informs
http://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Vol. 66, No. 5, May 2020, pp. 2213–2231
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Judgments Based on Stocks and Flows: Different Presentations of the Same Data Can Lead to Opposing Inferences

Stephen A. Spiller,[a] Nicholas Reinholtz,[b] Sam J. Maglio[c,d]

[a] Anderson School of Management, University of California, Los Angeles, Los Angeles, California 90095; [b] Leeds School of Business, University of Colorado Boulder, Boulder, Colorado 80309; [c] Rotman School of Management, University of Toronto, Toronto, Ontario M5S 3E6, Canada; [d] University of Toronto Scarborough, Toronto, Ontario M1C 1A4, Canada

Contact: stephen.spiller@anderson.ucla.edu, http://orcid.org/0000-0001-6951-6046 (SAS); nicholas.reinholtz@colorado.edu, http://orcid.org/0000-0003-0696-1972 (NR); sam.maglio@utoronto.ca, http://orcid.org/0000-0003-0691-1721 (SJM)

Received: August 30, 2017
Revised: June 20, 2018
Accepted: November 21, 2018
Published Online in Articles in Advance: September 20, 2019

https://doi.org/10.1287/mnsc.2019.3284

Abstract. Time-series data—measurements of a quantity over time—can be presented as *stocks* (the quantity at each point in time) or *flows* (the change in quantity from one point in time to the next). In a series of six experiments, we find that the choice of presenting data as stocks or flows can have a consequential impact on judgments. The same data can lead to positive or negative assessments when presented as stocks versus flows and can engender optimistic or pessimistic forecasts for the future. For example, when employment data from 2007 to 2013 are shown as flows (jobs created or lost), President Obama's impact on the economy during his first year in office is viewed positively, whereas when the same data are shown as stocks (total jobs), his impact is viewed negatively. The results highlight a challenge that accompanies the growing reliance on data and analytics for decision making within organizations: seemingly benign choices—such as that between two informationally equivalent data presentations—can substantively impact how data are interpreted and used, even though the underlying information is the same.

History: Accepted by Yuval Rottenstreich, decision analysis.
Funding: This research was supported by the UCLA Anderson School of Management, the Leeds School of Business, University of Colorado Boulder, and the Social Sciences and Humanities Research Council of Canada.
Supplemental Material: Data files and the online appendix are available at https://doi.org/10.1287/mnsc.2019.3284.

Keywords: judgments • forecasts • time series • stocks and flows • data visualization

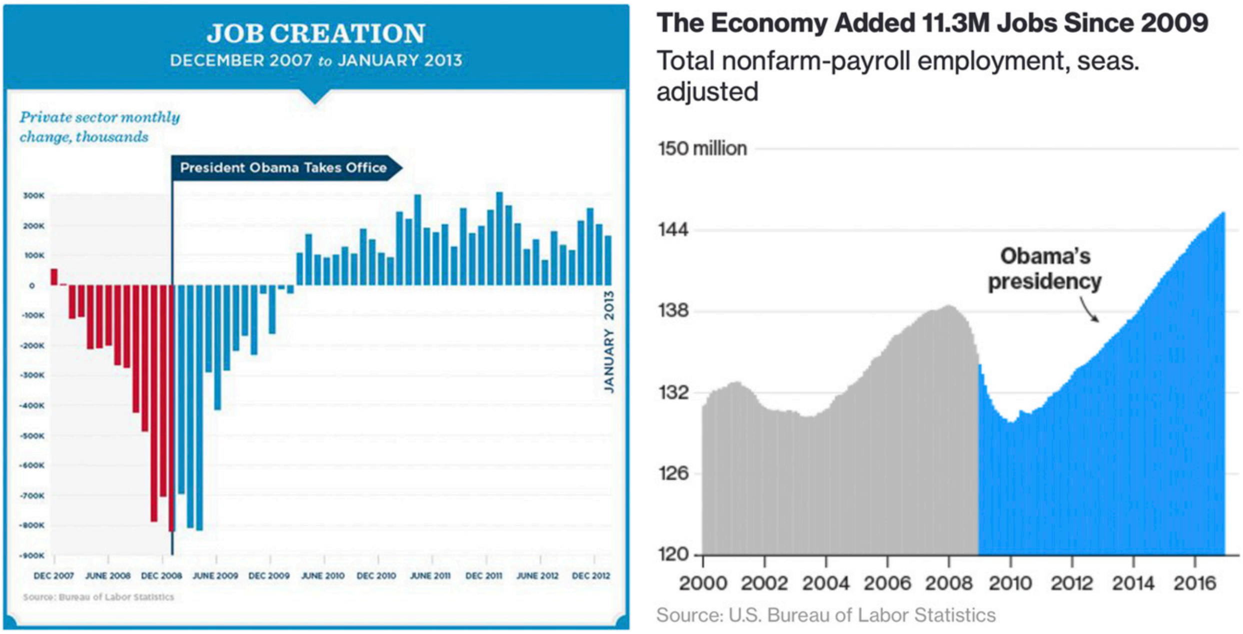
Introduction

Time-series data present measurements of the same quantity appraised at different times and are frequently used within organizations to make assessments and aid in decision making. For example, time-series employment data can provide insight into the health of the economy. The Bureau of Labor Statistics estimated that in 2015, 146 million people were employed in the U.S. in nonagricultural industries (Bureau of Labor Statistics 2017). In 2016 and 2017, those figures were 149 and 151 million, respectively. Describing the changing economy by the total number of jobs presents time-series data as a *stock*—the magnitude of the target quantity at each period. The same data could also be described as a *flow*—the change in magnitude of the target quantity between each period: from 2015 to 2016, the U.S. economy gained 3 million jobs. From 2016 to 2017, the U.S. economy gained 2 million jobs. Figure 1 reproduces two examples of figures intended for public consumption showing similar underlying jobs data, depicted as either flows or stocks.

Both stock and flow measures describe the same evolving quantity, and accordingly, each can be transformed to the other: the net flow is the mathematical derivative of a stock, and the stock is the integral of the net flow.[1] Despite this functional equivalence, we propose that presentation format—stock versus flow—can impact qualitative judgments made from the same data in consequential ways. Our focus on judgments departs from past research, which has focused on whether people can accurately translate between the two formats (e.g., Booth Sweeney and Sterman 2000, Cronin et al. 2009). Instead, we explore whether the format in which the same data are presented—either as stocks or flows—can systematically shift how people reason about the data and the conclusions they draw from them.

To understand this effect of format, consider again the opening example of U.S.-held jobs from 2015 to 2017. The stock of jobs at three intervals (146 million, 149 million, and 151 million) shows a pattern that increases over time and may well paint a particularly rosy picture of the U.S. economy. However, the flow in jobs between those same intervals (+3 million and +2 million) reflects a decreasing trend—a slowing in the rate of job growth to the tune of 1 million jobs—and may lead

Figure 1. (Color online) Two Presentations of Similar Jobs Data from the Bureau of Labor Statistics



Notes. The left panel, by Organizing for Action, the successor to Obama for America, presents the data as a flow, reflecting month-over-month gains or losses in the number of jobs (retrieved August 29, 2017, from https://web.archive.org/web/20171006225044/https://www.ofa.us/news/35-months-of-job-growth/). The right panel, by a data and graphics journalist for Bloomberg News, presents the data as a stock, reflecting the number of people employed (retrieved August 29, 2017, from https://web.archive.org/web/20170123093402/https://www.bloomberg.com/graphics/2017-obama-economic-legacy/).

people to see the U.S. economy in less bullish terms. Building directly from this insight, we use similar data to assess this possibility in Studies 1, 2, and 3.

The current investigation is motivated by the expanding role and frequently espoused importance of data and visualization in decision making. Managers and laypeople alike—often lacking technical training—seek data to evaluate past decisions and inform future decisions. When presenting data to such decision makers, the focus is often on organizing the information in a manner that is clear and succinct (e.g., maximizing "data-ink" and removing "chart-junk"; Tufte 2001). What may get overlooked in this effort is that theoretically equivalent representations of data—those differing by a simple transformation, as is the case with stocks and flows—might lead the reader to opposing conclusions, as they can make different aspects of the data more or less salient.

The focus on time-series data reflects, in part, their ubiquity in quantitative communication. A survey of newspapers and magazines reported by Tufte (2001, p. 28) suggests that more than 75% of graphics convey time-series data. Beyond the Bureau of Labor Statistics, both stocks and flows commonly describe any number of time-series values of importance for decision makers. For example, when facing a new disease outbreak, health officials may see reports on the total number of people who have contracted the disease (stock) or the number of new cases (flow). When evaluating their personal finances, consumers may consider their total assets by month (stock) or the net of their earnings and expenses over each month (flow). When launching a start-up, the founders may focus on capital on hand (stock) or burn rate (flow). When evaluating the performance of companies, investors may consider total holdings (stock) or cash flows (flow).

In the remainder of the introduction, we first describe prior work documenting the difficulty people—even well-educated people—have translating between stocks and flows. We then consider a broader perspective by reviewing how presentation format can systematically influence judgments made from data. We contend that there is no neutral representation of time-series data—the data have to take some form (e.g., a stock or flow) in order to be observed and processed. The presentation format chosen (stock or flow) will make different patterns and aspects of the data salient, and people will use these salient features when forming judgments, both because this is the information that is most available and because it is difficult for most people to transform the data between formats.[2]

Because time-series data are used to inform many types of decisions, we examine the effects of presenting data as stocks or flows on two different yet prominent judgments: evaluative assessments and quantitative forecasts. In each, we find that the same data presented as a stock or as a flow can lead people to draw qualitatively different conclusions. The same data can be seen as a good sign or a bad sign and can lead to a forecasted increase or decrease depending on the salient trend depicted in the given presentation format. These different judgments hold when data are presented graphically as well as when data are presented in a tabular (numeric) format.

Stocks and Flows

Previous research regarding people's understanding of stocks and flows has focused on translation between formats, questioning whether people understand the process of accumulation (Booth Sweeney and Sterman 2000, Sterman and Booth Sweeney 2007). In a typical study, participants are given time-series data on a quantity as a flow and asked questions about levels of the stock. For example, in the "department store task" (Sterman 2002, Cronin et al. 2009), respondents are shown the number of people entering and leaving a store at every minute (flow) and asked, among other questions, when the most people were in the store (stock). A typical example of the stimulus (from Sterman 2002) is reproduced in Figure 2. For the relatively simple versions of the task typically used, the answer can be obtained without any calculation. The number of people in the store increases whenever more people are entering the store than leaving it and decreases whenever more people are leaving the store than entering it. This means that the number of people in the store reaches a maximum in the last period that the inflow exceeds the outflow (i.e., minute 13 in Figure 2). Although most (more than 90%) participants quite accurately reported other aspects of the data (e.g., the times at which the most people were entering or leaving the store), fewer than half accurately identified when the most people were in the store. Instead, a prominent incorrect answer to this task is to report that the stock of shoppers reached its maximum at the time of the maximum inflow (i.e., minute 4 in Figure 2, reflecting the use of a pattern matching or correlation heuristic; Sterman and Booth Sweeney 2002, 2007; Cronin et al. 2009). Even highly educated, incentivized adults are prone to these types of stock-flow failures (Cronin et al. 2009).

The present investigation, motivated by the many instances and applications in which people fail "to determine how the quantity in a stock varies over time given the rates of flow into and out of the stock" (Booth Sweeney and Sterman 2000, p. 252), builds from this prior work in asking how presenting time-series data in one format or the other—stock or flow—directly impacts judgment. Thus, our intended contribution is not to lend greater credence or qualification to the important body of literature on the general difficulty people have with accumulation discussed above. Instead, we present people with time-series data as either stocks or flows and ask them to make either qualitative assessments of the situation or forecasts about the future. We note that these judgments do not necessarily require translating between formats, but we assess whether the ability to do so impacts our findings in Study 5.

In Figure 3, we show four possible relationships that can arise between the stock and the flow of a time series, relationships on which our methodology will rely. In panel A, the stock is increasing and positively accelerating, which corresponds to a positive and increasing flow. Similarly, in panel D, the stock is decreasing and negatively accelerating, which corresponds

Figure 2. A Simplified Version of the Department Store Task, Reproduced Based on Data from Figure 3 in Sterman (2002)

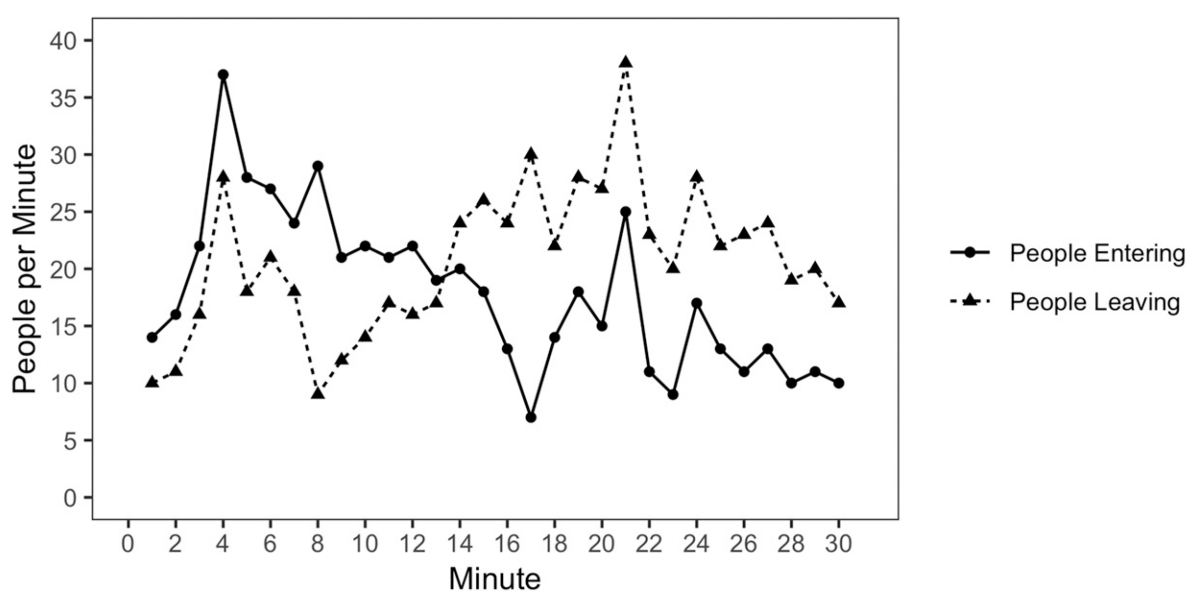
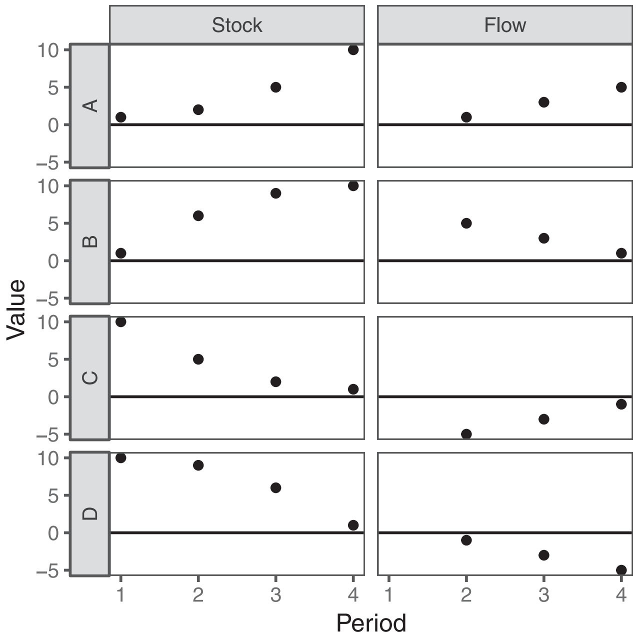


Figure 3. Four Examples of Stock and Flow Patterns



Notes. Flow levels are shown in the period following the change (e.g., the flow from period 1 to 2 is shown as period 2). In panels A and D, the stock and flow trends have the same sign. In panels B and C, the stock and flow trends have opposite signs.

to a negative and decreasing flow. In both of these situations, the stock and flow trends are in the same direction. Contrast this with panel B, in which the stock is increasing with negative acceleration, corresponding to a positive but decreasing flow. In this case, the two trends are moving in opposite directions. Panel C shows the converse situation: a decreasing stock with an increasing flow. We contend that the choice of presenting data as a stock or a flow in situations such as those in panels B and C might evoke different (opposing) representations of the positivity or negativity of the underlying data and consequently lead people to different judgments. By contrast, if people attend to the salient trend, presentations such as panels A and D are likely to evoke similar mental representations of positivity or negativity and lead people to relatively consistent judgments.

