants again had up to one week to complete the session consisting of COVID and non-COVID-related accumulation questions. Lastly, Session 4 occurred 11-12 weeks after Session 1. In this session we asked participants about their engagement in preventative behaviors targeted at reducing the spread of COVID, including mask-wearing, vaccination, and social distancing behaviors.

Materials

Session 1.

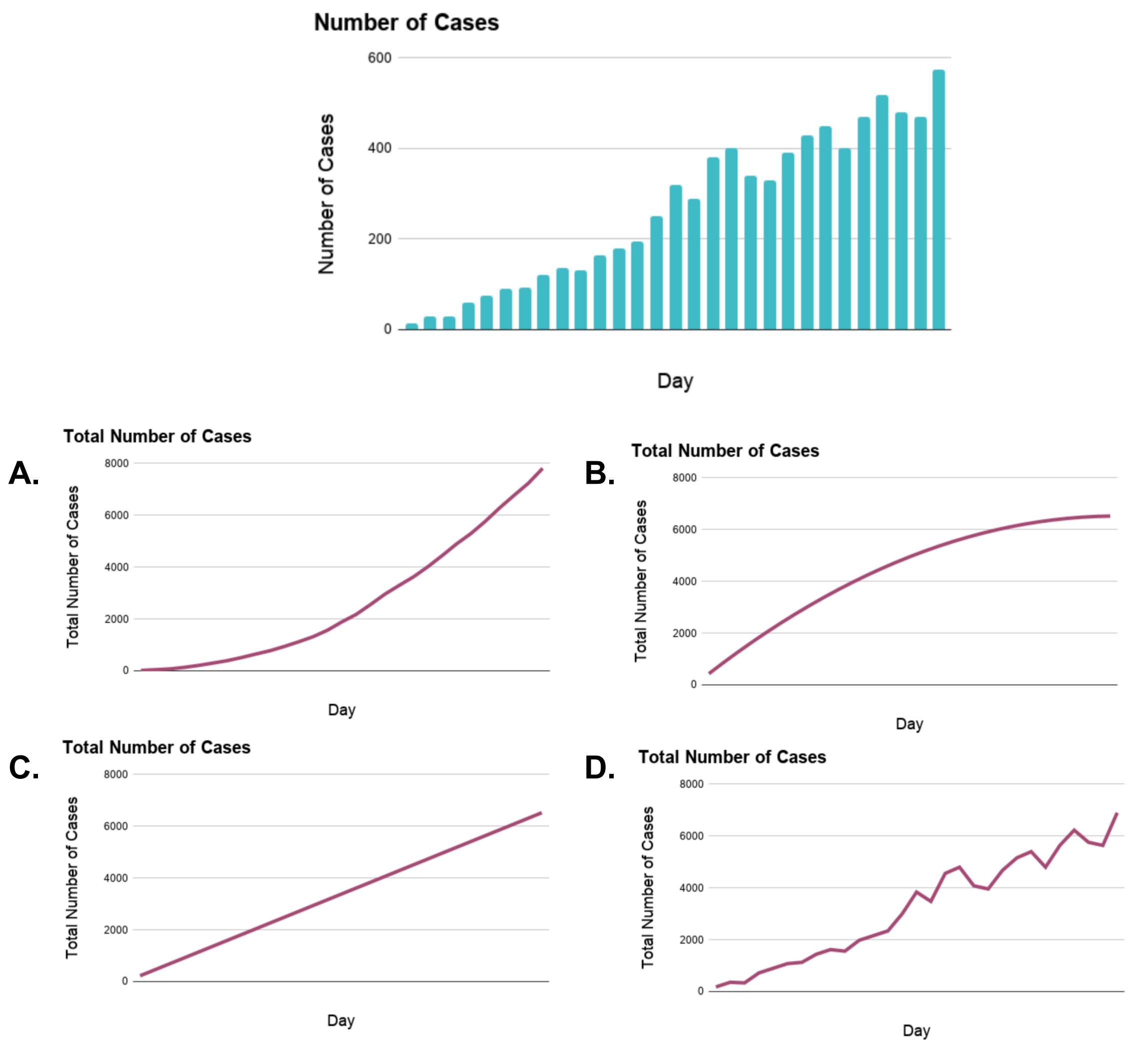
Pre-test Learning Assessment. Session 1 started with a 7-item assessment gauging participants' understanding of the relationship between cumulative and daily case curves. This same assessment is given to participants multiple times throughout the experiment, referred to as the "COVID-Related Assessment" on subsequent post-tests given immediately after the intervention in Session 1, and during Sessions 2 and 3.

For each of the seven items included in the assessment, participants were shown a daily case curve in the form of a bar graph and were asked to select the corresponding cumulative curve in the form of a line graph (see **Figure 9**). Each daily case graph was labelled as belonging to a "hypothetical country" (e.g., "Country A"). Each multiple-choice question had four answer choices, with the answer choices being correct, incorrect, incorrect but perceptually similar to the daily curve, or incorrect but perceptually most similar to the daily case curve (see **Fig. 9A-D** for an example). We included answer choices that were perceptually similar to the daily case graph presented because we wanted to examine whether participants used the correlation heuristic when making their judgements.

The seven items each showed a different daily case curve. Three items showed daily case curves that were truly linear (without noise), and these items showed either increasing, decreasing, or flattening functions. Three additional items also showed increasing, decreasing, or flattening functions; however, these three trends contained noise. The seventh item was a "challenge" question in which the daily cases increased, decreased, and then increased again (see **Appendix H** for exact stimuli).

Participants first viewed the three questions with daily cases that were linear functions without noise in random order, followed by the three linear functions with noise in random order, followed by the "challenge" question. For each question participants rated the importance of social distancing for the hypothetical country with a slider scale from 0 (not at all) to 100 (extremely important). They also recommended a social distancing policy for each hypothetical country with a slider scale from 0-100 with anchors at 0 (*allow essential services only*), 50 (*some social distancing policies),* and 100 (*return life back to normal*).

Figure 9*. Example Assessment Item*



Note. Pre-test item in which participants must choose the cumulative curve associated with a daily case curve (right) that is flat. Possible answer choices are correct (A), incorrect (B), incorrect but perceptually similar to the daily case curve (C), or incorrect but perceptually most similar to the daily case curve (D).

3.2.1 Intervention Condition

We created a short video intervention explaining the relationship between daily and cumulative functions in the context of COVID-19 cases. The intervention consisted of four video clips with comprehension questions between each of the videos. All the content was included as part of a Qualtrics survey. The content of the learning module included the following:

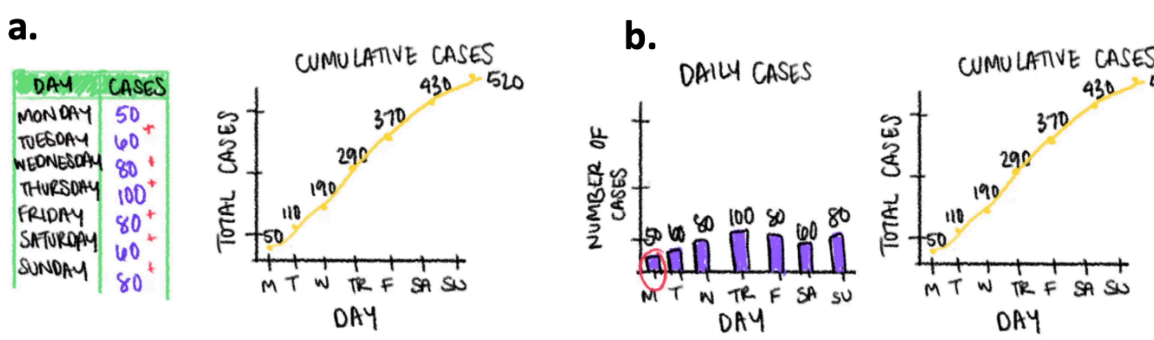
Introduction (1 min 45 sec). The first part of the intervention introduced participants to different COVID-19 graphs that they may have seen in the media. Specifically, it described and showed examples of the daily, cumulative, and active-case curves. The narrator explained that there are multiple interpretations of "flatten the curve", but both interpretations indicate a decreasing number of cases. The narrator explained how flattening the active-case curve reduces pressure on the healthcare system so that people may get treatment, and how the cumulative curve will only flatten when there are few to no new cases reported each day. Participants were then told that the tutorial would focus on flattening the cumulative curve, and that they would learn about the relationship between daily and cumulative curves in subsequent videos.

After the introductory video clip, a comprehension/attention check item appeared on the screen. Participants were shown a daily case curve with a linearly increasing slope and were asked to draw the corresponding cumulative case curve with their cursor. They were allowed to "clear" and redraw the curve as many times as they felt was necessary. After submitting their drawing, they were shown the curve that they should have drawn - an exponentially increasing cumulative curve. We added comprehension checks between video clips in that it is well-established that retrieving information from memory (i.e., "retrieval practice) improves knowledge retention (Karpicke & Blunt, 2011; see Roediger & Butler, 2011 for a review). Retrieving information from memory may help people identify holes in their reasoning, and the act of engaging in retrieval practice may lead to increased engagement with the material. Participants received feedback on all the comprehension checks included with the intervention because feedback on both incorrect and correct response improves learning (Butler et al., 2008).

Part 1: Introducing the Curves (1 min 32 sec). The primary goal of the second video clip was to explain the relationship between daily and cumulative cases. Example cumulative and daily case curves were shown on the screen. The narrator explained the labels and meanings of each of the axes. Next, participants were shown a blank table (see **Fig. 10a**) with day of the week in one column and daily cases in another. Animation was used to draw values into the daily case column. As each daily case count is added to the table, a point was added to a blank axis to the left of the table. This axis would eventually show the number of cumulative cases. We used the table and corresponding graph to illustrate how the number of daily cases are added to create the cumulative case curve. After the cumulative curve was drawn from the daily case table, we then translated the table into a daily case bar graph to show participants how the daily and cumulative curves looked side-by-side (see **Fig. 10b**). A meta-analysis of 61 studies revealed that animation enhanced learning compared to the use of static graphics, with animation being especially helpful when accompanied by audio commentary such as in our intervention (Berney & Bétrancourt, 2016). The incorporation of animation can increase motivation and facilitate transfer of knowledge in science and technology learning (Rosen, 2009). Moreso, the animations included in the videoclips follow Mayer & Moreno's (2002) seven principles of using animation in multimedia instruction, such as presenting animations simultaneously with narrated explanations and avoiding the use of descriptive text with animations.

After watching the video, a comprehension check question appeared: "Pop quiz! If there are 100 cases on day one, 30 cases on day two, and 50 cases on day three, what will the number of cumulative cases be on day 3?" Possible answer choices were (a) 50, (b) 130, (c) 180, and (d) I'm not sure. Participants had to continue answering the question until they got the answer correct (c). They were given hints with each incorrect answer and were given an explanation when they selected the correct answer. Providing elaborative feedback improves learning (Smith et al., 2019) and promotes transfer of knowledge to new contexts (Butler et al., 2013).

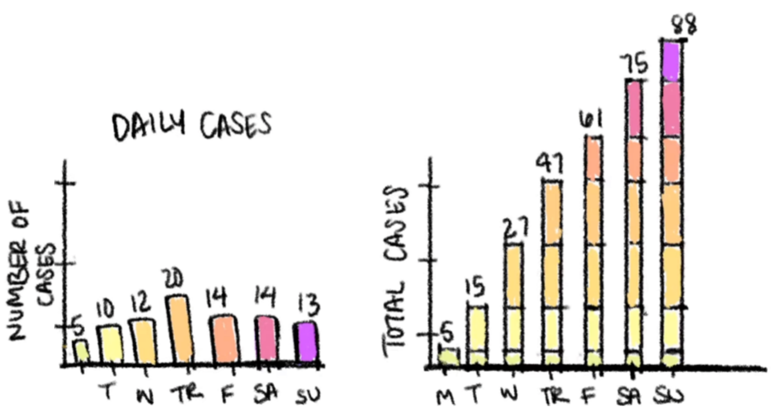
Figure 10. *Stimuli from the second video clip*



Note. Participants were shown a table with the number of cases over a weeklong time-period and the narrator explained how the table translates into a cumulative curve (a) The same data were then shown in graphical format (b)

Part 2: What does "flatten the curve" mean? (4 min 4 sec). In the third video clip, participants learned about how different daily trends translated to different cumulative trends, and how a flattened cumulative curve may be achieved. Participants were first shown a bar graph of increasing daily cases, where each bar was a different color. To illustrate that a linearly increasing daily curve is associated with an exponentially growing cumulative curve, the bars from the daily case graph were animated to "stack" on top of one another to create the cumulative curve (see **Fig. 11**). The same type of animation was then used to show that a daily case graph that increases and then flattens is associated with a cumulative curve that increases exponentially and then linearly. This specific example was provided to explain that a flat daily case curve is associated with a linearly increasing cumulative curve. Lastly, the animation technique was used to show that a decreasing daily case curve is associated with a flat cumulative curve.

Figure 11. *Stimuli from the third video clip*



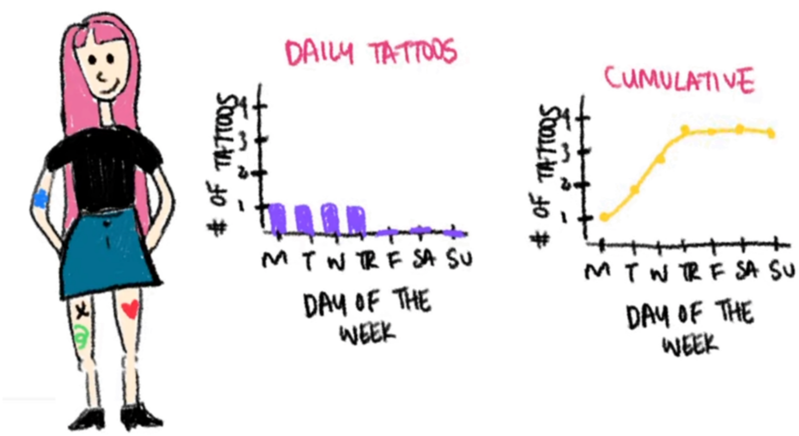
Note. Animated bar charts were shown for various trends to illustrate that the cumulative trend is a sum of the number of cases reported each day.

In the next portion of the clip, the narrator explained why the goal is to have a flattening and not decreasing cumulative curve. They explain that it's impossible for the cumulative curve to decrease since it is a sum of the daily cases. There would have to be a negative number of COVID-19 cases reported for the cumulative curve to decrease, which is impossible. Considering prior work has shown that people often fail to understand that the cumulative curve will never decrease (Fansher et al., 2022b), this point was elaborated further with two analogy examples as analogy enhances learning of scientific topics (Chen & She, 2020; Vosniadou &

Skopeliti, 2019) and more generally has been shown to help people create abstract representations of problems that may be transferred to new contexts (Gentner & Holyoak, 1997).

The first example presented participants with a scenario where a person had been getting new tattoos throughout the week. We explained that even though the person would eventually stop getting tattoos, their cumulative number of tattoos would never decrease (see **Fig. 12**). The second example presented a scenario in which a new movie had been released in theaters. A graph was shown with the number of movie-goers each day, as well as a graph showing the cumulative number of people who have seen the movie. The narrator explained that even though the number of people who go to see the movie each day decreased, the total number of people who had seen the movie would not decrease. The narrator described how it is the same thing with COVID-19 cases; just because the number of daily cases decrease, does not mean that the total number of people who have had COVID decreases.

Figure 12. *Analogy Example*



Note. An example in the context of cumulative number of tattoos was provided to illustrate that cumulative curves will never decrease.

After watching the video, two comprehension questions appeared. The first question asked, "When does the number of cumulative cases decrease?", with possible answer choices: (a) when the number of daily cases increase, (b) when the number of daily cases decrease, (c) when the number of daily cases flattens, and (d) never. The second question asked, "When does the cumulative curve flatten?" and had the same answer options as the first question. Participants had to select the correct answers before moving forward. They were again given hints with each incorrect answer and were provided with an explanation when they did select the correct answers (d and b, respectively). Lastly, participants were shown a picture of a cumulative curve that exponentially increased and then flattened. They were asked to use their cursor to draw the corresponding daily curve in the provided box. After submitting the graph, they were shown the correct response along with an explanation.

Part 3: Let's Recap (46 sec). In the final video clip, participants were given a brief recap of the information presented. Specifically, they were reminded that an increasing daily curve is associated with an exponentially increasing cumulative curve, that a flat daily curve is associated with a linearly increasing cumulative curve, and that a decreasing daily curve is associated with a flattening cumulative curve.

3.2.2 Control Condition

In the active control condition, participants learned about scientific reasoning skills. They specifically learned about sampling and selection in experiments. They were taught about the importance of random assignment and common sampling errors and biases. This material was framed to participants as relevant because "scientific reasoning skills are especially important for interpreting and understanding COVID-19 data". The control condition was matched to the intervention condition as closely as possible. Both interventions had four video clips of approximately the same length, used narration and drawn figures, and included the same number and type of comprehension checks between video clips. The exact videos are available on OSF.

Post-Test Learning Assessment. Immediately after interacting with the intervention or control materials, all participants repeated the seven-item assessment given at pre-test as a post-test measure of understanding of accumulation functions.

Session 2.

COVID-Related Assessment. After returning for Session 2, participants first completed the same seven-item assessment given at pre- and post-test in Session 1.

Theme Park-Related Assessment. To assess whether participants could transfer their knowledge of accumulation functions to new contexts, in Session 2 participants completed another seven-item assessment in the context of number of visitors to a theme park. They were shown a graph of daily visitors to a theme park and were asked to select the graph showing the cumulative number of visitors. The graphical trends included in the daily and cumulative curves were identical to the COVID-related assessment meaning we presented the exact same questions but in a different context. The axes of the graphs were re-labelled to align with the theme park scenario and the colors of the bar and line graphs were altered to decrease the likelihood that participants would realize that the graphs were the same as in the COVID-related assessment.

Individual Differences. A battery of individual-difference measures was collected for exploratory purposes to examine whether the intervention was more or less effective for individuals with specific characteristics. The individual-difference measures included variables thought to influence one's ability to interpret graphs and reason about COVID data. These variables include subjective numeracy, measured by the Subjective Numeracy Scale (Fagerlin et al., 2007), graph literacy measured by the Short Graph Literacy Scale (Okan et al., 2019), conservatism (Mehrabian, 1996), and working memory capacity (Baddeley, 2010; Cowan, 2010). Working memory capacity was measured by performance on a change detection task (Luck & Vogel, 1997).

Session 3.

The same COVID-related and theme-park-related assessments from Session 2 were also given to participants at the beginning of Session 3. We also added additional questions probing participants' understanding of the relationship between change and accumulation over time. These questions were in the same format as the previous questions; participants were presented with a graph and had to choose the corresponding graph from four multiple-choice options.

Reverse Items. Up until this point, participants had only selected the *cumulative* case curve corresponding to a presented *daily* case curve. For two new items, participants were asked to select the *daily* curve corresponding to a presented *cumulative* curve. We added these items to test the flexibility of the participants' knowledge of accumulation. One item presented a flattening cumulative curve, and another showed an impossibly decreasing cumulative curve. For these questions, the four answer choices included three daily case graphs, along with statement "none of these graphs, the cumulative curve in the question is impossible" as the fourth option.

U.S. Data Item. An additional item was added that had participants select the cumulative curve associated with the U.S. daily-case curve. We added this item because we were interested in whether training would directly translate to interpreting the current state of the pandemic. Participants were not told that the curves were actual U.S. data (from the Centers for Disease Control) to prevent them from looking up the correct response.

New Cumulative Gain Item. The following scenario was presented to participants:

> Amelia is a professional chess player who travels around the world competing in chess tournaments. The graph below shows the number of chess games Amelia has won at each tournament she attended over the last year. Which graph shows Amelia's cumulative or total number of wins?

Participants were shown a bar graph where the trend was mostly flat but with high variability. They were asked to select the graph showing the cumulative number of wins from four multiple choice options.

Cumulative Loss Item. Until this point participants had only been given items in the context of cumulative gain, so we added a novel cumulative loss scenario to see if the intervention would improve understanding of cumulative loss in addition to cumulative gain. Understanding cumulative loss may be a more difficult concept for people to grasp. Although the concept of "loss" is associated with a decrease, cumulative loss may be represented as an increasing trend. Participants were presented with the following scenario:

> Jordan's new year's resolution is to lose 20 pounds. The graph below shows the amount of weight Jordan has lost each day since he started dieting and exercising. Which graph shows Jordan's cumulative or total weight loss?

The graph of Jordan's daily weight loss showed a bar graph with a decreasing highly variable trend. Participants were tasked with selecting the line graph corresponding to Jordan's cumulative weight loss from four multiple-choice options.

Session 4.

The final session consisted of a questionnaire on preventative behaviors with the following items. All items required a slider scale response unless otherwise stated:

1. Have you received at least one dose of a COVID-19 vaccine? (*Y*/*N*)

   If *N*, how likely are you to get a COVID-19 vaccine once it becomes available to you? 0 (*not at all likely*) - 100 (*extremely likely*)

2. How important is it to social distance in the U.S.?

   0 (*not at all*) - 100 (*very important*)

3. How successful have you been in engaging in social distancing/isolation?

   0 (*very unsuccessful*) - 100 (*very successful*)

4. How often do you wear a mask when interacting with others at a close distance?

   0 (*never*) - 50 (*half the time*) - 100 (*always*)

Procedure

Across all four sessions, participants completed all the assessments, surveys, and intervention materials on Qualtrics. Participants provided informed consent before beginning each session. For Sessions 2-4, we collected responses for a seven-day period to retain as many participants as possible, with a reminder email being sent in the middle of the week.

In Session 1, Amazon Mechanical Turk workers residing in the United States with a HIT approval rate >= 90% were invited to participate in a survey about COVID-19 graphs. After providing basic demographic information, they were given the following instructions followed by the pre-test assessment:

> In the following section you will see a series of graphs that show the number of COVID-19 cases reported each day for different hypothetical countries. Your task will be to choose the graph that shows the total (or cumulative) number of cases based on the daily trend for each country. Please try your best when answering the questions.

After completing the pre-test, they were told the number of items to which they correctly responded on the pre-test but were not given item-specific feedback. They were incentivized to pay attention to the subsequent tutorial with the promise of a 20-cent bonus per correct question on a future assessment (up to $1.40 in bonus).

Participants were then randomly assigned to the intervention or active control condition. Those in the intervention condition were told that they would watch four short video clips that would teach them how to estimate the number of cumulative cases from the number of daily cases. For those in the active control condition, they were told that they would watch four short video clips about scientific reasoning principles in that “scientific reasoning skills are especially important for interpreting and understanding COVID-19 data”. After watching the video clips and completing the questions associated with their condition, participants were given the post-test assessment, were told the bonus amount they would receive based on their post-test performance (but were not given item-specific feedback), and they were dismissed and compensated $5 plus bonus. Two attention check items were embedded in this survey (Likert scale items) so inattentive participants could be removed from the analysis.

One week after Session 1 data collection, workers were invited back to participate in a follow-up survey for $3. In Session 2, participants first completed the COVID-related and theme-park-related assessments, with order counterbalanced across participants. There was no bonus incentive for correct answers as there was in Session 1. After completing the two assessments, participants were asked to complete the individual-difference scales and questionnaires, along with the working memory change-detection task. We again embedded two attention-check Likert-scale items in this session. Participants were also asked to report how well they remembered the prior session and whether they relied on outside sources during the assessments.

Six weeks after Session 1 and 4 weeks after the end of data collection for Session 2, participants were again sent an email inviting them back for another follow-up study for $3. They first completed the COVID-related and theme-park-related assessments, which were again counterbalanced across participants. This was followed by the two new questions in which participants had to select the correct daily curve associated with a cumulative curve, in which the order of the two questions was randomized. Lastly, they completed the questions about the U.S. data, the new cumulative gain item, and the new cumulative loss item, which were randomly presented. At the end of Session 3, participants were again asked to report how well they remembered the prior session and whether they relied on outside sources during the assessments. Two attention check items were included in this questionnaire, but due to technical error, these data were not collected and thus are not included in the exclusion criteria.

Lastly, participants were invited back 11 weeks after Session 1 where they completed the preventative behaviors questionnaire for $0.50 compensation. Compensation for all sessions was calculated by estimating the amount of time it would take participants to complete sessions (paying participants at a $10/hr rate).

3.3 Results

First, we present the result methods including outlining the impact of our exclusion criteria on sample size, providing the demographic characteristics of the participants returning for each session, and discussing the relationship between attrition and performance during Session 1. We then discuss the Bayesian modeling methods used to test our hypotheses in the subsequent analyses.

Next, we present the analyses used to test our hypotheses. First, we discuss the impact of intervention group on understanding accumulation functions in both COVID and novel contexts. We also discuss use of the correlation heuristic during the COVID-related assessments. Second, we examine the impact of intervention on rated importance of social distancing measures and recommended social distancing policies. We also compare preventative behaviors between intervention and control groups. Lastly, we present exploratory analyses examining the relationship between the various individual-difference measures collected and the impact of intervention on performance.

Exclusion Criteria and Sample Size

Participants were excluded from the Session 1 and 2 data analyses if they failed either of two attention checks embedded in each session (n = 32 excluded in Session 1, n = 77 excluded in Session 2). The number of returning participants varied between sessions, with a retention rate of 78.61% in Session 2, 50.12% in Session 3, and 27.61% in Session 4 when compared to the original sample. **Table 1** illustrates the characteristics of the samples included in the Session 1-4 analyses after applying the exclusion criteria.

Table 1. *Demographic characteristics of the participants included in the Session 1-4 analyses*

Session	**Sample Size**			**Gender**			**Age M(SD)**
	Intervention	*Control*	*Total*	*Female*	*Male*	*Other*	
1	394	378	772	46.11%	53.63 %	.003%	39.15(11.47)
2	286	269	555	47.03%	52.61 %	.004%	39.67(11.76)
3	208	195	403	46.65%	53.10 %	.002%	39.40(11.58)
4	125	97	222	49.55%	50%	.005%	40.82(12.10)

Attrition and Performance

To examine whether there was a relationship between attrition and performance, we compared performance on the COVID-related assessments between participants who dropped out versus remained for each session. Participants who returned for Session 2 did not significantly differ in pre-test (M_{diff} = .02, $t(212.47) = 0.54$, $p = 0.59$) or immediate post-test (M_{diff} = .01, $t(208.52) = 0.34$, $p = 0.73$) score when compared to those who dropped out of the study. Participants who returned for Session 3 had higher pre-test (M_{diff} = .07, $t(761.78) = 2.70$, $p = 0.007$) and immediate post-test scores(M_{diff} = .13, $t(768.7) = 4.46$, $p < .001$) when compared to those who didn't participate in Session 3. We also find that participants returning for Session 4 had higher pre-test (M_{diff} = .09, $t(353.37) = 2.88$, $p = 0.004$) and immediate post-test scores (M_{diff} = .23, $t(417.68) = 8.61$, $p < .001$) than those who did not return. These results suggest that people who performed better during Session 1 at both pre- and post-test were more likely to return for

Sessions 3 and 4. However, since we don't compare performance between sessions, this finding does not affect our interpretation of the subsequent analyses.

Modelling Methods

Data were modelled using the R-package brms: Bayesian Regression Models using 'Stan' (Bürkner, 2017). This package translates input models into the probabilistic programming language Stan, which enables approximate Bayesian inference over model parameters using Markov Chain Monte Carlo (MCMC) sampling (Bürkner, 2018). Data were modelled with the default priors provided by brms unless otherwise indicated (v2.14.4). After fitting the models, graphical posterior predictive checks using the R packages {bayesplot} (Carpenter et al., 2017) and {loo} (Gabry et al., 2019) were performed. In cases where model fit was poor, the family was reassigned from the default gaussian to better-fitting distribution (zero-one-inflated beta or skewed normal, as described below). To quantify uncertainty about the effects of interest, we computed 95% credible intervals (CI) 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).

Understanding of Accumulation in COVID Contexts

Model. To examine the effect of intervention on understanding of the relationship between daily and cumulative case graphs, we separately modelled performance on each 7-item assessment (immediate post-test, 1-2 weeks after intervention, and 6-7 weeks after intervention). An accuracy score was created for each assessment. by averaging performance on the 7 items. Internal consistency was considered good to excellent after computing the standardized Cronbach's alpha based upon the correlations of the seven items for each assessment ($\alpha \geq .86$). Accuracy data were best fit by zero-one-inflated beta regression models (see **Fig 13a**). Zero-one-inflated beta models model data with two components: a beta distribution for responses between 0 and 1, and a Bernoulli distribution for 0 and 1 responses (Liu & Kong, 2015; Vuorre, 2018). Bürkner (2017) describes the density of the zero-one-inflated family as follows:

$$f_{\alpha,\gamma}(y) = \alpha(1-\gamma), if\ y = 0$$

$$f_{\alpha,\gamma}(y) = \alpha\gamma, if\ y = 1$$

$$f_{\alpha,\gamma}(y) = (1-\alpha)f(y), if\ y \notin \{0,1\}$$

where α is the zero-one-inflation probability (i.e., the probability that zero or one occurs) and γ is the conditional one-inflation probability (i.e. the probability that one occurs rather than zero).

The models included Condition and pre-test score as covariates, with score on the assessment in question as the outcome variable (immediate post-test, 1–2-week follow-up, or 6–7-week follow-up). In addition, we model the conditional one-inflation (coi) for each of the predictors. Coi is a feature of zero-one-inflated beta models indicating the probability of a response being a 1 given that the response is equal to either 0 or 1.

