

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

Stephen A. Spiller<sup>1</sup>

Nicholas Reinholtz<sup>2</sup>

Sam J. Maglio<sup>3</sup>

<sup>1</sup> Stephen A. Spiller is an associate professor of marketing and behavioral decision making at the UCLA Anderson School of Management. E-mail: [stephen.spiller@anderson.ucla.edu](mailto:stephen.spiller@anderson.ucla.edu).

<sup>2</sup> Nicholas Reinholtz is an assistant professor of marketing at the University of Colorado Boulder, Leeds School of Business. E-mail: [nicholas.reinholtz@colorado.edu](mailto:nicholas.reinholtz@colorado.edu).

<sup>3</sup> Sam J. Maglio is an associate professor of marketing at the University of Toronto Scarborough and Rotman School of Management. E-mail: [sam.maglio@utoronto.ca](mailto:sam.maglio@utoronto.ca).

**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.

*Keywords:* judgments; forecasts; time series; stocks and flows; data visualization

### Judgments Based on Stocks and Flows:

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

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. In May of 2015, the Bureau of Labor Statistics estimated that 250 million American civilians were employed. In May of 2016 and 2017, those figures were 253 and 255 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 May 2015 to May 2016, the US economy gained 3 million jobs. From May 2016 to 2017, the US 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.

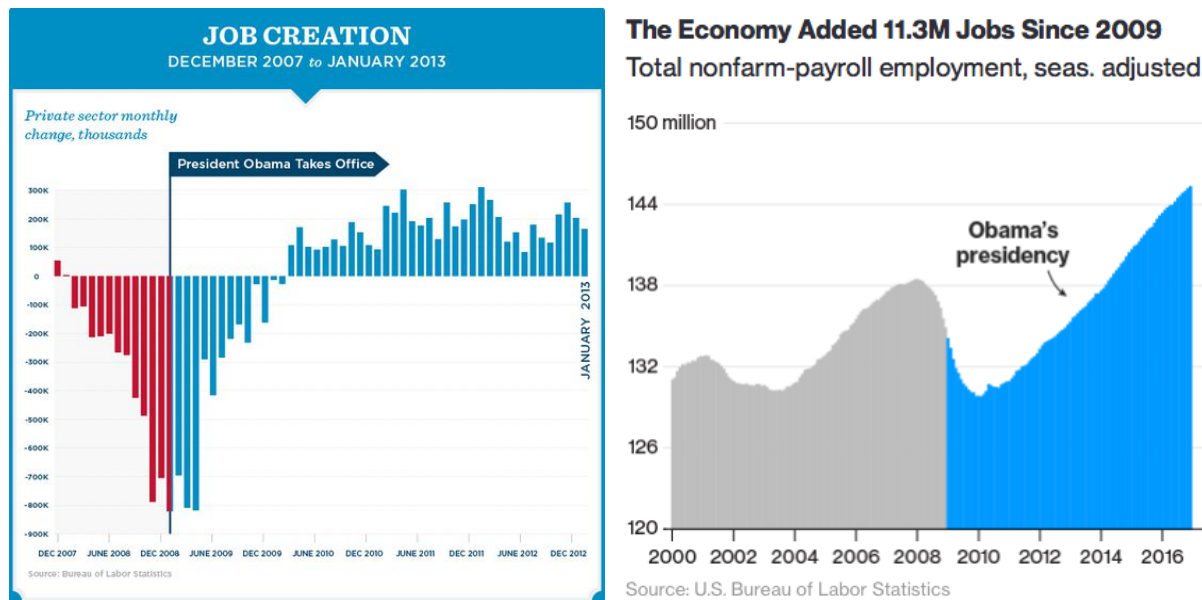


Fig. 1.

Two presentations of similar jobs data from the Bureau of Labor Statistics. 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 8/29/2017 from <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 number of people employed (retrieved 8/29/2017 from <https://www.bloomberg.com/graphics/2017-obama-economic-legacy/>).