We depart from the prior work discussed above in that our focus is not on the qualitative accuracy of conversion from one format to the other. Rather, we target systematic differences in inferences drawn from presenting the same data as a stock or as a flow. Accordingly, because our focus is on inferences rather than translation accuracy, we do not situate our predictions within the same tradition of Sterman and colleagues, who proposed that errors in converting from flows to stocks reflect an inherent difficulty in understanding accumulation. In the next section, we draw from the separate literature on inference and judgment deriving from presentation format to examine the means by which presentation format might systematically shift judgment regardless of whether conversion from one format to the other is successful.

Judgments Based on Presentation Format

Prior research has documented other inconsistencies in judgment caused by different representations of equivalent data formats. In cases where the presentation of numerical information differs by a multiplier—reflecting only a difference in scale—people often make different judgments. Consider the numerosity heuristic (Pelham et al. 1994), by which people infer that larger numbers correspond to greater magnitude. As a result, they gamble on worse odds (9/100 over 1/10; Pacini and Epstein 1999, Reyna and Brainerd 2008), spend as though prices are higher when dealing with less numerous currencies (Raghubir and Srivastava 2002), and prefer options that dominate alternatives on attributes presented with a more expansive scale (Burson et al. 2009). In each of these instances, the considered quantities can be equated using a scalar multiplier (from U.S. dollars to Canadian dollars or from millimeters to centimeters; Maglio and Trope 2011).

Given that judgments can diverge when made from data formats that differ by a linear transformation, it comes as no surprise that judgments—even those made by well-educated people—can diverge between formats that differ by a nonlinear transformation (de Langhe et al. 2017). Take, for example, the "MPG illusion"—a well-documented bias in people's judgments about vehicle efficiency (Larrick and Soll 2008). Although the fuel economy of an automobile is typically displayed as miles per gallon (MPG), the savings gained by switching to a more efficient car are realized in burning fewer gallons of gasoline per mile driven (the reciprocal of miles per gallon). The illusion arises as a result of the fact that people seem to assume that an increase in miles per gallon corresponds to a linear decrease in gallons per mile, whereas the actual relationship is reciprocal. Similar issues arise for common productivity statistics such as megabytes per second (internet speed) and pages per minute (printer speed; de Langhe and Puntoni 2016). More generally, it has been recognized that the metric used to communicate data can serve as an important choice architecture tool (Camilleri and Larrick 2014).

These prior investigations suggest that people do not represent information in a format-neutral manner. Instead, people focus on the most salient characteristics in the information as presented (Slovic 1972, Cleveland and McGill 1984, Lurie and Mason 2007, Kahneman 2011). For the present investigation, these findings raise two considerations. First, people are likely to reason with the data in the format in which it is presented (Andreassen 1988, Andreassen and

Kraus 1990): if given stock data, people are likely to think about the data in terms of stocks and make judgments accordingly. If given flow data, people are likely to think about the data in terms of flows and make judgments accordingly.[3] Second, whereas these prior findings allowed a determination of accuracy or optimality, such prescriptive consideration need not always arise. In the MPG illusion, there is an objective benchmark for good judgment, better facilitated by one format (gallons per mile) than the alternative. In the current context, an objective benchmark is not available for comparing judgments made from stocks and flows without additional assumptions about the data generation process, so we make no claims as to whether one format facilitates correct responding. Should accuracy diverge between the two presentation formats, it poses a challenge regarding how to best present the data to enable accurate conclusions—and raises the question of what an "accurate conclusion" even is. We explore this latter question in the general discussion. Here, we simply examine whether judgments differ between formats and, if so, what type of data patterns are likely to generate these inconsistencies.

Time-series data represent successive measures that describe a dynamic value over time. We propose that relative differences in magnitude between successive data points—trends as opposed to the absolute levels of the points—will likely take priority in informing judgments. This is consistent with past research showing that recent trends inform predictions about the future (e.g., Freyd and Finke 1984). Because people tend to take the presentation format as given (i.e., fail to represent the data in a neutral format), we propose that they will make judgments based on their interpretations of the salient trend in that presentation, leading to inconsistencies between presentation formats when the salient trends differ. Furthermore, ending trends should be particularly salient for forecasts, as people exhibit a tendency to linearize the most recent observations when forming judgments about the future (Wagenaar and Sagaria 1975, McKenzie and Liersch 2011, Thomson and Oppenheimer 2016).

Overview of Studies

The same data presented as a stock can present a qualitatively different trend than when presented as a flow: as shown in panels B and C of Figure 3, a decrease in one format (stock) may simultaneously reflect an increase in the other (flow). Given the lay tendency to appraise data (and make resulting judgments) based on perceived trends revealed from past change, we predict that depicting the same data as a stock or a flow will elicit different perceptions of trends and, in turn, different judgments based on that data. Studies 1, 2, and 3 examine judgments regarding the economy and effects on the economy based on presentation of real jobs data from the Bureau of Labor Statistics, reported as either the number of jobs (stock) or the change in the number of jobs (flow). Studies 1 and 2 establish the core effect: participants' judgments from the same data can differ substantially and qualitatively depending on whether the data are presented as stocks or flows. Study 3 provides a robustness test by considering different sections of the time series depicting different characteristics and examining whether the inconsistencies are caused by participants inferring the relative diagnosticities of stocks and flows from the presentation format. Studies 4, 5, and 6 use artificially generated data and examine whether the differences driven by presentation format extend to forecasts. Study 5 further examines whether the ability to read and understand the data account for these differences. Misunderstanding the data explains some but not all of the effect, suggesting the role of presentation format affects even savvy data readers. Finally, Study 6 extends the investigation beyond graphical displays, yielding evidence that the effects cannot be exclusively attributed to quirks of visual processing. Throughout, all sample sizes were determined in advance, and all manipulations, measures, and data exclusions are reported.

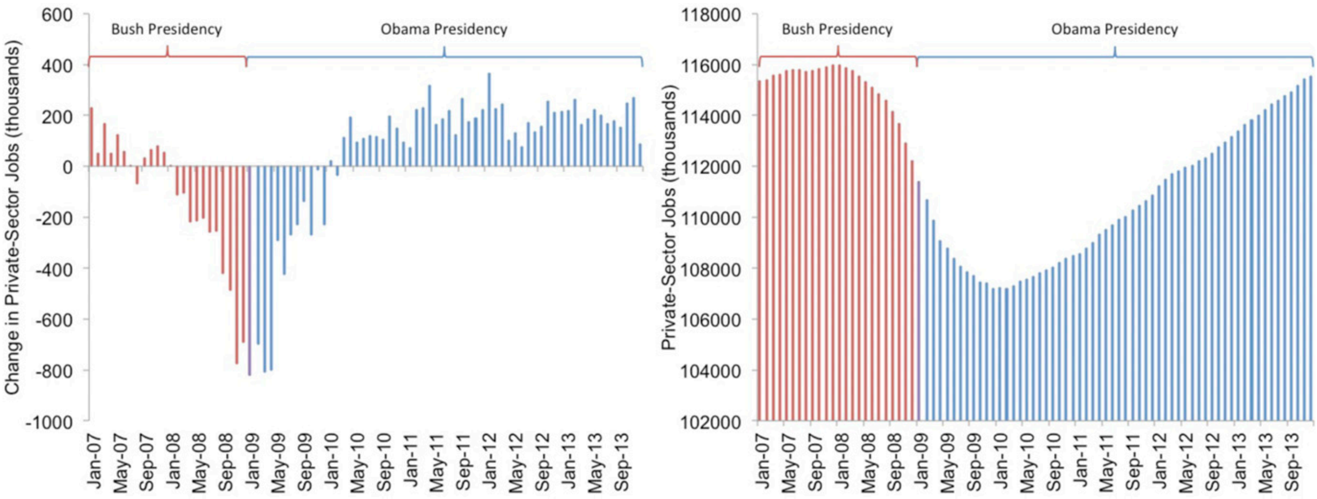
Study 1

Method

One hundred participants (32 women and 68 men; median age = 30) recruited from Amazon Mechanical Turk (AMT) completed Study 1.[4] They were randomly assigned to view one of two charts showing private-sector jobs in the United States from 2007 through 2013. In the flow condition, participants saw a chart that showed month-over-month changes in private-sector jobs in the United States. In the stock condition, participants saw a chart that showed the number of private-sector jobs each month (see Figure 4 for experimental stimuli; data from the Bureau of Labor Statistics). Months during the Bush presidency were shown in red; months during the Obama presidency were shown in blue (January 2009, when Obama was inaugurated, was shown in purple).

Participants were asked two sets of questions about the chart, with the order of the sets counterbalanced. One set of questions required reading information off the chart but no inferences beyond that. Specifically, participants were asked, "Based on the chart above, when did the rate at which the United States was losing private-sector jobs start to slow?" and "Based on the chart above, when did the number of private-sector jobs in the United States start to increase?" Participants responded by selecting a month and year from two drop-down menus. Options spanned from January 2007 to December 2014. These questions are analogous to those used in previous work (e.g., Cronin et al. 2009).

Figure 4. (Color online) Job Charts Used in Study 1



Notes. The flow chart on the left shows the flow of jobs (jobs gained or lost). The stock chart on the right shows the same data presented as the stock (total number of jobs).

The second set of questions required participants to make two judgments based on the chart. First, they were asked, "In your view, when did the recovery from the Great Recession begin?" and given dropdown menus to select a month and year from January 2007 to December 2014. Second, they were asked, "In your view, what effect did Barack Obama have on the American economy during his first year in office?" and responded on a seven-point scale completing the sentence, "During his first year in office, Barack Obama. . ." anchored by ". . .made it much worse" and ". . .made it much better," with five labels in between including ". . .made no difference" as the midpoint. Finally, participants reported sex, age, and ethnicity.

Results

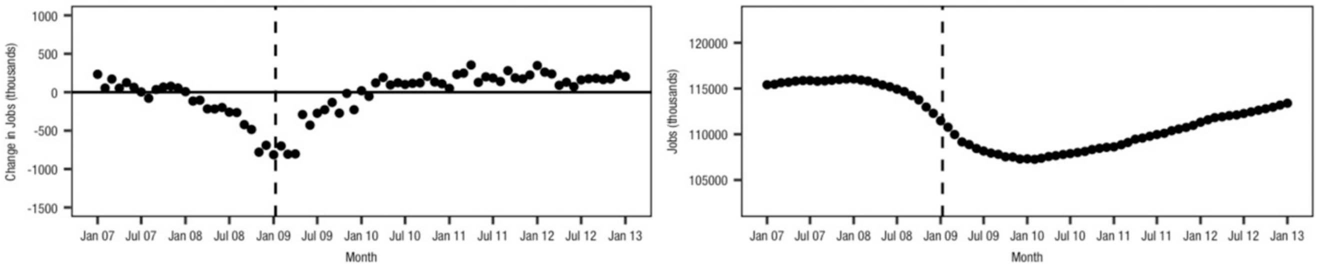
Reading the Data. Our main focus is on evaluative judgments, but we first describe participants' ability to accurately read the data. Participants showed high ability to assess when the number of jobs started to increase, given either chart. For stocks, 90% of responses were between September 2009 and May 2010. For flows, 82% of responses were between September 2009 and May 2010. There was more variability regarding judgment of when the rate of job loss began to slow. In the flow condition, 76% responded between January and September 2009, with 46% responding May 2009. In the stock condition, only 40% responding between January and September 2009; 40% responded sometime in 2008, and 18% responded between October 2009 and January 2010. Together, these results indicate that some values may have been difficult to read from the graph, but that was not as a result of misunderstanding of the graphs.

Evaluative Judgments. Order had no main nor interactive effects, so we exclude it from analysis for ease of exposition. There were two key dependent measures of interest: when the economic recovery was judged to have begun (coded as number of years since January 2007[5]) and what President Obama's influence on the economy during his first year in office was judged to have been.

First, judgments of when the recovery began varied depending on whether the data were presented as stocks or flows (M_{Stock} = 3.46 (June 2010), SD = 1.26; M_{Flow} = 2.85 (November 2009), SD = 0.94; $t(98) = 2.75$, $p = 0.007$): on average, viewing a chart of the number of jobs rather than the change in number of jobs led to a perception that the economy started recovering seven months later.

Second, judgments of Obama's influence on the economy during his first year in office also varied depending on the way the information was presented ($t(98) = 5.10$, $p < 0.001$). When jobs data were presented as a stock, participants evaluated President Obama's impact on the economy *negatively* (M = 3.32; less than the scale midpoint of 4 ("made no difference"), $p = 0.002$, with 60% reporting he made it worse and 24% reporting he made it better). However, when the jobs data were presented as a flow, participants evaluated President Obama's impact on the economy *positively* (M = 4.84; greater than the scale midpoint, $p < 0.001$, with 66% reporting he made it better and 8% reporting he made it worse). Thus, this simple difference in presentation format—number of jobs (stock) or change in jobs (flow)—led not just to a significant difference in evaluation but also to a qualitative reversal (bad to good) in judgments about President Obama's effect on the economy.