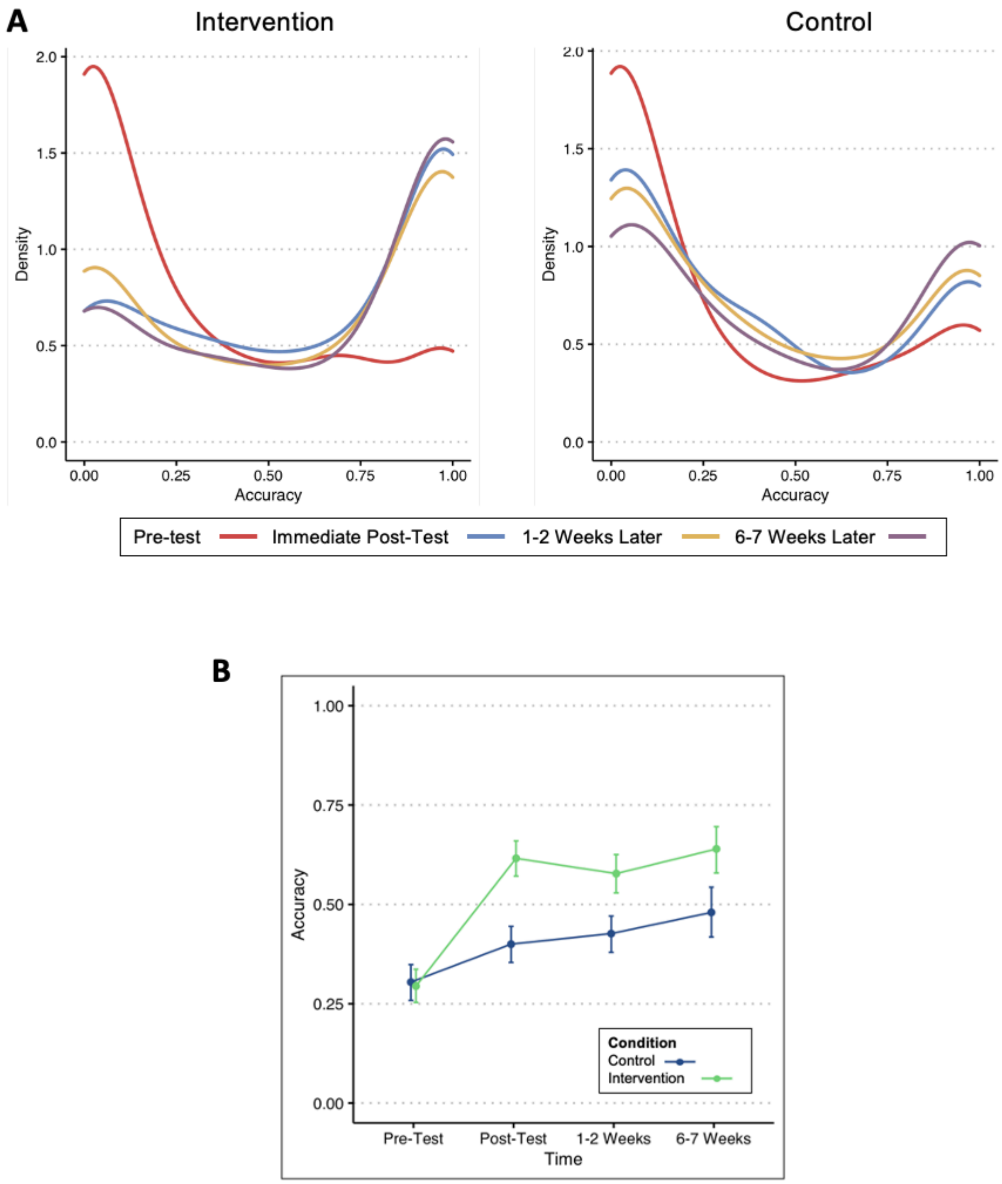
Results. Pre-test score significantly predicted accuracy on the assessments given immediately after the intervention ($\beta = 0.85$, $CI_{95\%} = [0.47, 1.21]$, $pd = 1$), 1-2 weeks after the intervention ($\beta = 0.98$, $CI_{95\%} = [0.59, 1.38]$, $pd = 1$), and 6-7 weeks after the intervention ($\beta = 1.08$, $CI_{95\%} = [0.60, 1.54]$, $pd = 1$). Critically, Condition predicted performance such that those in the intervention condition performed better than the active control group immediately after intervention ($\beta = 0.41$, $CI_{95\%} = [0.22, 0.60]$, $pd = 1$), 1-2 weeks later ($\beta = 0.39$, $CI_{95\%} = [0.15, 0.60]$, $pd = .999$), and 6-7 weeks later ($\beta = 0.35$, $CI_{95\%} = [0.06, 0.62]$, $pd = .99$), suggesting long-lasting effects of the intervention on understanding the relationship between daily and cumulative case functions (see **Fig 13b**).

Interpreting the conditional one-inflation covariates included in the regression models revealed that the higher one's pre-test score was the more likely it was that the participant would score a 100% than a 0% on the assessment given immediately after intervention ($\beta = 5.13$, $CI_{95\%} = [4.30, 6.03]$, $pd = 1$), 1-2 weeks later ($\beta = 3.98$, $CI_{95\%} = [3.20, 4.81]$, $pd = 1$), and 6-7 weeks later ($\beta = 3.88$, $CI_{95\%} = [3.04, 4.81]$, $pd = 1$). Critically, when compared to the control condition, those in the intervention condition were more likely to score a 100% than a 0% on the assessment given immediately after intervention ($\beta =1.86$, $CI_{95\%} = [1.28, 2.46]$, $pd = 1$), 1-2 weeks later ($\beta = 1.17$, $CI_{95\%} = [0.61, 1.77]$, $pd = 1$), and 6-7 weeks later ($\beta = 1.13$, $CI_{95\%} = [0.50, 1.74]$, $pd = 1$).

We would like to note that pre-test score was poor overall, with 44% of our participants getting zero items correct. In addition, we examined whether participants used the correlation heuristic when making their judgments. To do so, we fit a logistic regression model to the data with multiple choice selection (selecting the correct answer or the perceptually most similar curve) as the outcome variable and Session (pre-test, immediate posttest, 1–2-week follow-up, or

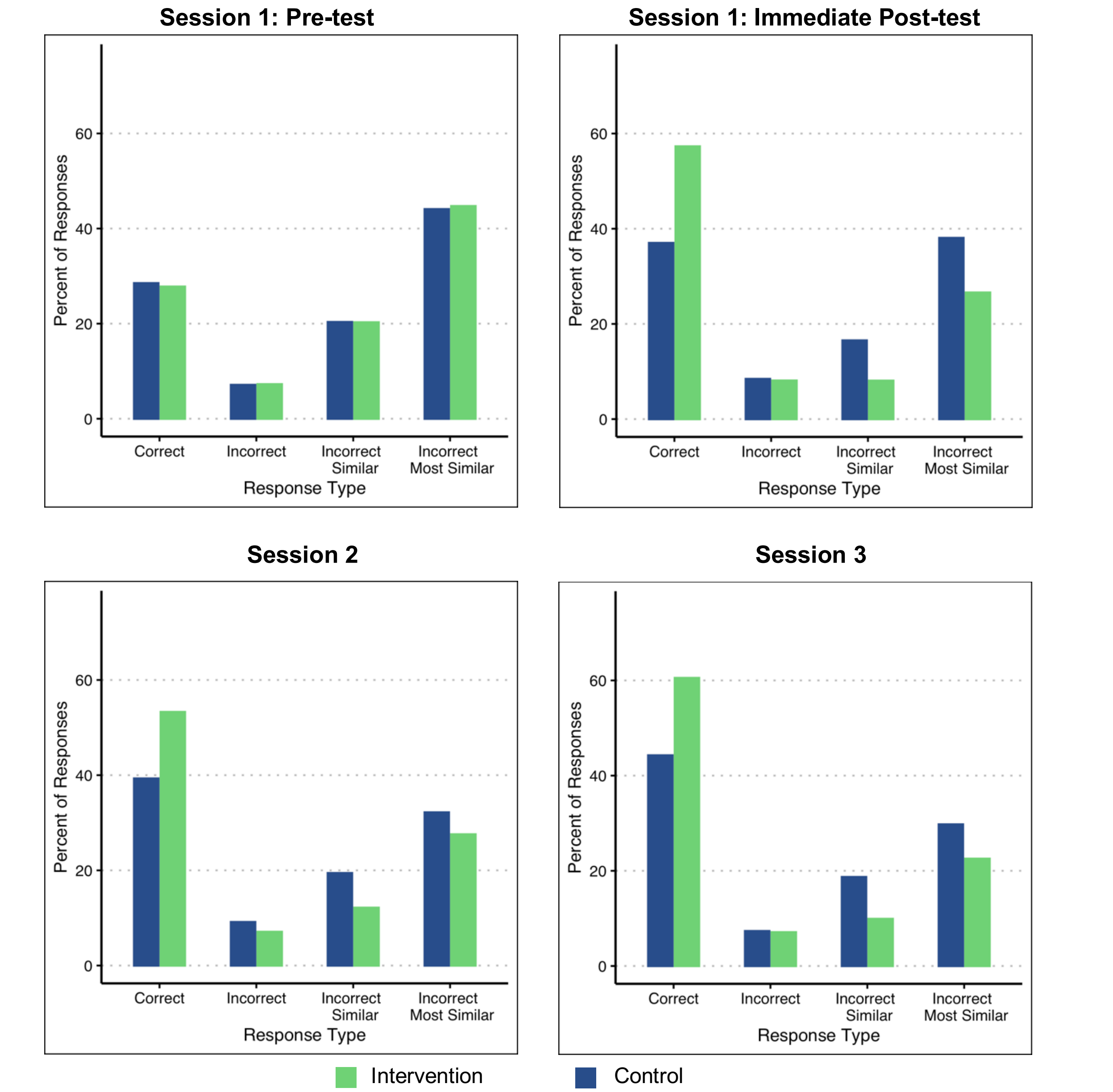
6-7 week follow) and Condition as predictors (see **Figure 14**). The outcome was coded such that the correct answer was 1 and demonstrating use of the correlation heuristic or choosing the perceptually most similar curve was coded as 0. For the Session variable, pre-test score is used as the reference group. Overall, there is evidence that all participants used the correlation less at immediate post-test (β = .41, $CI_{95\%}$ = [.28, .54], *pd* = 1), 1-2 week follow-up (β = .64, $CI_{95\%}$ = [.50, .77], *pd* = 1), and 6-7 week follow-up (β = .84, $CI_{95\%}$ = [.68, .99], *pd* = 1). Participants in the control group also used the correlation heuristic less overall than those in the control group (β = -.04, $CI_{95\%}$ = [-.16, .09], *pd* = 1). We also find evidence for significant Condition*Session interactions such that participants in the intervention group used the correlation heuristic less at immediate post-test (β = .83, $CI_{95\%}$ = [.65, 1.01], *pd* = 1), 1-2 week follow-up (β = .50, $CI_{95\%}$ = [.30, .69], *pd* = 1), and 6-7 week follow-up (β = .63, $CI_{95\%}$ = [.41, .84], *pd* = 1) in comparison to the control group.

Figure 13. *Performance on COVID-Related assessments by Condition*



Note. (A) Illustrates the distributions of accuracy scores when participants were assessed on their understanding of the relationship between daily and cumulative COVID-19 cases at pre-test (red), immediate post-test (blue), 1-2 weeks after intervention (yellow), and 6-7 weeks after intervention (purple). These data are generally bimodal, with higher densities at 0 and 1, which is why zero-one-inflated beta models were deemed appropriate. (B) Illustrates mean raw score on the learning assessments given at pre-test, immediate post-test, 1-2 weeks later, and 6-7 weeks later. Those in the intervention condition (green) show consistently better performance than the control condition (blue), after the intervention. Error bars represent 95% confidence intervals.

Figure 14. *Evidence for the use of the correlation heuristic*



Note. This figure illustrates the number of times each type of cumulative curve multiple choice option was selected, collapsed across the seven items. Each panel shows response selection on a different assessment (i.e., pre-test, immediate post-test, Session 2, and Session 3), and color represents Condition with the intervention condition in green and the control condition in blue. The x-axis on each panel shows the four multiple choice options: correct incorrect, incorrect but perceptually similar, and incorrect but perceptually most similar.

Understanding Accumulation in the Theme Park Context

Model. Participants were asked to solve accumulation problems in the context of number of visitors to a theme park during Sessions 2 and 3 of the experiment, 1-2 and 6-7 weeks after the intervention, respectively. To model the data, we created two accuracy scores, one was average performance on the seven items presented 1-2 weeks after intervention ($\alpha = .91$), and the second accuracy score was average performance on the seven items presented 6-7 weeks after the intervention ($\alpha = .92$). We again model performance with Condition and pre-test score as predictors. Since participants were not pre-tested on the theme park scenario, we use baseline understanding of the relationship between daily and cumulative COVID curves as a covariate to account for pre-existing understanding of accumulation functions. These data were again best fit with zero-one-inflated beta regression, thus, coi was also modelled for each of the two predictors.

Results. Pre-test score significantly predicted performance on the theme-park-related assessment both 1-2 ($\beta = .61$, $CI_{95\%} = [.15, 1.06]$, $pd = .996$) and 6-7 ($\beta = 1.09$, $CI_{95\%} = [.55, 1.61]$, $pd = 1$) weeks after intervention. Critically, those in the intervention group outperformed those in the control group 1-2 ($\beta = .30$, $CI_{95\%} = [.07, .53]$, $pd = .99$) and 6-7 ($\beta = .46$, $CI_{95\%} = [.18, .75]$, $pd = .999$) weeks later, suggesting that the intervention allowed participants to generalize their understanding to novel contexts (see **Fig 15A**). When interpreting the coi for each predictor, those with a higher pre-test score were more likely to score 100% than 0% on assessments given 1-2 ($\beta = 4.16$, $CI_{95\%} = [3.34, 4.98]$, $pd = 1$) and 6-7 ($\beta = 4.01$, $CI_{95\%} = [3.19, 4.93]$, $pd = 1$) weeks later. Those in the intervention group were more likely to score a 100% rather than a 0% on both assessments when compared to the control group (1-2 weeks: $\beta = 1.22$, $CI_{95\%} = [.60, 1.81]$, $pd = 1$; 6-7 weeks: $\beta = 1.12$, $CI_{95\%} = [.50, 1.80]$, $pd = 1$).

Understanding Accumulation in Other Contexts

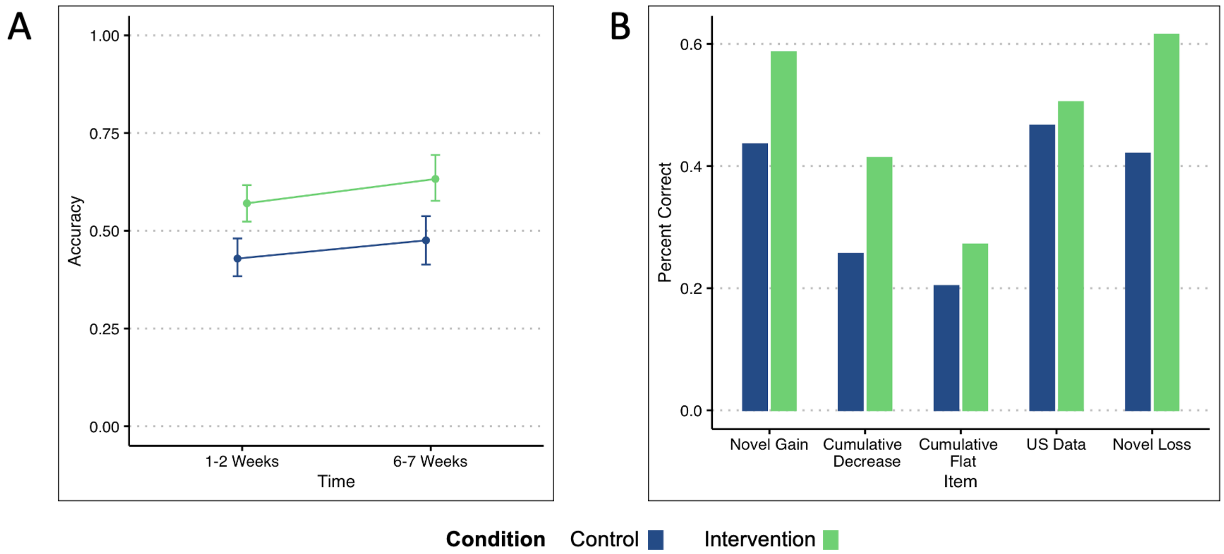
Modeling. One may wonder whether participants in the intervention group only performed better than the control group in the theme park context because the graphical trends presented were the same as those given in the COVID-related assessments. It is possible that participants realized that the trends were the same, and thus, those in the intervention group outperformed those in the control condition simply because they perform better on the COVID-related assessment, with this knowledge transferring over to the theme park context only because the graphs are identical. To provide further evidence that participants in the intervention group were better able to transfer their knowledge of accumulation to new contexts, at the 6–7-week follow-up, participants were given additional assessments of far transfer with graphs and contexts on which they had not been assessed previously. Five additional items were given, including answering questions about novel cumulative gain and loss scenarios, choosing the cumulative curve associated with U.S. data trends, and two reverse question items for which participants selected the *daily* curve associated with a presented *cumulative* curve. Performance was modeled separately for each of these five items with accuracy (1 or 0) as the outcome. Thus, we use logistic regression to model the data in which the outcome is modeled as part of the Bernoulli distributional family. Each model again includes pre-test score on the COVID-related pre-test and Condition as predictors. Model output is show in **Table 2**.

Results. Pre-test score was positively associated with accuracy on both the novel gain and loss scenarios. Those in the intervention group were more likely to get the novel gain and loss items correct than those in the control group. Pre-test score also predicted accuracy on the item in which participants were shown the daily curve associated with the U.S. data; however, condition did not predict accuracy on this item. Lastly, pre-test score and participating in the intervention were positively associated with performance on both reverse items in which participants were asked to choose the daily curve associated with a given cumulative curve.

Table 2. *Model output for the analyses examining the effect of pre-test score and Condition on performance on the novel items presented in Session 3*

Novel Cumulative Gain Context – Chess Games Won			
	β	$CI_{95\%}$	*pd*
Pre-test	3.42	[2.67, 4.21]	1
Condition	.80	[.33, 1.26]	1
Novel Cumulative Loss Context – Weight Lost			
	β	$CI_{95\%}$	*pd*
Pre-test	3.56	[2.80, 4.38]	1
Condition	1.06	[.58, 1.53]	1
United States Data			
	β	$CI_{95\%}$	*pd*
Pre-test	3.07	[2.36, 3.75]	1
Condition	.18	[-.26, .64]	.80
Reverse Item – Flat Cumulative Curve			
	β	$CI_{95\%}$	*pd*
Pre-test	3.05	[2.38, 3.74]	1
Condition	.62	[.16, 1.07]	.996
Reverse Item – Impossibly Decreasing Cumulative Curve			
	β	$CI_{95\%}$	*pd*
Pre-test	3.60,	[2.89, 4.33]	1
Condition	1.07,	[.53, 1.61]	1

Figure 15. *Performance on items in contexts other than COVID-19*



Note. (A) Illustrates raw mean performance by Condition on the theme park assessment given 1-2 and 6-7 weeks after intervention. Error bars represent 95% confidence intervals. (B) Illustrates the percent of participants in each condition who correctly answered each far transfer item correctly at the 6–7-week follow-up.

Importance of Social Distancing

Model. During the COVID-related assessments given at pre-test, immediate post-test, 1-2 weeks after intervention, and 6-7 weeks after intervention, participants were asked to rate the importance of social distancing for each hypothetical country based on the daily case trajectory. To examine whether participating in the intervention influenced perceived importance of social distancing, rated social distancing importance was averaged across the seven items from each assessment. Internal consistency was considered good to excellent after computing the standardized Cronbach's alpha based upon the correlations of the seven items for each assessment ($\alpha \geq .89$). Beliefs about the importance of social distancing at pre-test as a covariate to the subsequent models to control for pre-existing individual differences. Average perceived social distancing importance at each assessment was modelled with Bayesian linear regression best fit by the default Gaussian familial distribution.

Results. Pre-existing beliefs about the importance of social distancing were positively associated with perceived social distancing importance at immediate post-test ($\beta = .90$, $CI_{95\%} =$ [.86, .94], $pd = 1$), 1-2 weeks after the intervention ($\beta = .79$, $CI_{95\%} =$ [.73, .84], $pd = 1$), and 6-7 weeks after the intervention ($\beta = .78$, $CI_{95\%} =$ [.72, .85], $pd = 1$). Critically, those in the intervention group rated the importance of social distancing as higher than those in the control group at immediate post-test ($\beta = 2.33$, $CI_{95\%} =$ [.92, 3.73], $pd = .999$), 1-2 weeks later ($\beta = 4.20$,

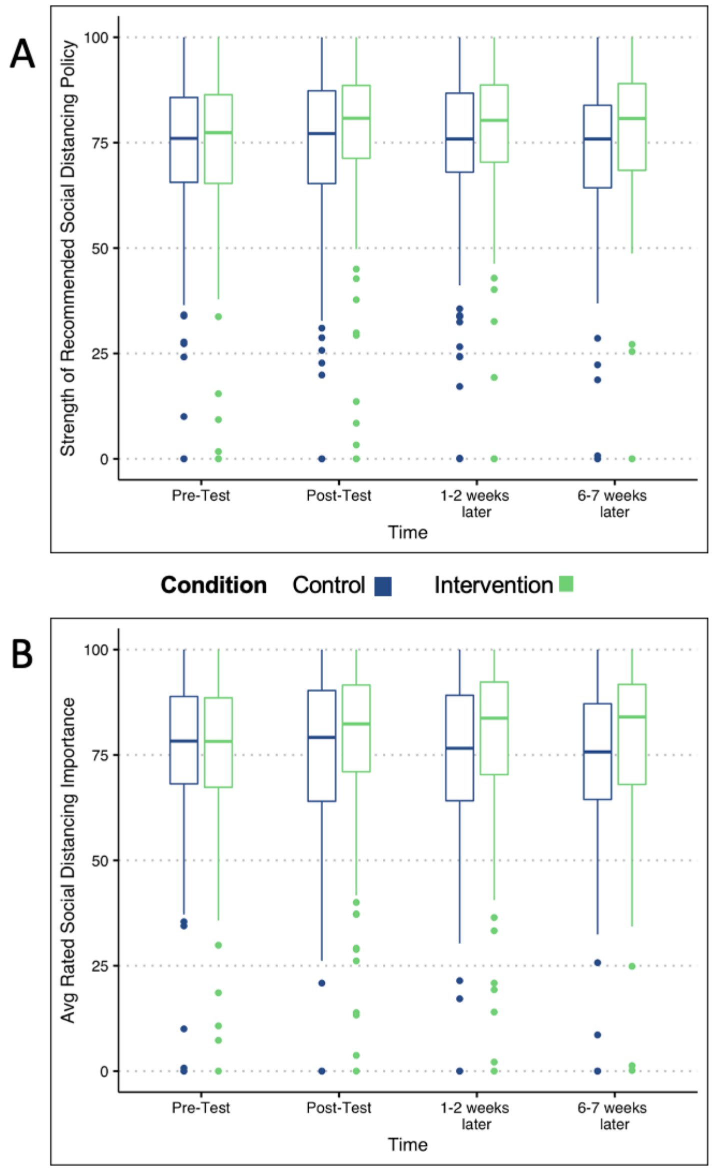
$CI_{95\%}$ = [2.17, 6.05], *pd* = 1), and 6-7 weeks later (β = 2.82, $CI_{95\%}$ = [.52, 5.29], *pd* = .99) (see **Fig 16A**).

Strength of Social Distancing Policy

Model. In addition to rating the importance of social distancing, participants were asked to recommend a social distancing policy for each of the hypothetical countries based on the COVID trajectory (0 indicating no social distancing policy and 100 indicating a restrictive policy). We again averaged across the seven items for each assessment and internal consistency was considered good to excellent after computing the standardized Cronbach's alpha based upon the correlations of the seven items for each assessment ($\alpha \geq .88$). We again controlled for pre-existing beliefs by adding average strength of the recommended social distancing policy from pre-test as a covariate. Again, these data were best fit by the default Gaussian distributional family.

Results. We find that pre-existing beliefs about the strength of social distancing policies predicted composite recommended social-distancing policy at immediate *post*-test (β = .88, $CI_{95\%}$ = [.84, .92], *pd* = 1), 1-2 weeks after intervention (β = .74, $CI_{95\%}$ = [.69, .80], *pd* = 1), and 6-7 weeks after intervention (β = .73, $CI_{95\%}$ = [.67, .80], *pd* = 1). Those in the intervention group recommended a stronger social-distancing policy than those in the control group at immediate post-test (β = 1.87, $CI_{95\%}$ = [.65, 3.11], *pd* = .999), 1-2 weeks later (β = 3.21, $CI_{95\%}$ = [1.46, 4.99], *pd* = 1), and 6-7 weeks later (β = 3.91, $CI_{95\%}$ = [1.51, 6.12], *pd* = 1) (see **Fig 16B**).

Figure 16. *Relationship between Condition and social distancing attitudes across sessions*



Note. (A) Average strength of the recommended social distancing policy and (B) average rated social distancing importance across items at pre-test, post-test, 1-2 weeks later, and 6-7 weeks later, separated by group.

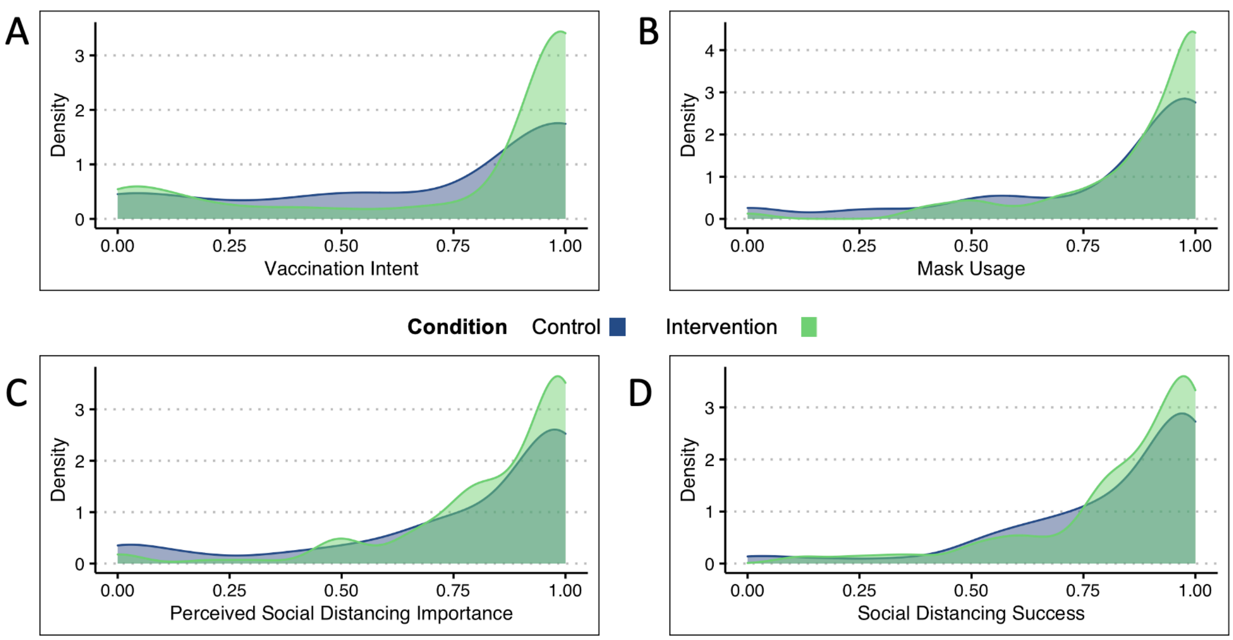
Self-Reported Preventative Behaviors

Vaccination intent, mask wearing, social distancing success, and perceived social distancing importance were again modeled with zero-one-inflated beta regression given that there were many extreme responses from participants (see **Fig. 17**). The measures were originally collected with slider scale items anchored from 0-100 and were rescaled to be from 0-1. There was little evidence suggesting that the intervention and control conditions differed on these measures ($.62 \leq pd \leq .90$), however it is worth noting that any possible effects are in the expected direction such that the intervention condition reported greater intent to vaccinate, more mask-wearing, more success with social distancing, and greater perceived importance of social distancing when compared to the control condition (see **Table 3**, **Fig 17**).

Table 3. *Modelling self-reported preventative behaviors as a function of Condition*

Vaccination Intent	*β*	*CI95%*	*pd*
Intermediate values	.08	[-.41, .54]	.63
Extremes (0 or 1)	.55	[-.30, 1.45]	.90
Mask Wearing	*β*	*CI95%*	*pd*
Intermediate values	.20	[-.11, .53]	.89
Extremes (0 or 1)	.74	[-.43, 1.92]	.89
Social Distancing Success	*β*	*CI95%*	*pd*
Intermediate values	.05	[-.26, .34]	.62
Extremes (0 or 1)	.60	[-.95, 2.19]	.79
Perceived Importance of Social Distancing	*β*	*CI95%*	*pd*
Intermediate values	.19	[-.15, .53]	.86
Extremes (0 or 1)	.63	[-.55, 1.87]	.85

Figure 17. *Relationship between Condition and self-reported preventative behaviors*



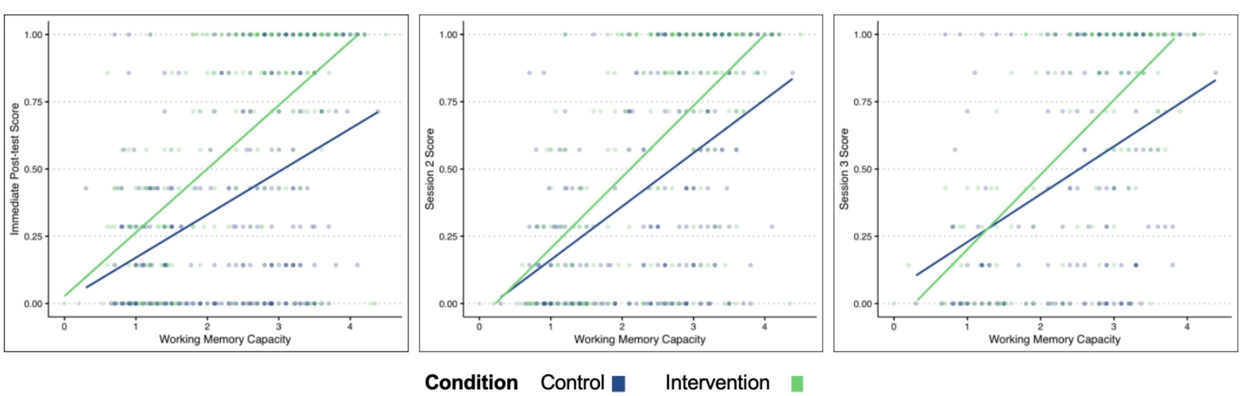
Note. Illustrates the density of responses between conditions for each of the preventative behavior items, including (a) vaccination intent, (b) mask wearing, (c) social distancing success, and (d) perceived importance of social distancing.