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.<sup>1</sup> 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 & Sterman, 2000; Cronin, Gonzalez, & Sterman 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 US-held jobs from 2015 to 2017. The stock of jobs at three intervals (250 M, 253 M, 255 M) shows a pattern that increases over time and may well paint a particularly rosy picture of the US economy. However, the flow in jobs between those same intervals (+3 M, +2 M) reflects a decreasing trend—a slowing in the rate of job growth to the tune of 1 M jobs—and may lead people to see the US 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 “chartjunk”; 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.

---

<sup>1</sup> Note that 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).

The focus on time-series data reflects, in part, its 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 startup, 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—it has 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.<sup>2</sup>

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

---

<sup>2</sup> For example, in a stock presentation, the velocity (the first derivative of stock) is salient. In contrast, flow presentation keeps the velocity information, but makes the acceleration (the second derivative of stock) more salient.

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 & Sterman, 2000; Sterman & 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" (Cronin et al. 2009; Sterman, 2002), 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). While most (> 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; Cronin et al. 2009; Sterman & Booth Sweeney 2002, 2007). 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" (p. 252, Booth Sweeney & Sterman, 2000), builds from this prior work in asking how presenting

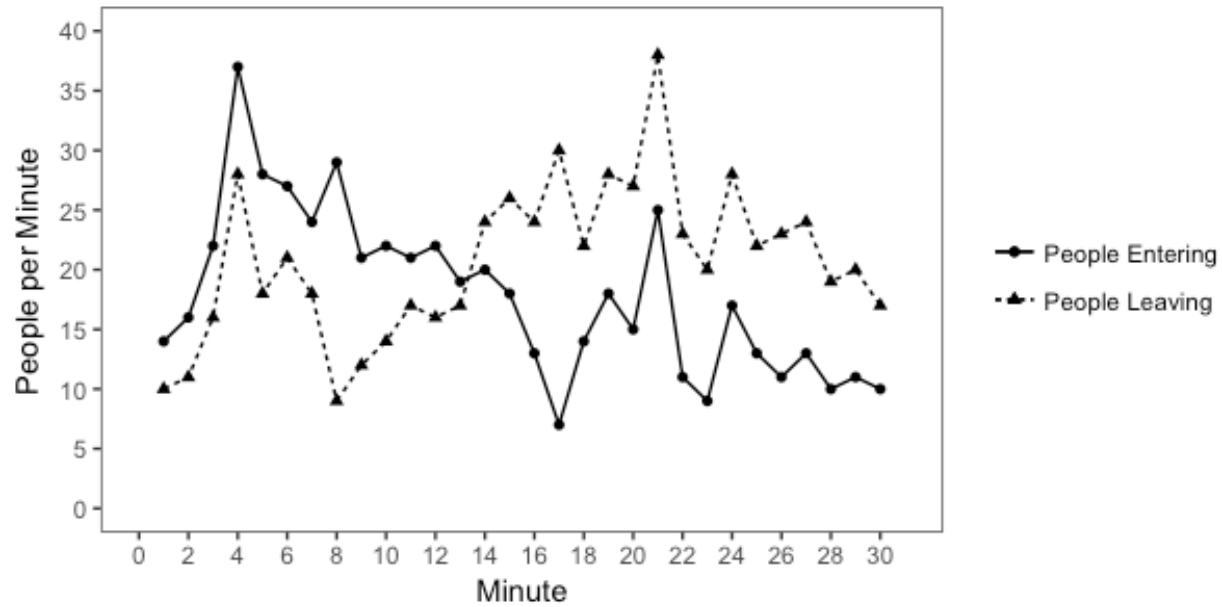


Fig. 2.

A simplified version of the department store task, reproduced based on data from Figure 3 in Sterman (2002).

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 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 like 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. In contrast, if people attend to the salient trend, presentations like Panels A and D are likely to evoke similar mental representations of positivity or negativity and lead people to relatively consistent judgments.