Figure 5. Job Charts Used in Study 2



Notes. The flow chart on the left shows the flow of jobs (jobs gained or lost). The stock chart on the right shows the same data presented as the stock (total number of jobs). The vertical dashed line indicates Barack Obama's first inauguration.

Discussion

Showing the same data as a stock or a flow affected what judgments people drew about the nature of the economy. When data were presented as a flow, the pattern revealed a minimum (i.e., maximal job losses) at the beginning of 2009 and the trend during Obama's first year as president was upward-sloping (becoming less negative). When the same data were presented as a stock, the pattern revealed a minimum at the beginning of 2010 (a year later, when the economy stopped losing jobs and started gaining jobs) and the trend during Obama's first year as president was downward-sloping (as the economy continued to lose jobs but at a slower rate). Despite being based on the same underlying data, judgments from stocks versus flows reflected these gestalt differences. When data were presented as flows, participants judged the recovery to have begun earlier than when the data were presented as stocks. Moreover, when data were presented as flows, Obama was judged to have had a positive effect on the economy during his first year as president, whereas when data were presented as stocks, Obama was judged to have had a negative effect on the economy during his first year as president. Different trends can be salient when viewing data as stocks or flows, leading to different inferences.

In Study 2, we sought to replicate and extend the results of Study 1. Beyond varying theoretically irrelevant characteristics of the stimuli in the interest of providing evidence for robustness, we examined whether the trends evoked from partial data sets presented as stocks or flows color the forecasts people make for the underlying data in the future.

Study 2

Method

Two hundred participants (80 women and 118 men; median age = 32) from AMT were recruited to participate in and completed Study 2.[6] Study 2 was a replication and extension of Study 1. It used similar methods, so we focus on the changes below.

First, the data were presented as points rather than bars to reduce any impact of a truncated axis in the stock condition but not the flow condition (see Figure 5). Previous investigations found that effects of mode of presentation (e.g., lines versus bars) on stock-flow reasoning were negligible (Cronin et al. 2009), but because position is more reliably assessed than length or area (Cleveland and McGill 1984), it is important to assess the robustness of our results.

Second, instead of being asked about the timing of the recovery, participants were asked directly about how the economy changed: "In your view, how did the economy change during Barack Obama's first year in office?" (seven-point scale anchored by "worsened a lot" to "improved a lot"). This was included first to measure perceptions of economic change while deemphasizing President Obama's impact. Following this, participants assessed President Obama's impact on the economy using the same measure as in Study 1.

Third, we included a set of forecasting questions, with the order of evaluation and forecasting question sets counterbalanced. For these questions, participants were only shown data from 2007 to 2009 (i.e., only the dots on the left side of the dashed line in Figure 5). Participants were asked to forecast the level of stock or flow (according to condition) they would expect in January 2010 based only on the data from January 2007 through January 2009 in the absence of any exogenous change (post-2009 data were not shown). They made this forecast by clicking on a point on the graph, the location of which was recorded (akin to Thomson and Oppenheimer 2016). They then evaluated how they expected the economy would have developed in 2009 based on the data they saw in 2007 and 2008 (seven-point scale anchored by "worsened a lot" to "improved a lot"). These questions were included to illuminate how presentation format influences expectations in the absence of realized changes.

Finally, participants completed a single-item measure of their political persuasion (from very conservative, coded as 1, to very liberal, coded as 5).

Results

Order did not affect the primary dependent variables of interest, so again, we exclude it for ease of exposition. The measure of Obama's effect on the economy replicated Study 1: participants who saw the jobs data as flows believed Obama had a positive impact on the economy during his first year in office (M = 5.21, greater than the scale midpoint of 4, $p <$ 0.001, with 73% reporting he made it better and 14% reporting he made it worse), whereas those who saw the same data as stocks believed he had a negative impact (M = 3.69, less than the scale midpoint of 4, $p =$ 0.037, with 49% reporting he made it worse and 29% reporting he made it better; difference between conditions: $t(198) = 7.31$, $p < 0.001$). Perceptions of how the economy changed largely tracked this measure and showed a similar difference ($t(197) = 8.40$, $p < 0.001$):[7] participants who saw the data as flows judged the economy to have improved (M = 5.31, greater than the scale midpoint, $p < 0.001$, with 78% reporting it improved and 18% reporting it worsened), whereas those who saw the data as stocks judged the economy to have worsened (M = 3.22, less than the scale midpoint, $p < 0.001$, with 70% reporting it worsened and 28% reporting it improved). For each of these two evaluative judgments, liberalism (versus conservatism) was correlated with more positive assessments (change in economy: $r = 0.20$, $p = 0.006$; Obama's effect: $r = 0.41$, $p < 0.001$), but in neither case did it moderate the effect of stock versus flow (p's > 0.18).[8] To put the effect size into context, presenting the data as stocks rather than flows creates a difference in the perceived effect of Obama on the economy (i.e., a decrease of 1.5 points) that is larger than the difference between participants who rated themselves "conservative" and those who rated themselves "liberal" (a difference of 1.2 points).

Participants' predictions and their evaluations of those predictions illuminate how they reason through stock and flow changes. In both conditions, the typical forecast followed the salient trend in the presentation format provided: in the flow condition, participants expected the negative trend in job losses to continue (with implied average monthly losses of 918,000 jobs). In the stock condition, participants expected the negative trend in total jobs to continue but with less severe consequences (with implied average monthly losses of 296,000 jobs). Despite this stark difference in forecasts between conditions, participants' subjective evaluations of these forecasts were similar and equally sour in both conditions (about one point below the midpoint).[9] This evidence is consistent with the notion that people interpret the same state of the world differently depending on whether it is presented as stocks or flows: a constant negative flow would be evaluated as nearly neutral in the flow condition (as the trend is flat) but as extremely negative in the stock condition (as the trend is extremely negative). Details regarding these analyses are presented in Online Appendix A.

Discussion

Study 2 replicated our key findings from Study 1: people reach qualitatively different evaluations from the same data depending on presentation format. In addition, we find that when equated to consistent units, people's forecasts also differ substantially between presentation formats, leading to divergent implications, a point to which we will return in Studies 4, 5, and 6. In short, participants reason about the data differently when considering stocks versus flows.

Study 3

Study 3 extends the results of Studies 1 and 2 in two key ways. First, we use the same data as in Studies 1 and 2 but have participants in different conditions make judgments regarding different time periods. This allows us to test an implicit assumption in the prior studies: that participants are forming judgments based on the specific time period in question rather than the entirety of the data presented. Second, we measure explicit beliefs about the importance of different inputs into what matters in terms of the economy. This allows us to assess whether changing the metric in which we present the data affects the subjective importance of possible signals about the economy (e.g., whether showing people jobs as flows makes people think the flow of jobs is a more important economic indicator).[10]

Method

Participants were recruited from AMT (n = 401; 184 women and 216 men; median age = 34).[11] Study 3 was similar to Study 2 with a few variations. First, and most important, Study 3 included a second factor in the design, resulting in a 2 (data presentation: stock versus flow) × 2 (time span: 2009 versus 2010) between-subjects design. Half of the sample was assigned to the 2009 condition. As in Studies 1 and 2, these participants evaluated the change in the economy and Obama's effect on the economy during 2009, his first year in office. The other half of the sample was assigned to the 2010 condition. These participants had the same task, except they evaluated the economy one year later during 2010, Obama's second year in office. Whereas the flow was increasing and the stock was decreasing during 2009, the flow was flat and the stock was increasing during 2010. The stimuli were adjusted to highlight the focal time span. See Figure 6.

Similar to Study 2, we measured perceived change in the economy and Obama's effect on the economy during the focal time span. Unlike Study 2, these were measured on three-point scales ("worsened," "did not change," and "improved" for how the economy changed; "made the economy worse," "made no difference," and "made the economy better" for Obama's effect on the economy). Unlike Study 2, we did not assess forecasts in Study 3. Instead, we assessed how important participants believed "the number of jobs," "the monthly growth rate (number of jobs gained or lost per month)," and "the change in the monthly growth rate" were to evaluating the overall state of the economy on five-point scales (from "not at all important" to "extremely important"). This allowed us to assess whether the type of data presentation had an impact on the perceived importance of stocks versus flows.

Prior to assessing the importance of each component, participants had the opportunity to provide an open-ended explanation regarding why they responded the way that they did regarding the economy and Obama's effect. Finally, we assessed to what extent participants believed presidents have the potential to impact the economy ("To what extent do you think presidents have the potential to impact the economy during their first/second year in office?"; from "not at all" to "to a great extent" on a four-point scale) and closed with measures of political liberalism, gender, and age.

Results

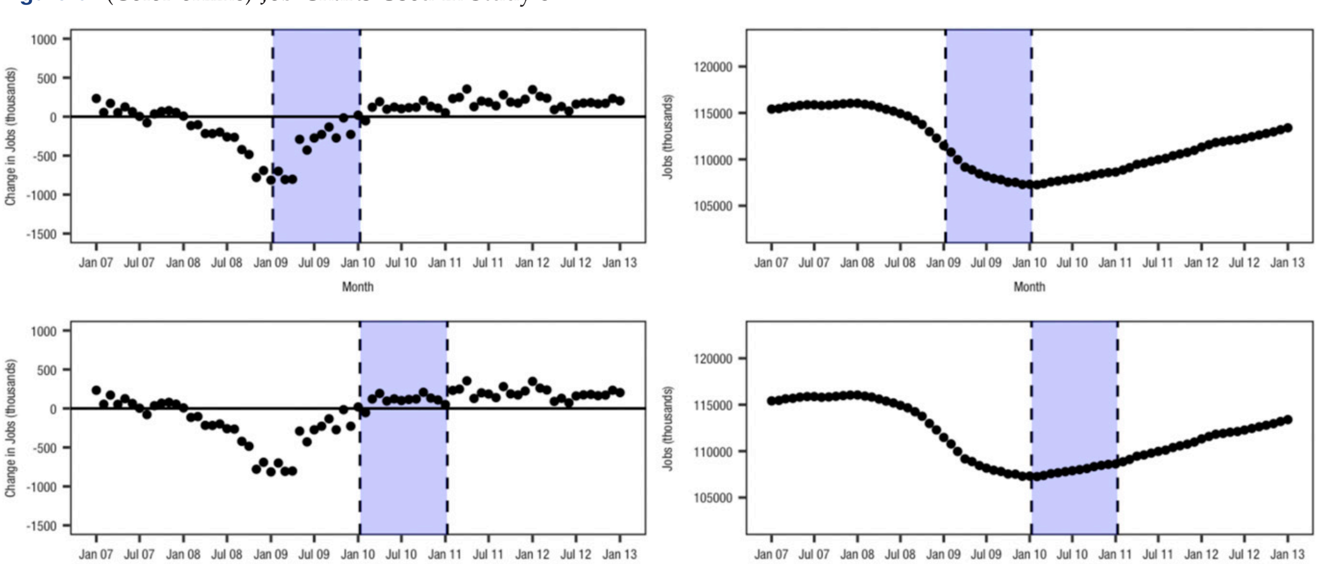
We conducted an ordered logistic regression, regressing each of the two dependent variables (change and attribution) on contrast codes for data presentation (stock = 1, flow = −1), year (2009 = 1, 2010 = −1), and their interaction.

The effect of data presentation on the qualitative evaluation of change in the economy varied depending on whether participants considered 2009 or 2010 (interaction $z = -9.23$, $p < 0.001$). For 2009, the data replicated Studies 1 and 2, such that most participants believed the economy worsened when considering stocks (84% said worsened and 8% said improved), whereas most participants believed the economy improved when considering flows (17% said worsened and 77% said improved; difference: $z = -10.23$, $p < 0.001$). For 2010, when the flow was positive but flat, the difference between conditions reversed. Participants were generally more positive about the economy when considering stocks (14% said worsened and 80% said improved) than when considering flows (2% said worsened and 58% said improved; difference: $z = 2.27$, $p = 0.023$).

The effect of data presentation on evaluations of Obama's effect on the economy also depended on whether participants considered 2009 or 2010 (interaction $z = -6.30$, $p < 0.001$). During 2009, more participants believed Obama had a negative effect than a positive effect when considering stocks (46% said made it worse and 22% said made it better), whereas most participants believed he had a positive effect when considering flows (9% said made it worse and 69% said made it better; $z = -7.52$, $p < 0.001$). During 2010, there was no significant difference in valence between conditions (stocks: 10% said made it worse and 71% said made it better; flows: 2% said made it worse and 57% said made it better; $z = 1.36$, $p = 0.174$), although twice as many participants said Obama had no effect when considering the (constant) flow trend rather than the (increasing) stock trend (41% versus 19%).

Regressing the three importance measures on time span, presentation format, and their interaction yielded only one statistically significant effect: the perceived importance of growth rate was higher for

Figure 6. (Color online) Job Charts Used in Study 3



Note. In contrast to the charts used in Study 2, these charts highlighted either 2009 (Obama's first year in office) or 2010 (his second year in office).

Table 1. Summary of Results from Studies 1–3

	Study 1 (2009)		Study 2 (2009)		Study 3 (2009)		Study 3 (2010)	
	Stock	Flow	Stock	Flow	Stock	Flow	Stock	Flow
Change in the economy								
Better			0.28	0.78	0.08	0.77	0.80	0.58
No change			0.02	0.04	0.09	0.06	0.06	0.39
Worse			0.70	0.18	0.84	0.17	0.14	0.02
Obama's effect on the economy								
Better	0.24	0.66	0.29	0.73	0.22	0.69	0.71	0.57
No change	0.16	0.26	0.23	0.13	0.32	0.22	0.19	0.41
Worse	0.60	0.08	0.49	0.14	0.46	0.09	0.10	0.02

Notes. Cells indicate proportion of participants in each condition who reported a positive, neutral, or negative evaluation for key dependent variables in Studies 1–3. In Studies 1 and 2, these were assessed on seven-point scales (with midpoints explicitly representing no change); in Study 3, they were assessed on three-point scales. Change in the economy was not measured in Study 1. Studies 1 and 2 examined only 2009 (where the stock trend was negative and the flow trend was positive); Study 3 also examined 2010 (where the stock trend was increasing and the flow trend was flat).

participants who evaluated 2010 data (M = 3.95) compared with those who evaluated 2009 data (M = 3.74; $t(397) = -2.33$, $p = 0.020$). The other eight possible effects were not significant (p's > 0.25). Thus, it seems unlikely that the differences in judgments we observe between presentation format conditions are caused by participants consciously inferring which aspects of the data are most diagnostic based on the choice of presentation format. As in Study 2, evaluations of the economy and of Obama's effect increased with political liberalism (p's < 0.015), but political views did not moderate either key effect (p's > 0.40).[12]

Discussion

Study 3 provides further evidence that people make judgments based on salient features of the data as they are presented. Furthermore, these judgments do not seem to be due to explicit reweighting of the importance of different aspects of the data (e.g., level, velocity, acceleration). This study also provides evidence that participants focus their assessments on the specific portion of the graph near the focal event, rather than the entirety of the data. Given the same data, some time spans can lead to more positive evaluations for stocks than flows, whereas other time spans can lead to more positive evaluations for flows than stocks. The full pattern of data across Studies 1, 2, and 3, summarized in Table 1 reveals a consistent story. In subsequent studies, we present participants with artificial data to enable more precise control over stock and flow trends.