Individual Differences and Intervention Effectiveness

All individual-difference variables (conservatism, graph literacy, subjective numeracy, education, and working memory capacity) were modelled individually, with the interaction between the individual difference and Condition as predictors, and performance on the COVID-related assessments as the outcome variables. Thus, for each independent variable, three models were run to model performance at immediate post-test in Session 1, during Session 2, and during Session 3. Each model controlled for pre-test score on the COVID-related assessment.

The data were again best fit by zero-one-inflated regression models, so the intermediate and extreme values (0 or 1) were modelled separately. The effects of subjective numeracy, graph literacy, and political ideology were either extremely small or inconsistent. However, working memory capacity consistently predicted performance on the COVID-related assessments, with working memory capacity being positively associated with performance at immediate post-test, during Session 2, and during Session 3. In addition, people with higher working memory capacity were more likely to get all the items correct at immediate posttest and both follow-up assessments as opposed to none of the items on each of the assessments ($pd \leq .94$). We also find evidence for a Condition by working memory capacity interaction at immediate posttest and both follow-up assessments ($pd \leq .97$), suggesting that people with high working memory capacity benefited most from the intervention. This was true for both immediate and extreme values (see **Figure 18**). For model output from all exploratory individual difference analyses, please refer to **Appendix I**.

Figure 18. *Relationship between Condition and Working Memory Capacity*



Note. Illustrates the relationship between working memory capacity and raw score on the immediate post-test and both follow-up assessments.

3.4 Discussion

At pre-test, participants were poor at stock-flow reasoning in the context of understanding accumulation of COVID-19 cases. We found that 44% of participants scored a zero on the seven-item assessment, even after excluding participants who failed our attention checks, suggesting that participants who scored zero didn't simply rush through the task. A short video intervention proved to effectively teach participants about the relationship between daily and cumulative cases, and subsequently lead to more positive attitudes towards social distancing policies. Moreso, the intervention was long-lasting and improved stock-flow reasoning in contexts beyond COVID-19. Here we discuss our results in detail along with some of the applications of this work to STEM instruction and other topics where knowledge of accumulation is relevant.

Intervention Effectiveness

We find strong evidence that our short video intervention effectively taught participants about the relationship between cumulative and daily cases. Participants in the intervention condition performed better than those in a content and length-matched active control group on all the COVID-related, and almost all the non-COVID-related assessments. This suggests that although the intervention taught participants about accumulation functions in the context of COVID-19 cases, they were able to generalize their knowledge to new contexts and types of stock-flow reasoning questions such as selecting the time series function associated with a cumulative function, and reasoning about both cumulative gain and loss scenarios. Moreso, we find evidence that the intervention was long-lasting, with participants in the intervention group outperforming those in the control group up to 6-7 weeks after the initial intervention. The actual length of the effect of the intervention could be much longer, as our last assessment was given at 6-7 weeks. This suggests that short evidence-based video interventions may be highly effective at teaching people about mathematical transformations related to data visualization (e.g., derivatives, accumulation). Future research could explore if similar interventions are effective at teaching people about other concepts that can be displayed graphically and animated – for example calculus concepts like acceleration and velocity or understanding risk.

Use of the Correlation Heuristic

At pre-test, participants erroneously applied the correlation heuristic (Cronin et al., 2009) when making their judgements (see **Figure 14**), as participants most often selected the cumulative curve that looked perceptually most like the daily case curve in question. At each assessment after pre-test, the intervention group consistently selected the correct answer instead of the most perceptually similar choice, suggesting that the intervention made participants realize that the heuristic is an ineffective strategy. Interestingly, the control group also selected the perceptually most similar curve less at the tests administered after the control intervention, even though they weren't given any training on stock-flow reasoning and didn't perform as accurately as the intervention group. There are many reasons why this may have occurred, but here we list a few possible explanations.

First, the control group participants may have improved from pre- to post- test because they were given feedback on their initial pre-test performance. Participants may have initially applied the correlation heuristic to all the questions at pre-test. When they received feedback that all their responses were incorrect, they may have realized that the correlation heuristic was an ineffective strategy, thus they were more likely to select cumulative graphs that weren't perceptually similar to the probed daily case curve, making it more likely they select the correct answer. Alternatively, after receiving feedback that they had gotten none of the items correct, they may have paid more attention to the instructions on subsequent assessments. Lastly, participants may have been more motivated in the immediate post-test because they were given monetary incentive for correct responses, however, this cannot explain why control participants used the correlation heuristic less in the Session 2 and Session 3 assessments when compared to pre-test, considering that they weren't offered a bonus as an incentive for correct responses in these sessions.

Influence of Training on Social Distancing Attitudes and Preventative Behaviors

Across sessions, we consistently found that participants who were trained on the relationship between daily and cumulative case curves had more positive attitudes towards social distancing and social distancing policies. This aligns with prior work suggesting that educating people about COVID-19 data visualizations leads to more favorable attitudes towards preventative behaviors such as social distancing (Lammers et al., 2020) and vaccination (Fansher et al., 2022c). Unfortunately, we did not find an effect of intervention on engagement in real-life preventative behaviors, although the direction of the effects of intervention on social distancing attitudes, social distancing success, mask usage, and vaccination were in the hypothesized direction when participants were questioned 11-12 weeks after the intervention.

The exact mechanism by which understanding stock-flow reasoning in the context of accumulation of COVID cases influences social distancing attitudes is unknown. For example, perhaps realizing that the only way to flatten the cumulative curve is to significantly decrease the number of daily cases may make participants realize the importance of decreasing the spread – thus leading them to have more favorable attitudes towards social distancing. To ensure that the differences observed between conditions were not due to pre-existing differences between groups, we compared mean pre-test attitudes towards social distancing importance and social distancing policies between conditions. We find that the two groups reported virtually identical attitudes (SD Importance: Intervention ($M = 75.49$), Control ($M = 75.22$); SD Policy Attitudes: Intervention ($M = 74.33$), Control ($M = 74.07$) suggesting that the effect of intervention on social distancing attitudes is due to training.

Individual Differences and Training Effectiveness

We found it surprising that there were no reliable relationships between subjective numeracy and performance or graph literacy and performance. It may be the case that our measures of numeracy and graph literacy were less accurate as they were based on subjective judgements of one's own skills. It is also likely that understanding of accumulation goes beyond basic numeracy and graph skills and is a specific mathematical transformation skill about which people must be formally instructed. This would explain why other researchers have found the people exhibit stock-flow failure even if they have high motivation, graph literacy, cognitive capacity, and education (Cronin et al., 2009; Sweeney & Sterman, 2000).

The only consistent individual difference we observed was that people with high working memory capacity performed better on the COVID-related assessments and benefitted most from the training. This is perhaps unsurprising as working memory capacity has been shown to be correlated with a multitude of performance outcomes such as intelligence (Conway et al., 2003), math ability (Raghubar et al., 2010), and academic achievement (Swanson & Alloway, 2012). People with higher working memory capacity may have been better able to absorb the information presented in the intervention than those with lower capacity, considering that our intervention did not display text descriptions of the narrations. High working memory capacity may have allowed participants to keep track of the narrators' audio commentary while simultaneously paying attention to the provided visualizations and animations. Low working memory capacity participants may have benefitted from "pause" and "rewind" capabilities that our videos did not implement.

3.5 Conclusions

While this research was conducted in the context of understanding COVID-19 graphs, understanding accumulation functions is important in contexts beyond what we present, such as financial literacy and climate change. For example, understanding the accumulation of interest or accumulation of greenhouse gases. Understanding mathematical concepts like accumulation and how they may be depicted graphically is also important for health-based decision-making. In this study we show that a brief 8-minute video intervention improved understanding of accumulation in COVID and other contexts, that the effects of the training were long-lasting, and that knowledge of accumulation impacted attitudes towards real life behaviors like social distancing. This work provides evidence that it is possible to instruct people to avoid the use of the correlation heuristic and to overcome stock-flow failure. We also provide additional evidence that educating people about data may influence their behaviors in a positive manner. Lastly, we encourage learning researchers in other areas to consider implementing similar evidence-based video interventions when teaching people about mathematical and statistical concepts.

Chapter 4 Icon Arrays Reduce Concern Over COVID-19 Vaccine Side Effects: A Randomized Control Study

4.1 Introduction

On April 13, 2021, the CDC paused administration of Johnson & Johnson's (J&J) COVID-19 vaccine to review six reports of a serious blood clotting condition out of the ~6.8 million doses that had been administered (CDC, 2021). People generally struggle to comprehend probabilistic risk information when it is depicted numerically (Peters, 2012; Slovic et al., 2000) and often overestimate the occurrence of consequential but unlikely events, including those associated with vaccination (Reyna, 2004). Such risks may evoke high dread when viewed by non-experts, socially amplifying small risks to society-level problems (e.g., Slovic & Weber, 2002). It is possible that the CDC's announcement increased vaccine hesitancy due to these psychological biases (Slovic & Weber, 2010) especially considering that of those who are hesitant to be vaccinated for COVID-19, 72% cite concern over side effects as the main contributor (Funk & Tyson, 2021). Two days after the CDC's announcement, we investigated how probability language and data visualizations incorporated into the announcement might have alleviated potential increases in aversion towards both the J&J and *all* COVID-19 vaccines.

People interpret risk differently depending on how it is presented (see Reyna & Brainerd, 2008 for a review). Thus, in Experiment 1 we examine the influence of language (i.e., expressing the probability as a ratio, percentage, or single number) on changes in vaccine aversion. We also tested whether viewing an icon array depicting the small risk of experiencing the blood-clotting side effect would prevent increases in vaccine aversion. Prior work suggests that understanding of risk may be improved with the use of such displays (Tait et al., 2010; Waters et al., 2007a; see **Figure 19**). Graphical depictions of risk in the form of icon arrays are thought to be beneficial because they highlight both the numerator (the number of times X has happened) and denominator (the number of time X *could* have happened) (for a review, see Garcia-Retamero & Cokely, 2013). People often neglect the information presented in the denominator when interpreting risk information, thus overestimating the occurrence of risks (Garcia-Retamero &

Galesic, 2009; Reyna, 2004). Icon arrays have been shown to be especially helpful with communicating risks to people with low numeracy (see Galesic et al., 2009). The effectiveness of icon arrays is usually tested in hypothetical scenarios in which participants compare treatment benefits and side-effects (see Galesic et al., 2009; Garcia-Retamero & Galesic, 2010; Hawley et al., 2008). The literature on whether real-world and hypothetical decisions differ provides mixed evidence, usually in the context of risky decision-making (Kühberger et al., 2002). One novel contribution of the current investigation is that we examine the influence of icon arrays on risk perception in a real-world context, which is particularly important because of the immediate public health implications of vaccination. Another unique contribution of this investigation is that we use icon arrays to illustrate a very small risk (~ 1 in 1 million). Typically, in prior investigations the focus has been on much higher side-effect risks. For example, Tait et al. (2010) discussed a 5% side effect risk.

In Experiment 2, we further explore how different types of icon arrays influence vaccine attitudes by adding a condition in which participants viewed the relative risk of experiencing side effects to lives saved by the vaccine. Across both studies, we found evidence that viewing icon arrays prevented increases in aversion to the J&J vaccine and possibly to *all* COVID-19 vaccines.

4.2 Experiment 1

Experiment 1 examined how probability language would influence changes in aversion to the J&J and all COVID-19 vaccines. The experiment also examined whether the presence of an icon array illustrating side-effect-risk would prevent increases in vaccine hesitancy.

4.2.1 Methods

Participants. Data were collected from 1,143 participants from Amazon MTurk. Ninety participants were excluded from the analyses for inattentiveness, leaving 1,052 participants. See demographics in **Table 4**.

Table 4. *Demographic characteristics of participants in Experiment 1.*

Age M (SD)	Gender		Education	
38.81 (14.37)	Female	61.31%	Some High School	.48%
	Male	38.02%	High School	7.7%
	Other	.7%	Some College	12.07%

2-yr degree	9.31%
4-yr degree	55.22%
Advanced degree	15.21%

Design and Materials. Experiment 1 used a 3 (probability expression) by 2 (visualization presence) between-subjects design. Participants were randomly assigned to read the probability of incurring the J&J side effect as a percentage (.0001% of people), ratio (6 in 6.8 million people), or single number (6 people). As an example, the following vignette was shown to those assigned to the single number condition:

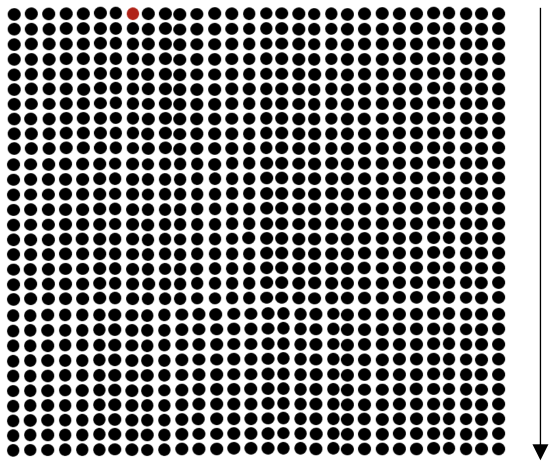
> The US Centers for Disease Control and Prevention and the US Food and Drug Administration are recommending that the United States pause the use of Johnson & Johnson's Covid-19 vaccine over six reported US cases of a "rare and severe" type of blood clot.

Participants were also assigned to view either an icon array depicting the risk of experiencing the blood clotting side effect, or no icon array. The icon array contained one million dots, one of which was red, representing the .0001% probability of experiencing the side effect reported by the CDC. The icon array had labels on the left side of the image, breaking up the visualization into of multiples of 100,000 (e.g., “100,000”, “200,000”, etc.). All of the dots were large enough that they were clearly visible to participants (see OSF for materials).

Participants read the following description:

> In the chart presented below, we illustrate the proportion of people who experience the blood clotting side effect after getting the Johnson & Johnson vaccine. Each dot represents a single person who received the vaccine. One of these dots is red. The red dot represents a person who experiences the blood clotting side effect. Out of all the dots below, only one will experience the side effect.

Figure 19. *Example icon array from Experiment 1*



Note. Figure 19 is an icon array illustrating the 1 (red dot) in 900 chance of experiencing a side effect due to a treatment. The icon array in Experiment contained 1 million dots, one of them red, that participants had to scroll through if assigned to a visualization condition. The arrow on the right represents how participants had to scroll through the array of dots, but this arrow wasn't part of the original figure.

Procedure. Participants provided informed consent, reported their vaccination status, and were shown one of three vignettes about the CDC's new guidelines for the J&J vaccine (depending on condition). If assigned to the icon array condition, the participants viewed this information after reading the vignette. Participants then self-reported their change in attitudes towards the J&J and *all* COVID-19 vaccines with slider scales from 0-100, totaling 6 items:

1. This announcement would make me more hesitant to get (the J&J/any COVID-19) vaccine (shown only to vaccinated participants)
2. This announcement has made me more hesitant to get (the J&J/any COVID-19) vaccine (shown only to unvaccinated participants)
3. I'm more concerned about the safety of (the J&J/any COVID-19) vaccine after this announcement
4. Compared to yesterday, I'm less likely to recommend that my friends and family get (the J&J/any COVID-19) vaccine

Lastly, participants completed the subjective numeracy scale (Fagerlin et al., 2007), to be used as a covariate in the modelling of the data. Participants were compensated $1, and all procedures were determined to be exempt by the University of Michigan IRB. Readers may access our surveys, data, and code at https://osf.io/psvmw/?view_only=7a63dae90fb34411b49a9ffaa7e0d8e4.

Modeling Methods. Slider scale responses to increases in vaccine hesitancy, safety concern, and reluctance to recommend vaccination items were rescaled from 0-100 to 0-1. These items were highly correlated ($r > .8$) and were averaged to create two composite changes in vaccine aversion scores, one for the J&J vaccine and one for *all* COVID-19 vaccines. It is reasonable to assume that the announcement may influence perceptions of the J&J vaccine, however, it is unknown whether the announcement would influence change in attitudes towards other vaccines that were not associated with the reported side effects. Thus, we modelled change in aversion to the J&J vaccine and all COVID-19 vaccines separately, even though they were moderately correlated ($r = .48$ in Exp. 1, $r = .61$ in Exp. 2). When interpreting the composite scores, 1 indicates a large increase in aversion and 0 indicates no increase in aversion toward the vaccine(s).

The two dependent variables were modelled using zero-one-inflated Beta-distributional regression models given that the data were not normally distributed and could only take on values between (and including) zero and one (see **Figure 21**). The zero-one-inflated Beta distribution is a mixture of a beta distribution (for intermediate values between 0 and 1) and a Bernoulli distribution (for extreme values, 0 and 1) via a mixing parameter $\gamma \in [0, 1]$. Intermediate scores between 0 and 1 were described using a beta distribution parameterized with mean (μ) and precision (φ). For scores equal to 0 (no change in aversion) or 1 (large increase in aversion), the probability that the response equals 1 is described by a Bernoulli distribution with a probability parameter (α).

Models for Experiment 1 included the following covariates: vaccination status (vaccinated – unvaccinated), framing condition (percent – number, ratio – number), visualization condition (icon array – none), z-scored subjective numeracy, and the interaction between framing and visualization. Regression formulae for location parameters (μ and α) included all covariates listed above, however regression formulae for the auxiliary parameters (φ and γ) omitted numeracy and interactions between framing and visualization. We implemented the model using the R-package brms: Bayesian Regression Models using 'Stan' (Bürkner, 2017, 2018). Brms translates input models into the probabilistic programming language Stan, enabling approximate Bayesian inference over model parameters using Markov Chain Monte Carlo (MCMC) sampling (Carpenter et al., 2017). We assigned weakly-informative Normal (0,1) priors to regression coefficients and used the default priors provided by brms for all other parameters (v2.14.4).

The model passed all convergence and efficiency diagnostic tests (see Vehtari et al., 2021 for more information). 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 computed 95% credible intervals (CI) 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 ease of interpretation, we replicate the findings below with Factorial ANOVA and report these results in the **Appendix J.** See **Table 5** for descriptive statistics.

4.2.2 Results

First, we examine the influence of condition on increases in aversion toward the J&J vaccine. Our main finding in Experiment 1 is that participants reported lower increases in aversion towards the J&J vaccine if they viewed an icon array (M(SD) = .53(.36)) compared to no visualization (M(SD) = .66(.31)) (β = -0.34, CI = [-0.59, -0.08], pd = 1). After viewing an icon array, participants were also more likely to report no increase in aversion (0) rather than a large increase in aversion (1) toward the J&J vaccine (β = -0.99, CI = [-1.89, -0.02], pd = .98). In contrast to the noticeable effect of visualization, there was no evidence for effects of probability expression (all pd $\leq$ .59) nor interactions between probability expression and the presence of an icon array for intermediate values (all pd $\leq$ .76). There was some evidence that participants were more likely to report a large increase in aversion (1) than no change increase in aversion (0) toward the J&J vaccine if risk was presented as a single number rather than a ratio (β= 1.04, CI = [-0.17, 2.26], pd = .96) and participants were more likely to report no increase in aversion (0) rather than a large increase in aversion (1) if risk was presented as a percentage rather than a ratio (β= -0.7, CI = [-1.75, 0.35], pd = .91) (see **Figure 21a**).

Next, we examined the influence of condition on changes in aversion toward *all* COVID-19 vaccines. After viewing an icon array, participants were more likely to report no increase in aversion (0) rather than a large increase in aversion (1) toward *all* COVID-19 vaccines (β = -1.02, CI = [-2.26, 0.07], pd = .96). However, icon array presence did not affect increases in aversion for those reporting intermediate vaccine aversion scores between 0 and 1 (β = -0.004, CI = [-0.27, 0.26], pd = .51). There was little evidence for effects of probability expression (all pd ≤ .85) or interactions between probability expression and the presence of an icon array (all pd ≤ .72) (see Figure **21b**).

4.2.3 Discussion

Experiment 1 found little evidence for an effect of probability expression on increases in aversion towards vaccination. There was strong evidence that viewing an icon array prevented increases in aversion towards the J&J vaccine and some evidence that such visualizations prevented increases in aversion towards *all* COVID-19 vaccines. These results suggest that viewing an icon array illustrating the potential *risks* of vaccination prevented large increases in aversion toward vaccination. In Experiment 2 we examine whether aversion could be further prevented by viewing an icon array showing both the risks and *potential benefits* of vaccination.

Table 5. *Change in aversion toward vaccination by Condition for Experiments 1 and 2*

Experiment 1

	Change in Aversion to J&J Vaccine				Change in Aversion to *All* COVID-19 Vaccines			
Probability Expression	No Icon Array		Icon Array		No Icon Array		Icon Array	
	M(SD)	N	M(SD)	N	M(SD)	N	M(SD)	N
Number-Only	.69(.30)	179	.53(.36)	163	.42(.36)	179	.38(.38)	163
Ratio	.66(.32)	196	.54(.36)	161	.43(.36)	196	.37(.36)	161
Percentage	.63(.31)	158	.51(.36)	195	.47(.36)	158	.35(.36)	195

Experiment 2

Change in Aversion to J&J Vaccine						Change in Aversion to *All* COVID-19 Vaccines					
No Icon Array		Icon Array (Side Effect)		Icon Array (Relative Risk)		No Icon Array		Icon Array (Side Effect)		Icon Array (Relative Risk)	
M(SD)	N	M(SD)	N	M(SD)	N	M(SD)	N	M(SD)	N	M(SD)	N
.63(.34)	278	.52(.38)	293	.50(.36)	280	.42(.38)	278	.34(.38)	293	.36(.36)	280

4.3 Experiment 2

Interpretation of risks is context-dependent, so viewing the relative risk between vaccine and disease consequences may improve decision making (Reyna, 2008). Thus, in Experiment 2 we included another visualization condition showing the expected lives saved by the vaccine in addition to the risk of incurring the blood clotting side effect (1 million dots with 1 red dot representing risk of side effect and 10,000 green dots representing lives saved, assuming that 1 in 10 unvaccinated people contract COVID-19 and that 1 in 100 of those who contract COVID-19 die (Philip Bump, 2021)).

4.3.1 Methods

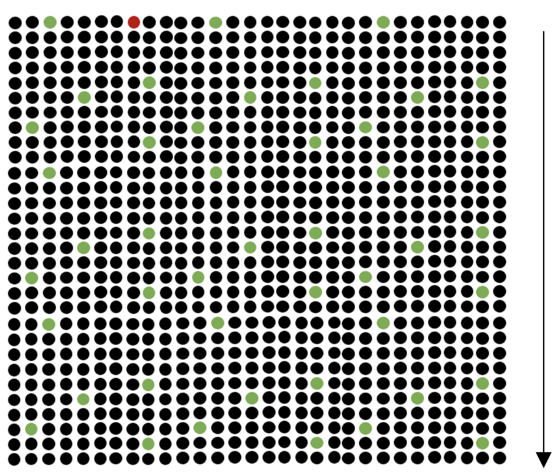
Participants. Data were collected from 903 participants from Amazon MTurk. Fifty-two participants were excluded from the analyses for failing an attention check, leaving 851 participants. See demographics in **Table 6**.

Table 6. *Demographic characteristics of participants in Experiment 2*

Age M (SD)	Gender		Education	
38.81 (14.37)	Female	61.31%	Some High School	.11%
	Male	38.02%	High School	6.58%
	Other	.7%	Some College	13.87%
			2-yr degree	8.70%
			4-yr degree	47.83%
			Advanced degree	22.91%

Design and Materials. Experiment 2 was a between-subjects design where participants were randomly assigned to view one of three visualizations: no visualization, the side effect-only icon array from Experiment 1, or the relative risk icon array illustrating both disease and vaccine risk. All participants viewed the probability expressed as a ratio since there was little evidence for an effect of probability expression in Experiment 1.

Figure 20. *Example icon array from Experiment 2*



Note. Figure 20 illustrates relative risk, where 1 (red dot) in 900 experience a side effect and 1 (green) in 20 lives are saved by the treatment. The relative risk icon array in Experiment 2 contained 1 million dots that participants had to scroll through if assigned to a visualization condition. The arrow on the right represents how participants had to scroll through the array of dots, but this arrow wasn't part of the original figure.

Modelling Methods. Models for Experiment 2 included only vaccination status, visualization condition, and z-scored subjective numeracy as covariates.

4.3.2 Results

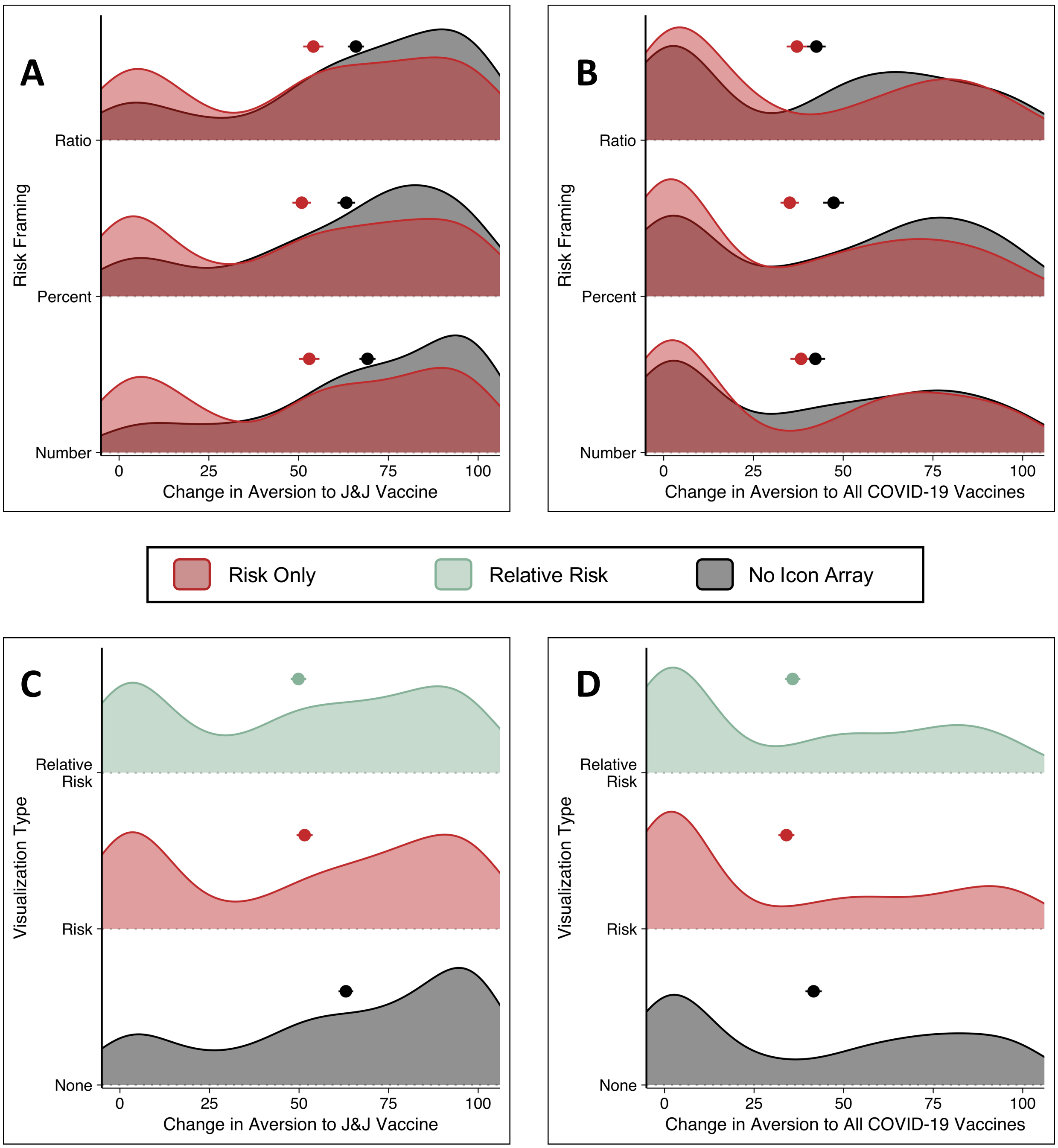
In Experiment 2, we successfully replicated the key results of Experiment 1. Participants self-reported lower increase in aversion to the J&J vaccine if they viewed an icon array illustrating probability of side effect (M(SD) = .52(.38)) compared to no visualization (M(SD) = .63(.33)). Viewing this icon array also prevented increases in aversion for those with intermediate scores (β = -0.24, CI = [-0.43, -0.06], pd = .98). Participants were again more likely to report no increase in aversion (0) rather than a large increase in aversion (1) after viewing the icon array (β = -1.48, CI = [-2.05, -0.89], pd = 1). Viewing an icon array of relative risk was also associated with lower increases in vaccine aversion when compared to the no-visualization condition (M(SD) = .49(.36)) (β = -0.20, CI = [-0.37, -0.02], pd = .96). Participants viewing the relative risk visualization were also more likely to report no increase in aversion rather than a large increase in aversion (β = -1.75, CI = [-2.39, -1.13], pd = 1). The relative risk and side-effect-only icon arrays appear to be equally effective in preventing increases in vaccine aversion (see SI; **Figure 21c**).