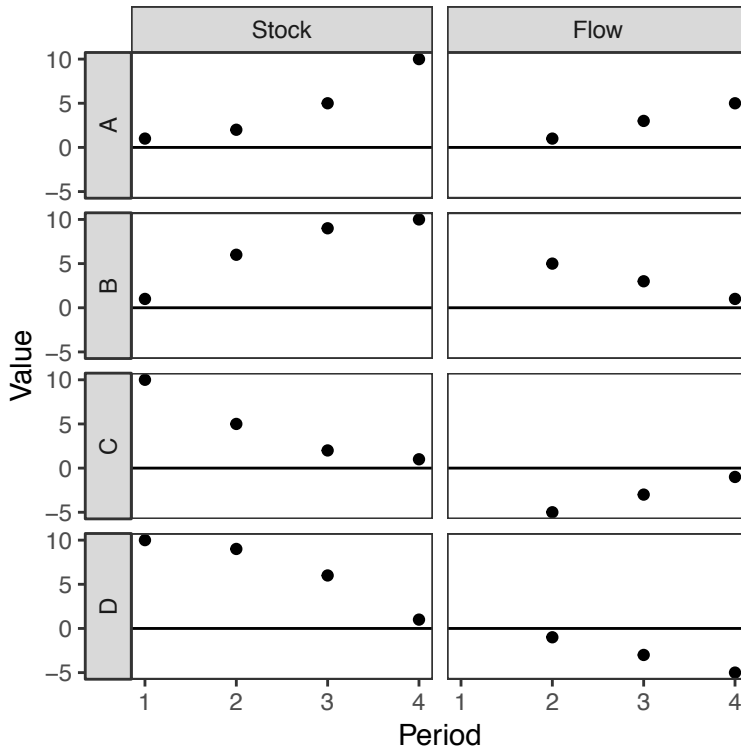


Fig. 3. Four examples of stock and flow patterns. Flow levels are shown in the period following the change (e.g., the flow from period 1 to 2 is shown as period 2). In Panels A and D, the stock and flow trends have the same sign. In Panels B and C, the stock and flow trends have opposite signs.

We depart from the work of Sterman and colleagues 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, Sumarta, & Myaskovsky, 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 & Epstein, 1999; Reyna & Brainerd, 2008), spend as though prices are higher when dealing with less-numerous currencies (Raghubir & Srivastava, 2002), and prefer options that dominate alternatives on attributes presented with a more expansive scale (Burson, Larrick, & Lynch, 2009). In each of these instances, the considered quantities can be equated using a scalar multiplier (from US dollars to Canadian dollars or from millimeters to centimeters; Maglio & 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 non-linear transformation (de Langhe, Puntoni, & Larrick 2017). Take, for example, the “MPG illusion”—a well-documented bias in people’s judgments about vehicle efficiency (Larrick & Soll, 2008). While 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, while the actual relationship is reciprocal. Similar issues arise for common productivity statistics like megabytes-per-second (internet speed) and pages-per-minute (printer speed; de

Langhe & Puntoni, 2016). More generally, it has been recognized that the metric used to communicate data can serve as an important choice architecture tool (Camilleri & 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 (Cleveland & McGill, 1984; Kahneman, 2011; Lurie & Mason, 2007; Slovic, 1972). 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 & 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.<sup>3</sup> Second, while 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 & Finke, 1984).

---

<sup>3</sup> Note that while 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.

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. Further, 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 (McKenzie & Liersch, 2011; Thomson & Oppenheimer 2016; Wagenaar & Sagaria, 1975).

### **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 either as 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, 68 men; median age = 30) recruited from Amazon Mechanical Turk (AMT) completed Study 1.<sup>4</sup> 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 by Sterman and colleagues in previous work.

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 drop-down menus to select a month and year from January 2007 to December 2014. Second, they were asked

---

<sup>4</sup> Seven additional participants consented to participate but did not complete the study.

“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.