Study 4

Studies 1, 2, and 3 examined evaluations of a real data set and found stark differences depending on whether the data were presented as stocks or flows. In Studies 4, 5, and 6, we extend the initial forecasting results of Study 2 to more comprehensively examine how qualitative properties of forecasts (signed changes) depend on different stock and flow trends. In doing so, we use artificial smooth trends to reduce extraneous noise and maximize power.

Method

Four hundred two participants (155 women and 247 men; median age = 30) were recruited from AMT and completed Study 4.[13] We generated nine time series data sets in a full factorial design such that the flow was decreasing, constant, or increasing and the stock on average was decreasing, constant, or increasing.[14] Each pattern represented four years of data, with 49 data points corresponding to either the stock at a given month or the flow between each successive pair of months. The overall patterns can be seen in Figure 7.

To make the data more concrete, one of nine different scenarios was used for each pattern of data (e.g., money in a bank account, employees at a firm, gallons of water in a reservoir). Each scenario had a unique metric and was thus given a unique end value from which the other values were determined by the stock and flow conditions.[15] The scenarios were designed to encompass different domains (money, people, or objects), different subjects (companies, governments, or people), and different magnitudes. More detail on the scenarios is provided in Online Appendix C.

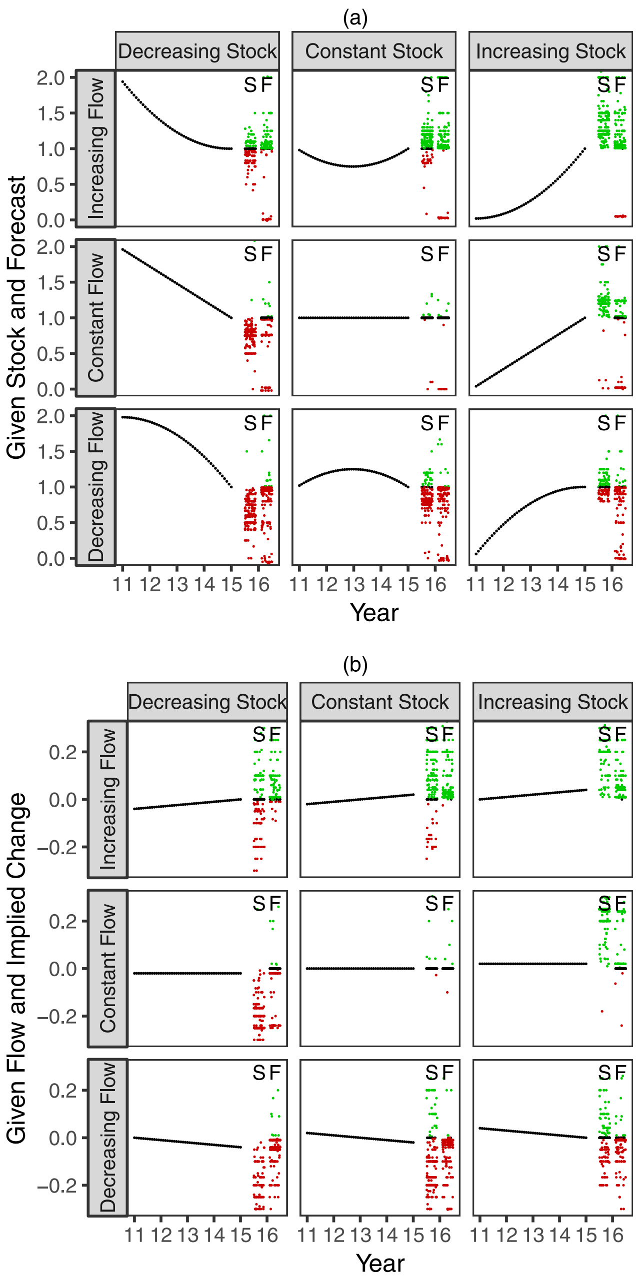
Participants were randomly assigned to a presentation format condition (stock or flow). Participants responded to each of the nine data patterns in a random order, with one of the nine scenarios randomly assigned to each pattern. For each pattern, participants were shown four years of data, as either a stock or flow (depending on condition). Participants were given the (stock) value at the end of the data period (January 1, 2015) to equate the two conditions in terms of information and then asked to forecast the (stock) value exactly one year in the future (January 1, 2016).

After making forecasts for the nine scenarios, participants were asked how well they felt like they understood the data as presented on the charts (−4: very poorly; +4: very well), a question assessing a basic level of understanding of the data (whether they recognized they were being shown stocks or flows), demographic information (self-perceived math ability, sex, age, and education), and any additional comments.

Results

Analysis Strategy. Each of the 402 participants made 9 forecasts, for 3,618 total observations. To combine across scenarios in different units and on different

Figure 7. (Color online) Study 4 Forecasts with Displayed Stock Trend (Panel (a)) and Displayed Flow Trend (Panel (b))



Notes. S, stock condition; F, flow condition. Forecasts in black indicate no forecasted change (1 in panel (a) and 0 in panel (b)); shaded forecasts indicate a change (positive change greater than 1 in (a) and greater than 0 in (b); negative change less than 1 in (a) and less than 0 in (b)). Panels contain the same data on different scales, where panel (b) highlights a more limited range of responses.

scales, we divided forecasts by the ending stock of the observed trend. Thus, for all domains, 1 represents a forecast equal to the ending stock of the observed sequence, 2 represents a forecast equal to twice the ending stock, etc.[16] We first present the data visually and describe the overall patterns. We then present formal statistical tests.

Depicting the Data. Figure 7 presents the data twice, with different scales. Figure 7(a) depicts the observed stock sequence (only observed by those in the stock condition), along with responses from *both* conditions, depicted separately for each of nine sequences: three stock trends (decreasing, constant, or increasing) × three flow trends (decreasing, constant, or increasing). The 82 responses (2%) that were below 0 or above 2 after rescaling are excluded from this panel. Forecasts representing no change are black; forecasts representing changes are shaded accordingly.

Figure 7(b) depicts the same data in a different way, presenting the observed flow sequence (only observed by those in the flow condition), along with responses from *both* conditions depicted for each of nine sequences. Unlike Figure 7(a), the y axis in Figure 7(b) represents the implied change from the end of the observed sequence, so a 0 in Figure 7(b) is equivalent to a 1 in Figure 7(a). In Figure 7(b), the y axis displays a narrower range to highlight the flow patterns better, meaning more data are excluded.

The prevailing patterns in the data are stark. The typical forecast in the stock condition follows a linear extension of the ending stock trend. The typical forecast in the flow condition follows a linear extension of the ending flow trend. This yields qualitatively different forecasts for some of the scenarios in terms of whether the average participant expects the stock to increase, decrease, or stay the same over the following year. The proportion of participants making each type of forecast (represented by different shades in Figure 7) is shown in Table 2.

Qualitatively Different Inferences from Stocks vs. Flows. Visual inspection of Figure 7(a) reveals that the mean may not be a good summary statistic for these data and that assumptions of normal errors and homoscedastic variance may be problematic. To formally test the effect of stock versus flow presentation, we analyze the qualitative outcome (decrease, no change, or increase) as a function of condition in an ordered logistic regression.[17] Our primary interest is with the effect of data presentation (stock versus flow) separately for each trend, so we analyze each pattern separately for expositional ease.[18] In all six cases where the stock trend and flow trend were qualitatively different (i.e., one was more positive or less negative than the other), forecasts followed the salient trend. That is, when the stock trend was more positive than the flow trend (increasing stock with a decreasing flow, constant stock with a decreasing flow, or increasing stock with a constant flow), presenting the data as stocks rather than flows led to forecasts that were significantly more positive (see Table 2; each $p < 0.001$). Similarly, when the stock trend was more negative than the flow trend (decreasing stock with an increasing flow, decreasing

Table 2. Study 4 Forecast Trends

		Decreasing stock		Constant stock		Increasing stock	
	Forecast	Stock	Flow	Stock	Flow	Stock	Flow
Increasing flow	Increase	0.23	0.81	0.79	0.90	1.00	0.91
	No change	0.47	0.05	0.08	0.01	0.00	0.01
	Decrease	0.30	0.15	0.14	0.09	0.00	0.08
Constant flow	Increase	0.01	0.05	0.04	0.04	0.98	0.33
	No change	0.00	0.55	0.94	0.90	0.00	0.53
	Decrease	0.99	0.40	0.02	0.06	0.02	0.14
Decreasing flow	Increase	0.01	0.10	0.15	0.05	0.32	0.08
	No change	0.00	0.00	0.05	0.01	0.41	0.10
	Decrease	0.99	0.90	0.80	0.94	0.28	0.82

Note. Cells indicate proportion of participants in each condition (stock, flow) who reported a forecast that represented an increase, no change, or a decrease relative to the ending stock for each of nine combinations of stock trends (decreasing, constant, increasing) and flow trends (decreasing, constant, increasing).

stock with a constant flow, or constant stock with an increasing flow), presenting the data as stocks rather than flows led to forecasts that were significantly more negative (see Table 2; each $p < 0.005$). In the other three cases, the stock trend was directionally equivalent to the flow trend (increasing stock with an increasing flow, constant stock with a constant flow, or decreasing stock with a decreasing flow). In these cases, there is little variation in responses between conditions, as nearly everyone made a forecast that followed the predominant trend, which did not differ between conditions.[19]

Whereas the previous analyses examined sensitivity to a given pattern across conditions, it is also possible to examine sensitivity to different stock and flow patterns within condition. Supplementary analyses indicate that participants presented with stock trends are more sensitive to differences among stock trends than differences among flow trends, whereas participants presented with flow trends are more sensitive to differences among flow trends than differences among stock trends. Details of these analyses are presented in Online Appendix E.

Subjective Understanding. Participants reported understanding the data better in the stock condition ($M = +2.76$, $SD = 1.29$) than in the flow condition ($M = +0.67$, $SD = 2.29$, $t(400) = 11.36$, $p < 0.001$). This is potentially problematic, as this difference in understanding could underlie the differences in judgment we observe. However, when we enter subjective understanding as a moderator and estimate the simple effects for the maximum value of subjective understanding (following the approach outlined by Spiller et al. 2013), all results remain consistent, suggesting the difference in subjective understanding does not account for the difference in forecasts. We further address this possibility in the subsequent studies.

Discussion

By systematically and independently varying the stock trend and the flow trend, we find further evidence that the choice of presenting data as stocks or flows has systematic and substantial effects on the judgments people make. The results across the different combinations of stock and flow trends are consistent with participants using linear extrapolation from the end of the observed sequence (Wagenaar and Sagaria 1975, McKenzie and Liersch 2011). Because the trends differ depending on whether the data are presented as stocks or flows, these linear extrapolations of (potentially) nonlinear trends lead to systematically and qualitatively different forecasts, differing not only in degree but also in type (i.e., forecasted increase versus decrease).

Study 5

The previous studies demonstrate that, in specific situations, presenting data as a stock or a flow can lead to substantial differences in judgments (evaluations and forecasts). We contend these differences are caused by people relying on the salient features of the data as presented. However, another potential explanation is that people incorrectly, and superficially, interpret flow data as stock data. The explanation is consistent with prior research showing people have considerable difficulty in understanding what stock and flow data represent (Booth Sweeney and Sterman 2000, Cronin et al. 2009). In Study 5, we examine two variants on this explanation by explicitly assessing participants' ability to read the data, much as in prior research. By examining the subset of participants who can accurately read the data, we assess whether the results are driven by a *lack of ability*. By examining whether the effect on forecasts differs depending on whether participants describe the data before or after making a forecast, we also assess whether the results are driven by a *lack of consideration*.

Method

Six hundred five participants (301 women and 302 men; median age = 33) were recruited from AMT and completed Study 5.[20] This study uses a similar design as in Study 4, so we detail only the changes below.

First, in addition to the stock and flow conditions, this study also included a third combined stock and flow condition in which participants saw both the stock data and the flow data presented side by side.

Second, the stimuli were simplified to present data at the yearly (rather than monthly) level. Given that we asked participants to make a prediction one year out, providing monthly data as we did in Study 4 added one additional level of complexity beyond the factors of interest, leading some participants in the flow condition to make forecasts that appeared to be based on a single month's flow.[21]

Third, we only included four conditions from Study 4 for which the ending trends of the stock and flow graphs diverged (increasing stock with a decreasing flow, increasing stock with a constant flow, decreasing stock with a constant flow, and decreasing stock with an increasing flow) to reduce the load on participants and focus on cases that make qualitatively distinct predictions. Accordingly, we also reduced the number of scenario frames from nine to four.

Fourth, and most important, we included an additional set of measures that assess whether each participant could correctly read and interpret each data set as presented. Participants in each condition were asked to state the final stock level (which was explicitly given in all conditions), the penultimate stock level (shown on the graph in the stock and combined conditions but not the flow condition), and the change between the last two stock levels (shown on the graph in the flow and combined conditions but not the stock condition). In the stock condition, accurate answers required reading the last two stocks and calculating the flow. In the flow condition, accurate answers required reading the stock, reading the last flow, and calculating the second-to-last stock. In the combined condition, accurate answers did not require any calculations. To allow us to assess whether explicit consideration of the graph's meaning contributes to differences in forecasts between conditions, we varied the task order such that some participants first described the data and then made forecasts and others first made forecasts and then described the data.

Finally, we dropped two measures from Study 4: the basic comprehension measure and the measure of self-perceived math ability.

Results

Analysis Strategy. Each of the 605 participants reported descriptions and forecasts for each of the four patterns, for a total of 2,420 observations. Unlike Study 4, there was not a systematic cluster of outliers after transforming the data.

We follow a similar analysis approach to that of Study 4, with a few changes. The key changes are that we focus our analysis on the subset of participants who could accurately describe the data, and after considering the basic analysis, we consider the effect of order. Here, we define accuracy in describing the data as (1) correctly reporting the ending stock, (2) correctly reporting that the second-to-last stock plus the change equals the ending stock, and (3) reporting the second-to-last value and the change within 5% of the true ending stock value.