Viewing the side-effect-only icon array was associated with lower increases in vaccine aversion for intermediate values (β = -0.27, CI = [-0.46, -0.07], pd = .98), but the presence of an icon array did not affect the probability of reporting large increases in aversion rather than no increase in aversion (β = -0.40, CI = [-1.00, 0.20], pd = .86). Increases in vaccine aversion after viewing the relative risk icon array were no different from viewing no visualization (β = -0.06, CI = [-0.25, -0.12], pd = .69). After viewing the relative risk icon array, people were more likely to report no increase in aversion, rather than a large increase in aversion (β = -0.77, CI = [-1.48, -0.01], pd = .96) (see **Figure 21d**).

4.3.3 Discussion

Experiment 2 replicates the main finding from Experiment 1 that viewing icon arrays of small side-effect risk prevented increase in aversion toward the J&J vaccine. There was also some evidence that viewing these icon arrays prevented increased aversion toward *all* COVID-19 vaccines more generally. There was no evidence suggesting that viewing the relative-risk icon array was more beneficial than viewing a side-effect-only icon array.

Figure 21. *Change in aversion toward the J&J and all COVID-19 vaccines by Experiment and Condition*



Note. Panels a and b illustrate mean and standard error change in vaccine aversion by condition in Experiment 1. Notice that the data are displayed as overlapping distributions. Point color indicates probability expression group (see legend). Panels c and d illustrate mean and standard error change in vaccine aversion by condition in

Experiment 2. Note that while the y-axes above range from 0.25 to 0.75, the full range was 0 to 1 and that the data are displayed as stacked distributions.

4.4 General Discussion

The main takeaway from this research is that presenting icon arrays illustrating the very small risk of experiencing side effects in tandem with the announcement from the CDC could have minimized increases in vaccine hesitancy to both the J&J and possibly *all* COVID-19 vaccines. These results provide evidence that icon arrays are effective at communicating risk information outside of the lab, in a real-world context with real-world consequences. We are optimistic that our findings contribute to the literature on risk-perception more generally, as other work shows icon arrays to similarly improve decision-making in many different contexts (e.g., Galesic et al., 2009; Garcia-Retamero et al., 2010; Okan et al., 2012; Walker et al, 2022; Waters et al., 2007a; Zikmund-Fisher et al., 2008), although some evidence is mixed (e.g., Recchia et al., 2022; Ruiz et al., 2013; Waters et al., 2007b). Given that much of the prior work on icon arrays has been in the context of hypothetical scenarios, while the current study was in the context of real-world decision-making, we also provide evidence that icon arrays are effective in more than just hypothetical decision-making.

Another contribution of our work is the finding that icon arrays can effectively communicate very small risks (.0001%). However, it is possible that the presence of the single red dot in the array did not matter, and that the visualization prevented increases in vaccine hesitancy by helping participants understand the magnitude of 1 million. Prior work shows that it is difficult for everyday people to conceptualize very large numbers, such as 1 million (see Landy et al. 2013). The icon array provides a concrete representation of an abstract idea by showing participants 1 million icons. By scrolling through the icon array, this may help participants understand just how large 1 million is. This could also explain why we find no difference between the side-effect only and relative risk icon arrays in Experiment 2. Alternatively, the main reason why icon arrays are thought to be beneficial in reasoning about probabilities is that they highlight the denominator (Garcia-Retamero & Cokely, 2013). If providing this concrete representation helps people better understand the magnitude of 1 million, it may also help them understand the magnitude of the denominator. Thus, it is possible that the icon array both helped participant conceptualize the magnitude of 1 million and overcome denominator neglect. Future research should disentangle these possibilities.

Conceptually, scrolling through an icon array of 1 million icons may help people understand risk magnitude through other cognitive mechanisms. Padilla et al. (2018) present a dual model of visualization processing for decision-making, where Type I processing is heuristic-based and open to perceptual biases, while Type II processing is more effortful and is associated with higher levels of accuracy in graph-based reasoning. Scrolling through the icon array displaying very small risk may help people engage with the visualization through a Type II pathway as the visualization provides viewers with both a temporally coded and visually coded risk estimate.

One alternative explanation for the findings is that viewing the visualization made the data appear more trustworthy, resulting in lower increases in vaccine hesitancy. Some prior work has found that other types of data visualization, such as bar graphs (Tal and Wansink, 2016), increase the perceived credibility of data. However, more recent work has cast doubt on the validity of these findings (see Dragicevic and Jansen, 2018; Fansher et al., 2022a). Future work could explore if including icon arrays influences the perceived trustworthiness of data.

Limitations

One limitation of the current study is that we did not compare the effectiveness of icon arrays to other types of data visualizations. It is possible that icon arrays were more effective because they repeated the information given in the vignette graphically. However, we have reason to believe that icon arrays helped participants understand risk magnitude beyond repetition given that other studies that have compared icon arrays to other types of data visualizations (without controlling for repetition) have found icon arrays to be most effective (e.g., Waters et al., 2007a; Tait et al., 2010). Another limitation is that participants self-reported their changes in attitudes towards vaccination. Ideally, we would have measured vaccine hesitancy both before and after the announcement (which, of course, was logistically not possible). One alternative explanation, and possible limitation, of the finding that there was no difference between the side-effect-only icon array and relative-risk icon array in Experiment 2, is that our participants were not tested for red/green colorblindness. To test this possibility. since colorblindness is a sex-linked trait, we reran the Experiment 2 analysis with only the females in our sample, and still found no difference between groups ($p >= .42$). This suggests that possible red/green colorblindness in our participants did not significantly influence our results. Lastly, it is possible that the high complexity of the language we used (i.e., “more hesitant”) introducing construct-irrelevant variance because the instructions may not have been understood equally well by all participants.

4.5 Conclusion

Regardless of these limitations, we believe our results suggest that icon arrays can prevent large increases in vaccine hesitancy from small risks. Future work could examine if such techniques would also be beneficial at communicating small probabilities in contexts other than side effect risk and vaccine hesitancy. For example, in the context of COVID, other potential side effect risks beyond the blood-clotting side effect could be examined. Caution should be taken when communicating information about such side effects to the public, especially given that people tend to take no action if the action is perceived to potentially cause harm, even if there is a greater risk of inaction (i.e., abstaining from vaccination, (Bond & Nolan, 2011).

Chapter 5 A Review of COVID Visualization Research

The goal of my dissertation was to examine how everyday Americans understood COVID data and visualizations, whether explanations of data through visualization would improve understanding of COVID-related concepts, and whether understanding of data would be related to attitudes towards preventative behaviors and policies. While the series of experiments reported in the previous chapters were occurring, other groups of social scientists were conducting their own research on these topics. I extensively reviewed this body of literature and identified three key areas of research: (1) how people understood COVID data visualizations, (2) the influence of data visualizations on COVID-related risk perception, and (3) how graphs were used to mislead the public.

In the current chapter I synthesize this body of literature with my own findings to provide a broader perspective on how people understood data during the pandemic, and how this understanding shaped risk perception. Another goal of my dissertation was to examine whether work conducted in the context of COVID replicated key findings from past psychological research that was usually conducted in the context of hypothetical scenarios. As such, I also discuss whether the key findings from this work align with past research and theory.

5.1 Understanding of Common COVID Data Visualizations

While graphs are often useful for making quantitative information easier to understand, they are not always properly understood (Franconeri et al., 2021; Glazer, 2011; Shah et al., 2005). As such, researchers have studied how people understood common COVID visualizations throughout the pandemic, namely graphs depicting time-series data. In this section I summarize how people understood data in the context of exponential growth functions, logarithmic transformations of data, and incident to cumulative transformations of data. I also discuss how these visualizations influenced risk perception in the form of attitudes towards social distancing, worry about the virus, etc.

Understanding Exponential Functions

People Exhibit Exponential Growth Bias. At the beginning of the pandemic, COVID was rapidly spread throughout the population as no preventative actions were in place. In order to understand how the number of cases could grow so quickly, one needs to understand the concept of exponential growth. This understanding may allow for proper appreciation of COVID's threat leading to greater adherence to social distancing measures. Unfortunately, it is well documented that people misunderstand exponential growth and often exhibit exponential growth bias (Levy & Tasoff, 2015). Heyd-Metzuyanim et al. (2021) examined mathematical literacy in a group of over 1,100 Jewish Israelis in the context of COVID-related topics commonly covered in the media, and found that less than half of their participants were able to answer basic questions about exponential functions (i.e., is 3, 6, 9, 12, 15 an exponential sequence?). Given this finding, it is of no surprise that people generally misunderstood visualizations depicting exponential functions.

A number of researchers examined whether people underestimated the growth of exponential trends (i.e., exhibited exponential growth bias) when asked to forecast the future number of COVID cases. In Chapter 2 (Fansher et al., 2022b) I discuss a series of three studies examining how people understood exponential growth as it related to disease incidence. In Study 1, conducted at the start of the pandemic, we found that participants underestimated the growth of a rapidly accelerating exponential function. In contrast, in Study 2 we found that participants tended to overestimate the growth of COVID cases when the growth curve appeared to be less exponential (more linear). In a third study we replicated these findings and concluded that participants tend to underestimate more exponential functions and overestimate more linear growth. Prior literature shows that underestimation of exponential growth trends increases with an increasing exponent, supporting this assertion (Wagenaar & Sagaria, 1975). Lammers et al. (2020) and Banerjee et al. (2021) conducted similar studies at the start of the pandemic in March 2020 (see below), and found that participants underestimated the growth of rapidly accelerating functions similar to the stimuli we used in Study 1.

Lammers et al. (2020) asked participants to estimate the number of cases over the *past 5* days. They found that participants exhibited exponential growth bias and generally predicted more linear trends when asked to estimate the past number of cases. However, a simple intervention merely explaining the concept of exponential growth corrected these misperceptions ("the number of corona patients doubles and keeps doubling every three days" pg. 16265). In a separate study, participants were given an estimate of the current number of COVID cases and were asked to predict the number of cumulative cases in 15 days, given that the number of cases doubles every three days. They found that instructing participants to generate their forecasts at 3-day intervals decreased exponential growth bias when compared to a control condition who was not given these instructions. In another line of work, Banerjee et al. (2021) had participants view three data points (weeks 1-3) and were asked to predict the number of cases that would be reported on weeks 4 and 5. Participants underestimated the growth and the researchers determined that they made their estimates based on a linear rather than exponential model.

Understanding Exponential Growth is Related to Risk Perception. All three of the empirical lines of work described above examined the relationship between understanding exponential growth and risk perception. In Chapter 2 I find mixed evidence on whether understanding of exponential growth was related to social distancing attitudes. In my Study 1 there was a positive correlation between forecast size and the anticipated amount of time that would pass before we could stop social distancing, as well as past and future social distancing intentions. In Study 2 we only replicated the finding that forecast size was correlated with estimated time to stop social distancing orders in the United States. Lammers et al., (2020) found that participating in an intervention that decreased exponential growth bias was associated with increased support for social distancing, and Banerjee et al. (2021) found that the greater one's exponential growth prediction bias, the less participants reported complying with safety measures and the more they endorsed the violation of safety norms, even after correcting for individual difference factors like education.

Viewing Raw Data Decreases Exponential Growth Bias. In Chapter 2 I explore the role of data visualization type on understanding of exponential growth. I find that viewing data in the form of tables, rather than graphs, improved forecasting accuracy. This finding was true regardless of whether participants generally over- or underestimated the time series data. In addition, participants who viewed graphs were falsely more confident in their predictions than those who viewed tables. Banerjee et al., (2021) explored similar ideas by randomly assigning participants to view the data as a line graph or as raw numbers. They found that viewing the data as raw numbers significantly reduced exponential growth prediction bias. While the topic of data presentation was not explicitly covered in Lammers et al., (2020), I suggest that their work may also support the idea that viewing the data in numeric form decreases exponential growth bias.

They found that having participants forecast intermediate timepoints led to more accurate forecasting. It could be the case that viewing all of the data together, in numeric form, improved forecasting accuracy. However, follow-up studies would need to be conducted to confirm this proposal.

Summary. In sum, researchers found that participants exhibited exponential growth bias at the start of the pandemic. This aligns with prior work suggesting that people tend to underestimate exponential growth (Levy & Tasoff, 2015). Chapter 2 provides promising evidence that the degree to which participants underestimate exponential growth is influenced by the size of the exponent, an idea that has been discussed in past research (Wagenaar & Sagaria, 1975) but hardly explored. This work also provides evidence that understanding of exponential growth was related to perception of COVID-related risks as well as intentions to engage in preventative behaviors such as social distancing. Lastly, this body of work contributes to the sparse literature on how data presentation format influences exponential growth forecasting, suggesting that viewing graphs may promote exponential growth bias and lead to false confidence. The reason why tables or raw data points improve forecast accuracy should be explored in future work. While Banerjee et al. (2021) do not provide possible explanations for this phenomenon, in Chapter 2 I suggest that viewing the raw numbers may allow participants to extract patterns from the data that would otherwise go unnoticed. For example, people may use the visual heuristic of "adding a digit to each row of numbers" to generate exponential growth forecasts or they may spend more time computing the next number in a sequence of numbers as opposed to visually extrapolating a graph.

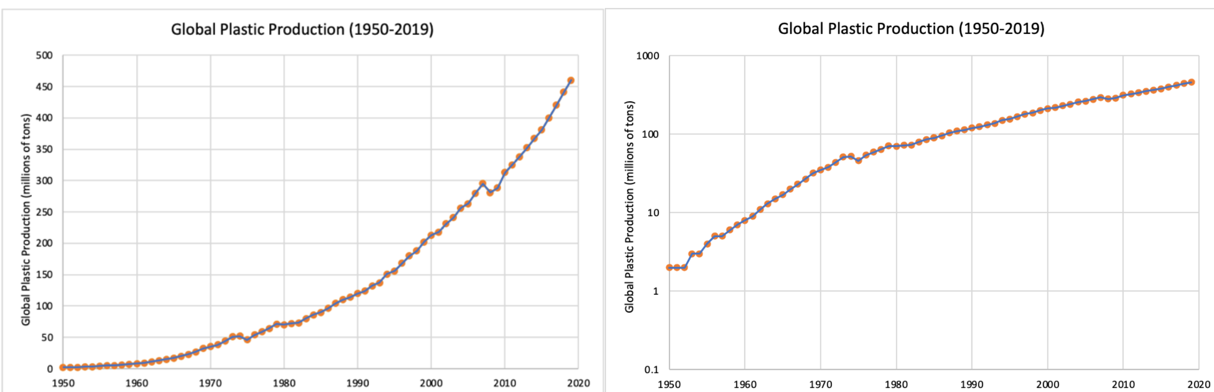
Understanding Logarithmic Functions

At the height of the pandemic, COVID death and case data were often presented to the public in logarithmic form. Several groups of researchers examined whether the public understood logarithmic functions and how the scale of time series graphs impacted risk perception. Ryan and Evers (2020) found evidence that people misinterpreted logarithmic graphs of COVID data and that viewing logarithmic graphs was associated with less worry. Participants indicated that they would be less likely to wear a mask and less likely to socially distance when shown the logarithmic data in comparison to the linear data. Educating people about logarithmic functions decreased the effect of the y-axis scale on risk perception but did not eliminate the effects. In a similar study, Romano et al. (2020) randomly assigned participants to view graphs depicting COVID death data in either linear or logarithmic form. They found that participants frequently misunderstood the data when plotted logarithmically and reported less worry about the virus. Viewing the data in linear form was also associated with increased worry about the virus. While these two studies support the idea that viewing logarithmic functions was associated with decreased risk perception, a study conducted by Sevi et al. (2020) failed to replicate this result. A group of 2,500 Canadian participants were randomly assigned to view COVID case data in linear or logarithmic form. They were also asked whether they supported stay at home orders and when they estimated people would be expected to return to work. They found no differences in risk perception between the two groups. What could have caused the discrepancy between these findings?

Romano et al. (2020) suggest that the shape of the graph may influence worry - an upwards linear trend indicates that there is no sign of the virus stopping while a flatter logarithmic trend may appear less ominous. Research from Padilla et al. (2022) supports this claim. In an initial experiment they found that viewing graphs of cumulative COVID deaths was associated with greater perceived risk when compared to participants who viewed graphs of incident (daily) deaths. In a second experiment they tested the hypothesis that the cumulative graph was associated with greater risk perception because the cumulative curve showed an increasing trend, while the incident graph showed a decreasing trend. To test this hypothesis participants were assigned to view a cumulative curve with an increasing trend or an incident curve with an increasing trend. Here, the authors found no difference in risk perception between groups, suggesting that the graph's slope influences risk perception. This finding supports the idea that logarithmic graphs evoke less worry than linear graphs of the same data because they have a less steep slope. Given this information, it's possible that Sevi et al., (2020) may have failed to find a relationship between y-axis scale and risk perception because the two graphs used in their study both showed increasing trends. However, Ryan and Evers (2020) found that risk perception was lower in the logarithmic scale group regardless of graph slope, suggesting that characteristics of the logarithmic scale other than slope may be influencing perceived threat.

Summary. In sum, this work replicated previous findings that people often misunderstand logarithmic graphs (Heckler et al., 2013; Menge et al., 2018), which is especially concerning given that a recent review found that the media rarely explained the meaning and significance of the logarithmic curve (Hammes et al., 2021). This work also suggests that risk perception may be shaped by the physical shape of the graph, with increasing slopes leading to increased concern. Future work should examine this possibility, and if true, examine whether the perceptual characteristics of time-series graphs may influence risk perception in other contexts. For example, viewing the data on global plastic production in **Figure 22** in logarithmic form may mask that global plastic production is increasing exponentially (data from Rose & Ritchie, 2022). This work also suggests that media outlets should use linear/cumulative scales rather than logarithmic/incident scales when presenting COVID data to the public, as linear functions are more likely to be understood and linear/cumulative functions promote appropriate cautionary behavior.

Figure 22 *Illustrates global plastic production (1950-2019) in linear (left) and logarithmic (right) form*



Understanding Incident and Cumulative COVID Functions

In Chapter 2 participants were tasked with extrapolating cumulative COVID case graphs. We found that a large percentage of our participants generally misunderstood the concept of cumulative growth, predicting that the cumulative curve could *decrease*. This indicated a general misunderstanding of accumulation functions in the general population, which was troubling given that graphs showing cumulative data were common visualizations used to communicate disease prevalence with the public. As such, in Chapter 3 I discuss a longitudinal study examining whether people understood the relationship between daily and cumulative case curves. We found that people did not understand the relationship between daily and cumulative curves, and that people tended to rely on the correlation heuristic when making their judgements (i.e., assume that the daily and cumulative curves will look similar to one another). These findings align with prior work on stock-flow failure (Cronin & Gonzalez, 2007, Sweeney &

Sterman, 2000). However, participating in a brief video intervention improved understanding of the relationship between daily and cumulative cases, with the effects lasting a minimum of 6-7 weeks after intervention. Participating in the intervention also was associated with more favorable attitudes towards social distancing and social distancing policies.

Summary. People generally misunderstand the concept of accumulation functions and the relationship between incident and cumulative graphs, often relying on simple visual heuristics (i.e., the correlation heuristic) when making their judgements. Understanding of the relationship between daily and cumulative functions is associated with increased risk perception. Relatively simple interventions can help the public understand these topics, with research suggesting that the effects of these interventions are long-lasting.

5.2 Data Visualizations Influence COVID-Related Risk Perception

There is a vast literature on how data visualizations influence risk perception, especially in the context of medical decision-making (for reviews see Ancker et al., 2006; Garcia-Retamero & Cokely, 2017; Lipkus, 2007; Padilla et al., 2018). In the previous section I highlight research showing that graph understanding is related to COVID risk perception in multiple contexts. Understanding exponential growth is associated with greater risk perception, viewing logarithmic functions of COVID data is associated with less worry than viewing linear functions, cumulative graphs evoke greater risk perception than incident graphs, and understanding the concept of accumulation is associated with greater support for mitigation measures. The second key theme that emerged from research on COVID data interpretation is that data visualizations can evoke different levels of risk perception depending on their features. In this section I further expand upon this finding through a series of examples.

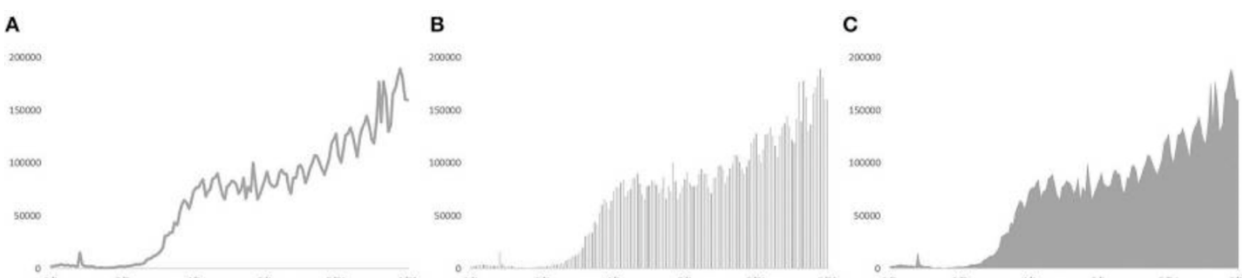
Icon Arrays Decrease Vaccine Hesitancy

People often demonstrate probability neglect (i.e., ignoring the denominator) resulting in overestimation of risk (Reyna, 2004). After the CDC announced that they would be recalling the Johnson & Johnson vaccine in April 2021 due to a rare side effect (Centers for Disease Control and Prevention, 2021), we examined whether icon arrays could be useful for reducing possible increases in vaccine hesitancy caused by this announcement. In Chapter 4 (Fansher et al., 2022c) we discuss two studies consisting of ~2,500 MTurk workers. We show that viewing an icon array illustrating the small probability of experiencing the side effect (~ 1 in 1 million) significantly reduced possible increases in vaccine aversion towards both the Johnson & Johnson vaccine as well as all COVID vaccines. These findings replicate prior work showing that icon arrays effectively illustrate probabilities and help reduce risk perception (Tait et al., 2010; Waters et al., 2007a). This suggests that media outlets should consider using icon arrays when communicating probability information to the public to decrease risk aversion, and that icon arrays are effective at communicating very small probabilities. In addition, icon arrays may affect real-life decision-making in contexts outside of the lab.

Graph Area Affects Risk Perception

Previously we have discussed the possibility that people interpret graphs based on simple visual heuristics (e.g., slope of the line, use of the correlation heuristic). Another possible graph interpretation heuristic explored in Luo et al. (2022) is that people based their concerns about COVID on the contrast between the data and background of time series graphs. It could be the possibility that graphs with greater area could lead to increased perception. To study this idea, participants were randomly assigned to view graphs of new cases either as line graphs, bar charts or stacked bar charts (see **Figure 23**). They found that participants viewing the stacked bar charts reported the highest levels of COVID anxiety and greater social distancing intentions than the other two groups. This suggests that the area of such visualizations influences risk perception. This interesting novel finding has many potential applications. For example, what is the role of foreground when interpreting other types of charts (i.e., scatter plots)? Does merely increasing the y-axis scale to decrease the foreground area affect risk perception? Does this finding apply to contexts outside of COVID incidence?

Figure 23 *Figure 1 from Luo et al. (2022)*

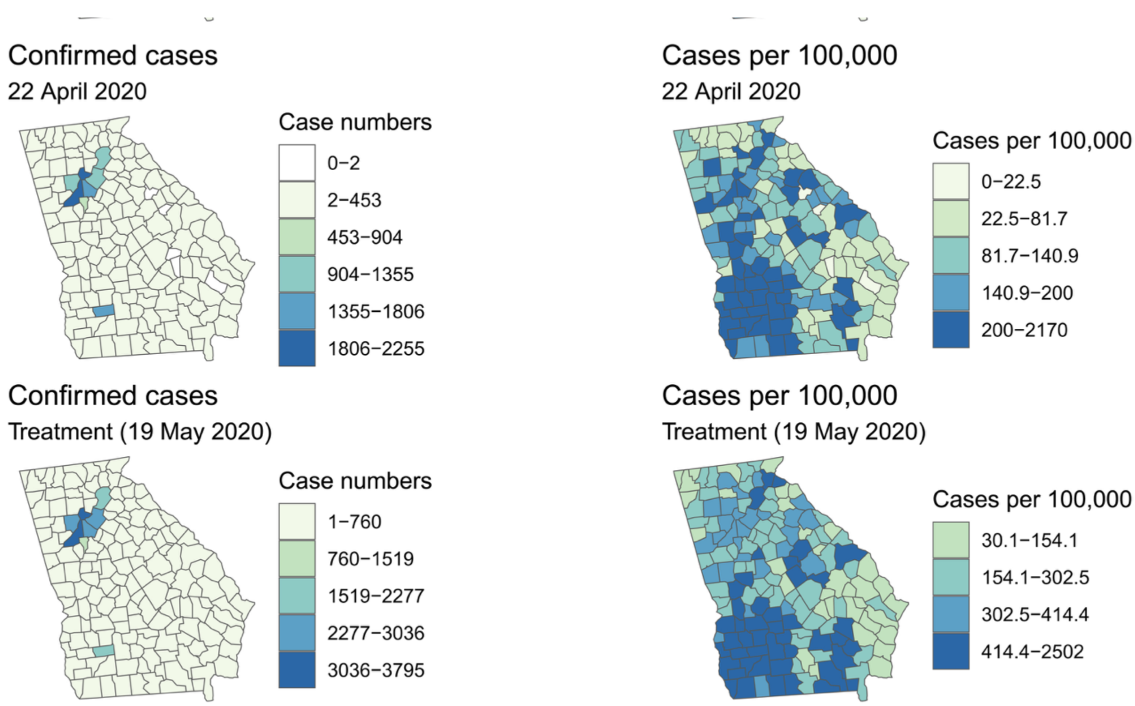


Note: (a) line graph, (b) bar chart, (c) stacked bar chart

Case Rate Maps Improve Understanding

One common method used throughout the pandemic to illustrate COVID data was heat map or geospatial visualizations (Fan et al., 2022; Zhang et al., 2021). While these types of visualizations are useful for identifying areas where transmission risk is high, unfortunately, the media often fails to correct for population size in these types of visualizations (Hammes et al., 2021). This makes it difficult to compare the rate of infection between different regions, inflating risk in high population areas and deflating risk in low population areas. Engel et al. (2022) investigated how people understood choropleth maps presented by the Georgia Department of Health to the public in June 2021. They found that correcting for population size significantly changed perceived threat from COVID. Participants viewing case rate (right of **Figure 24**) maps expressed greater hesitation about reopening and greater concern about the virus in comparison to participants who were shown maps of raw numbers (left of **Figure 24**). This suggests that geospatial visualizations presented to the public should always correct for population size for more accurate decision making.

Figure 24 *Choropleth maps used in Engel et al. (2022) showing raw cases (left) compared to case rates (right)*



Process Data Visualizations Increase Risk Perception

Visualizations used to communicate COVID data are often static. A study by Witt et al. (2022) suggests that process visualizations may be useful at helping people understand exponential growth and the effects of social distancing on COVID spread. Their participants viewed either static visualizations illustrating COVID spread with varying degrees of social distancing in the population, or the process visualizations presented in a popular Washington Post article (Stevens, 2020). In this article, the author walks the reader through various visualizations simulating COVID spread throughout a population that is engaging in varying degrees of social distancing. People are represented by colored dots that bounce around the screen, when they touch one another this can be considered a social interaction. The visualization shows how COVID can be spread throughout a population exponentially as infected people (orange dots) interact with healthy (blue dots) and recovered (pink dots). Witt et al., (2022) found that interacting with the process visualizations was associated with increased intentions to engage in social distancing and more favorable attitudes towards social distancing in comparison to participants who viewed static images. This suggests that process visualizations were effective at helping people understand the importance of social distancing early in the pandemic and that process visualizations could be useful in contexts outside of COVID as well. Such visualizations should be used in the future to help the public understand exponential growth.

5.3 COVID Graphs Can Be Misleading

COVID graphs significantly influenced risk perception in a variety of contexts and risk perception influenced intended social distancing behaviors. One perhaps unsurprising but disheartening finding from research on COVID visualizations is that misleading graphs were sometimes used in public communications. For example, as previously mentioned, failing to correct choropleth maps for population size could have led to inaccurate comparisons of COVID severity between different counties in Georgia (Engel et al., 2022). When creating visualizations, specific design choices can change how one perceives the data, and graphic designers do not always follow the best principles for creating data visualizations. For example, one common finding is that a truncated y-axis can be used to potentially mislead (e.g., Yang et al., 2021). Poorly designed graphs contributed to the intentional or unintentional spread of misinformation as the public adjusted their perceived threat from COVID based on misleading graphs.

Several groups of researchers have examined the visualizations used in media communications to identify cases where misleading graphs were used. Doan (2021) highlights three examples of when design choices were used to intentionally mislead the public. For example, visualizations used by Fox 31 downplayed the growth of the virus by using a truncated y-axis and inconsistent scaling on the y-axis (see **Figure 25**). Similarly, Engledowl and Weiland (2021) discuss two other examples of misleading COVID visualizations. The first example highlights how graphs with two vertical axes make the information difficult to understand, especially when the axes are on different scales. **Figure 26** (bottom) shows how multiple inconsistent scales were used to give the impression that counties with mask mandates (red) had fewer COVID cases than counties without mask mandates (blue). In the top panel, the data is presented on a single vertical axis and paints a completely different picture. The second example shows how a misleading x-axis can lead to misinformation. It appears that the Georgia Department of Public Health intentionally presents the dates on the x-axis in non-chronological order to give the impression that COVID cases were consistently decreasing over time (see **Figure 27**).

Figure 25 *An example of a misleading visualization provided by Doan (2021)*

Dare Obasanjo's tweet from April 4th, 2020 shows a line graph taken from Fox 31. The graph is titled "New Cases Per Day" with the "Total cases 3,342" listed in the upper right corner. The x-axis includes a label for each day between March 18 and April 1; the y-axis starts at 30 and ends at 400. Although the vertical space between numbers on the y-axis stays constant, misrepresenting the numerical difference between numbers.