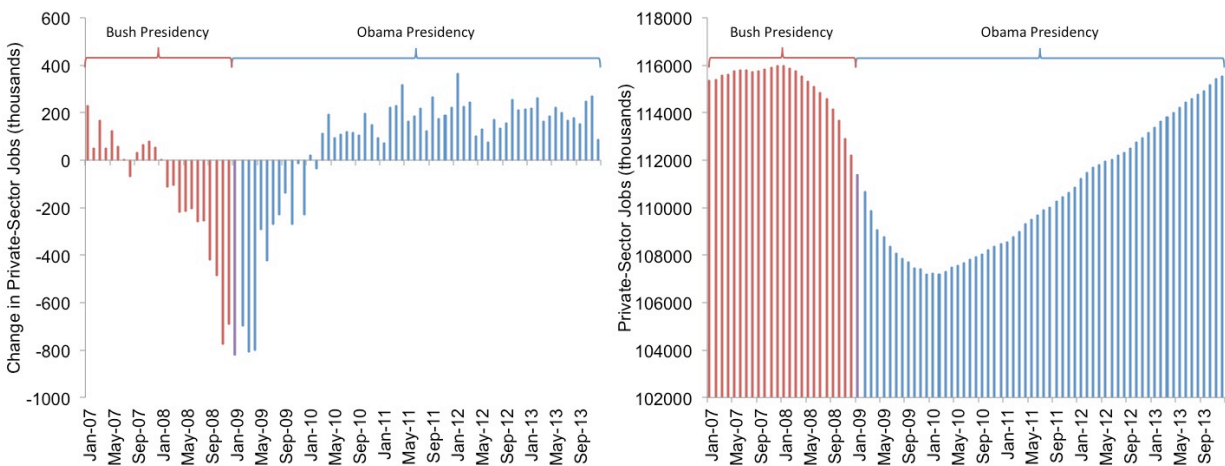


Fig. 4.

Job charts used in Study 1. 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).

## 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 due to 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<sup>5</sup>) 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 ( $M_{\text{Stock}} = 3.46$  (June 2010),  $SD = 1.26$ ;  $M_{\text{Flow}} = 2.85$  (November 2009),  $SD = 0.94$ ;  $t(98) = 2.75$ ,  $p = .007$ ): 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 7 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 ( $t(98) = 5.10$ ,  $p < .001$ ). 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"),  $p = .002$ , 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,  $p < .001$ , 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 to a qualitative reversal (bad to good) in judgments about President Obama's effect on the economy.

## 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 (that is, 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

---

<sup>5</sup> 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, etc.

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 datasets 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, 118 men; median age = 32) from AMT were recruited to participate in and completed Study 2.<sup>6</sup> 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 vs. bars) on stock-flow reasoning were negligible (Cronin et al. 2009),

---

<sup>6</sup> As in Study 1, an additional 8 partial responses were excluded from analysis.

but since position is more reliably assessed than length or area (Cleveland & 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?” (7-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.

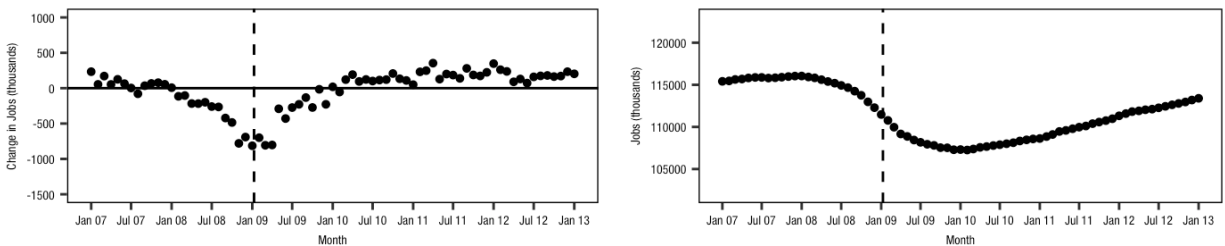


Fig. 5. Job charts used in Study 2. The flow chart on the left shows the flow of jobs (jobs gained or lost). The stock chart on the right shows the same data presented as the stock (total number of jobs). The vertical dashed line indicates Barack Obama’s first inauguration.