Assessing Whether Participants Understood the Data. We describe the pattern of accuracy here and reserve more thorough analytic details for Online Appendix F. Participants were less likely to accurately describe the data when presented as flows rather than stocks (39% versus 48%), with the combined condition resulting in similar accuracy to stocks-only condition (54%). Increasing stock trends were more likely to be accurately described than decreasing stock trends (64% versus 30%, although see the caveat regarding the magnitude of effect in Online Appendix F), and constant flow trends were more likely to be accurately described than varying flow trends (53% versus 41%).

Depicting the Data. Figure 8 presents the data twice, analogous to Figure 7. These data only include participants who accurately described the data. The prevailing pattern is clear. When the flow was constant, there was little variation across participants; the vast majority of all participants in all conditions extrapolated that the change in the next period would be the same as the change in the last period. By contrast, there was more variability when the flow varied over time. In these cases, the stock prediction followed the stock trend, whereas the flow prediction was more evenly spread with a median of no change. The proportion of informed participants making each kind of qualitative forecast (indicated by different shades in Figure 8) is given in Table 3 for both the accurate subsample and the whole sample.

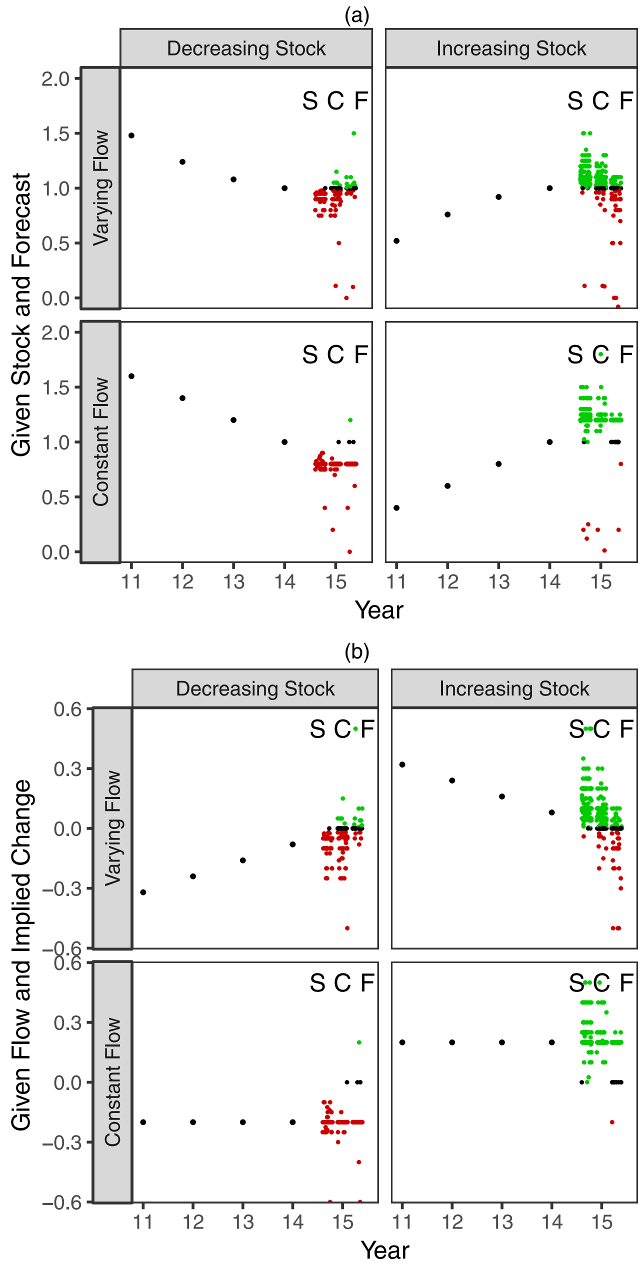
Qualitatively Different Forecasts from Stocks vs. Flows. As in Study 4, we analyze the signed outcome (decrease, no change, or increase) as a function of condition in an ordered logistic regression; the text reports the statistical tests as the effect sizes are more directly given by the proportions in Table 3 than coefficients. As there were three conditions, we created a set of linear and quadratic contrast codes: linear contrast-coded stock = 1, combined = 0, flow = −1; and quadratic contrast-coded stock = −1, combined = 2, flow = −1. The linear contrast represents the comparison of stock versus flow; the quadratic contrast represents the comparison of combined versus the midpoint of stock and flow. We again examine the different sequences independently. We focus on the subset of responses reflecting an accurate understanding. Results for the full data set are generally stronger than those for the accurate subsample, as it is partially driven by misunderstanding the data. Both sets of proportions are given in Table 3.

For an increasing stock with a decreasing flow, forecasts in the stock condition almost exclusively indicated an increase, whereas those in the flow condition were more evenly divided. The difference was significant (z = 8.33, p < 0.001), and the combined condition did not differ from the midpoint ($z = -1.31$, $p = 0.19$). For an increasing stock with a constant flow, nearly all forecasts indicated an increase, although there was a significant difference between the stock and flow conditions (z = 2.37, p = 0.018), with the combined condition lying marginally significantly above the average of stock and flow (z = 1.91, p = 0.056) such that it differed from the flow condition (z = −2.81, p = 0.005) but not the stock condition (z = −0.89, p = 0.38). For a decreasing stock with an increasing flow, forecasts in the stock condition almost exclusively indicated a decrease, whereas those in the flow condition were more evenly divided. The difference was significant (z = −5.67, p < 0.001), and again, the combined condition did not differ from the midpoint (z = 1.04, p = 0.30). For a decreasing stock trend with a constant flow, nearly all forecasts indicated a decrease; the model did not converge.

Follow-up analyses allowing for interactions with whether participants described the data first or made forecasts from the data first did not significantly qualify any of these results. The key differences (when stock and flow trends were of opposite signs) were statistically significant even when participants accurately described the data before making a forecast (p's < 0.001), suggesting that the effect persists in the presence of active and accurate consideration of the meaning of the underlying data.

Subjective Understanding. Across all participants (including incorrect responses), subjective understanding varied between conditions ($F(2, 602)$ = 34.54, p < 0.001) such that it was higher for the stock (M = +2.55, SD = 1.65) than for combined (M = +2.12, SD = 1.60; $t(602) = 2.316$, $p = 0.021$) and higher for combined than for the flow (M = +1.05, SD = 2.26; $t(602)$ = 5.798, p < 0.001). We repeated the forecast analyses on the subset of responses with accurate descriptions while also allowing for moderation by subjective understanding. Even at the highest value of subjective understanding, the estimated simple effects between the stock and flow conditions were consistent, substantively and significantly.

Figure 8. (Color online) Study 5 Forecasts with Displayed Stock Trend (Panel (a)) and Displayed Flow Trend (Panel (b))



Notes. S, stock condition; C, combined condition; F, flow condition. Black points indicate no change (1 in panel (a) and 0 in panel (b)); shaded points indicate a change (positive change greater than 1 in panel (a) and greater than 0 in panel (b); negative change less than 1 in panel (a) and less than 0 in panel (b)). Panels contain the same data on different scales. Only forecasts associated with accurate descriptions of the data are included.

Discussion

Study 5 replicates the key results from Study 4 and provides four key extensions. First, we find that, consistent with prior research, people do have difficulty

Table 3. Study 5 Forecast Trends

		Decreasing stock			Increasing stock		
		Stock	Combined	Flow	Stock	Combined	Flow
Accurate responses							
Varying flow	Increase	0.04	0.08	0.21	0.97	0.68	0.25
	No change	0.02	0.34	0.60	0.01	0.24	0.34
	Decrease	0.94	0.58	0.19	0.01	0.08	0.41
Constant flow	Increase	0.00	0.01	0.03	0.97	0.99	0.90
	No change	0.00	0.01	0.03	0.01	0.00	0.09
	Decrease	1.00	0.98	0.95	0.02	0.01	0.02
All responses							
Varying flow	Increase	0.03	0.08	0.53	0.96	0.72	0.15
	No change	0.05	0.22	0.30	0.02	0.20	0.24
	Decrease	0.92	0.70	0.16	0.02	0.08	0.61
Constant flow	Increase	0.02	0.02	0.05	0.98	0.98	0.63
	No change	0.00	0.00	0.26	0.00	0.00	0.32
	Decrease	0.98	0.98	0.70	0.02	0.01	0.05

Notes. Cells indicate the proportion of participants in each condition (stock, combined, flow) who reported a forecast that represented an increase, no change, or a decrease relative to the ending stock. When the flow varied, its trend was opposite the stock trend.

assessing stocks from flows, especially when that flow varies over time or is negative. Second, we find that whereas the inability to accurately read the data does contribute to the effect (illustrated by the comparison of proportions in the upper and lower portions of Table 3), it does not fully account for the effect, given that the effect persists strongly even when participants could accurately describe the data. Third, we find that these effects are not due to a lack of consideration. One possible alternative was that people are able to describe the data but only experience an "aha" moment once they are prompted to do so. We find that even when people accurately describe the data before making a forecast, their forecasts are still affected by presentation format. Fourth, and finally, we find that it is not merely that one format is sufficient and the other is merely an imperfect substitute. If that were the case, we would expect forecasts in the combined presentation format to be equivalent to either the stock or the flow presentation format. Instead, the combined format results lie in between the stock results and the flow results, suggesting participants may use both when making forecasts and further supporting our previous assertion that there may be no neutral way to present time-series data.

Study 6

Thus far, we have focused on graphical displays. Although this is a common format in which people encounter time-series data, it is, of course, not the only one. Some perceptual effects of trends over time are limited to graphical presentation (e.g., Duclos 2015), whereas others are robust to presentation formats (e.g., graph, tabular, text; Cronin et al. 2009). In Study 6, we examine whether our findings extend to tabular formats or, alternatively, whether they rely primarily on visual extrapolation or difficulty in reading precise values from the graph axes.

Method

Four hundred and one participants (194 women and 206 men; median age = 32) were recruited from AMT and completed Study 6.[22] Participants responded to only one pattern of data that led to opposing inferences in previous studies: increasing stock with a decreasing flow. Participants first provided a description of the data (same measures as Study 5) and then made a forecast (same measures as Studies 4 and 5).

As in previous studies, we randomly assigned participants to view the data as either a stock or a flow. However, in Study 6, we included a second, orthogonal manipulation: whether the data were presented as a graph (as in previous studies) or as a table depicting either the stock or flow for each year numerically. After completing the description and forecasting measures, participants rated their subjective understanding of the information and provided basic demographic data as in Study 5.

Results

Assessing Whether Participants Understood the Data. Using a logistic regression analyzing accurate descriptions using the same criteria as in Study 5 as a function of trend type (contrast coded stock versus flow), presentation type (contrast coded table versus graph), and their interaction, we find that the odds of an accurate description are again greater when data are presented as stocks rather than flows ($z = 6.87$, $p < 0.001$). We also find that the odds of an accurate description are greater when data are presented as a table rather than a graph ($z = 7.45$, $p < 0.001$); this may be due to either comprehension or precision. The ratio of the odds of accurately describing the data given flow versus stock did not depend on table versus graph format (interaction: $z = -0.91$, $p = 0.36$). Cell proportions are given in Table 4.

Table 4. Study 6 Accuracy

	Stock	Flow
Graph	0.65	0.20
Table	0.91	0.68

Note. Cells indicate proportion of participants in each condition who accurately described the data.

Qualitatively Different Forecasts from Stocks vs. Flows. Forecasts associated with accurate descriptions are depicted in Figure 9. An ordinal regression on the informed sample revealed that for this sequence (increasing stock with a decreasing flow), presenting the data as stocks (versus flows) induced more positive forecasts ($z = 6.50$, $p < 0.001$), and this did not vary depending on whether the data were presented as graphs ($z = 4.24$, $p < 0.001$) or text ($z = 5.80$, $p < 0.001$; interaction $z = -0.89$, $p = 0.37$). There was no main effect of presentation format ($z = -1.58$, $p = 0.11$). Cell proportions are given in Table 5.[23]

Subjective Understanding. This study also included a report of subjective understanding. In the full sample, subjective understanding was higher for the table than the graph ($t(397) = 8.16$, $p < 0.001$) and higher for the stock than the flow ($t(397) = 8.27$, $p < 0.001$), and the difference between the stock and flow was greater for the graph ($M_{stock} = +2.41$, $M_{flow} = +0.04$, $t(397) = 8.89$, $p < 0.001$) than for the table ($M_{stock} = +3.09$, $M_{flow} = +2.39$, $t(397) = 2.71$, $p = 0.007$; interaction $t(397) = 4.50$, $p < 0.001$). Even among the subset of participants who accurately described the data, subjective understanding was higher for the table than for the graph ($t(246) = 8.68$, $p < 0.001$) and higher for the stock than the flow ($t(246) = 7.58$, $p < 0.001$), and the difference between the stock and flow was not only greater for the graph ($M_{stock} = +2.67$, $M_{flow} = -0.61$, $t(246) = 8.14$, $p < 0.001$), it was small and nonsignificant for the table ($M_{stock} = +3.20$, $M_{flow} = +2.92$, $t(246) = 1.15$, $p = 0.25$; interaction $t(246) = 6.41$, $p < 0.001$). Among those who accurately described the data, subjective understanding did not moderate the effects of the stock versus flow presentation format, table versus graph format, or their interaction (p's > 0.1). The estimated simple effect at the highest level of subjective understanding remained significant for a table format ($z = 4.94$, $p < 0.001$), although it was not significant for a graph format ($z = 0.81$, $p = 0.42$), in large part because the smaller cell size and lower subjective understanding led to larger standard errors (0.61 for the graph and 0.23 for the table).

Discussion

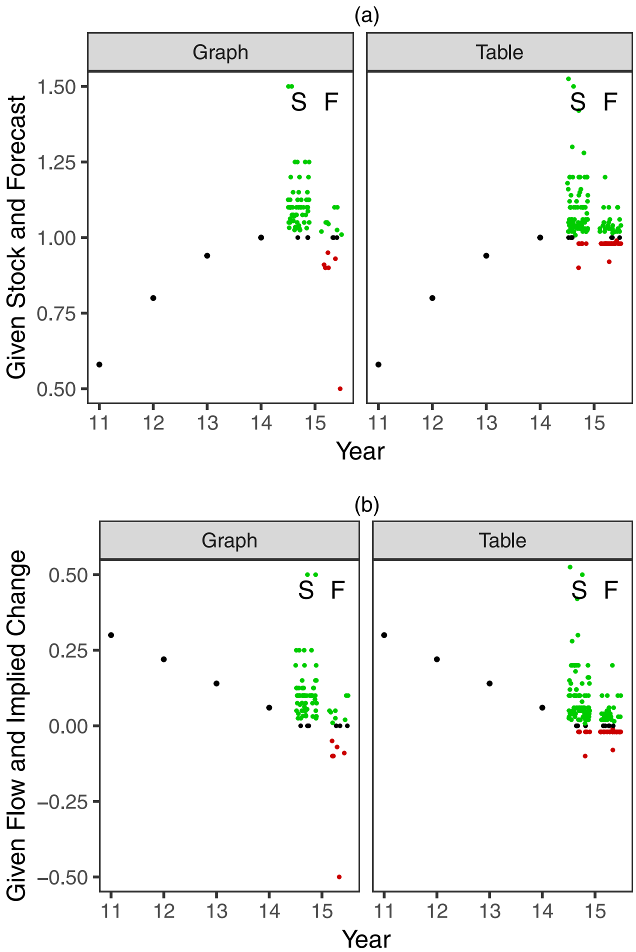
In Study 6, we find that the results observed throughout extend to a case in which the results are presented numerically rather than graphically. In this case, not only was the effect as strong but participants were also more likely to accurately describe the data, providing further evidence that the differences in forecasts are not primarily caused by inability to accurately read the data. Furthermore, as all participants described the data prior to making a forecast, a lack of consideration is unlikely to account for the effect.