Figure 26 *Example 1 provided by Engledowl et al. (2021)*

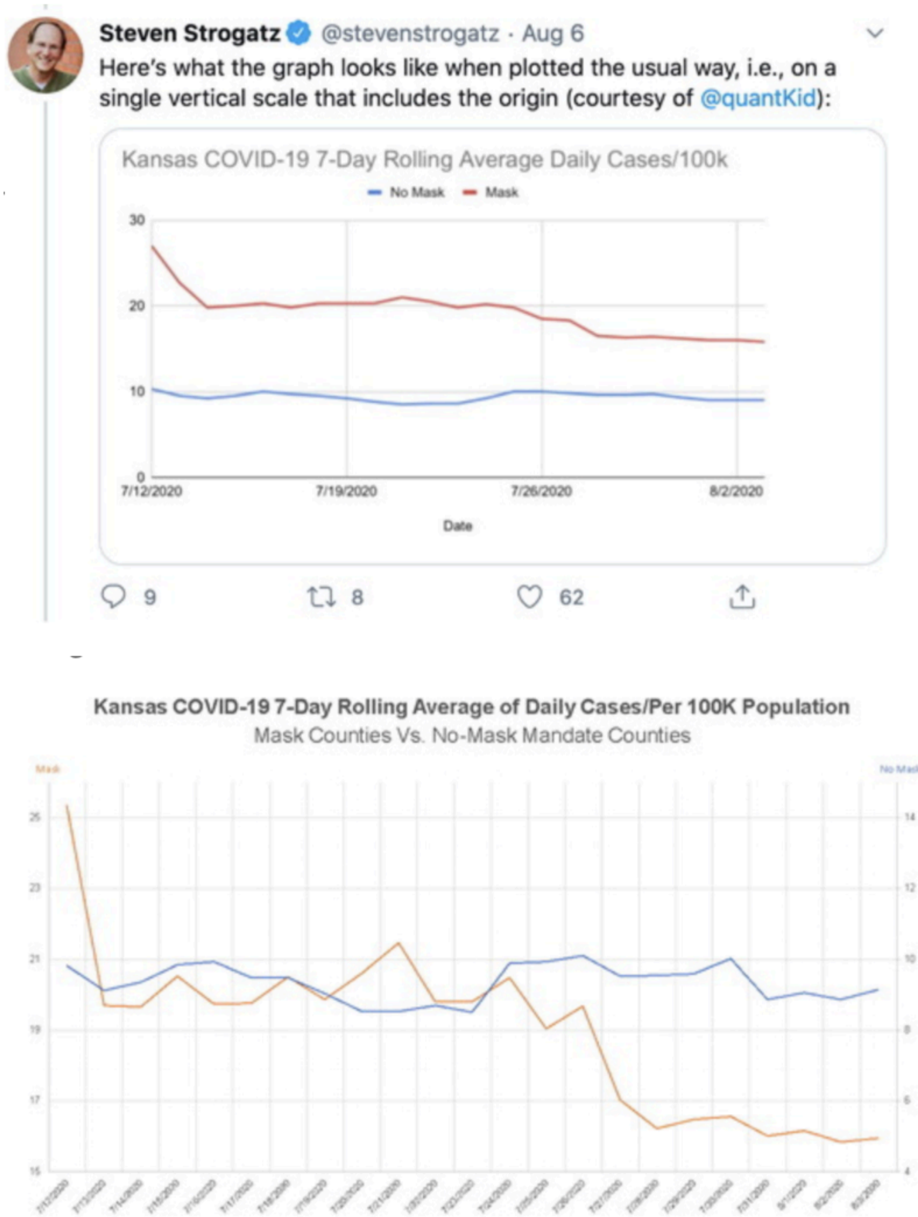
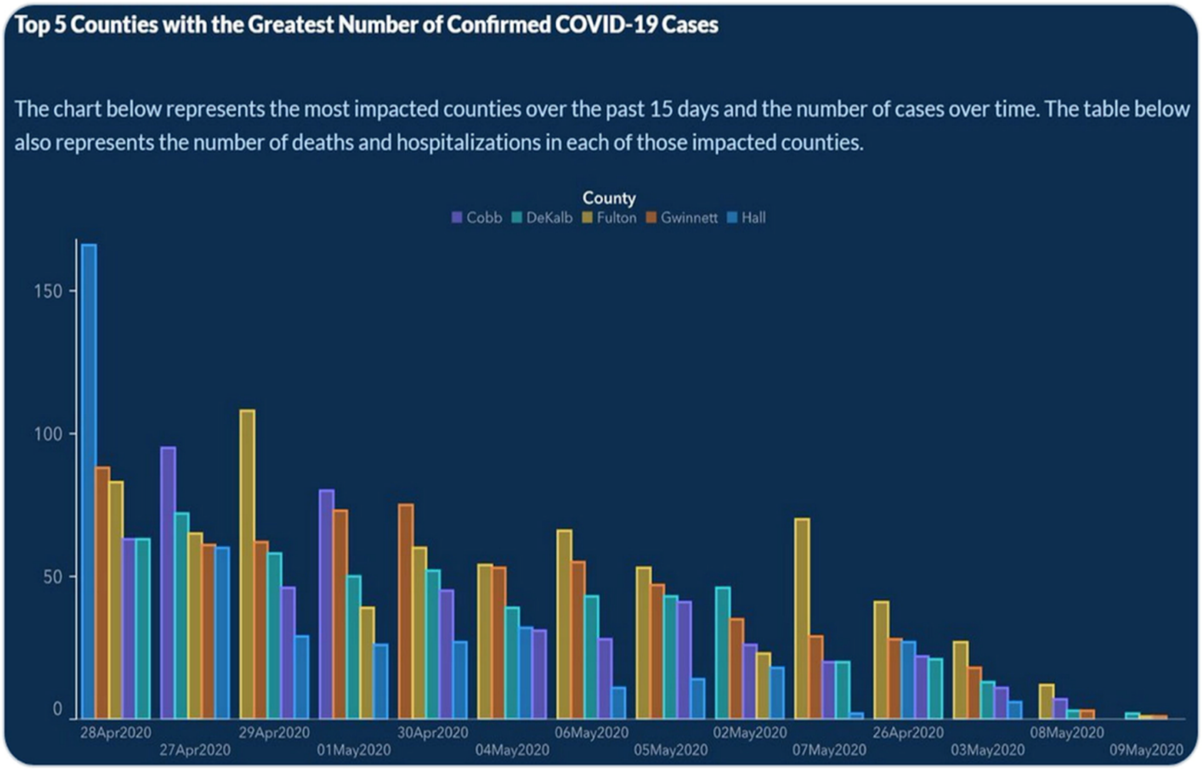
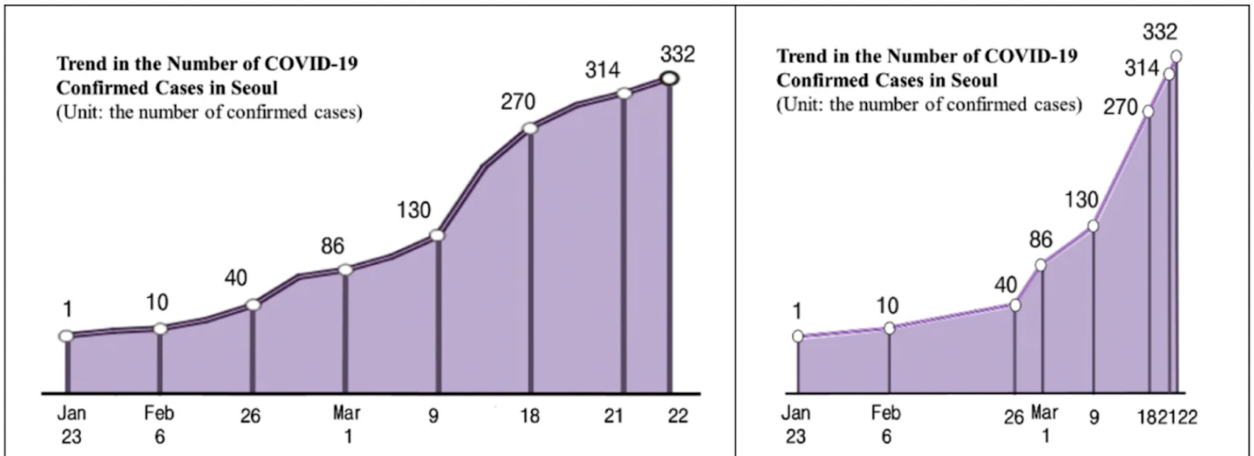


Figure 27 *Example 2 provided by Engledowl et al. 2021*



Kwon et al. (2021) examined all the COVID news stories including graphs that were published in common Korean media outlets from January to April 2020. They found that graphs commonly used inaccurate x- or y-axis scaling. They use **Figure 28** as an example. Here we see that the incorrect y-axis scaling of the graph used in a media report (left) downplayed the exponential growth of COVID in Seoul. The right panel of **Figure 28** shows the actual growth of confirmed cases with correct scaling.

Figure 28 *Figure 3 in Kwon et al. (2021)*



Excerpt from B.W. Hwang (2020) | Corrected scale

In sum, we find evidence that graphs were used to mislead the public during the pandemic. This highlights the importance of teaching graph literacy skills to people so that they can appropriately interpret misleading visualizations.

5.4 Conclusion

In Chapter 5 I synthesize the work on how people understood COVID visualizations and how COVID visualizations were presented to the public throughout the pandemic. Some of the key findings from this systematic review are that people often misinterpreted COVID data when presented graphically, and that their interpretations of data influenced their risk perception. Additionally, misleading visualizations were presented to the public which could have led to inaccurate assessments of COVID risk. This body of literature suggests that best practices for data visualizations and visualizations of risk are applicable to real-life scenarios.

Chapter 6 General Discussion

This dissertation provides evidence that people misunderstood commonly used data visualizations during the pandemic, and that one's understanding of COVID data was associated with risk perception. In this chapter I discuss the major theoretical contributions and broader impacts of the presented research.

6.1 Summary of Main Findings

In Chapter 2 I show that people over- or under-estimated exponential growth trends depending on the linearity of the data, that viewing tables of data improved forecasting accuracy compared to graphs, that viewing graphs was associated with false confidence in one's forecasts, and some evidence that attitudes towards social distancing was positively correlated with the magnitude of participants' forecasts. In Chapter 3 I show that people misunderstood the relationship between daily and cumulative case curves and that participating in a brief video intervention improved understanding of accumulation. The effects of the intervention were long lasting and transferred to contexts outside of COVID. Participating in the intervention was also associated with more favorable attitudes towards social distancing and social distancing policies. In Chapter 4 I show that viewing icon arrays illustrating the 1 in 1 million chance of experiencing the reported side effect from the Johnson & Johnson vaccine prevented significant increases in aversion towards the Johnson & Johnson vaccine as well as all COVID vaccines. Lastly, in Chapter 5 I provide a synthesis of the literature conducted during the pandemic on how people understood COVID visualizations and described three main findings: (1) people misunderstood commonly used COVID visualizations, (2) data visualizations influenced risk perception, and (3) graphs were sometimes used to mislead the public during the pandemic.

Theoretical Contributions

This body of research provides several general contributions to the fields of cognitive and educational psychology as well as media communications. First, in Chapter 2 I thoroughly investigate how data visualization impacts how one forecasts exponential growth. This area of research has largely been neglected. I find evidence aligning with other COVID researchers

(Banerjee et al., 2021) suggesting that viewing data in tables improves forecasting. I also find evidence suggesting that people misunderstood exponential growth as it related to COVID spread, aligning with prior research showing that people exhibit exponential growth bias (Lammers et al., 2020; Levy & Tasoff, 2015). I also provide evidence that viewing graphs may lead to false confidence – a novel finding in this body of work. In Chapter 3 I contribute to the literature on stock-flow reasoning by replicating previous findings that people misunderstand accumulation and tend to use the correlation heuristic when reasoning about accumulation (Cronin & Gonzalez, 2007, Sweeney & Sterman, 2000). I also present a novel intervention that helped people understand the concept of accumulation, suggesting that although stock-flow failure is a pervasive issue, it can be corrected through simple intervention. In Chapter 4 I replicate previous findings that icon arrays are effective at helping people overcome denominator neglect (Tait et al., 2010; Waters et al., 2007a). This study adds to the literature on health visualizations by showing that icon arrays are effective at illustrating very small probabilities as small as 1 in 1 million.

6.2 Broader Impacts

The findings of this work are applicable to many different contexts. For example, the findings from Chapter 2 are applicable to any situation in which it is important to understand exponential growth. While the most obvious context is the spread of disease throughout a population, other relevant examples include financial contexts like understanding accumulation of interest on savings, health contexts like how cancer is spread throughout the body, and environmental contexts like growth of human or animal populations. Chapter 3 finds evidence that people misunderstand the concept of accumulation, thus journalists should refrain from using public health messaging like "flatten the curve" that relies on an understanding of accumulation. This work suggests that a simple video intervention is effective at helping people understand complex mathematical concepts and suggests that similar interventions could be used when trying to communicate statistical information with the public. Chapter 4 provides evidence that icon arrays are effective at improving risk perception even in highly politicized contexts. It also provides evidence that such arrays could be used to help the public understand very small probabilities more generally.

6.3 Conclusion

In sum, this work suggests that there is a need for better statistical and graph literacy skills among the public. It's possible that thousands of lives could have been saved if people properly understood the threat of COVID. Misunderstanding common data visualizations could have led to improper assessment of risk among the public. Educators should consider teaching these skills to students so that they may be better consumers of science as adults. This work also suggests that media outlets should consider the implications of using different types of visualization design on risk assessment. For example, at a time where COVID spread is high, using linear rather than logarithmic y-axes may have helped people better understand the magnitude of the virus's spread. The media should consider providing explanations of data visualizations when presenting data to the public, given individual differences in graph literacy among the public. Lastly, this work suggests that studies of data visualization and risk perception that are usually conducted in the context of hypothetical scenarios are relevant to everyday reasoning.

Bibliography

Ainsworth, S. (2008). The Educational Value of Multiple-representations when Learning Complex Scientific Concepts. In J. K. Gilbert, M. Reiner, & M. Nakhleh (Eds.), *Visualization: Theory and Practice in Science Education* (pp. 191–208). Springer Netherlands. https://doi.org/10.1007/978-1-4020-5267-5_9

Allcott, H., Boxell, L., Conway, J., Gentzkow, M., Thaler, M., & Yang, D. (2020). Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic. *Journal of Public Economics*, *191*, 104254. https://doi.org/10.1016/j.jpubeco.2020.104254

Amidon, T. R., Nielsen, A. C., Pflugfelder, E. H., Richards, D. P., & Stephens, S. H. (2021). Visual Risk Literacy in “Flatten the Curve” COVID-19 Visualizations. *Journal of Business and Technical Communication*, *35*(1), 101–109. https://doi.org/10.1177/1050651920963439

Ancker, J. S., Senathirajah, Y., Kukafka, R., & Starren, J. B. (2006). Design features of graphs in health risk communication: A systematic review. *Journal of the American Medical Informatics Association*, *13*(6), 608–618. https://doi.org/10.1197/jamia.M2115

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

Baddeley, A. (2010). Working memory. *Current Biology*, *20*(4), R136–R140. https://doi.org/10.1016/j.cub.2009.12.014

Banerjee, R., Bhattacharya, J., & Majumdar, P. (2021). Exponential-growth prediction bias and compliance with safety measures related to COVID-19. *Social Science & Medicine (1982)*, *268*, 113473. https://doi.org/10.1016/j.socscimed.2020.113473

Berney, S., & Bétrancourt, M. (2016). Does animation enhance learning? A meta-analysis. *Computers & Education*, *101*, 150–167. https://doi.org/10.1016/j.compedu.2016.06.005

Bond, L., & Nolan, T. (2011). Making sense of perceptions of risk of diseases and vaccinations: A qualitative study combining models of health beliefs, decision-making and risk perception. *BMC Public Health*, *11*(1), 943. https://doi.org/10.1186/1471-2458-11-943

Brewer, N. T., Chapman, G. B., Gibbons, F. X., Gerrard, M., McCaul, K. D., & Weinstein, N. D. (2007). Meta-analysis of the relationship between risk perception and health behavior: The example of vaccination. *Health Psychology*, *26*(2), 136–145. https://doi.org/10.1037/0278-6133.26.2.136

Brunstein, A., Gonzalez, C., & Kanter, S. (2010). Effects of domain experience in the stock-flow failure: Domain Experience and Stock-Flow Failure. *System Dynamics Review*, *26*(4), 347–354. https://doi.org/10.1002/sdr.448

Bump, P. (2021, April 13). The risk-reward calculus of the Johnson & Johnson vaccine, visualized. *The Washington Post*. https://www.washingtonpost.com/politics/2021/04/13/risk-reward-calculus-johnson-johnson-vaccine-visualized/

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

Butler, A. C., Godbole, N., & Marsh, E. J. (2013). Explanation feedback is better than correct answer feedback for promoting transfer of learning. *Journal of Educational Psychology*, *105*(2), 290–298. https://doi.org/10.1037/a0031026

Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 918–928. https://doi.org/10.1037/0278-7393.34.4.918

Calvillo, D. P., Ross, B. J., Garcia, R. J. B., Smelter, T. J., & Rutchick, A. M. (2020). Political ideology predicts perceptions of the threat of COVID-19 (and susceptibility to fake news about it). *Social Psychological and Personality Science*, *11*(8), 1119–1128. https://doi.org/10.1177/1948550620940539

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

Castro-Alonso, J. C., Ayres, P., & Sweller, J. (2019). Instructional Visualizations, Cognitive Load Theory, and Visuospatial Processing. In J. C. Castro-Alonso (Ed.), *Visuospatial Processing for Education in Health and Natural Sciences* (pp. 111–143). Springer International Publishing. https://doi.org/10.1007/978-3-030-20969-8_5

Centers for Disease Control and Prevention. (2021, April 13). *J&J/Janssen Update*. https://www.cdc.gov/coronavirus/2019-ncov/vaccines/safety/JJUpdate.html

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

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

Chen, S.-C., & She, H.-C. (2020). Effects of analogical learning approaches and presentation modalities on ninth graders' learning outcome and eye movements: A preliminary study. *Journal of Science Education and Technology*, *29*(4), 547–560. https://doi.org/10.1007/s10956-020-09835-7

Christensen, S. R., Pilling, E. B., Eyring, J. B., Dickerson, G., Sloan, C. D., & Magnusson, B. M. (2020). Political and personal reactions to COVID-19 during initial weeks of social distancing in the United States. *PLOS ONE*, *15*(9), e0239693. https://doi.org/10.1371/journal.pone.0239693

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

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, *7*(12), 547–552. https://doi.org/10.1016/j.tics.2003.10.005

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, *19*(1), 51–57. https://doi.org/10.1177/0963721409359277

Cronin, M. A., & Gonzalez, C. (2007). Understanding the building blocks of dynamic systems. *System Dynamics Review*, *23*(1), 1–17. https://doi.org/10.1002/sdr.356

Cronin, M. A., Gonzalez, C., & Sterman, J. D. (2009). Why don't well-educated adults understand accumulation? A challenge to researchers, educators, and citizens. *Organizational Behavior and Human Decision Processes*, *108*(1), 116–130. https://doi.org/10.1016/j.obhdp.2008.03.003

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

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

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

De La Maza, C., Davis, A., Gonzalez, C., & Azevedo, I. (2019). Understanding cumulative risk perception from judgments and choices: An application to flood risks. *Risk Analysis*, *39*(2), 488–504. https://doi.org/10.1111/risa.13206

DeSanctis, G., & Jarvenpaa, S. L. (1985). An investigation of the "tables versus graphs" controversy in a learning environment. *ICIS 1985 Proceedings.*

Doan, S. (2021). Misrepresenting COVID-19: Lying with charts during the second golden age of data design. *Journal of Business and Technical Communication*, *35*(1), 73–79. https://doi.org/10.1177/1050651920958392

Doyle, J. K. (1997). Judging cumulative risk. *Journal of Applied Social Psychology*, *27*(6), 500–524. https://doi.org/10.1111/j.1559-1816.1997.tb00644.x

Dragicevic, P., & Jansen, Y. (2017). Blinded with science or informed by charts? a replication study. *IEEE transactions on visualization and computer graphics*, *24*(1), 781-790.

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

Engel, C., Rodden, J., & Tabellini, M. (2022). Policies to influence perceptions about COVID-19 risk: The case of maps. *Science Advances*, *8*(11), eabm5106. https://doi.org/10.1126/sciadv.abm5106

Engledowl, C., & Weiland, T. (2021). Data (mis)representation and COVID-19: Leveraging misleading data visualizations for developing statistical literacy across grades 6–16. *Journal of Statistics and Data Science Education*, *29*(2), 160–164. https://doi.org/10.1080/26939169.2021.1915215

Fagerlin, A., Zikmund-Fisher, B. J., Ubel, P. A., Jankovic, A., Derry, H. A., & Smith, D. M. (2007). Measuring numeracy without a math test: Development of the subjective numeracy scale. *Medical Decision Making*, *27*(5), 672–680. https://doi.org/10.1177/0272989X07304449

Fagerlin, A., Zikmund-Fisher, B. J., & Ubel, P. A. (2011). Helping patients decide: Ten steps to better risk communication. *JNCI Journal of the National Cancer Institute*, *103*(19), 1436–1443. https://doi.org/10.1093/jnci/djr318

Fan, S., Han, L., Demartini, G., & Sadiq, S. (2022). Exploring data literacy levels in the crowd – the case of COVID-19. *Proceedings of the International AAAI Conference on Web and Social Media*, *16*, 1398–1403. https://doi.org/10.1609/icwsm.v16i1.19395

Fansher, M., Adkins, T. J., & Shah, P. (2022a). Graphs do not lead people to infer causation from correlation. *Journal of Experimental Psychology: Applied*, *28*(2), 314-328. https://doi.org/10.1037/xap0000393

Fansher, M., Adkins, T. J., Lewis, R. L., Boduroglu, A., Lalwani, P., Quirk, M., Shah, P., & Jonides, J. (2022b). How well do ordinary Americans forecast the growth of COVID-19? *Memory & Cognition*, *50*(7), 1363-1380. https://doi.org/10.3758/s13421-022-01288-0

Fansher, M., Adkins, T. J., Lalwani, P., Boduroglu, A., Carlson, M., Quirk, M., Lewis, R. L., Shah, P., Zhang, H., & Jonides, J. (2022c). Icon arrays reduce concern over COVID-19 vaccine side effects: A randomized control study. *Cognitive Research: Principles and Implications*, *7*(1), 38. https://doi.org/10.1186/s41235-022-00387-5

Fiorella, L., & Mayer, R. E. (2018). What works and doesn't work with instructional video. *Computers in Human Behavior*, *89*, 465–470. https://doi.org/10.1016/j.chb.2018.07.015

Franconeri, S. L., Padilla, L. M., Shah, P., Zacks, J. M., & Hullman, J. (2021). The science of visual data communication: What works. *Psychological Science in the Public Interest*, *22*(3), 110–161. https://doi.org/10.1177/15291006211051956

Funk, C., & Tyson, A. (2021, March 5). Growing Share of Americans Say They Plan To Get a COVID-19 Vaccine – or Already Have. *Pew Research Center*. https://www.pewresearch.org/science/2021/03/05/growing-share-of-americans-say-they-plan-to-get-a-covid-19-vaccine-or-already-have/

Fyfield, M., Henderson, M., & Phillips, M. (2019). 25 principles for effective instructional video design. *ASCILITE Publications*, 418-423.

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

Galesic, M., Garcia-Retamero, R., & Gigerenzer, G. (2009). Using icon arrays to communicate medical risks: overcoming low numeracy. *Health Psychology*, *28*(2), 210.

Galesic, M., & Garcia-Retamero, R. (2011). Graph literacy: A cross-cultural comparison. *Medical Decision Making*, *31*(3), 444–457. https://doi.org/10.1177/0272989X10373805

Garcia-Retamero, R., & Cokely, E. T. (2017). Designing visual aids that promote risk literacy: A systematic review of health research and evidence-based design heuristics. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *59*(4), 582–627. https://doi.org/10.1177/0018720817690634

Garcia-Retamero, R., Galesic, M., & Gigerenzer, G. (2010). Do icon arrays help reduce denominator neglect?. *Medical Decision Making*, *30*(6), 672-684.

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

Gentner, D., & Holyoak, K. J. (1997). Reasoning and learning by analogy: Introduction. *American Psychologist*, *52*(1), 32–34. https://doi.org/10.1037/0003-066X.52.1.32

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction*, *15*(4), 313–331. https://doi.org/10.1016/j.learninstruc.2005.07.001

Glazer, N. (2011). Challenges with graph interpretation: A review of the literature. *Studies in Science Education*, *47*(2), 183–210. https://doi.org/10.1080/03057267.2011.605307

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? *University of Chicago, Becker Friedman Institute for Economics Working Paper,* (2020-26).

Hammes, L. S., Rossi, A. P., Pedrotti, L. G., Pitrez, P. M., Mutlaq, M. P., & Rosa, R. G. (2021). Is the press properly presenting the epidemiological data on COVID-19? An analysis of newspapers from 25 countries. J*ournal of Public Health Policy*, *42*(3), 359–372. https://doi.org/10.1057/s41271-021-00298-7

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

Hawley, S. T., Zikmund-Fisher, B., Ubel, P., Jancovic, A., Lucas, T., & Fagerlin, A. (2008). The impact of the format of graphical presentation on health-related knowledge and treatment choices. P*atient Education and Counseling*, *73*(3), 448–455. https://doi.org/10.1016/j.pec.2008.07.023

Heckler, A. F., Mikula, B., & Rosenblatt, R. (2013). Student accuracy in reading logarithmic plots: The problem and how to fix it. *2013 IEEE Frontiers in Education Conference (FIE)*, 1066–1071. https://doi.org/10.1109/FIE.2013.6684990

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

Heyd-Metzuyanim, E., Sharon, A. J., & Baram-Tsabari, A. (2021). Mathematical media literacy in the COVID-19 pandemic and its relation to school mathematics education. *Educational Studies in Mathematics*, *108*(1–2), 201–225. https://doi.org/10.1007/s10649-021-10075-8

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, *331*(6018), 772–775. https://doi.org/10.1126/science.1199327

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

Korzilius, H., Raaijmakers, S., Rouwette, E., & Vennix, J. (2014). Thinking aloud while solving a stock-flow task: Surfacing the correlation heuristic and other reasoning patterns. *Systems Research and Behavioral Science*, *31*(2), 268-279.

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

Kwon, O. N., Han, C., Lee, C., Lee, K., Kim, K., Jo, G., & Yoon, G. (2021). Graphs in the COVID-19 news: A mathematics audit of newspapers in Korea. *Educational Studies in Mathematics*, *108*(1–2), 183–200. https://doi.org/10.1007/s10649-021-10029-0

Kühberger, A., Schulte-Mecklenbeck, M., & Perner, J. (2002). Framing decisions: Hypothetical and real. *Organizational Behavior and Human Decision Processes*, *89*(2), 1162-1175.

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

Landy, D., Silbert, N., & Goldin, A. (2013). Estimating large numbers. *Cognitive science*, *37*(5), 775-799.

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

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

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

Levy, M. R., & Tasoff, J. (2015). Exponential-growth bias and overconfidence. *Journal of Economic Psychology, 58,* 1-14.

Lipkus, I. M. (2007). Numeric, verbal, and visual formats of conveying health risks: Suggested best practices and future recommendations. *Medical Decision Making*, *27*(5), 696–713. https://doi.org/10.1177/0272989X07307271

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

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281. https://doi.org/10.1038/36846

Luo, J., Zhang, Y., & Song, Y. (2022). Design for pandemic information: Examining the effect of graphs on anxiety and social distancing intentions in the COVID-19. *Frontiers in Public Health*, *10*. https://www.frontiersin.org/articles/10.3389/fpubh.2022.800789

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

Mayer, R. E., & Moreno, R. (2002). Animation as an aid to multimedia learning. *Educational Psychology Review*, *14*(1), 87–99. https://doi.org/10.1023/A:1013184611077

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

Mehrabian, A. (1996). Relations among political attitudes, personality, and psychopathology assessed with new measures of libertarianism and conservatism. *Basic and Applied Social Psychology*, *18*(4), 469–491. https://doi.org/10.1207/s15324834basp1804_7

Menge, D. N. L., MacPherson, A. C., Bytnerowicz, T. A., Quebbeman, A. W., Schwartz, N. B., Taylor, B. N., & Wolf, A. A. (2018). Logarithmic scales in ecological data presentation may cause misinterpretation. *Nature Ecology & Evolution*, *2*(9), 1393–1402. https://doi.org/10.1038/s41559-018-0610-7

Okan, Y., Garcia-Retamero, R., Cokely, E. T., & Maldonado, A. (2012). Individual differences in graph literacy: Overcoming denominator neglect in risk comprehension. *Journal of Behavioral Decision Making*, *25*, 390–401.

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

Okan, Y., Janssen, E., Galesic, M., & Waters, E. A. (2019). Using the short graph literacy scale to predict precursors of health behavior change. *Medical Decision Making*, *39*(3), 183–195. https://doi.org/10.1177/0272989X19829728

Oudhoff, J. P., & Timmermans, D. R. M. (2015). The effect of different graphical and numerical likelihood formats on perception of likelihood and choice. *Medical Decision Making*, *35*(4), 487–500. https://doi.org/10.1177/0272989X15576487

Padilla, L., Hosseinpour, H., Fygenson, R., Howell, J., Chunara, R., & Bertini, E. (2022). Impact of COVID-19 forecast visualizations on pandemic risk perceptions. *Scientific Reports*, *12*, 2014. https://doi.org/10.1038/s41598-022-05353-1

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

Peters, E. (2012). Beyond Comprehension: The role of numeracy in judgments and decisions. *Current Directions in Psychological Science*, *21*(1), 31–35. https://doi.org/10.1177/0963721411429960

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].

Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences*, *20*(2), 110–122. https://doi.org/10.1016/j.lindif.2009.10.005

Recchia, G., Lawrence, A. C., & Freeman, A. L. (2022). Investigating the presentation of uncertainty in an icon array: A randomized trial. *PEC Innovation*, *1*, 100003.

Reyna, V. F. (2004). How people make decisions that involve risk: a dual-processes approach. *Current Directions in Psychological Science*, *13*(2), 60–66. https://doi.org/10.1111/j.0963-7214.2004.00275.x

Reyna, V. F. (2008). A theory of medical decision making and health: Fuzzy trace theory. *Medical Decision Making*, *28*(6), 850–865. https://doi.org/10.1177/0272989X08327066

Reyna, V. F., & Brainerd, C. J. (2008). Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. *Learning and Individual Differences*, *18*(1), 89–107. https://doi.org/10.1016/j.lindif.2007.03.011

Rodríguez, V., Andrade, A. D., García-Retamero, R., Anam, R., Rodríguez, R., Lisigurski, M., Sharit, J., & Ruiz, J. G. (2013). Health literacy, numeracy, and graphical literacy among veterans in primary care and their effect on shared decision making and trust in physicians. *Journal of Health Communication*, *18*, 273–289. https://doi.org/10.1080/10810730.2013.829137

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*(1), 20–27. https://doi.org/10.1016/j.tics.2010.09.003

Romano, A., Sotis, C., Dominioni, G., & Guidi, S. (2020). The scale of COVID-19 graphs affects understanding, attitudes, and policy preferences. *Health Economics*, *29*(11), 1482–1494. https://doi.org/10.1002/hec.4143

Roozenbeek, J., Schneider, C. R., Dryhurst, S., Kerr, J., Freeman, A. L. J., Recchia, G., van der Bles, A. M., & van der Linden, S. (2020). Susceptibility to misinformation about COVID-19 around the world. *Royal Society Open Science*, *7*(10), 201199. https://doi.org/10.1098/rsos.201199

Rose, H. & Ritchie, M. (2022, April). *Plastic pollution*. Our world in data. https://ourworldindata.org/plastic-pollution

Rosen, Y. (2009). The effects of an animation-based on-line learning environment on transfer of knowledge and on motivation for science and technology learning. *Journal of Educational Computing Research*, *40*(4), 451–467. https://doi.org/10.2190/EC.40.4.d

Ruiz, J. G., Andrade, A. D., Garcia-Retamero, R., Anam, R., Rodriguez, R., & Sharit, J. (2013). Communicating global cardiovascular risk: Are icon arrays better than numerical estimates in improving understanding, recall and perception of risk?. *Patient Education and Counseling*, *93*(3), 394-402.

Ryan, W. H., & Evers, E. R. K. (2020). Graphs with logarithmic axes distort lay judgments. *Behavioral Science & Policy*, *6*(2), 13–23. https://doi.org/10.1353/bsp.2020.0011

Sallam, M. (2021). COVID-19 vaccine hesitancy worldwide: A concise systematic review of vaccine acceptance rates. *Vaccines*, *9*(2), 160. https://doi.org/10.3390/vaccines9020160

Schlager, T., & Whillans, A. V. (2022). People underestimate the probability of contracting the coronavirus from friends. *Humanities and Social Sciences Communications*, *9*(1), 59. https://doi.org/10.1057/s41599-022-01052-4

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

Sevi, S., Aviña, M. M., Péloquin-Skulski, G., Heisbourg, E., Vegas, P., Coulombe, M., Arel-Bundock, V., Loewen, P. J., & Blais, A. (2020). Logarithmic versus Linear Visualizations of COVID-19 Cases Do Not Affect Citizens' Support for Confinement. *Canadian Journal of Political Science. Revue Canadienne De Science Politique*, 1–6. https://doi.org/10.1017/S000842392000030X

Shah, P., Freedman, E. G., & Vekiri, I. (2005). The Comprehension of Quantitative Information in Graphical Displays. In P. Shah & A. Miyake (Eds.), *The Cambridge Handbook of Visuospatial Thinking* (1st ed., pp. 426–476). Cambridge University Press. https://doi.org/10.1017/CBO9780511610448.012

Slovic, P., Monahan, J., & MacGregor, D. G. (2000). Violence risk assessment and risk communication: The effects of using actual cases, providing instruction, and employing probability versus frequency formats. *Law and Human Behavior*, *24*(3), 271–296. https://doi.org/10.1023/A:1005595519944

Slovic, P., & Weber, E. (2010). Perception of risk posed by extreme events. In *Regulation of Toxic Substances and Hazardous Waste (2nd edition). Ed. J. S. Applegate, J. G. Laitos, J. M. Gaba, and N. M. Sachs.* Foundation Press.

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1978). Accident probabilities and seat belt usage: A psychological perspective. *Accident Analysis & Prevention*, *10*(4), 281–285. https://doi.org/10.1016/0001-4575(78)90030-1

Smith, A. C., Ralph, B. C. W., MacLeod, C. M., & Smilek, D. (2019). Test feedback and learning: Student preferences and perceived influence. *Scholarship of Teaching and Learning in Psychology*, *5*(4), 255–264. https://doi.org/10.1037/stl0000140

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

Stevens H. (2020, March 14). *Why outbreaks like coronavirus spread exponentially, and how to "flatten the curve"*. The Washington Post. https://www.washingtonpost.com/graphics/2020/world/coronasimulator/

Swanson, H. L., & Alloway, T. P. (2012). Working memory, learning, and academic achievement. In K. R. Harris, S. Graham, T. Urdan, C. B. McCormick, G. M. Sinatra, & J. Sweller (Eds.), *APA educational psychology handbook, Vol 1: Theories, constructs, and critical issues.* (pp. 327–366). American Psychological Association. https://doi.org/10.1037/13273-012

Sweeney, L. B., & Sterman, J. D. (2000). Bathtub dynamics: Initial results of a systems thinking inventory. *System Dynamics Review*, *16*(4), 249–286. https://doi.org/10.1002/sdr.198

Tabak, I., & Dubovi, I. (2021). Changes in the Media Landscape in the Wake of COVID-19 as a Catalyst for Data Literacy Development thru Life Routines. *In Proceedings of the 15th International Conference of the Learning Sciences-ICLS 2021*. International Society of the Learning Sciences.

Tait, A. R., Voepel-Lewis, T., Zikmund-Fisher, B. J., & Fagerlin, A. (2010). The effect of format on parents' understanding of the risks and benefits of clinical research: A comparison between text, tables, and graphics. *Journal of Health Communication*, *15*(5), 487–501. https://doi.org/10.1080/10810730.2010.492560

Tal, A., & Wansink, B. (2016). Blinded with science: Trivial graphs and formulas increase ad persuasiveness and belief in product efficacy. *Public Understanding of Science*, *25*(1), 117-125.

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., Roozenbeek, J., & Compton, J. (2020). Inoculating Against Fake News About COVID-19. *Frontiers in Psychology*, *11*, 566790. https://doi.org/10.3389/fpsyg.2020.566790 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

Vosniadou, S., & Skopeliti, I. (2019). Evaluating the effects of analogy enriched text on the learning of science: The importance of learning indexes. *Journal of Research in Science Teaching*, *56*(6), 732–764. https://doi.org/10.1002/tea.21523

Wagenaar, W. 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

Walker, A. C., Stange, M., Dixon, M. J., Fugelsang, J. A., & Koehler, D. J. (2022). Using icon arrays to communicate gambling information reduces the appeal of scratch card games. *Journal of Gambling Studies*, 1-20.

Waters, E. A., Weinstein, N. D., Colditz, G. A., & Emmons, K. M. (2007a). Reducing aversion to side effects in preventive medical treatment decisions. *Journal of Experimental Psychology: Applied*, *13*(1), 11–21. https://doi.org/10.1037/1076-898X.13.1.11

Waters, E. A., Weinstein, N. D., Colditz, G. A., & Emmons, K. M. (2007b). Aversion to side effects in preventive medical treatment decisions. *British journal of health psychology*, *12*(3), 383-401.

Wise, T., Zbozinek, T. D., Michelini, G., Hagan, C. C., & Mobbs, D. (2020). Changes in risk perception and self-reported protective behaviour during the first week of the COVID-19 pandemic in the United States. *Royal Society Open Science*, *7*(9), 200742.

Witt, J., Hao, C., & Shah, P. (2022). The impact of visualizing the process of disease spread on social distancing intentions and attitudes. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *66*(1), 2026–2030. https://doi.org/10.1177/1071181322661172

Yang, B. W., Vargas Restrepo, C., Stanley, M. L., & Marsh, E. J. (2021). Truncating bar graphs persistently misleads viewers. *Journal of Applied Research in Memory and Cognition*, *10*(2), 298–311. https://doi.org/10.1016/j.jarmac.2020.10.002

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, May 12). Reading the pandemic data. Editor's Blog. https://blogs.sciencemag.org/editors-blog/2020/05/12/reading-the-pandemic-data/

Zhang, Y., Sun, Y., Padilla, L., Barua, S., Bertini, E., & Parker, A. G. (2021). Mapping the Landscape of COVID-19 Crisis Visualizations. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–23. https://doi.org/10.1145/3411764.3445381

Zikmund-Fisher, B. J., Ubel, P. A., Smith, D. M., Derry, H. A., McClure, J. B., Stark, A., ... & Fagerlin, A. (2008). Communicating side effect risks in a tamoxifen prophylaxis decision aid: the debiasing influence of pictographs. *Patient education and counseling*, *73*(2), 209-214.

Zipkin, D. A., Umscheid, C. A., Keating, N. L., Allen, E., Aung, K., Beyth, R., Kaatz, S., Mann, D. M., Sussman, J. B., Korenstein, D., Schardt, C., Nagi, A., Sloane, R., & Feldstein, D. A. (2014). Evidence-based risk communication: A systematic review. *Annals of Internal Medicine*, *161*(4), 270. https://doi.org/10.7326/M14-0295

Appendices

Appendix A News Articles

Figure 29. *Study 1- Table Group*

Scientific Citizen

Home Local World Latest Subscribe About Us

An Update on COVID-19 in the United States

Author: **Richard Owens**
Published March 27, 2020

The number of confirmed cases of the novel coronavirus has increased dramatically in the United States during the past two weeks. Many state governments and private businesses are taking precautionary measures to slow the spread of the disease. Some of these measures include cancelling large events such as music festivals, sporting events, and parades, as well as closing public K-12 schools and Universities.

The first case of coronavirus was reported on December 31, 2019 in Wuhan, China, since then there has been rapid growth in the number of cases, with nearly every country being affected by the virus. The table below shows the growth of cases in the U.S.

United States COVID-19 Data		
Date	Confirmed Cases	Deaths
Feb 29	68	1
Mar 6	319	15
Mar 13	2183	48
Mar 20	19,367	255
Mar 27	96,968	1,477

The first case of COVID-19 in the United States was reported on January 21, 2020. Since the first confirmed case in the United States, the number of cases has changed dramatically, especially in densely populated areas such as New York, New Jersey, California, and Washington. Many states are doing their part to "flatten the curve" by enforcing curfews on residents and closing sit-in restaurants and bars.

Government officials hope that encouraging citizens to take preventative measures, such as frequent hand washing and limiting close contact, will reduce the spread of the disease.

Figure 30. *Study 1 - Graph Group*

Scientific Citizen

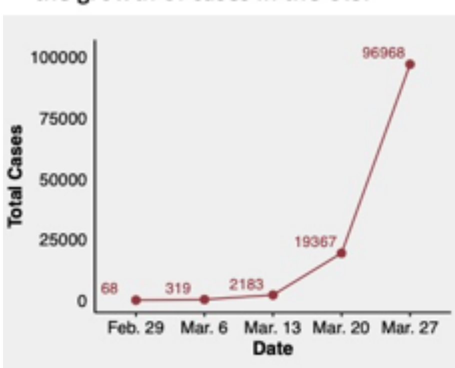
Home | Local | World | Latest | Subscribe | About Us

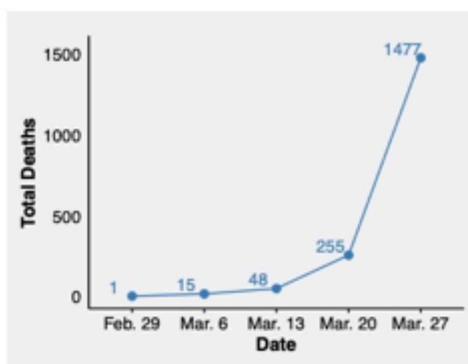
An Update on COVID-19 in the United States

Author: **Richard Owens**
Published March 27, 2020

The number of confirmed cases of the novel coronavirus has increased dramatically in the United States during the past two weeks. Many state governments and private businesses are taking precautionary measures to slow the spread of the disease. Some of these measures include cancelling large events such as music festivals, sporting events, and parades, as well as closing public K-12 schools and Universities.

The first case of coronavirus was reported on December 31, 2019 in Wuhan, China, since then there has been rapid growth in the number of cases, with nearly every country being affected by the virus. The graphs below show the growth of cases in the U.S.





The first case of COVID-19 in the United States was reported on January 21, 2020. Since the first confirmed case in the United States, the number of cases has changed dramatically, especially in densely populated areas such as New York, New Jersey, California, and Washington. Many states are doing their part to "flatten the curve" by enforcing curfews on residents and closing sit-in restaurants and bars.

Government officials hope that encouraging citizens to take preventative measures, such as frequent hand washing and limiting close contact, will reduce the spread of the disease.

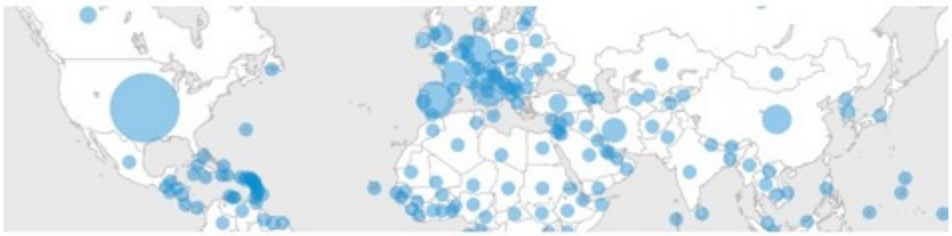
Figure 31. *Study 2 - Table Group*

Scientific Citizen

Home Local World Latest Subscribe About Us

An Update on COVID-19 in the United States

Author: **Richard Owens**
Published April 7, 2020



The number of confirmed cases of the novel coronavirus has increased dramatically in the United States during the past two weeks. Many state governments and private businesses are taking precautionary measures to slow the spread of the disease. Some of these measures include cancelling large events such as music festivals, sporting events, and parades, as well as closing public K-12 schools and Universities.

The first case of coronavirus was reported on December 31, 2019 in Wuhan, China, since then there has been rapid growth in the number of cases, with nearly every country being affected by the virus. The table below shows the growth of cases in the U.S.

United States COVID-19 Data		
Date	Confirmed Cases	Deaths
Mar 11	1,301	38
Mar 18	9,197	150
Mar 25	68,211	1,027
Apr 1	215,003	5,102
Apr 7	393,782	12,692

The first case of COVID-19 in the United States was reported on January 21, 2020. Since the first confirmed case in the United States, the number of cases has changed dramatically, especially in densely populated areas such as New York, New Jersey, California, and Washington. Many states are doing their part to “flatten the curve” by enforcing curfews on residents and closing sit-in restaurants and bars.

Government officials hope that encouraging citizens to take preventative measures, such as frequent hand washing and limiting close contact, will reduce the spread of the disease.

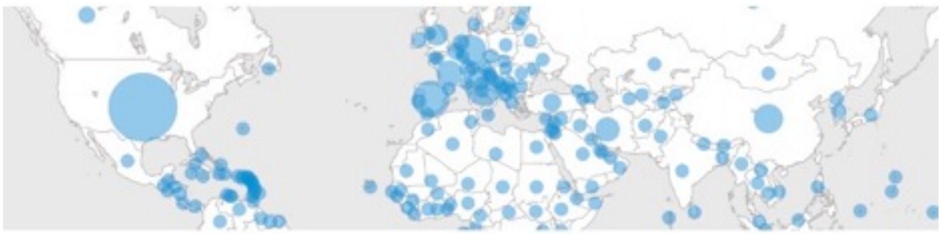
Figure 32. *Study 2 - Graph Group*

Scientific Citizen

Home Local World Latest Subscribe About Us

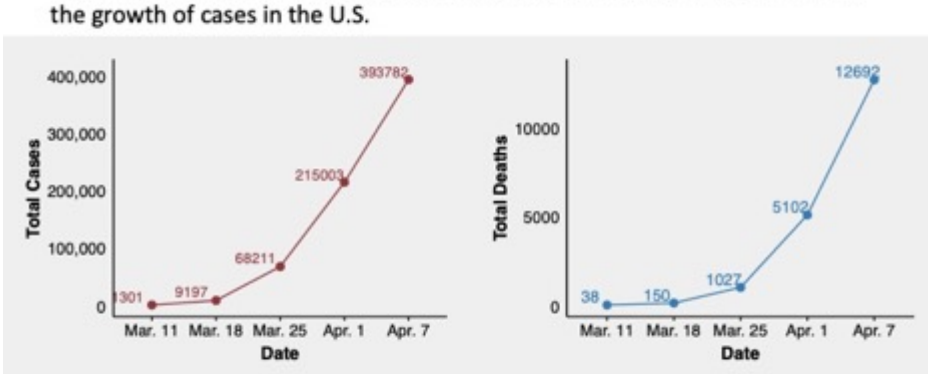
An Update on COVID-19 in the United States

Author: **Richard Owens**
Published April 7, 2020



The number of confirmed cases of the novel coronavirus has increased dramatically in the United States during the past two weeks. Many state governments and private businesses are taking precautionary measures to slow the spread of the disease. Some of these measures include cancelling large events such as music festivals, sporting events, and parades, as well as closing public K-12 schools and Universities.

The first case of coronavirus was reported on December 31, 2019 in Wuhan, China, since then there has been rapid growth in the number of cases, with nearly every country being affected by the virus. The graphs below show the growth of cases in the U.S.



The first case of COVID-19 in the United States was reported on January 21, 2020. Since the first confirmed case in the United States, the number of cases has changed dramatically, especially in densely populated areas such as New York, New Jersey, California, and Washington. Many states are doing their part to "flatten the curve" by enforcing curfews on residents and closing sit-in restaurants and bars.

Government officials hope that encouraging citizens to take preventative measures, such as frequent hand washing and limiting close contact, will reduce the spread of the disease.

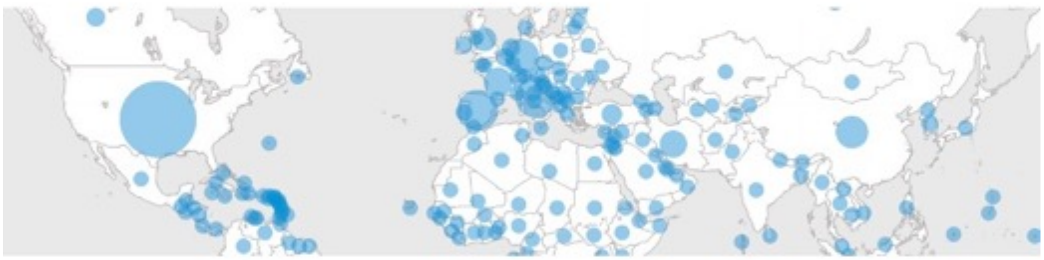
Figure 33. *Study 3.1 - Table Group*

Scientific Citizen

Home | Local | World | Latest | Subscribe | About Us

An Update on COVID-19

Author: **Richard Owens**
Published March 27, 2020



The number of confirmed cases of the novel coronavirus has increased dramatically during the past two weeks. Many governments and private businesses are taking precautionary measures to slow the spread of the disease. Some of these measures include cancelling large events such as music festivals, sporting events, and parades, as well as closing public K-12 schools and Universities.

The first case of coronavirus was reported on December 31, 2019 in Wuhan, China, since then there has been rapid growth in the number of cases, with nearly every country being affected by the virus. The table below shows the growth of cases in one of the countries affected. Government officials hope that encouraging citizens to take preventative measures, such as frequent hand washing and limiting close contact, will reduce the spread of the disease.

COVID-19 Data		
Date	Confirmed Cases	Deaths
Feb 29	68	1
Mar 6	319	15
Mar 13	2183	48
Mar 20	19,367	255
Mar 27	96,968	1,477

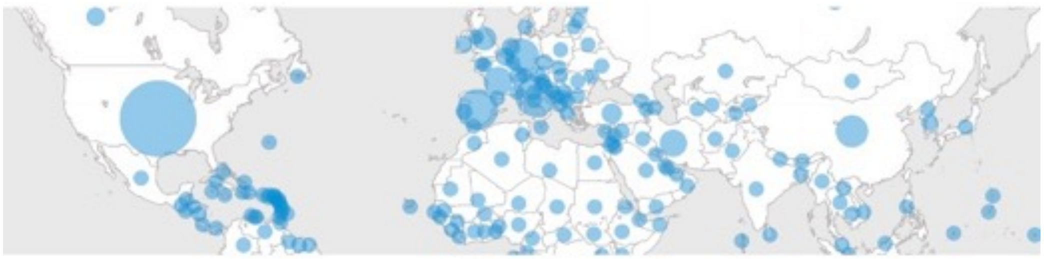
Figure 34. *Study 3.1 - Graph Group*

Scientific Citizen

Home | Local | World | Latest | Subscribe | About Us

An Update on COVID-19

Author: **Richard Owens**
Published March 27, 2020



The number of confirmed cases of the novel coronavirus has increased dramatically during the past two weeks. Many governments and private businesses are taking precautionary measures to slow the spread of the disease. Some of these measures include cancelling large events such as music festivals, sporting events, and parades, as well as closing public K-12 schools and Universities.

The first case of coronavirus was reported on December 31, 2019 in Wuhan, China, since then there has been rapid growth in the number of cases, with nearly every country being affected by the virus. The graphs below show the growth of cases in one of the countries affected. Government officials hope that encouraging citizens to take preventative measures, such as frequent hand washing and limiting close contact, will reduce the spread of the disease.

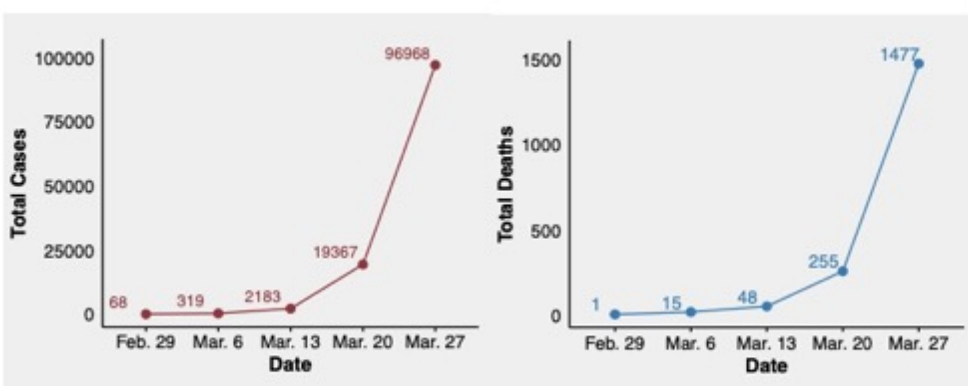


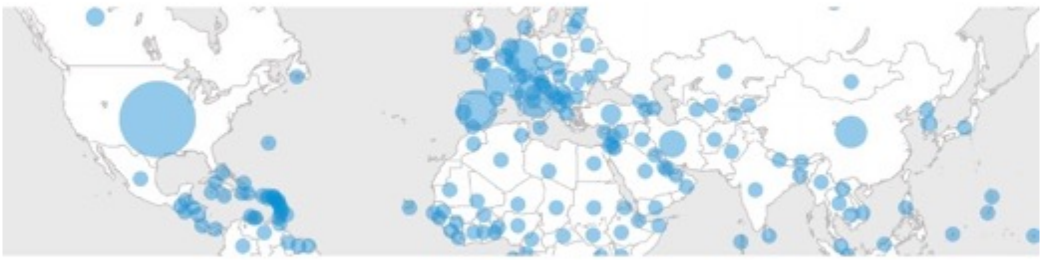
Figure 35. *Study 3.2 - Table Group*

Scientific Citizen

Home | Local | World | Latest | Subscribe | About Us

An Update on COVID-19

Author: **Richard Owens**
Published April 7, 2020



The number of confirmed cases of the novel coronavirus has increased dramatically during the past two weeks. Many governments and private businesses are taking precautionary measures to slow the spread of the disease. Some of these measures include cancelling large events such as music festivals, sporting events, and parades, as well as closing public K-12 schools and Universities.

The first case of coronavirus was reported on December 31, 2019 in Wuhan, China, since then there has been rapid growth in the number of cases, with nearly every country being affected by the virus. The table below shows the growth of cases in one of the countries affected. Government officials hope that encouraging citizens to take preventative measures, such as frequent hand washing and limiting close contact, will reduce the spread of the disease.

COVID-19 Data		
Date	**Confirmed Cases**	**Deaths**
Mar 11	1,301	38
Mar 18	9,197	150
Mar 25	68,211	1,027
Apr 1	215,003	5,102
Apr 7	393,782	12,692

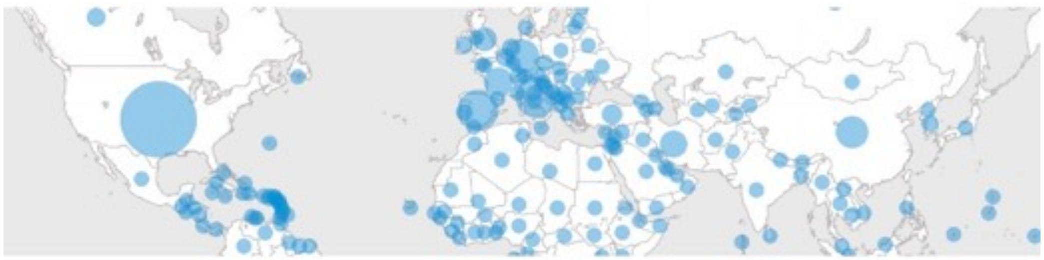
Figure 36. *Study 3.2 - Graph Group*

Scientific Citizen

Home | Local | World | Latest | Subscribe | About Us

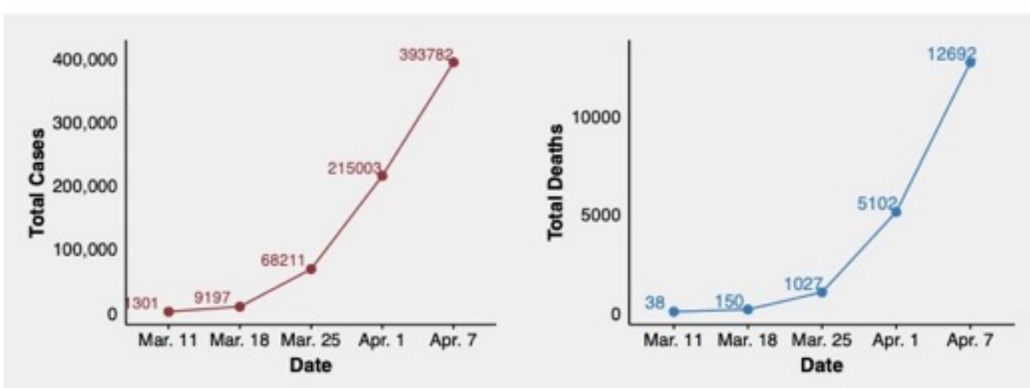
An Update on COVID-19

Author: **Richard Owens**
Published April 7, 2020



The number of confirmed cases of the novel coronavirus has increased dramatically during the past two weeks. Many governments and private businesses are taking precautionary measures to slow the spread of the disease. Some of these measures include cancelling large events such as music festivals, sporting events, and parades, as well as closing public K-12 schools and Universities.

The first case of coronavirus was reported on December 31, 2019 in Wuhan, China, since then there has been rapid growth in the number of cases, with nearly every country being affected by the virus. The graphs below show the growth of cases in one of the countries affected. Government officials hope that encouraging citizens to take preventative measures, such as frequent hand washing and limiting close contact, will reduce the spread of the disease.



Appendix B Individual Differences Measures

Individual Difference Measures

Because of the unique opportunity to collect data at the beginning of the pandemic, we collected other individual difference and demographic measures that are not analyzed in the current work:

1. Estimate the probability that you will have contracted the illness within 9 days (0-100%)
2. How anxious are you about the current virus situation? (Slider scale from 0 (Not at all) to 100 (Extremely Anxious))
3. Estimate the maximum number of new cases per day that will be reported in the US. (Free response)
4. Estimate the probability that you will have been hospitalized because of the illness within 9 days (0-100%)
5. Estimate the probability that you will have died from the illness within 9 days (0-100%)
6. How closely have you been following the news on the coronavirus? (Slider scale from 0 (Not at all) to 100 (Very closely))
7. Age
8. Gender
9. Zip code
10. Highest level of education
11. Their mother's highest level of education
12. Native language
13. Risk aversion (Mandrik & Bao, 2005)
14. Conservatism (Mehrabian, 1996)
15. Numeracy (Fagerlin et al., 2007)
16. Health status (Eriksson et al., 2001).

Appendix C Impact of Exclusion Criteria

Impact of exclusion criteria

In the main text, we report the results obtained from filtered data. The data were filtered according to several exclusion criteria. Some exclusion criteria were immutable, such as age > 18, valid zip code, and the 'no impossible forecasts' rule. However, it may be argued that the other exclusion criteria were not strictly necessary. For example, perhaps our results should generalize to people who were not paying attention. So, we examined the robustness of our results to variation in the following optional exclusion criteria:

failed the basic attention check trial ("Please select option 6")

failed to identify the President using a string containing 'don' or 'trump'.

reported investing effort of less than 5 out of 10

took less than 30 seconds to complete the task

forecast greater than 10x the last datum provided

We initially examined the impact of criteria 4 – 8. We considered all combinations of these criteria and reanalyzed the forecasting and confidence data under each combination. The results of this multiverse analysis demonstrated that the key results regarding graph vs. table reported in the main text did not depend on these exclusion criteria (Table S1). Our second analysis looked at the sensitivity of our results to different thresholds in criterion 8. We considered 10 different thresholds (10x – 20x, by 1) and reanalyzed the forecasting and confidence data under each combination. The results of this multiverse analysis demonstrated that the key results regarding graph vs. table reported in the main text did not depend much on the threshold in this exclusion criterion (Table S2).