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 (data post-2009 were not shown). They made this forecast by clicking on a point on the graph, the location of which was recorded (akin to Thomson & 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 (7-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,  $p < .001$ , 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,  $p = .037$ , with 49% reporting he made it worse and 29% reporting he made it better; difference between conditions:  $t(198) = 7.31$ ,  $p < .001$ ). Perceptions of how the economy changed largely tracked this measure and showed a similar difference ( $t(197) = 8.40$ ,  $p < .001$ )<sup>7</sup>: Participants who saw the data as flows judged the economy to have improved ( $M = 5.31$ , greater than the scale midpoint,  $p < .001$ , 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,  $p < .001$ , with 70% reporting it worsened and 28% reporting it improved). For each of these two evaluative judgments, liberalism (vs. conservatism) was correlated with more positive assessments (change in economy:  $r = 0.20$ ,  $p = .006$ ; Obama's effect:  $r = 0.41$ ,  $p < .001$ ), but in neither case did it moderate the effect of stock vs. flow ( $ps > .18$ ).<sup>8</sup> 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

<sup>7</sup> One data point is missing for this measure due to a coding error.

<sup>8</sup> Politics did not vary by condition ( $p > .40$ ).

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).<sup>9</sup> 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 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.

---

<sup>9</sup> Within-condition, evaluations and forecasts were consistent such that lower forecasts were associated with more negative evaluations.

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).<sup>10</sup>

## Method

Participants were recruited from AMT ( $N=401$ ; 184 women, 216 men; median age = 34).<sup>11</sup> Study 3 was similar to Study 2 with a few variations. First and most importantly, Study 3 included a second factor in the design, resulting in a 2 (data presentation: stock vs. flow)  $\times$  2 (timespan: 2009 vs. 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 timespan. See Figure 6.

Like Study 2, we measured perceived change in the economy and Obama's effect on the economy during the focal timespan. 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

---

<sup>10</sup> 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 timespan (March 2010 to March 2011) compared to Studies 1 and 2, and show results qualitatively consistent with those of the 2010 condition in Study 3.

<sup>11</sup> An additional 36 participants consented to participate but did not complete the study.

monthly growth rate” were to evaluating the overall state of the economy on 5-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 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.

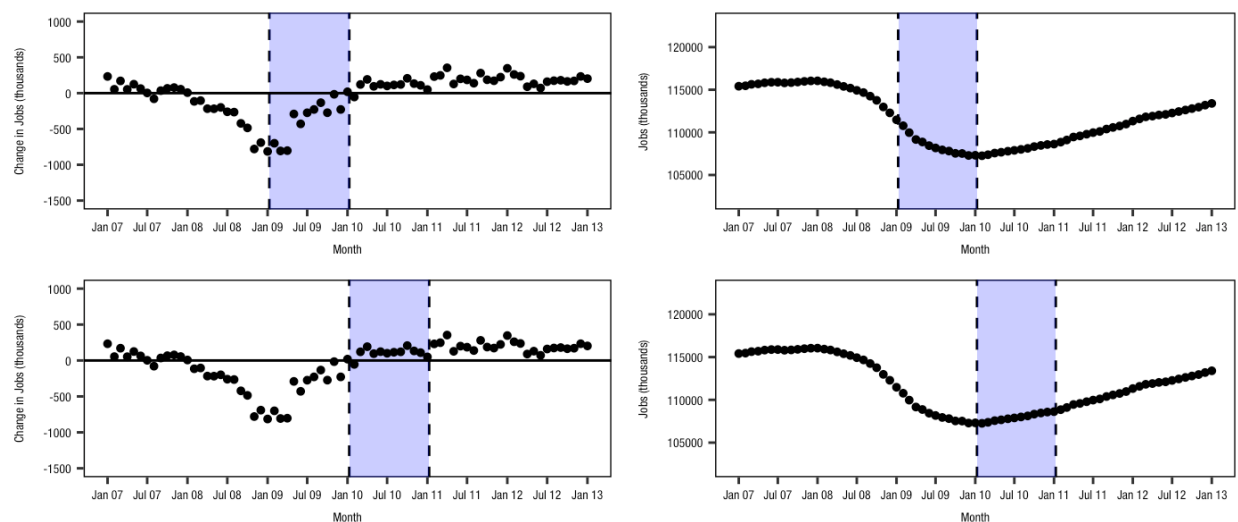


Fig. 6.