Because the comprehension rates were high and the effect remained strong, the tabular conditions also allow us to provide evidence against an alternative explanation: selection differences between conditions causing the observed effect. The effect of interest on the reduced sample is the effect of stock versus flow among people who can accurately describe the data presented to them. It is possible that some participants in the flow condition who accurately described the data could not have accurately described the stock data, or that some participants in the stock condition who accurately described the data could not have accurately described the flow data (especially because the accuracy rates varied across conditions). Is it possible that there is no effect of presentation format as stock or flow among those who could accurately describe both data sets, and we are merely detecting differential selection? No. Considering just the tabular conditions (where the accuracy rates were relatively high), if we assume that everyone who did not accurately describe the flow data would have reported an increasing forecast and everyone who did not accurately describe the stock data would have reported a decreasing forecast (thus testing the minimum possible difference between conditions, a conservative test), the difference remains significant in the same direction ($z = 2.72$, $p = 0.007$).

General Discussion

At a time when the amount of digital data in the world is doubling every two years (IDC 2014), it is imperative to distill information to the simplest, most straightforward means possible in order to facilitate interpretation and action for both managers and consumers. The comprehension of accumulation is one input into performance in dynamic decisions that involve feedback between the decision maker and the environment (e.g., Sterman 1987, 1989; Paich and Sterman 1993). The present investigation underscores the fact that data are never presented in a vacuum, transmitting information in a purely abstracted, neutral

Figure 9. (Color online) Study 6 Forecasts with Displayed Stock Trend (Panel (a)) and Displayed Flow Trend (Panel (b))



Notes. S, stock condition; F, flow condition. Black points indicate no change (1 in panel (a) and 0 in panel (b)); shaded points indicate a change. Panels contain the same data on different scales. Only forecasts associated with accurate descriptions of the data are included. Participants in the table condition did not see the past data depicted graphically.

form. Rather, to serve a communicative function, information must take some particular format, which includes the choice of presenting time-series data as a stock or as a flow. The present investigation suggests this choice can matter a great deal, as one presentation format may lead the viewer to a qualitatively different conclusion than the other.

We document inconsistencies in judgments arising from stock versus flow presentations across many domains and patterns of data. Perhaps most strikingly, using real jobs data, we show that people can draw opposing inferences about President Obama's impact on the U.S. economy: people believed he had a positive impact in his first year when viewing the data as a flow but a negative impact when viewing the data as a stock (Studies 1, 2, and 3). We find this example particularly compelling because the result occurs in a consequential domain where people's prior beliefs are likely strong (e.g., predisposition to view President Obama's actions favorably or unfavorably).

Systematic inspection of different data patterns (Study 4) reveals that the greatest inconsistencies in judgment emerge when the stock and flow presentations yield opposing ending trends. For example, when presented with a flow pattern that ends in an upward trend—even when the net flow is still negative—people typically make optimistic forecasts. However, when presented with the implied stock pattern, people's forecasts tend toward greater pessimism. This happens even after people have correctly read and interpreted the data (Studies 5 and 6) and does not appear to reflect some differential weighting of the importance of flow versus stock information (Study 3). In fact, when given both types of information, judgments tended toward the average of those from the two presentation formats in isolation, suggesting that, in aggregate, people find both types of information informative even when their perceived implications were opposed (Study 5).

In all but one of the experiments, we asked participants to appraise the data visually, as a graph of either stocks or flows. However, we still observed inconsistencies when the data were presented numerically (Study 6). This suggests that even nonvisual means may be sufficient to engender a difference between evaluation of data presented as a stock or as a flow (consistent with the potential to evoke trends using nongraphical means; Cronin et al. 2009, Maglio and Polman 2016; cf. Larkin and Simon 1987).

Although our results show that the choice of presenting data as a stock or a flow can have a substantive impact on judgments, we remain agnostic as to whether one presentation format reflects a more faithful portrayal (e.g., of the economy) or yields more

Table 5. Study 6 Forecast Trends

		Stock	Flow
Accurate responses			
Graph	Increase	0.94	0.44
	No change	0.05	0.17
	Decrease	0.02	0.39
Table	Increase	0.86	0.42
	No change	0.04	0.08
	Decrease	0.10	0.50
All responses			
Graph	Increase	0.95	0.23
	No change	0.04	0.08
	Decrease	0.01	0.70
Table	Increase	0.86	0.42
	No change	0.04	0.05
	Decrease	0.10	0.53

Note. Cells indicate proportion of participants in each condition who reported a forecast that represented an increase, no change, or a decrease relative to the ending stock.

accurate judgments than another (cf. Larrick and Soll, 2008, de Langhe and Puntoni 2016). Nevertheless, we note that this is a variable that could be used by marketers or other choice architects to influence opinions and decisions (Thaler and Sunstein 2008). In fact, a version of the "flow" graph used in Study 1—which we find leads people toward relatively optimistic assessments of economic recovery—was used heavily in Democratic messaging during President Obama's reelection campaign. By the same token, perhaps presenting a monthly bank statement that reflects not an increasing stock of savings but a decelerating rate of wealth accumulation would cause investors to see their financial planning as off-track. These potentially persuasive pursuits may go undetected, as normative expectations of communication lead people to infer that communicators situate the information that they share within the most important or relevant frame (Grice 1975). In this sense, the choice of presenting data as a stock or a flow might represent another avenue for "how to lie with statistics" (Huff 1954).

With data comes the need to comprehend and act on those data. With the rising tide of big data has come the compulsion to integrate, analyze, and share information at an ever-accelerating rate. Heralding this call have been celebrity statisticians (e.g., Nate Silver), courses on data presentation (e.g., those offered by Edward Tufte), and software to facilitate the translation from facts to figures. These tools and tacticians often emphasize the presentation of information in the most efficient manner possible. What might get lost in this push to present is that any format creates and operates within a specific context, instantiating a unique set of expectations and patterns in the eyes of observers that create the potential for systematic shifts in interpretation. Our work gives pause to show that the *inputs* to the presentation process—specifically, time-series data presented as stocks and flows—can distort how people view the data and thus merit consideration in their own right.

Acknowledgments

The authors thank Ryan Hamilton, Jeff Parker, Sara Dommer, Hal Hershfield; seminar participants at Wharton, the University of Colorado, and the UCLA Behavioral Decision Making lab; and the editor, associate editor, and three anonymous reviewers for constructive comments and valuable feedback.

Endnotes

[1] Note that the level of the stock is lost in the translation. To make the two types of information completely equivalent, a value of the stock must be provided with the flow (e.g., the stock level at the first period).

[2] For example, in a stock presentation, the velocity (the first derivative of stock) is salient. By contrast, flow presentation keeps the velocity information but makes the acceleration (the second derivative of stock) more salient.

[3] Note that whereas people struggle to translate between stocks and flows, this ability is not required to understand data in a given format. For example, in the "department store task" described above, people shown flow data could accurately interpret questions about flows, even though many of these people could not answer the stock (accumulation) questions.

[4] Seven additional participants consented to participate but did not complete the study.

[5] This was calculated as months since January 2007 divided by 12, such that January 2007 was coded as 0, July 2007 was coded as 0.5, December 2007 was coded as 0.92, and so forth.

[6] As in Study 1, an additional eight partial responses were excluded from analysis. Number of women plus number of men does not always equal the total sample due to non-responses and non-binary responses.

[7] One data point is missing for this measure because of a coding error.

[8] Politics did not vary by condition ($p > 0.40$).

[9] Within conditions, evaluations and forecasts were consistent such that lower forecasts were associated with more negative evaluations.

[10] For completeness and transparency, three additional studies are included in the Online Appendix B. Similar to Study 3, these additional studies focus on a different time span (March 2010 to March 2011) compared with Studies 1 and 2, and they show results qualitatively consistent with those of the 2010 condition in Study 3.

[11] An additional 36 participants consented to participate but did not complete the study.

[12] Politics did not vary by condition (p's > 0.20).

[13] An additional 36 partial responses were excluded from analysis.

[14] Notice that an increasing flow with a stock that is constant on average means a U-shaped stock, whereas a decreasing flow with a stock that is constant on average means an inverted U-shaped stock.

[15] Change in yearly flow corresponded to either −1%, 0%, or +1% of the scenario's end value with the average monthly flow equal to −2%, 0%, or +2% of the scenario's end value.

[16] The data included a collection of systematic outliers. For one (and only one) scenario, results were elicited "in millions," with the intent that a response of "100 million" would be entered as "100." In all other cases, results were elicited as unqualified numeric values. For that scenario (and only that scenario), approximately 20% of responses (86 out of 402) fell between 500,000 and 2,000,000 *after* rescaling. No other scenario had any responses in that range (more than 97% of responses fell between 0 and 2). This indicates that a large proportion of responses for that scenario were reported as unqualified numbers rather than in millions (e.g., as 100,000,000 rather than 100). For that scenario, we rescaled all transformed values between 500,000 and 2,000,000 (inclusive) by dividing by 1,000,000.

[17] Comparisons with parametric results for Studies 4, 5, and 6 are given in Online Appendix D.

[18] All conclusions remain the same when using a single model with either clustered standard errors or random effects to account for nonindependence.

[19] Of six cells, representing three cases for each of two conditions, the smallest mode was 90%. The coefficient was significant when both trends were decreasing, was not significant when both trends were increasing, and did not converge when both trends were increasing.

[20] An additional 142 partial responses were excluded.

[21] Note that although this may be a concern with interpreting differences in magnitude in the previous study, it should not affect

differences in sign (positive versus negative), which was the focus of our analysis.

[22] An additional 52 partial responses were recorded.

[23] When using all of the data, including inaccurate responses, there was an interaction ($z = 3.16$, $p = 0.002$), but the simple effect of stock versus flow was large and significant whether the data were presented as a graph ($z = 8.15$, $p < 0.001$) or as a table ($z = 6.51$, $p < 0.001$). This interaction can largely be attributed to the fact that the effect is larger among those who are inaccurate, and participants were more inaccurate with the graph than the table.

References

Andreassen PB (1988) Explaining the price-volume relationship: The difference between price changes and changing prices. *Organ. Behav. Human Decision Processes* 41(3):371–389.

Andreassen PB, Kraus SJ (1990) Judgmental extrapolation and the salience of change. *J. Forecasting* 9(4):347–372.

Booth Sweeney L, Sterman JD (2000) Bathtub dynamics: Initial results of a systems thinking inventory. *System Dynam. Rev.* 16(4):249–286.

Bureau of Labor Statistics (2017) Labor force statistics from the current population survey. Accessed June 15, 2017, https://www.bls.gov/cps.

Burson KA, Larrick RP, Lynch JG Jr (2009) Six of one, half dozen of the other: Expanding and contracting numerical dimensions produces preference reversals. *Psych. Sci.* 20(9):1074–1078.

Camilleri AR, Larrick RP (2014) Metric and scale design as choice architecture tools. *J. Public Policy Marketing* 33(1):108–125.

Cleveland WS, McGill R (1984) Graphical perception: Theory, experimentation, and application to the development of graphical methods. *J. Amer. Statist. Assoc.* 79(38):531–554.

Cronin MA, Gonzalez C, Sterman JD (2009) Why don't well-educated adults understand accumulation? A challenge to researchers, educators, and citizens. *Organ. Behav. Human Decision Processes* 108(1):116–130.

de Langhe B, Puntoni S (2016) Productivity metrics and consumer's misunderstanding of time savings. *J. Marketing Res.* 53(3):396–406.

de Langhe B, Puntoni S, Larrick RP (2017) Linear thinking in a nonlinear world. *Harvard Bus. Rev.* 95(3):130–139.

Duclos R (2015) The psychology of investment behavior: (De)biasing financial decision-making one graph at a time. *J. Consumer Psych.* 25(2):317–325.

Freyd JJ, Finke RA (1984) Representational momentum. *J. Experiment. Psych. Learn. Memory Cognition* 10(1):126–132.

Grice HP (1975) Logic and conversation. Cole P, Morgan JL, eds. *Syntax and Semantics, Volume 3: Speech Acts* (Academic Press, New York), 41–58.

Huff D (1954) *How to Lie with Statistics* (W.W. Norton & Company, New York).

IDC (2014) The digital universe of opportunities: Rich data and the increasing value of the internet of things. Accessed July 5, 2017, https://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm.

Kahneman D (2011) *Thinking, Fast and Slow* (Farrar, Straus and Giroux, New York).

Larkin JH, Simon HA (1987) When a diagram is (sometimes) worth ten thousand words. *Cognitive Sci.* 11(1):65–99.

Larrick RP, Soll JB (2008) The MPG illusion. *Science* 320(5883): 1593–1594.

Lurie NH, Mason CH (2007) Visual representation: Implications for decision making. *J. Marketing* 71(1):160–177.

Maglio SJ, Polman E (2016) Revising probability estimates: Why increasing likelihood means increasing impact. *J. Personality Soc. Psych.* 111(2):141–158.

Maglio SJ, Trope Y (2011) Scale and construal: How larger measurement units shrink length estimates and expand mental ho rizons. *Psych. Bull. Rev.* 18(1):165–170.

McKenzie CRM, Liersch MJ (2011) Misunderstanding savings growth: Implications for retirement savings behavior. *J. Marketing Res.* 48(Special Issue):S1–S13.

Pacini R, Epstein S (1999) The relation of rational and experiential information processing styles to personality, basic beliefs, and ratio-bias phenomenon. *J. Personality Soc. Psych.* 76(6):972–987.

Paich M, Sterman JD (1993) Boom, bust, and failures to learn in experimental markets. *Management Sci.* 39(12):1439–1458.

Pelham BW, Sumarta TT, Myaskovsky L (1994) The easy path from many to much: The numerosity heuristic. *Cognitive Psych.* 26(2): 103–133.