Next, we turn to the impact of our exclusion criteria on the sample size of the data analyzed in Studies 1 and 2 (Table S3). The number of subjects remaining after applying our mandatory criteria (age, zip, no decreases) was 1140 participants. The sample size decreased by 7 participants after applying the attention criterion, by 8 after applying the president criterion, by 9 after applying the effort criterion, by 0 after applying the RT criterion, and by 5 after applying

the extreme outlier criterion. Since the extreme outlier exclusion was applied pointwise, we also note that the number of trials decreased by 69 from a prior sample size of 3308 trials.
The exclusion criteria with the greatest impact on sample size, given the order with which we applied the criteria, was the “No decreasing forecasts” rule. Figure S1B above shows the forecasts of total cases of the individuals who violated this rule. The results show that most of these participants forecasted extremely small numbers (often in single digits). This further suggests that these participants did not understand the forecasting task.

Appendix D Multiverse Analysis

Table 7. *Multiverse Analysis 1 Results*

Here we report the posterior mean and standard deviation of the Table-graph effect on forecasts (F) and confidence (C) in Study 1 (S1) and Study 2 (S2) under all combinations of exclusion criteria 4 - 8 above. These results show that our key results in the main text were not affected by our choices among these exclusion criteria.

Exclusions	FS1	FS2	CS1	CS2
4,5,6,7	0.05 (0.02)	-0.06 (0.01)	-0.16 (0.05)	-0.11 (0.05)
4,5,7	0.04 (0.02)	-0.07 (0.01)	-0.16 (0.05)	-0.1 (0.05)
5,6,7	0.05 (0.02)	-0.06 (0.01)	-0.16 (0.05)	-0.11 (0.05)
5,7	0.04 (0.02)	-0.07 (0.02)	-0.15 (0.05)	-0.11 (0.05)
4,6,7	0.05 (0.02)	-0.06 (0.02)	-0.14 (0.05)	-0.1 (0.05)
4,7	0.04 (0.02)	-0.07 (0.01)	-0.14 (0.05)	-0.1 (0.05)
6,7	0.05 (0.02)	-0.06 (0.01)	-0.14 (0.05)	-0.1 (0.05)
7	0.04 (0.02)	-0.07 (0.01)	-0.14 (0.05)	-0.1 (0.05)
4,5,6	0.05 (0.02)	-0.06 (0.01)	-0.16 (0.05)	-0.11 (0.05)
4,5	0.04 (0.02)	-0.07 (0.01)	-0.16 (0.05)	-0.1 (0.05)
5,6	0.05 (0.02)	-0.06 (0.01)	-0.16 (0.05)	-0.11 (0.05)
5	0.04 (0.02)	-0.07 (0.01)	-0.15 (0.05)	-0.11 (0.05)
4,6	0.05 (0.02)	-0.06 (0.01)	-0.15 (0.05)	-0.1 (0.05)
4	0.04 (0.02)	-0.07 (0.01)	-0.14 (0.05)	-0.1 (0.05)
6	0.05 (0.02)	-0.06 (0.01)	-0.14 (0.05)	-0.11 (0.05)

Table 8. *Multiverse Analysis 3 Results*

Here we report the posterior mean and standard deviation of the Table-graph effect on forecasts (F) and confidence (C) in Study 1 (S1) and Study 2 (S2) for 10 different extreme outlier thresholds. In the main text, the outlier threshold was 10 times the last data point provided. Below we examine more liberal thresholds up to 20 times the last datum. These results show that our key results in the main text did not depend much on the extreme outlier exclusion criterion.

Threshold	FS1	FS2	CS1	CS2
10x	0.05 (0.02)	-0.06 (0.02)	-0.16 (0.05)	-0.11 (0.05)

11x	0.05 (0.03)	-0.07 (0.02)	-0.16 (0.05)	-0.11 (0.05)
12x	0.05 (0.03)	-0.07 (0.02)	-0.16 (0.05)	-0.1 (0.05)
13x	0.06 (0.03)	-0.07 (0.02)	-0.16 (0.05)	-0.1 (0.05)
14x	0.06 (0.03)	-0.07 (0.02)	-0.16 (0.05)	-0.1 (0.05)
15x	0.06 (0.03)	-0.07 (0.02)	-0.16 (0.05)	-0.1 (0.05)
16x	0.07 (0.03)	-0.07 (0.02)	-0.15 (0.05)	-0.1 (0.05)
17x	0.07 (0.03)	-0.07 (0.02)	-0.15 (0.05)	-0.1 (0.05)
18x	0.07 (0.03)	-0.07 (0.02)	-0.15 (0.05)	-0.1 (0.05)
19x	0.07 (0.03)	-0.08 (0.02)	-0.15 (0.05)	-0.1 (0.05)
20x	0.07 (0.03)	-0.08 (0.02)	-0.15 (0.05)	-0.1 (0.05)

Table 9. *Impact of Exclusion Criteria on Sample Size*

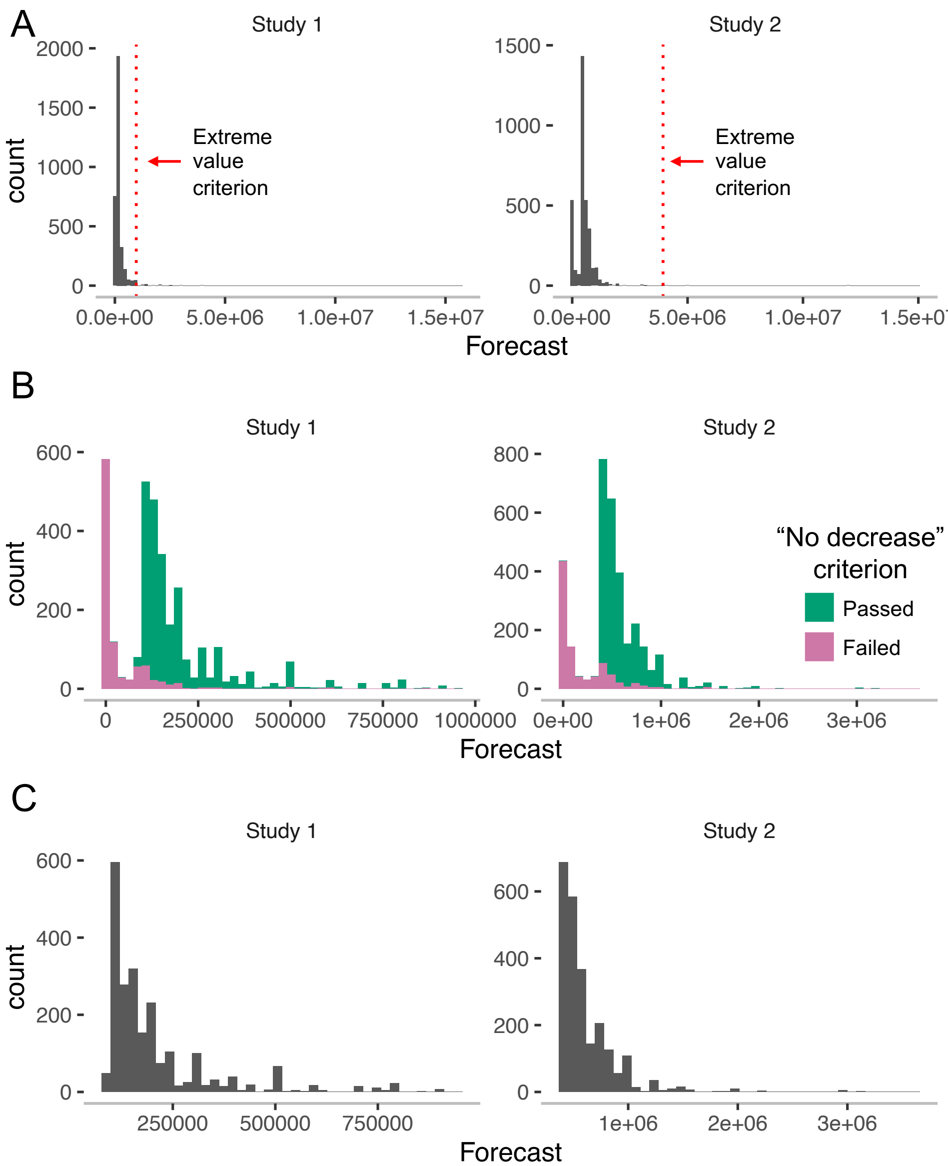
Here we report the number of participants remaining after applying the following exclusion criteria: (1) not over the age of 18, (2) did not provide a valid zip code, (3) reported a decreasing forecast, (4) failed the basic attention check, (5) failed the free response attention check to name the president of the U.S., (6) self-reported investing effort of less than 5 out of 10, and (7) took less than 30 seconds to complete the task. We also report the number of trials (and participants) remaining after applying our extreme outlier exclusion criterion (8).

Study	Participants Remaining								Trials Remaining	
	Initial	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*Initial*	*8*
1	1198	1195	1129	792	788	777	774	774	2322	2243 (770)
2	1180	1177	1139	816	813	810	803	803	2409	2397 (802)
3	803	797	760	471	470	468	468	468	1404	1355 (459)

Appendix E Data Distributions

Prior to applying our exclusion criteria, the distribution was unwieldy, with an extremely long right tail—even longer than shown here, as the plots were truncated at 1.5e7 for visualization purposes. To address this issue, we excluded forecasts that exceeded 10x the last datum provided; these thresholds are depicted with red dotted lines (about 1mil for Study 1, 4mil for Study 2). B. After removing extreme outliers, we have a much better view of the data, but we notice that the responses appear to be drawn from two separate distributions corresponding to participants who forecasted a decrease at some point and participants who did not. We therefore exclude participants who reported such impossible forecasts. C. After applying our exclusion criteria, we are left with distributions that are well approximated by the gamma distributional family. These are the forecasting data that we analyzed and reported in the main text.

Table 10. *Distributions of Total Case Forecasts from Study 1 and Study 2*


A
Study 1
Study 2
count
Extreme value criterion
Extreme value criterion
Forecast
B
Study 1
Study 2
count
"No decrease" criterion
Passed
Failed
Forecast
C
Study 1
Study 2
count
Forecast

Appendix F Other Forecasts

Other forecasts

We also collected participants' forecasts about deaths and 'actual' cases (as opposed to officially 'confirmed'). While we omit these data from the main text for brevity, we paste the results below. However, these results are largely consistent with the confirmed cases results, with the table group forecasting closer-to-truth, both groups showing gross underestimation of highly exponential curves (i.e. exponential growth bias), and graphs consistently leading to higher confidence than tables.

Table 11. *Number of Deaths – Forecasts*

Description of Posterior Distributions

Parameter	Median	Mean	MAP	95% CI	pd	Rhat	ESS
Intercept	1.646	1.645	1.649	[1.545, 1.743]	1.000	1.001	9379.280
day2M1	-0.229	-0.229	-0.226	[-0.351, -0.111]	1.000	1.000	26052.728
day3M2	-0.152	-0.151	-0.155	[-0.269, -0.026]	0.992	1.000	24821.876
grouptaMg	-0.173	-0.173	-0.172	[-0.274, -0.073]	1.000	1.000	34464.246
day2M1.grouptaMg	0.005	0.006	0.004	[-0.235, 0.240]	0.519	1.000	27405.993
day3M2.grouptaMg	-0.009	-0.010	-0.029	[-0.251, 0.233]	0.529	1.000	28243.358

Description of Posterior Distributions

Parameter	Median	Mean	MAP	95% CI	pd	Rhat	ESS
Intercept	1.633	1.632	1.633	[1.531, 1.734]	1.000	1.000	9351.938
day2M1	-0.218	-0.218	-0.218	[-0.337, -0.102]	1.000	1.000	24920.271
day3M2	-0.181	-0.181	-0.179	[-0.302, -0.068]	0.998	1.000	23387.764
grouptaMg	-0.170	-0.171	-0.164	[-0.265, -0.071]	1.000	1.000	30504.594
day2M1.grouptaMg	-0.017	-0.017	-0.027	[-0.261, 0.206]	0.557	1.000	25326.560
day3M2.grouptaMg	0.038	0.038	0.038	[-0.194, 0.274]	0.624	1.000	25079.943

Figure 37. *Number of Deaths – Forecasts*

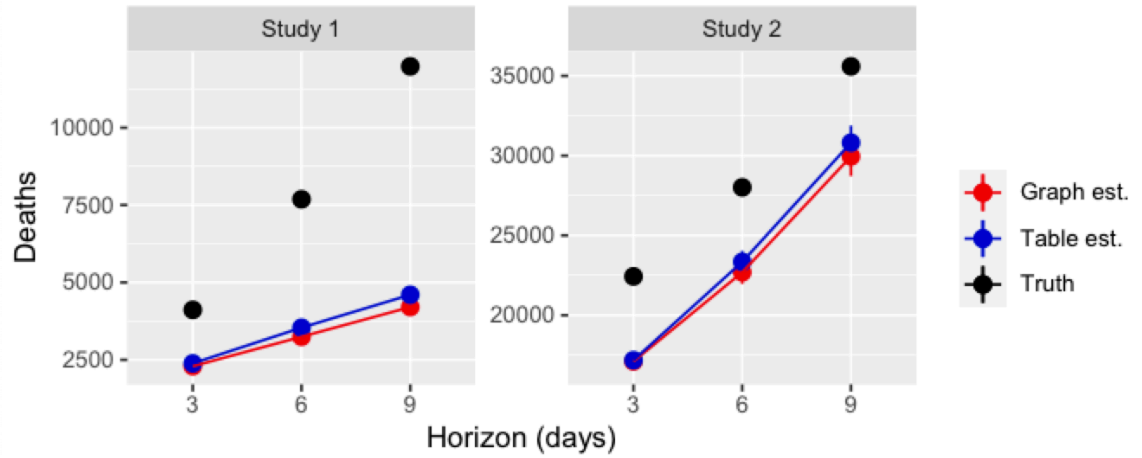


__Table 12.__ Number of Deaths - Confidence

Description of Posterior Distributions

Parameter	Median	Mean	MAP	95% CI	pd	Rhat	ESS
Intercept	1.646	1.645	1.649	[1.545, 1.743]	1.000	1.001	9379.280
day2M1	-0.229	-0.229	-0.226	[-0.351, -0.111]	1.000	1.000	26052.728
day3M2	-0.152	-0.151	-0.155	[-0.269, -0.026]	0.992	1.000	24821.876
grouptaMg	-0.173	-0.173	-0.172	[-0.274, -0.073]	1.000	1.000	34464.246
day2M1.grouptaMg	0.005	0.006	0.004	[-0.235, 0.240]	0.519	1.000	27405.993
day3M2.grouptaMg	-0.009	-0.010	-0.029	[-0.251, 0.233]	0.529	1.000	28243.358

Description of Posterior Distributions

Parameter	Median	Mean	MAP	95% CI	pd	Rhat	ESS
Intercept	1.633	1.632	1.633	[1.531, 1.734]	1.000	1.000	9351.938
day2M1	-0.218	-0.218	-0.218	[-0.337, -0.102]	1.000	1.000	24920.271
day3M2	-0.181	-0.181	-0.179	[-0.302, -0.068]	0.998	1.000	23387.764
grouptaMg	-0.170	-0.171	-0.164	[-0.265, -0.071]	1.000	1.000	30504.594
day2M1.grouptaMg	-0.017	-0.017	-0.027	[-0.261, 0.206]	0.557	1.000	25326.560
day3M2.grouptaMg	0.038	0.038	0.038	[-0.194, 0.274]	0.624	1.000	25079.943

__Figure 38.__ Number of Deaths – Confidence

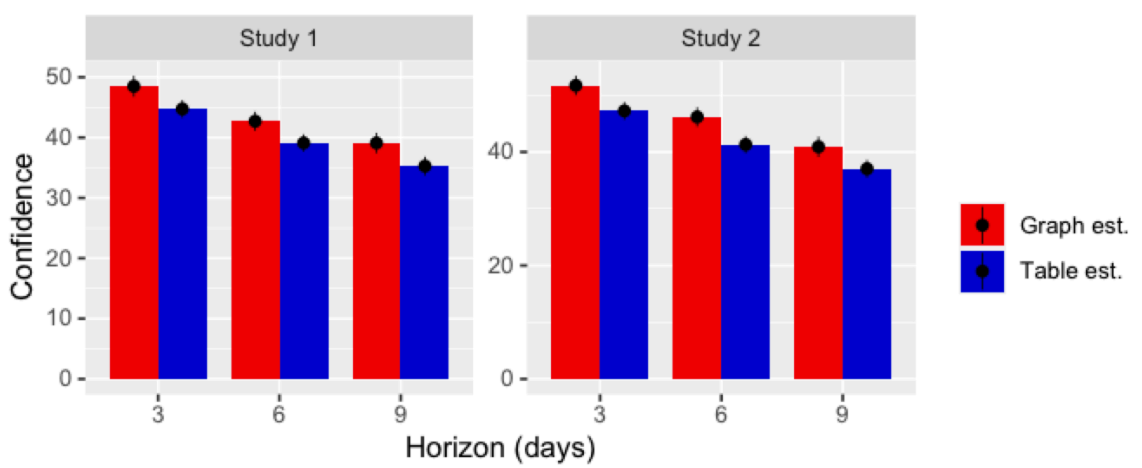


__Table 13.__ Actual Cases – Forecasts

Description of Posterior Distributions

Parameter	Median	Mean	MAP	95% CI	pd	Rhat	ESS
Intercept	0.363	0.363	0.368	[0.303, 0.417]	1.000	1.000	9484.888
day2M1	0.337	0.337	0.333	[0.267, 0.409]	1.000	1.000	21607.781
day3M2	0.141	0.142	0.145	[0.067, 0.217]	1.000	1.000	22207.485
grouptaMg	0.107	0.107	0.105	[0.046, 0.168]	1.000	1.000	25155.383
day2M1.grouptaMg	0.031	0.031	0.025	[-0.112, 0.174]	0.665	1.000	20053.447
day3M2.grouptaMg	-3.994e-04	-9.736e-04	-1.074e-04	[-0.149, 0.147]	0.503	1.000	19928.641

Description of Posterior Distributions

Parameter	Median	Mean	MAP	95% CI	pd	Rhat	ESS
Intercept	0.474	0.474	0.474	[0.409, 0.541]	1.000	1.001	5608.452
day2M1	0.233	0.233	0.232	[0.175, 0.291]	1.000	1.000	25996.017
day3M2	0.168	0.168	0.170	[0.110, 0.226]	1.000	1.000	27882.098
grouptaMg	-0.061	-0.061	-0.060	[-0.111, -0.010]	0.991	1.000	35265.783
day2M1.grouptaMg	-0.015	-0.015	-0.014	[-0.135, 0.101]	0.601	1.000	26569.955
day3M2.grouptaMg	0.025	0.025	0.021	[-0.092, 0.143]	0.662	1.000	27882.862

Figure 39. *Actual Cases - Forecasts*

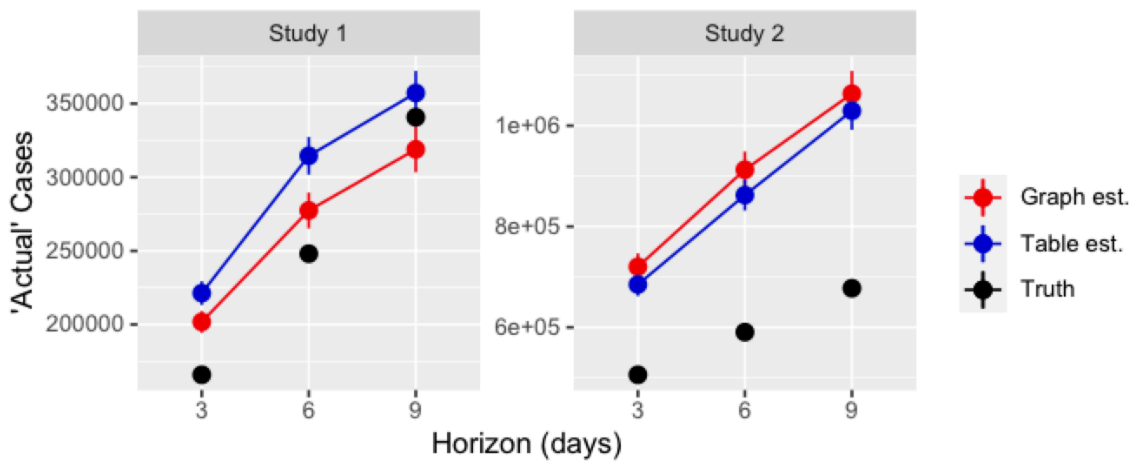


Table 14. *Actual Cases – Confidence*

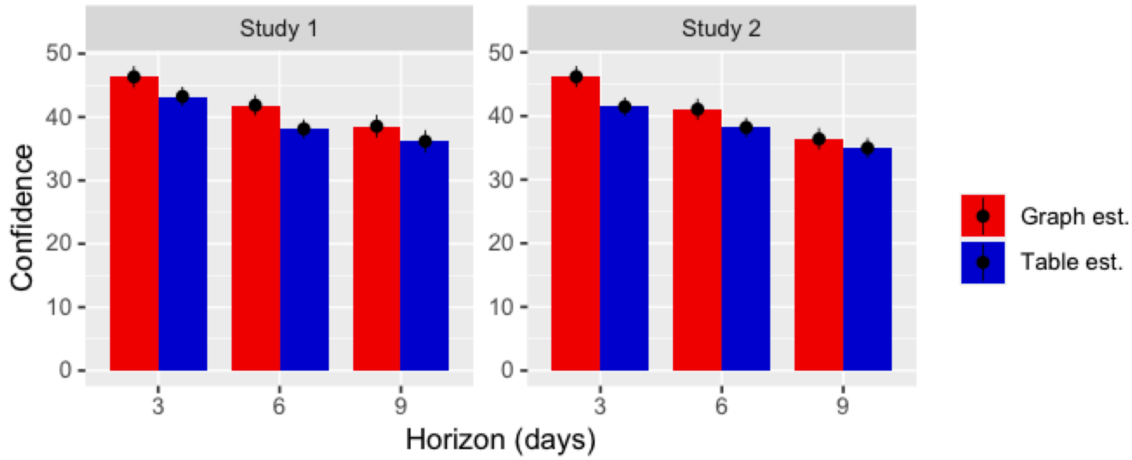
Description of Posterior Distributions

Parameter	Median	Mean	MAP	95% CI	pd	Rhat	ESS
Intercept	1.598	1.597	1.595	[1.482, 1.715]	1.000	1.000	7582.287
day2M1	-0.191	-0.190	-0.199	[-0.315, -0.073]	0.999	1.000	26773.737
day3M2	-0.103	-0.103	-0.104	[-0.223, 0.027]	0.949	1.000	27638.408
grouptaMg	-0.140	-0.141	-0.137	[-0.245, -0.034]	0.996	1.000	31691.123
day2M1.grouptaMg	-0.020	-0.020	-0.017	[-0.272, 0.217]	0.566	1.000	23927.810
day3M2.grouptaMg	0.051	0.051	0.040	[-0.199, 0.305]	0.659	1.000	23310.852

Description of Posterior Distributions

Parameter	Median	Mean	MAP	95% CI	pd	Rhat	ESS
Intercept	1.499	1.498	1.497	[1.398, 1.594]	1.000	1.000	11315.646
day2M1	-0.165	-0.165	-0.169	[-0.285, -0.048]	0.997	1.000	27810.867
day3M2	-0.155	-0.156	-0.147	[-0.272, -0.031]	0.995	1.000	27234.838
grouptaMg	-0.119	-0.119	-0.122	[-0.220, -0.018]	0.988	1.000	30897.031
day2M1.grouptaMg	0.076	0.076	0.077	[-0.174, 0.308]	0.731	1.000	27133.212
day3M2.grouptaMg	0.057	0.057	0.064	[-0.193, 0.293]	0.678	1.000	25158.869

Figure 40. *Actual Cases - Confidence*


Study 1
Study 2
Confidence
50
40
30
20
10
0
3
6
9
Horizon (days)
Graph est.
Table est.

Appendix G Forecasting Error

An alternative measure of forecasting error

In our studies, we present five data points and ask participants to make forecasts for future dates. In the main text, we compare participants' forecasts to the *true number of cases* for the future days, which revealed substantial misestimation. But what if we compared participants' forecasts to the extrapolated *trend* of the initial five data points, rather than the true data (which may diverge from suggested trend)? Are participants predictions still inaccurate with respect to the extrapolated trends? We address this question by fitting exponential models ($y \sim a * e^{bx}$) and using the fitted models to predict the number of cases at future dates. Next, we convert participants' forecasts to errors, using the following formula:

$$\epsilon = \frac{|y - y_{\text{pred}}|}{y_{\text{pred}}} * 100,$$

where y is a participant's forecast for a future day and y_{pred} is the predicted number of cases for that day from a fitted exponential model.

We calculated average forecasting error for both Study 1 and Study 2, and for each day (3, 6, and 9-day forecasts). We find that participants still dramatically misestimated the prevalence of COVID-19 even when their forecasts are compared to the predictions of fitted exponential trends.

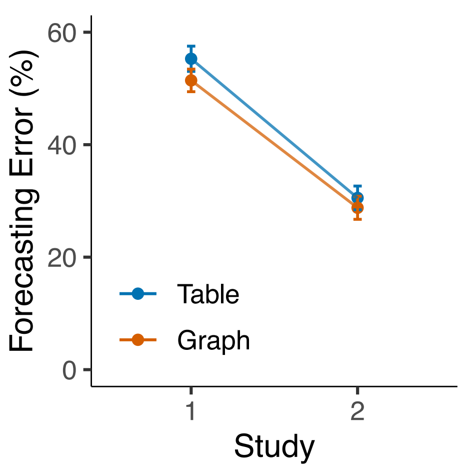
Table 15. *Forecasting Error*

Here we report the mean and standard deviation of forecasting error when estimates are compared to the values obtained by extrapolating the curve shown to participants. These data are shown for each of the three estimates for Study 1 and Study 2. In this table we collapse across text and table conditions.

Study	Day	*M*(*SD*) Forecasting Error %
1	3	36.79(21.68)
1	6	53.87(19.48)
1	9	66.55(22.13)
2	3	19.55(11.99)
2	6	30.59(18.10)
2	9	41.35(23.79)

Figure 41. *Forecasting Error*

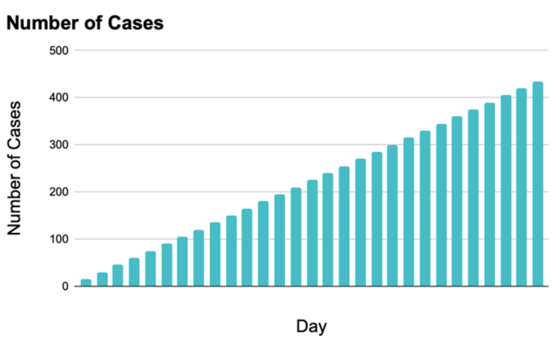
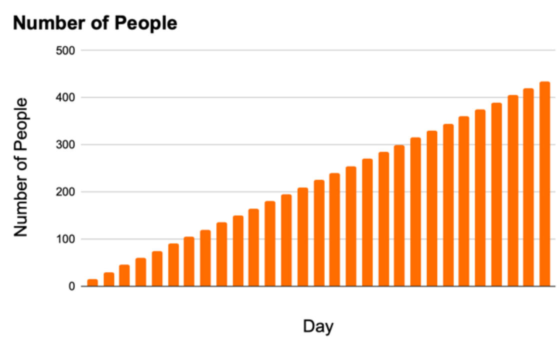
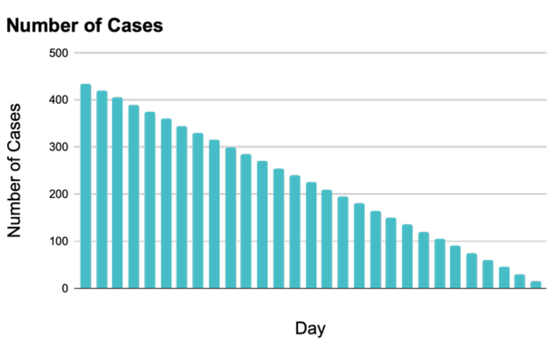
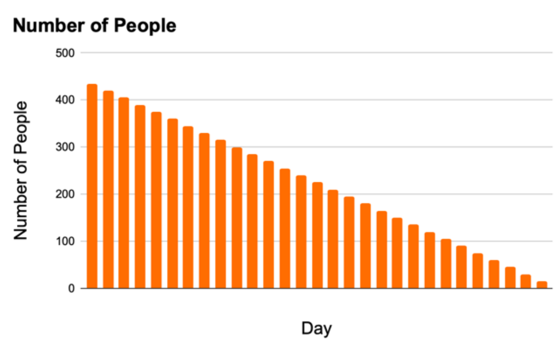
Forecasting Error. Here we show the average % error of participants forecasts with respect to the predicted values from exponential models fit to the data presented in the stimuli. These results show that whether error is measured with respect to actual or predicted values, forecasting error is quite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

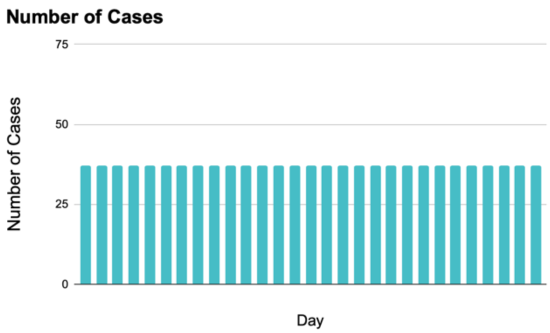


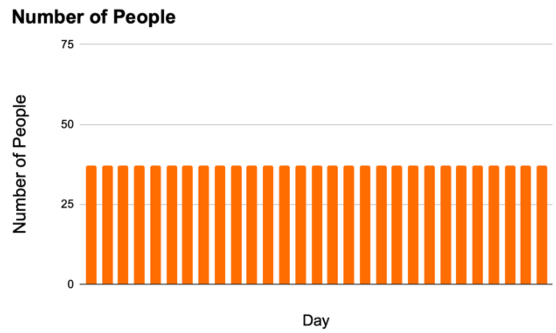
Appendix H Assessment Stimuli

Here we show the daily case curves for each of the seven questions for COVID-related and theme park scenarios. For each of the graphs below, there with four possible cumulative curves for participants to choose from.