Job charts used in Study 3. 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).

## 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 qualitative evaluation of change in the economy varied depending on whether participants considered 2009 or 2010 (interaction  $z = -9.23, p < .001$ ). For 2009, the data replicated Studies 1 and 2, such that most participants believed the economy worsened when considering stocks (84% said worsened, 8% said improved) whereas most participants believed the economy improved when considering flows (17% said worsened, 77% said improved; difference:  $z = -10.23, p < .001$ ). 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, 80% said improved) than when considering flows (2% said worsened, 58% said improved; difference:  $z = 2.27, p = .023$ ).

The effect of data presentation on evaluations of Obama's effect on the economy also depended on whether participants considered 2009 or 2010 (interaction  $z = -6.30, p < .001$ ). During 2009, more participants believed Obama had a negative effect than a positive effect when considering stocks (46% said made it worse, 22% said made it better), whereas most participants believed he had a positive effect when considering flows (9% said made it worse, 69% said made it better;  $z = -7.52, p < .001$ ). During 2010 there was no significant difference in valence between conditions (Stocks: 10% said made it worse, 71% said made it better; Flows: 2% said made it worse, 57% said made it better;  $z = 1.36, p = .174$ ), though twice as many participants said Obama had no effect when considering the (constant) flow trend rather than the (increasing) stock trend (41% vs. 19%).

Regressing the three importance measures on timespan, presentation format, and their interaction yielded only one statistically significant effect: The perceived importance of growth rate was higher for participants who evaluated 2010 data ( $M = 3.95$ ) compared to those who evaluated 2009 data ( $M = 3.74$ ;  $t(397) = -2.33, p = .020$ ). The other eight possible effects were not significant ( $ps > .25$ ). Thus, it seems unlikely that the difference 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 ( $ps < .015$ ), but political views did not moderate either key effect ( $ps > .40$ ).<sup>12</sup>

## Discussion

Study 3 provides further evidence that people make judgments based on salient features of the data as they are presented. Further, these judgments do not seem to be due to by 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 timespans can lead to more positive evaluations for stocks than flows while other timespans 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 reveal a consistent story. In subsequent studies, we present participants with artificial data to enable more precise control over stock and flow trends.

Table 1. For key dependent variables in Studies 1 through 3, proportion of participants in each condition who reported a positive, neutral, or negative evaluation. 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 only examined 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).

Change in the Economy								
	Study 1 (2009)		Study 2 (2009)		Study 3 (2009)		Study 3 (2010)	
	Stock	Flow	Stock	Flow	Stock	Flow	Stock	Flow
Better			.28	.78	.08	.77	.80	.58
No Change			.02	.04	.09	.06	.06	.39
Worse			.70	.18	.84	.17	.14	.02
Obama's Effect on the Economy								
	Study 1 (2009)		Study 2 (2009)		Study 3 (2009)		Study 3 (2010)	
	Stock	Flow	Stock	Flow	Stock	Flow	Stock	Flow
Better	.24	.66	.29	.73	.22	.69	.71	.57
No Change	.16	.26	.23	.13	.32	.22	.19	.41
Worse	.60	.08	.49	.14	.46	.09	.10	.02

<sup>12</sup> Politics did not vary by condition ( $ps > .20$ ).

### Study 4

Studies 1, 2, and 3 examined evaluations of a real dataset 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, 247 men; median age = 30) were recruited from AMT and completed Study 4.<sup>13</sup> We generated nine time series datasets in a full factorial design such that the flow was decreasing, constant, or increasing and the stock on average was decreasing, constant, or increasing.<sup>14</sup> 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.<sup>15</sup> 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.