Raghubir P, Srivastava J (2002) Effect of face value on product valuation in foreign currencies. *J. Consumer Res.* 29(3):335–347.

Reyna VF, Brainerd CJ (2008) Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. *Learn. Individual Differences* 18(1):89–107.

Slovic P (1972) *From Shakespeare to Simon: Speculations—And Some Evidence—About Man's Ability to Process Information* (Oregon Research Institute, Eugene).

Spiller SA, Fitzsimons GJ, Lynch JG Jr, McClelland GH (2013) Spotlights, floodlights, and the magic number zero: Simple effects tests in moderated regression. *J. Marketing Res.* 50(2): 277–288.

Sterman JD (1987) Testing behavioral simulation models by direct experiment. *Management Sci.* 33(12):1572–1592.

Sterman JD (1989) Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment. *Management Sci.* 35(3):321–339.

Sterman JD (2002) All models are wrong: Reflections on becoming a systems scientist. *System Dynam. Rev.* 18(4):501–531.

Sterman JD, Booth Sweeney L (2002) Cloudy skies: Assessing public understanding of global warming. *System Dynam. Rev.* 18(2): 207–240.

Sterman JD, Booth Sweeney L (2007) Understanding public complacency about climate change: Adults' mental models of climate change violate conservation of matter. *Climate Change* 80(3–4): 213–238.

Thaler RH, Sunstein CR (2008) *Nudge: Improving Decisions About Health, Wealth, and Happiness* (Yale University Press, New Haven, CT).

Thomson KS, Oppenheimer DM (2016) Cognitive reflection and non-linear reasoning errors. Presentation, International Conference on Thinking 2016, August 4, Brown University, Providence, RI.

Tufte ER (2001) *The Visual Display of Quantitative Information* (Graphics Press LLC, Cheshire, CT).

Wagenaar WA, Sagaria SD (1975) Misperception of exponential growth. *Perception Psychophys.* 18(6):416–422.

Supplementary Materials for Spiller, Reinholtz, & Maglio

"Judgments Based on Stocks and Flows:

Different Presentations of the Same Data Can Lead to Opposing Inferences"

Online Appendix A: Study 2 Forecasts

Given the data from January 2007 to January 2009 (the data in Figure 4 before the dashed vertical line), the mean forecast in the stock condition was a stock of 107.9 million jobs in January 2010 (median = 106.2 million). The mean forecast in the flow condition was a flow of -1.022 million jobs in January 2010 (median = -1.181 million).[1] Thus, participants in both conditions reported values numerically lower than the values in January 2009 (in terms of job stock and net job flow, respectively), but the units were different (total jobs versus change in total jobs). If we assume linear extrapolations on the graph from the last observed datapoint to the forecast, and thereby translate the point estimates of the forecasts into the same scale, the forecasts imply very different paths that the economy would take. Given that 814,000 jobs were lost in January 2009 and the average forecast in the flow condition was that 1.022 million jobs would have been lost in January 2010, a linear extrapolation implies that the forecast monthly loss during 2009 averaged 918,000 jobs. In contrast, given that there were 111.5 million jobs in January 2009 and the average forecast in the stock condition was that there would be 107.9 million jobs in January 2010, a linear extrapolation implies that the forecast average monthly loss during 2009 (i.e., the total loss during 2009 divided by 12) implies an average mean loss of only 296,000 jobs ($t(127) = 7.63$, $p < .001$).

Despite implying very different changes in the economy (average monthly job losses of 918,000 vs. 296,000), participants' subjective evaluations of how the economy would have changed did not differ between the two conditions ($M_{\text{Stock}} = 3.01$, $SD_{\text{Stock}} = 1.54$; $M_{\text{Flow}} = 3.15$, $SD_{\text{Flow}} = 1.35$; $t(127) = 0.51$, $p = 0.61$; each mean significantly below the midpoint of 4, ps $< .001$). Although this could merely reflect the

[1] These analyses exclude 12 participants who did not click within a pre-defined region around the time of interest, as this was indicative of not following instructions and could reflect a different judgment than was asked, and 59 participants for whom there was no record of a click.

fact that people are insensitive to large numbers, we find strong correspondence between the quantitative forecast and evaluation of that forecast within condition. Regressing evaluation of the forecasts on a contrast-coded condition variable (stock = 1, flow = -1), the forecast average change (in thousands), and their interaction revealed no significant interaction ($t(125) = 1.23$, $p = .223$), but a large positive coefficient on forecast ($b = 0.0024$, $SE = 0.0002$, $t(125) = 13.33$, $p < .001$), indicating that an increase in monthly job change of 100,000 jobs was associated with a .24 increase on the 7-point scale.

The estimated mean evaluation for each condition depends on the value of the forecast. To better understand the relationship between the forecasts and the evaluations, we first consider the estimated mean evaluations for each condition if the monthly job change were equal to the average implied forecast in the stock condition (i.e., a monthly loss of 296,000 jobs). We then consider the estimated mean evaluations for each condition if the monthly job change were equal to the average implied forecast in the flow condition (i.e., a monthly loss of 918,000 jobs). Whereas an average monthly loss of 296,000 jobs in the stock condition corresponds to an estimated evaluation of 3.01 (corresponding to the mean listed above, significantly below the midpoint, $t(125) = -8.73$, $p < .001$), that same monthly loss in the flow condition corresponds to an estimated evaluation of 4.50, significantly above the midpoint ($t(125) = 2.40$, $p = .018$). Analogously, whereas an average of monthly loss of 918,000 jobs in the flow condition corresponds to an estimated evaluation of 3.15 (corresponding to the mean listed above, significantly below the midpoint, $t(125) = -7.28$, $p < .001$), that same monthly loss in the stock condition corresponds to an estimated evaluation of 1.39, far below the midpoint ($t(125) = -14.24$, $p < .001$).

In short, because the different data presentations lead to different quantitative forecasts but similar qualitative evaluations, they also suggest starkly different qualitative evaluations for the same quantitative forecasts.

Online Appendix B: Studies A1, A2, A3

Studies A1, A2, and A3 were conducted after Studies 1, 2, 4, 5, and 6, but prior to Study 3. The motivation for these studies was similar to that of Study 3: to examine whether the effect of stock vs. flow we observed in Studies 1 and 2 was due to other characteristics of the graph. If it was due to irrelevant graph characteristics (not aspects of the time series), we would not expect it to replicate when considering other spans of the data which do not show an increasing but negative flow. Instead, we would expect the result to be eliminated or reversed.

To test this, we examined the year following the passage of the Affordable Care Act, reflecting both a different cause and a different stretch of data. To preview the results, we found a qualitatively different pattern of results for Studies A1–A3 compared to Studies 1 and 2: Participants no longer gave more positive economic assessments in the flow (vs. stock) condition. This is consistent with our claim that the results are due to properties of the data series during the appropriate timespan. But we hesitate to over-interpret these (mostly) null effects. In particular, it could be that participants believed the Affordable Care Act could not have impacted the economy within its first year, or it could be that the effect we observed in Studies 1 and 2 simply failed to replicate. Thus, rather than relying on Studies A1, A2, and A3, we then conducted Study 3 which we included in the paper. For transparency and to attempt to reduce publication bias, we include Studies A1, A2, and A3 in this online appendix.

Study A1

In Study A1, we extend Studies 1 and 2 in two ways. First, we use the same data as in Studies 1 and 2, but ask participants to make judgments regarding a different focal region (2010; corresponding to the passage of the Affordable Care Act, ACA). This provides three benefits: (i) This region features a different pattern of stock and flow trends and thus allows us to assess whether some other aspect of the data as a whole (2007-2013) may have been contributing to the previously found effects. (ii) Because of the difference in the patterns of stock and flow in this region—the stock trend is increasing while the flow trend is flat—we now expect the stock presentation to lead to more positive judgments about economic changes, in contrast to Studies 1 and 2. (iii) The flow trend in the focal region does not cross the x-axis, relieving concern that the effects may be a reflection of this reference point.

As a second extension from the previous studies, we also consider whether the different data presentation formats (stock or flow) affect how participants consciously weight aspects of the data (e.g., absolute levels, velocity of level changes, acceleration of level changes). This allows us to assess an alternative account of the previous findings: That differences in judgment between formats are caused by differences in the perceived importance or diagnosticity of the given presentation format.

Method

One hundred twenty-one participants (49 women, 72 men; median age = 30) were recruited from AMT and completed Study A1.[2] Study A1 used the same basic context as Studies 1 and 2, except rather than using President Obama's inauguration as the beginning of the focal period, it used the date of the passage of the ACA (March 23, 2010) as the beginning of the focal period. The stimuli were the same as those used in Study 2, except that in Study A1, the reference point (vertical dashed line) was 14 months later (see Figure A1). In this case, the kinks in both the stock and the flow trends are less dramatic and qualitatively different from the prior studies: The flow changes from increasing (before ACA) to flat (after ACA), whereas the stock shifts from flat (before ACA) to increasing (after ACA). This allows us to assess a different stock/flow relationship as well as to test whether participants focus on the subset of data following the target event (versus making a gestalt assessment from the entirety of the data).

[2] An additional 8 participants consented to participate but did not complete the study.

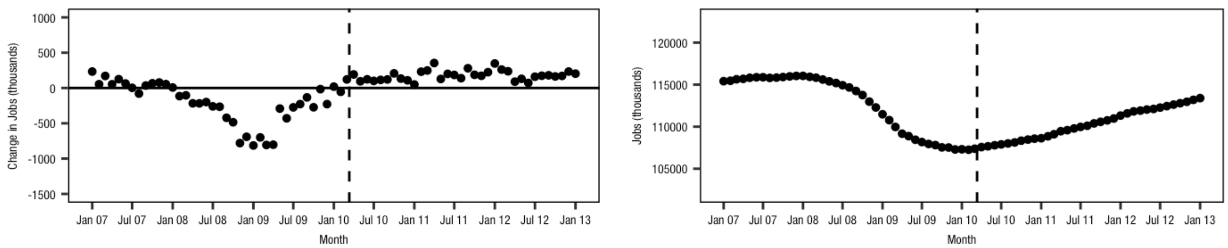


Fig. A1.
Job charts used in Study A1. The flow chart on the top shows the flow of jobs (jobs gained or lost). The stock chart on the bottom shows the same data presented as the stock (total number of jobs). The vertical dashed line indicates the passage of the Affordable Care Act.

Analogous to Study 2, participants were asked to rate how the economy changed during the first year of the ACA and what effect the ACA had on the economy during its first year. In addition, participants rated the importance of three possible measures of the economy: the number of jobs, the monthly growth rate in the number of jobs, and the change in the monthly growth rate of the number of jobs. Order was held constant for all participants. Finally, we assessed political leaning, gender, and age.[3]

Results

Consistent with the hypothesis that people interpret the local trend in the given presentation format, the stock graph did not lead to a more negative assessment of economic change as it did in Studies 1 and 2. Instead, it led to a marginally significant more positive assessment of economic change during the first year of the ACA (M = 4.88, 83% improved, 14% worsened) than the flow graph (M = 4.43, 63% improved, 13% worsened; $t(119) = -1.95$, $p = .053$). These results suggest that participants indeed attend to the focal parts of the graphs, as these results substantively differ from those in Studies 1 and 2.

Unexpectedly, there was no effect of presentation on attribution to the ACA of whether it made the economy better or worse (M_{Stock} = 4.45, 57% made it better, 19% made it worse; M_{Flow} = 4.38, 46% made it better, 13% made it worse; $t(119) = 0.30$, $p = .764$). While this again substantively differs from

[3] We also collected an open-ended measure regarding the impact of the ACA, but do not discuss those results here.

the findings in Studies 1 and 2, we do not observe a reversal. This may be because the difference in local trends was not as dramatic for the ACA (March 2010) as it was for President Obama's first inauguration (January 2009) or, possibly, because people's opinions of the ACA are more concrete and strongly held.

A possible explanation for the previously observed effects is that participants merely inferred which trend (stocks or flows) is the more normatively important indicator of economic health based on the graph we presented. Perhaps they are able to translate between metrics, but infer that if someone had decided to show them one trend over the other, that action communicates information in and of itself. In contrast to this alternative account, none of the importance ratings varied by conditions: Participants did not report giving subjectively more weight to the number of jobs, the change in jobs, or the rate of change in the change in jobs in one condition versus the other (*p*s > .2). This suggests that participants are not differentially making inferences about what dimensions are more important based on the information presented to them. In general, they rated the number of jobs as more important than the change in jobs ($M_{\text{Number}} = 4.04$, $M_{\text{Change}} = 3.79$, $t(120) = 3.25$, $p = .002$) and the change in jobs as more important than the rate of change in the change in jobs ($M_{\text{Change}} = 3.79$, $M_{\text{Acceleration}} = 3.62$, $t(120) = 2.63$, $p = .010$).

We also note that we observed a significant difference in self-reported political leaning, such that people in the flow condition reported being more liberal than those in the stock condition ($M_{\text{Flow}} = 3.62$ vs. $M_{\text{Stock}} = 3.19$, $t(119) = 2.12$, $p = .036$). We did not observe differences on this measure in any other study, and the observed results for this study are similar controlling for self-reported political liberalism.

Study A2

Method

Three hundred and one participants (136 women, 163 men; median age = 31) completed Study A2.[4] This was a direct replication of Study A1 with the key change that we added an additional measure

[4] An additional 30 participants consented to participate but did not complete the study.

assessing causal potency: “To what extent do you think major health care laws (like the Affordable Care Act) have the potential to impact the economy shortly after they become law?” (not at all, very little, somewhat, to a great extent).

Results

Contrasting with Studies 1 and 2, though not replicating Study A1, evaluations of how the economy changed did not significantly differ between conditions (M_{Stock} = 4.50, 66% improved, 25% worsened; M_{Flow} = 4.65, 61% improved, 16% worsened; $t(299) = 0.98$, $p = .329$).

Contrasting with Studies 1 and 2, and consistent with Study A1, evaluations of attribution did not significantly differ between conditions (M_{Stock} = 4.28, 52% made it better, 26% made it worse; M_{Flow} = 4.41, 48% made it better, 23% made it worse; $t(298) = 0.84$, $p = .401$).

Subjective importance ratings did not vary by condition (ps > .4). 70% of participants responded with top two responses (“somewhat” or “to a great extent”) that major health care laws have the potential to impact the economy shortly after they become law.

Study A3

Method

Two hundred ninety-nine participants (142 women, 156 men; median age = 31) completed Study A3.[5] This was a direct replication of Study A2, except that we further emphasized when the Affordable Care Act became law by adding red text (“Affordable Care Act becomes law”) with a red arrow pointing at March 2010

[5] An additional 36 participants consented to participate but did not complete the study.