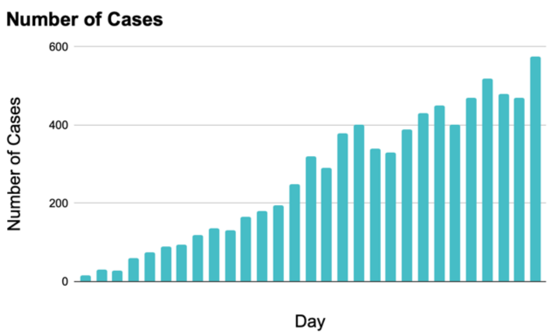
Figure 42. *COVID-Related Assessment Stimuli*

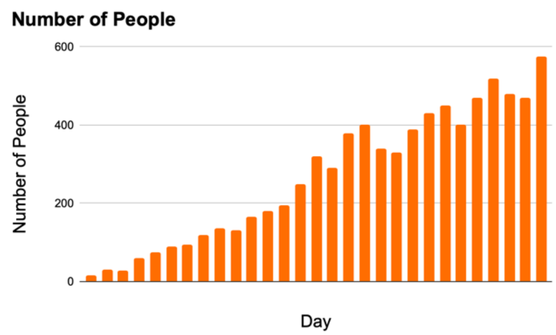
COVID-related	Theme Park
Increasing without noise	
	
Decreasing without noise	
	
Flat without noise	



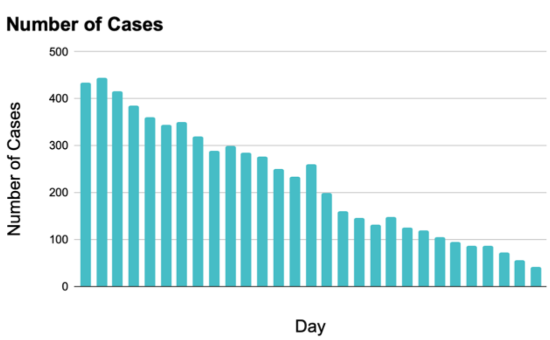


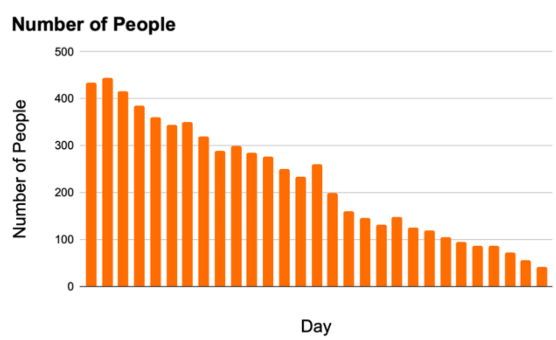
Increasing with noise



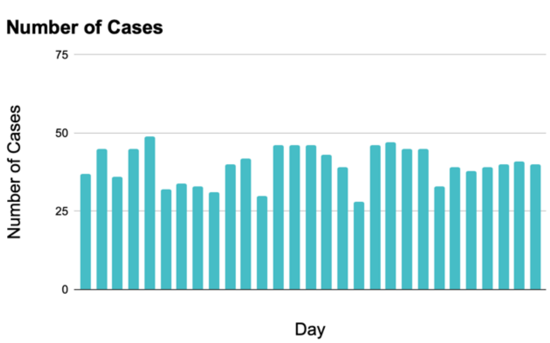


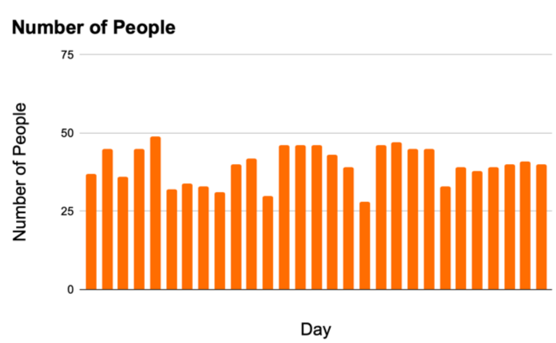
Decreasing with noise



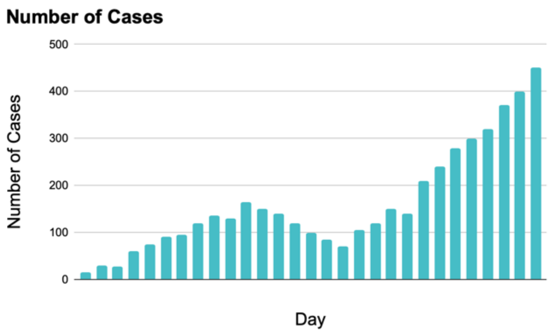


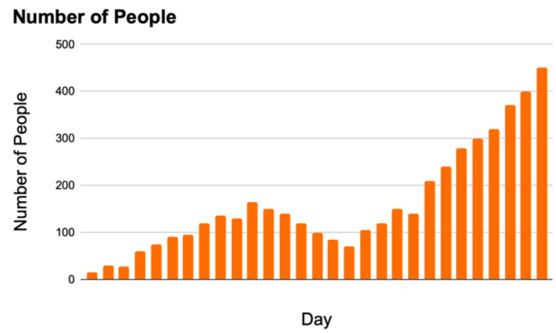
Flat with noise





Challenge



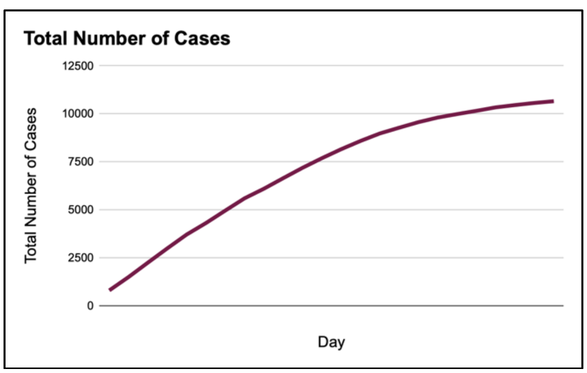


The following questions were presented to participants as additional measures of far transfer:

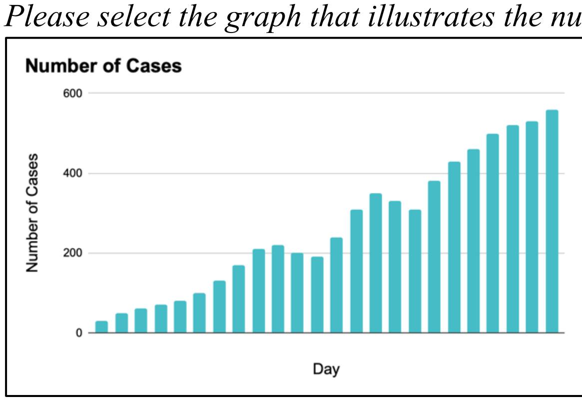
Up until this point you have selected the graph showing the cumulative number of cases based on the daily case graph. For the next two questions you will do the <u>reverse</u> of that task. That is, you will be shown a cumulative curve and have to choose the daily curve.

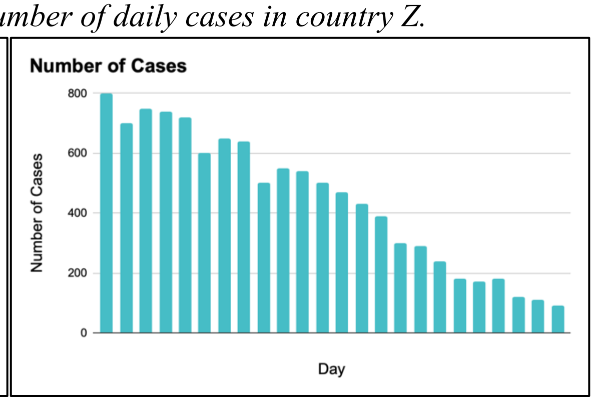
The graph below shows the number of cumulative COVID-19 cases for hypothetical country Z.

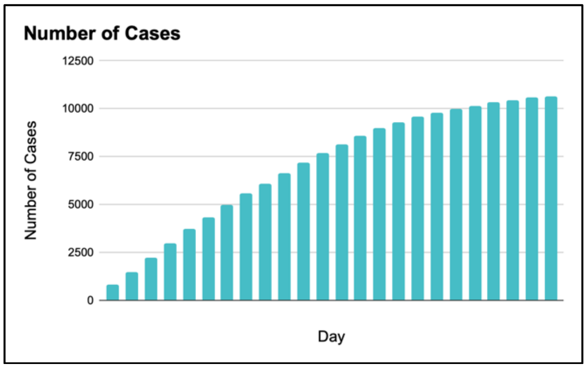
Figure 43. *Session 3 - Novel Stimuli*



Please select the graph that illustrates the number of daily cases in country Z.

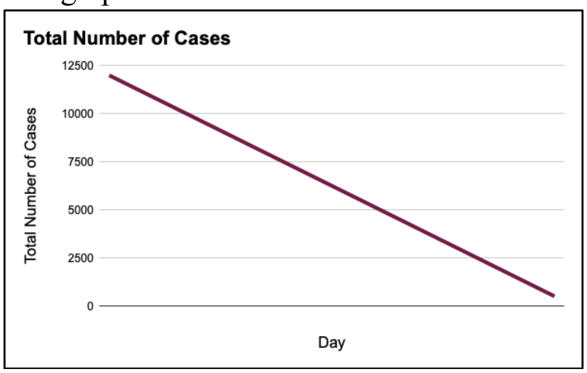




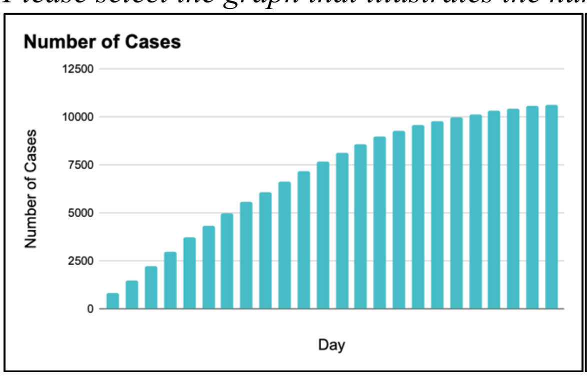


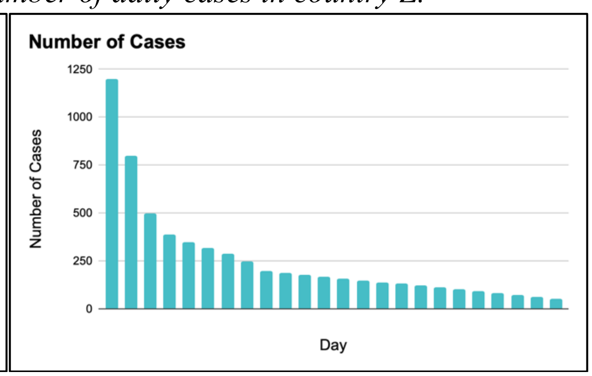
None of these graphs, the cumulative curve in the question is impossible

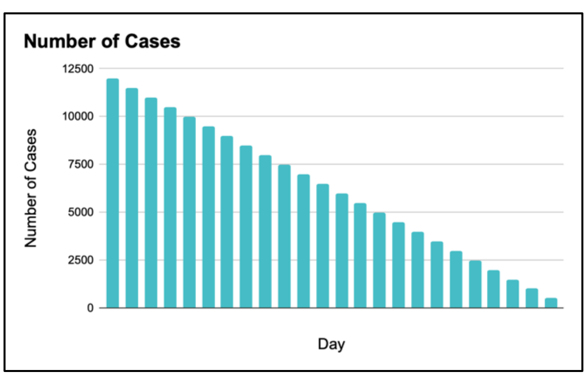
The graph below shows the number of cumulative COVID-19 cases for hypothetical country X.



Please select the graph that illustrates the number of daily cases in country Z.



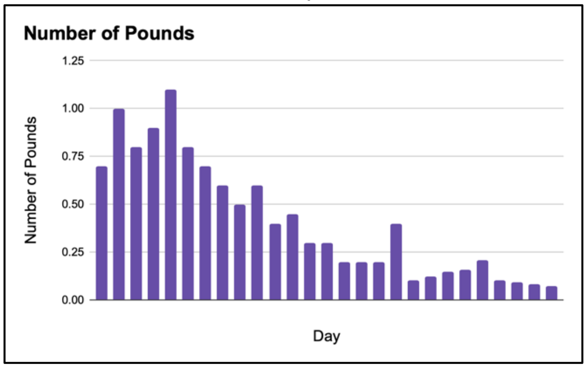




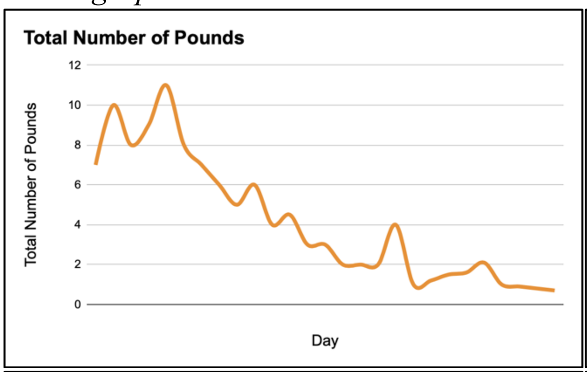
None of these graphs, the cumulative curve in the question is impossible

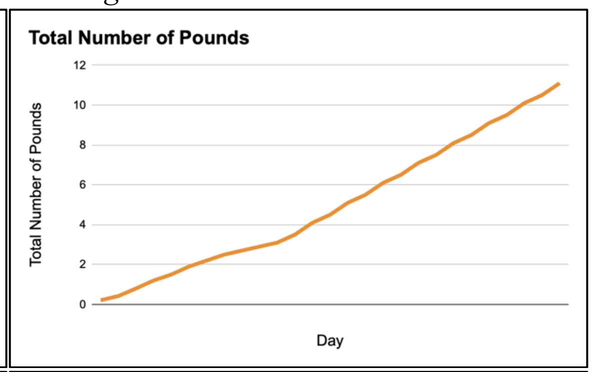
Let's test your knowledge in a few different types of scenarios.

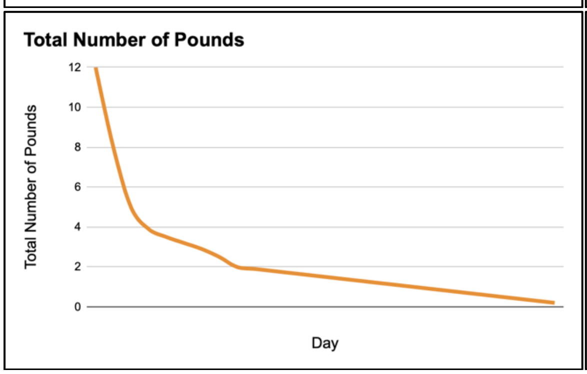
Jordan's new years resolution is to lose 20 pounds. The graph below shows the amount of weight Jordan has lost **each day** since he started dieting and exercising.

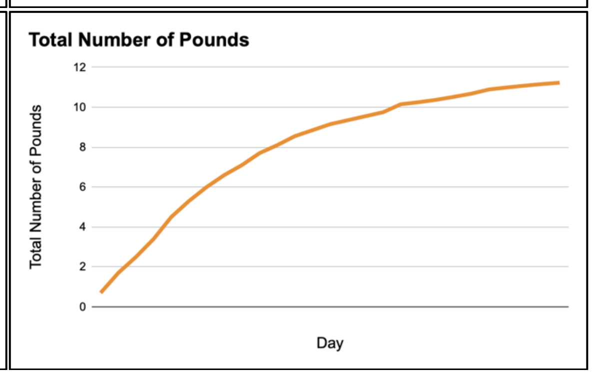


Which graph shows Jordan's cumulative or total weight loss?

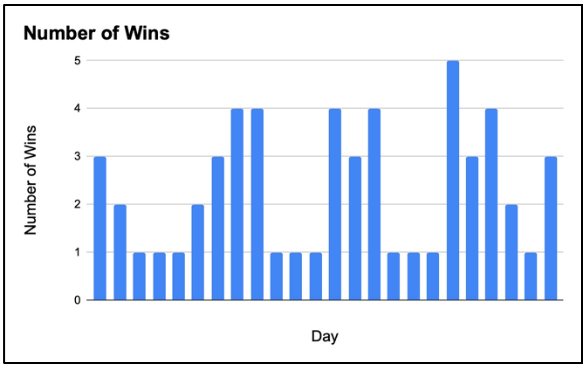




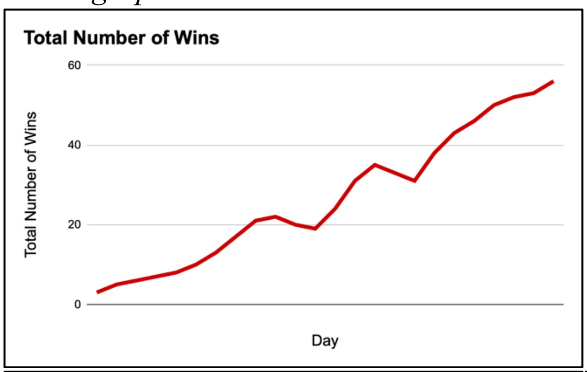


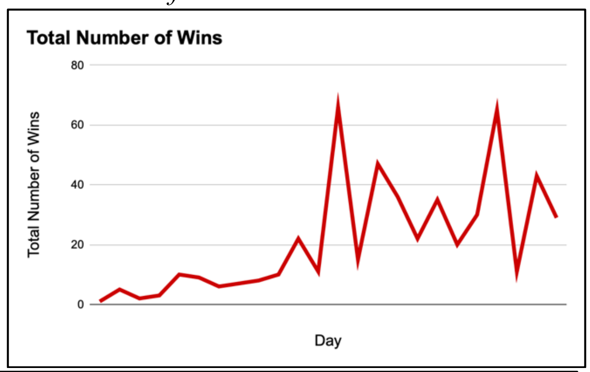


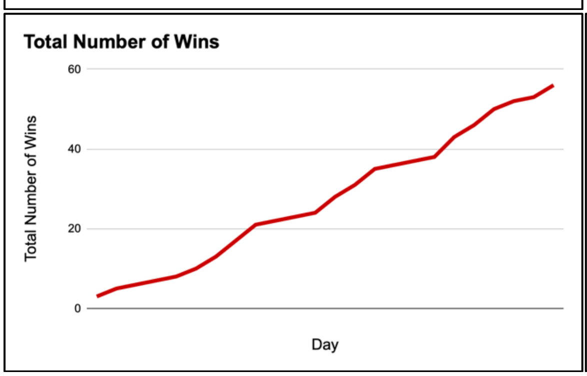
Amelia is a professional chess player who travels around the world competing in chess tournaments. The graph below shows the number of chess games Amelia has won at each tournament she attended over the last year.

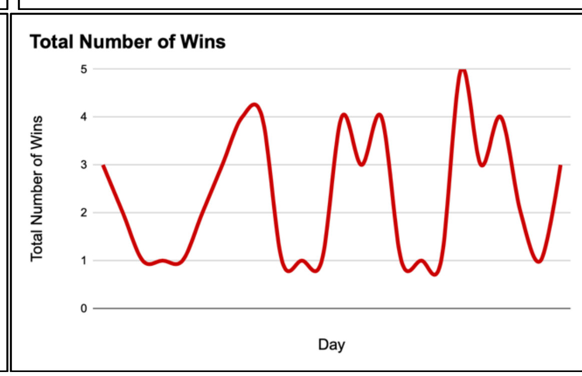


Which graph shows Amelia's cumulative or total number of wins?









Appendix I Individual Differences Analysis

Table 16. *Exploratory Individual Differences Analysis Model Output*

Working Memory (WM)				
		β	*CI95%*	*pd*
Session 1	WM (intermediate values)	0.21	[0.06, 0.35]	.99
Immediate post-	Condition (intermediate values)	-0.41	[-0.91, 0.08]	.95
test as dv	Condition*WM (intermediate values)	0.46	[0.25, 0.67]	.999
	WM (extremes (0 or 1))	1.11	[0.57, 1.67]	1
	Condition (extremes (0 or 1))	0.32	[-1.18, 1.79]	.66
	Condition*WM (extremes (0 or 1))	0.66	[0.09, 1.26]	.99
		β	*CI95%*	*pd*
Session 2	WM (intermediate values)	0.32	[0.15, 0.48]	1
	Condition (intermediate values)	-0.04	[-0.64, 0.51]	.55
	Condition*WM (intermediate values)	0.21	[-0.01, 0.45]	.97
	WM (extremes (0 or 1))	1.77	[1.22, 2.35]	1
	Condition (extremes (0 or 1))	0.23	[-.1.29, 1.70]	.62
	Condition*WM (extremes (0 or 1))	0.47	[-0.11, 1.06]	.94
		β	*CI95%*	*pd*
Session 3	WM (intermediate values)	0.17	[-0.04, 0.39]	.94
	Condition (intermediate values)	-0.41	[-1.16, 0.35]	.86
	Condition*WM (intermediate values)	0.35	[0.06, 0.64]	.99
	WM (extremes (0 or 1))	1.48	[.93, 2.08]	1
	Condition (extremes (0 or 1))	-0.50	[-2.04, 1.03]	.74
	Condition*WM (extremes (0 or 1))	0.73	[0.09, 1.36]	.99
Subjective Numeracy				
		β	*CI95%*	*pd*
Session 1	Numeracy (intermediate values)	-0.12	[-0.33, 0.10]	.86
Immediate post-	Condition (intermediate values)	1.29	[0.20, 2.41]	.99
test as dv	Condition*Numeracy (intermediate values)	-0.21	[-0.47, 0.05]	.94
	Numeracy (extremes (0 or 1))	-1.18	[-1.83, -0.53]	.99
	Condition (extremes (0 or 1))	1.08	[-0.71, 2.86]	.88
	Condition*Numeracy (extremes (0 or 1))	0.22	[-0.23, 0.67]	.84
		β	*CI95%*	*pd*
Session 2	Numeracy (intermediate values)	-0.20	[0.15, 0.48]	.95
	Condition (intermediate values)	0.45	[-0.64, 0.51]	.78
	Condition*Numeracy (intermediate values)	-0.01	[-0.01, 0.45]	.54
	Numeracy (extremes (0 or 1))	-1.06	[-1.68, -0.47]	.99
	Condition (extremes (0 or 1))	1.15	[-0.61, 2.94]	.90
	Condition*Numeracy (extremes (0 or 1))	0.05	[-0.39, 0.48]	.59
		β	*CI95%*	*pd*
Session 3	Numeracy (intermediate values)	0.51	[-0.48, 0.08]	.92
	Condition (intermediate values)	-0.20	[-0.82, 1.85]	.78
	Condition*Numeracy (intermediate values)	-0.03	[-0.36, 0.29]	.58
	Numeracy (extremes (0 or 1))	-0.92	[-1.52, -0.32]	.99
	Condition (extremes (0 or 1))	0.64	[-1.14, 2.42]	.76
	Condition*Numeracy (extremes (0 or 1))	0.16	[-0.29, 0.60]	.76
Graph Literacy				
		β	*CI95%*	*pd*
	Graph Literacy (intermediate values)	0.20	[-0.12, 0.16]	.62

Session 1	Condition (intermediate values)	1.34	[0.53, 2.15]	.99
Immediate post-	Condition* Graph Literacy (intermediate values)	-0.21	[-0.39, -0.03]	.99
test as dv	Graph Literacy (extremes (0 or 1))	-0.30	[-0.71, -0.11]	.93
	Condition (extremes (0 or 1))	1.62	[0.01, 3.27]	.98
	Condition* Graph Literacy (extremes (0 or 1))	0.05	[-0.32, 0.42]	.60
		β	$CI_{95\%}$	*pd*
Session 2	Graph Literacy (intermediate values)	-0.06	[-0.21, 0.09]	.77
	Condition (intermediate values)	0.55	[-0.33, 1.42]	.89
	Condition* Graph Literacy (intermediate values)	-0.04	[-0.24, 1.16]	.65
	Graph Literacy (extremes (0 or 1))	-0.36	[-0.74, 0.01]	.97
	Condition (extremes (0 or 1))	1.66	[-0.07, 3.29]	.98
	Condition* Graph Literacy (extremes (0 or 1))	-0.09	[-0.45, 0.27]	.69
		β	$CI_{95\%}$	*pd*
Session 3	Graph Literacy (intermediate values)	-0.03	[-0.21, 0.15]	.63
	Condition (intermediate values)	0.41	[-0.57, 1.41]	.80
	Condition* Graph Literacy (intermediate values)	-0.01	[-0.24, 0.22]	.54
	Graph Literacy (extremes (0 or 1))	-0.27	[-0.62, 0.09]	.93
	Condition (extremes (0 or 1))	1.19	[-0.46, 2.83]	.92
	Condition* Graph Literacy (extremes (0 or 1))	0	[-0.37, 0.37]	.51
Political Ideology (Higher Score = More Conservative)				
		β	$CI_{95\%}$	*pd*
Session 1	Political Ideology (intermediate values)	0	[-0.08, 0.08]	.50
Immediate post-	Condition (intermediate values)	0.41	[0.19, 0.62]	1
test as dv	Condition* Political Ideology (intermediate values)	0.03	[-0.10, 0.15]	.67
	Political Ideology (extremes (0 or 1))	0.08	[-0.21, 0.36]	.71
	Condition (extremes (0 or 1))	1.76	[1.12, 2.42]	1
	Condition* Political Ideology (extremes (0 or 1))	-0.06	[-0.41, 0.29]	.63
		β	$CI_{95\%}$	*pd*
Session 2	Political Ideology (intermediate values)	0.06	[-0.01, 0.14]	.95
	Condition (intermediate values)	0.39	[0.15, 0.62]	.99
	Condition* Political Ideology (intermediate values)	-0.08	[-0.21, 0.05]	.88
	Political Ideology (extremes (0 or 1))	-0.11	[-0.34, 0.14]	.79
	Condition (extremes (0 or 1))	1.27	[0.66, 1.88]	1
	Condition* Political Ideology (extremes (0 or 1))	0.19	[-0.13, 0.51]	.87
		β	$CI_{95\%}$	*pd*
Session 3	Political Ideology (intermediate values)	0.12	[0.03, 0.22]	.99
	Condition (intermediate values)	0.42	[0.14, 0.72]	.99
	Condition* Political Ideology (intermediate values)	-0.06	[-0.20, 0.09]	.79
	Political Ideology (extremes (0 or 1))	-0.05	[-0.32, 0.22]	.64
	Condition (extremes (0 or 1))	1.16	[0.54, 1.81]	1
	Condition* Political Ideology (extremes (0 or 1))	0.06	[-0.28, 0.42]	.64

Appendix J ANOVA Replication

While zero-one-inflated Beta regression is arguably the most suitable approach to analyzing slider scale data, we sought to replicate our findings using a simpler analysis of variance (ANOVA). Experiment 1 data were analyzed with a factorial ANOVA with probability expression and visualization condition as factors. Data for Experiment 2 were analyzed with a one-way ANOVA with Tukey HSD post-hoc tests and visualization condition as the predictor. Significance was assessed at the .05 level and all tests were two-tailed with aversion toward the J&J vaccine or all COVID-19 vaccines as the dependent variable. Data were analyzed with the stat package in R.

In Experiment 1, probability expression did not impact aversion toward the J&J vaccine ($F(2,1046) = 2.27$, $p = 0.1$), nor all COVID-19 vaccines ($F(2,1046) = .02$, $p = 0.98$). However, participants who viewed the icon array depicting side effect risk reported significantly less aversion towards the J&J ($F(1,1046) = 42.19$, $p < .001$) and all COVID-19 vaccines ($F(1,1046) = 10.31$, $p = .001$) when compared to participants who did not view an icon array. There was no interaction between probability expression and presence of an icon array on aversion towards the J&J vaccine ($F(2,1046) = 0.43$, $p = 0.65$) nor all COVID-19 vaccines ($F(2,1046) = 1.27$, $p = .28$).

In Experiment 2 we find an overall effect of visualization condition on aversion towards the J&J ($F(2, 848) = 10.9$, $p <.001$) and all COVID-19 vaccines ($F(2, 848) = 3.21$, $p = .04$). Tukey post-hoc tests reveal that those who viewed the icon array depicting side effect risk were significantly less averse toward the J&J ($M_{diff} = .12$, 95%CI = [-.19, -.04], $p < .001$) and all COVID-19 vaccines ($M_{diff} = .08$, 95%CI = [-.15, -.002], $p = .04$) than those who viewed no icon array. Those who viewed the relative risk icon array were significantly less averse than those who viewed no visualization for the J&J vaccine ($M_{diff} = .13$, 95%CI = [-.20, -.06], $p < 0.001$) but not for all COVID-19 vaccines ($M_{diff} = .06$, 95%CI = [-.13, .02], $p < 0.15$). There was no difference in aversion toward the J&J vaccine ($M_{diff} = .02$, 95%CI = [-.08, .05], $p = .83$) or all

COVID-19 vaccines (M_{diff} = .02, 95%CI = [-.06, .09], p = .84) when comparing the side effect-only and relative risk icon array conditions, suggesting the visualizations are equally effective.