---

<sup>13</sup> An additional 36 partial responses were excluded from analysis.

<sup>14</sup> 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.

<sup>15</sup> 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.

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 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.<sup>16</sup> 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. Panel 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,

---

<sup>16</sup> 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.

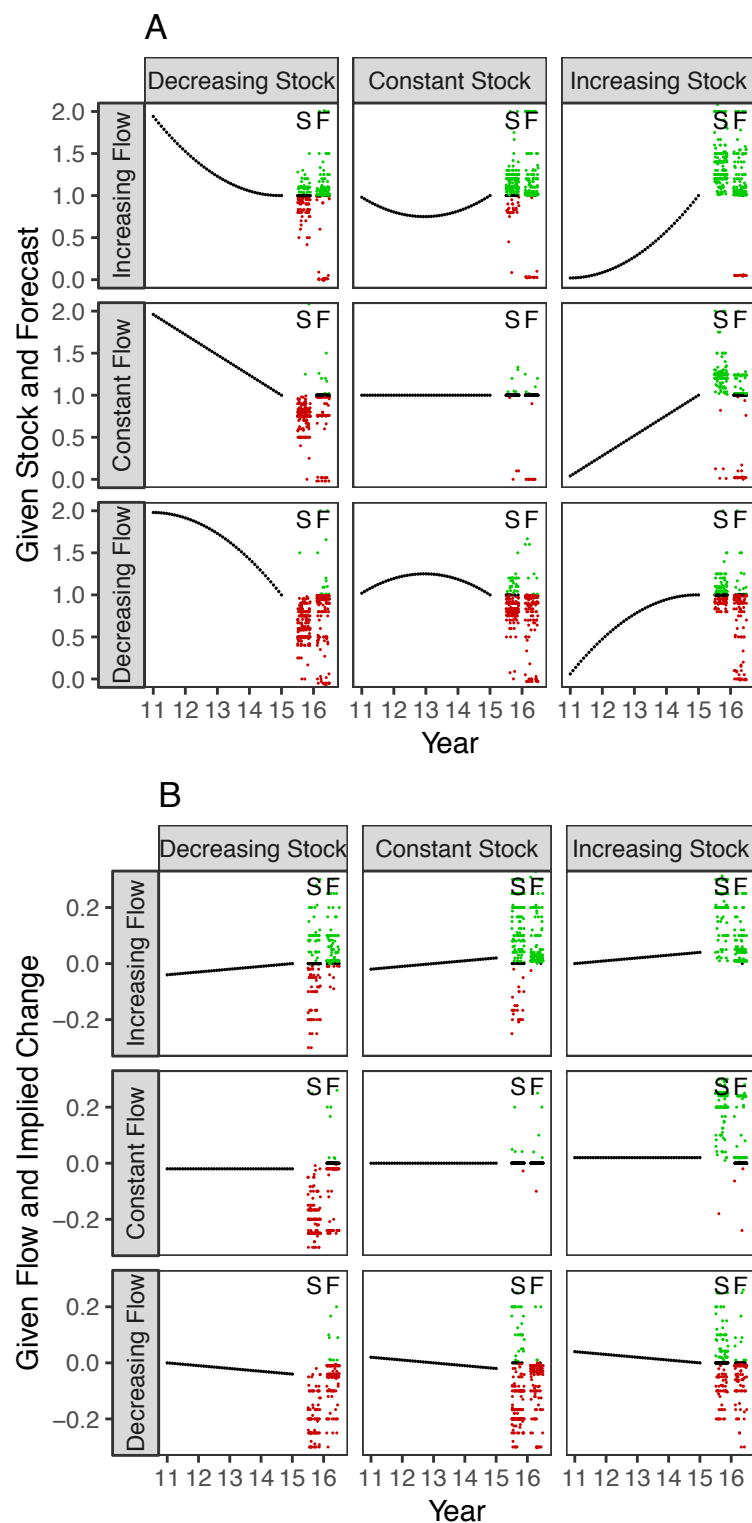


Fig. 7.