Results

Contrasting with Studies 1 and 2, and as in Study A2, evaluations of how the economy changed did not significantly differ between conditions (M_{Stock} = 4.90, 82% improved, 13% worsened; M_{Flow} = 4.83, 64% improved, 11% worsened; $t(297) = 0.54$, $p = .589$).

Contrasting with Studies 1 and 2, and consistent with Studies A1 and A2, evaluations of the ACA's effect on the economy did not significantly differ between conditions (M_{Stock} = 4.66, 69% made it better, 13% made it worse; M_{Flow} = 4.56, 54% made it better, 16% made it worse; $t(297) = 0.70$, $p = .486$).

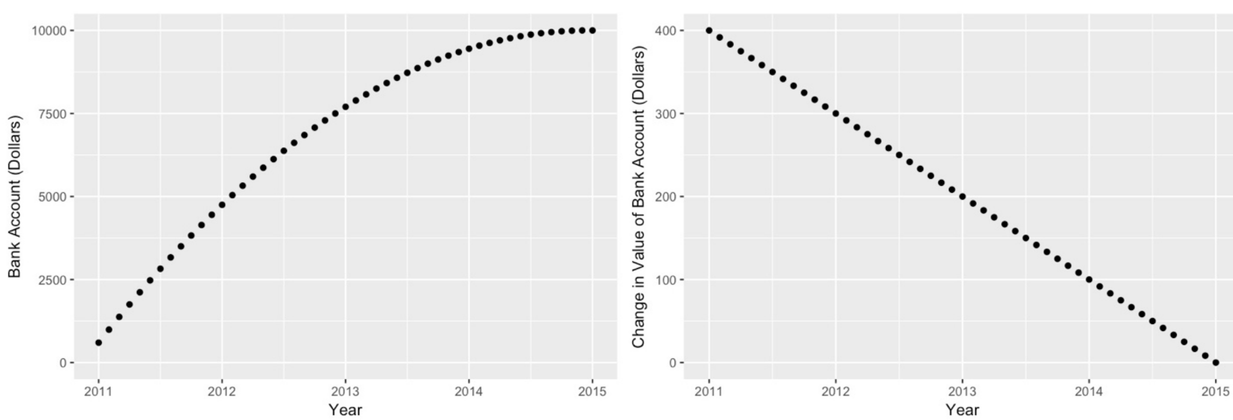
Subjective importance ratings did not significantly vary by condition (ps > .05). 73% of participants responded with top two responses ("somewhat" or "to a great extent") that major health care laws have the potential to impact the economy shortly after they become law.

Online Appendix C: Scenarios Used in Study 4

1) Money in Sam's bank account (end value = $10,000)
2) Valuation of NormaTech (end value = $250,000)
3) Donaldo City's municipal savings (end value = $1,000,000)
4) Number of Steve's facebook friends (end value = 1,200 friends)
5) Number of employees working at GeneriWare (end value = 2,500 people)
6) Number of residents of Hooperburg (end value = 10,000 residents)
7) Number of books Nick owns (end value = 1,200 books)
8) Number of shippable units in ProsaiCo's inventory (end value = 10,000 units)
9) Gallons of water in Weavertown's reservoir (end value = 100 million gallons)

Example of question prompt, corresponding to scenario 1 with a positive stock trend and negative flow trend. Participants in the stock condition would see just the image on the left. Participants in the flow condition would see just the image on the right. Question wording was the same in both conditions.

Below is a chart showing how Sam's bank account has changed from the beginning of 2011 to the beginning of 2015.



On January 1, 2015, Sam had $10,000 in his bank account. How much do you think Sam will have in his bank account on January 1, 2016?

Fig A2. Sample stimuli used in Study 4. Participants saw one of the two panels, each of which reflect the same data, a quantity that is increasing at a decreasing rate, that is, a quantity with a positive but decreasing flow. The panel on the left reflects the stock trend; the panel on the right reflects the flow trend.

Online Appendix D: Study 4, 5, 6 Parametric Results

In the main text, our analyses in Studies 4, 5, and 6 focus on qualitative shifts (decreases, no change, or increases). This is for three reasons. First, these qualitative differences (vs. quantitative differences) in forecasts are important in and of themselves and suggest it is not merely being more or less sensitive to different magnitudes of change. Second, in several cases the mean is not a good representation of the distribution due to focal values that elicit a large proportion of responses (e.g., in Study 4, for a constant stock with a constant flow, upwards of 90% of responses are exactly equal to the ending value). Third, due to the unbounded nature of the scale, there are some cases of extreme outliers with no clear exclusion thresholds.

Nonetheless, here we replicate the main analyses using linear models. Each analysis is conducted using two sets of thresholds to trim outliers: the first threshold only includes observations between 0.5 and 1.5 (inclusive) on the transformed scale; the second threshold only includes observations between 0 and 2 (inclusive) on the transformed scale.[6] In Study 4, we first transform the noted set of systematic outliers that were systematically off by a factor of one million. In each case we include random intercepts and slopes for participants. In Studies 5 and 6, we focus on the subset of responses for which participants accurately described the data.

Study 4. Of the 3,618 total observations, the narrow subset (from 0.5 to 1.5) includes 3,280, or 91%, and the broad subset (from 0 to 2) includes 3,536, or 98%. Table A1 compares summarized results from the ordinal category analysis described in the text and the linear models as described above. In each case, we specify whether the coefficient indicated higher values for stocks (S) or flows (F) and its level of significance. (While this overemphasizes statistical significance, it enables qualitative comparisons across models that are assessed using different metrics.)

[6] In each case in Studies 4 and 5, we consider separate analyses for the separate time series, but all patterns lead to substantively and statistically similar conclusions when analyses are conducted using a unified model with clustered standard errors or random effects to account for non-independence where possible. All analyses in Study 6 are between-subject.

In four cases (increasing or decreasing stocks with constant or decreasing flows), all three models lead to the same conclusions. The discrepancies in the remaining cells are attributable to two factors, both of which are observable in Figure 7. First, a minority of participants in the flow condition reported values close to 0, possibly reflecting forecasted flows rather than forecasted stocks based on flows. These are only included in the Linear [0, 2] model as they are excluded as outliers in the Linear [0.5, 1.5] model. They primarily have the effect of artificially decreasing the flow estimates. Second, a substantial portion of participants in the flow condition reported values that appear to be one month's adjustment from the ending stock rather than twelve months' adjustments from the ending stock. This primarily has the effect of artificially dragging the flow estimates towards 1. Note that these effects on the ordinal analysis reported in the main text are muted: the first is relatively constant across cells, and the second is immaterial once the forecasts are converted to signed changes compared to the ending flow.

Table A1.
Comparison of Study 4 results across three models. Ordinal Categories represents the ordered logistic regression reported in the main text. Linear [0.5, 1.5] represents linear model with outliers less than 0.5 or greater than 1.5 trimmed. Linear [0, 2] represents linear model with outliers less than 0 or greater than 2 trimmed.

		Decreasing Stock	Constant Stock	Increasing Stock
Increasing Flow	Ordinal Categories	S < F ***	S < F **	S > F did not converge
	Linear [0.5, 1.5]	S < F ***	^S > F ns	^S > F ***
	Linear [0, 2]	^S > F ns	^S > F *	^S > F ***
Constant Flow	Ordinal Categories	S < F ***	S > F ns	S > F ***
	Linear [0.5, 1.5]	S < F ***	S > F ns	S > F ***
	Linear [0, 2]	S < F ***	^S > F *	S > F ***
Decreasing Flow	Ordinal Categories	S < F **	S > F ***	S > F ***
	Linear [0.5, 1.5]	S < F ***	^S < F **	S > F ***
	Linear [0, 2]	S < F ***	^S < F ns	S > F ***

Note. * $p < .05$. ** $p < .01$. *** $p < .001$. ^ indicates different conclusion compared to ordinal result.

Study 5. In Study 5, the narrow subset included 1112 of 1142 accurate observations (97%) and the broad subset included 1131 of 1142 accurate observations (99%). Unlike Study 4, there were no notable response distortions in the flow condition. As seen in Table A2, the results are comparable (we just focus on the linear contrast, as the combined condition generally led to results between the stock and the flow conditions). The only points of difference from the main results were slight and of magnitude rather than of sign. These results are essentially the same under any of the three analysis plans.

Study 6. The narrow restriction in Study 6 included 242 of 250 accurate responses (97%) and the broad restriction included 249 of 250 accurate responses (>99%). As in Study 5, the results lead to the same conclusions across analysis plans. These results are shown in Table A3.

Table A2.
Comparison of Study 5 results across three models among responses with accurate descriptions. Ordinal Categories represents the ordered logistic regression reported in the main text. Linear [0.5, 1.5] represents linear model with outliers less than 0.5 or greater than 1.5 trimmed. Linear [0, 2] represents linear model with outliers less than 0 or greater than 2 trimmed.

		Decreasing Stock	Increasing Stock
Varying Flow	Ordinal Categories	S < F ***	S > F ***
	Linear [0.5, 1.5]	S < F ***	S > F ***
	Linear [0, 2]	S < F †	S > F ***
Constant Flow	Ordinal Categories	S < F did not converge	S > F *
	Linear [0.5, 1.5]	S < F ns	S > F ***
	Linear [0, 2]	S > F ns	S > F †

Note. † $p < .1$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Table A3.
Comparison of Study 6 results across three models among responses with accurate descriptions. Ordinal Categories represents the ordered logistic regression reported in the main text. Linear [0.5, 1.5] represents linear model with outliers less than 0.5 or greater than 1.5 trimmed. Linear [0, 2] represents linear model with outliers less than 0 or greater than 2 trimmed.

	Graph	Table
Ordinal Categories	S > F ***	S > F ***
Linear [0.5, 1.5]	S > F ***	S > F ***
Linear [0, 2]	S > F ***	S > F **

Note. † $p < .1$. * $p < .05$. ** $p < .01$. *** $p < .001$. ^ indicates difference relative to ordinal result.

Online Appendix E: Study 4 Sensitivity to Differences in Stocks vs. Flows

The design of Study 4 enables one additional analysis of interest. In particular, we can examine the within-subject rank ordering across the nine data patterns to examine the relative sensitivity to stock difference and flow differences. Whereas the previous analyses examined sensitivity to a given pattern across conditions, this analysis examines sensitivity to different patterns within conditions. Every participant has nine observations. After scaling responses, we rank-ordered them within-participant from 1 to 9, with higher numbers reflecting higher forecasts and ties set equal to the average rank. We regressed rank on stock trend (-1, 0, 1) and flow trend (-1, 0, 1), nested within condition, with clustered standard errors.[7] Participants in the stock condition were highly sensitive to stock trend (b = 2.14, SE = 0.036, $t(401) = 59.62$, $p < .001$) and less sensitive to flow trend (b = 1.58, SE = 0.053, $t(401) = 29.39$, $p < .001$; difference: $t(401) = 7.50$, $p < .001$), whereas those in the flow condition were highly sensitive to flow trend (b = 2.23, SE = 0.064, $t(401) = 34.57$, $p < .001$) and much less sensitive to stock trend (b = 0.60, SE = 0.066, $t(401) = 8.99$, $p < .001$; difference: $t(401) = 17.27$, $p < .001$). The differences between condition are also significant, such that those in the stock condition are more sensitive to stock trend than those in the flow condition ($t(401) = 20.51$, $p < .001$) and those in the flow condition are more sensitive to flow trend than those in the stock condition ($t(401) = 7.79$, $p < .001$).

[7] All conclusions are substantively and statistically the same from analyses using random effects.

Online Appendix F: Study 5 Accuracy

We use repeated measures logistic regression to examine accuracy as a function of presentation condition (contrast coded via a linear contrast, coded stock = 1, combined = 0, flow = -1, and a quadratic contrast, coded stock = -1, combined = 2, flow = -1), order (describe first = 1, forecast first = -1), stock trend (1 = increasing stock, -1 = decreasing stock), varying flow trend (1 = varying, -1 = constant), and all interactions, allowing for clustered standard errors.[8] Regression results are given in Table A4.

Table A4.
Repeated-measures logistic regression results for Study 5 accuracy. Three- and four-way interactions are included in analysis but excluded from table for space as none were significant.

	Estimate	SE	*z*	*p*
Intercept	-0.12	0.05	-2.70	.007
Presentation (Linear)	0.18	0.06	3.30	<.001
Presentation (Quadratic)	0.15	0.03	4.73	<.001
Order	0.15	0.05	3.30	<.001
Stock Trend	0.74	0.05	16.38	<.001
Varying Flow	-0.27	0.05	-5.97	<.001
Pres (Lin) × Order	0.07	0.06	1.32	.187
Pres (Lin) × Stock	0.36	0.06	6.35	<.001
Pres (Lin) × Flow	0.21	0.06	3.72	<.001
Pres (Quad) × Order	0.04	0.03	1.24	.216
Pres (Quad) × Stock	0.02	0.03	0.49	.627
Pres (Quad) × Flow	-0.03	0.03	-1.02	.308
Order × Stock	-0.02	0.05	-0.53	.593
Order × Flow	0.03	0.05	0.72	.473
Stock × Flow	-0.01	0.05	-0.20	.838
…				

Across participants, accuracy was higher for stocks than flows (linear contrast), with combined lying above the midpoint (quadratic contrast; stock = 48%, combined = 54%, flow = 39%), and slightly higher when participants described the graphs before making forecasts (describe first = 50%, forecast first = 44%). Across trend type, accuracy was higher for increasing stocks than decreasing stocks, especially when the stock trend was salient (stock: increasing = 73%, decreasing = 24%; combined: increasing = 71%, decreasing = 36%; flow: increasing = 48%, decreasing = 31%). Similarly, accuracy was higher for

[8] All conclusions are substantively and statistically the same from analyses using random intercepts.

constant flows than varying flows, especially when the flow trend was salient (stock: varying = 48%, constant = 49%; combined: varying = 47%, constant = 60%; flow: varying = 29%, constant = 50%).

These accuracy rates may seem low but are broadly consistent with recent findings on deriving calculations from flow data (Cronin et al. 2009). However, they may also be mildly artificially depressed. It appears some participants likely reported the *magnitude* of change rather than *signed* change: 15% of responses (365 out of 2420) reported the correct second-to-last value, and the magnitude of the change was accurate, but the sign was reversed. Counting these responses as accurate raises the proportion correct from 47% to 62%, almost entirely for decreasing stocks in the stock and combined conditions (stock: increasing = 73%, decreasing = 73%; combined: increasing = 72%, decreasing = 66%; flow: increasing = 49%, decreasing = 39%, thereby reversing the Presentation (Linear) × Stock interaction from significantly positive to significantly negative, z = -2.54, p = .011). Conservatively, we exclude these participants from further analyses (since they may represent true misunderstandings), but we note that these responses may reflect misunderstanding the question rather than the data. Including these observations in the main analyses does not change any substantive or statistical conclusions.