Observed stock (Panel A) and flow (Panel B) trends with participant forecasts by condition (S = stocks, F = flow, repeated in each panel). Forecasts in black indicate no forecasted change (1 in A, 0 in B), shaded forecasts indicate a change (positive change  $> 1$  in A,  $> 0$  in B; negative change  $< 1$  in A,  $< 0$  in B). Panels contain the same data on different scales, with Panel B highlighting a more limited range of responses.

increasing)  $\times$  three flow trends (decreasing, constant, 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.

Panel 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 Panel A, the y-axis in Panel B represents the implied change from the end of the observed sequence, so a 0 in Panel B is equivalent to a 1 in Panel A. In Panel 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) are shown in Table 2.

Table 2.

Proportion of respondents in each condition (stock, flow) who forecast an increase, no change, or a decrease compared to the last observed stock for each of nine combinations of stock trends (decreasing, constant, increasing) and flow trends (decreasing, constant, increasing).

		Decreasing Stock		Constant Stock		Increasing Stock	
	Forecast	Stock	Flow	Stock	Flow	Stock	Flow
Increasing Flow	Increase	.23	.81	.79	.90	1.00	.91
	No Change	.47	.05	.08	.01	.00	.01
	Decrease	.30	.15	.14	.09	.00	.08
Constant Flow	Increase	.01	.05	.04	.04	.98	.33
	No Change	.00	.55	.94	.90	.00	.53
	Decrease	.99	.40	.02	.06	.02	.14
Decreasing Flow	Increase	.01	.10	.15	.05	.32	.08
	No Change	.00	.00	.05	.01	.41	.10
	Decrease	.99	.90	.80	.94	.28	.82

**Qualitatively Different Inferences from Stocks versus Flows.** Visual inspection of Figure 7 Panel 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 vs. flow presentation, we analyze the qualitative outcome (decrease, no change, increase) as a function of condition in an ordered logistic regression.<sup>17</sup> Our primary interest is with the effect of data presentation (stock vs. flow) separately for each trend, so we analyze each pattern separately for expositional ease.<sup>18</sup> In all six cases where the stock trend and flow trend were qualitatively different (that is, 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, 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  $p < .001$ ). Similarly, when the stock trend was more negative than the flow trend (decreasing stock with an increasing flow, decreasing stock with a constant flow, 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  $p < .005$ ). 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; 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.<sup>19</sup>

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

---

<sup>17</sup> Comparisons with parametric results for Studies 4, 5, and 6 are given in Online Appendix D.

<sup>18</sup> All conclusions remain the same when using a single model with either clustered standard errors or random effects to account for non-independence.

<sup>19</sup> 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, not significant when both trends were increasing, and did not converge when both trends were increasing.

more sensitive to differences among flow trends than differences among stock trends. Details for 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 the flow condition ( $M = +0.67$ ,  $SD = 2.29$ ,  $t(400) = 11.36$ ,  $p < .001$ ). 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 (McKenzie & Liersch, 2011; Wagenaar & Sagaria, 1975). Because the trends differ depending on whether the data are presented as stocks or flows, these linear extrapolations of (potentially) non-linear trends lead to systematically and qualitatively different forecasts, differing not only in degree, but also in type (i.e., forecasted increase vs. 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 & 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 Sterman and colleagues. By examining the subset of participants who can accurately read the data, we assess whether the results are driven by *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 *lack of consideration*.

## Method

Six hundred five participants (301 women, 302 men; median age = 33) were recruited from AMT and completed Study 5.<sup>20</sup> 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.<sup>21</sup>

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

---

<sup>20</sup> An additional 142 partial responses were excluded.

<sup>21</sup> Note that while this may be a concern with interpreting differences in magnitude in the previous study, it should not affect differences in sign (positive vs. negative), which was the focus of our analysis.

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 importantly, we included an additional set of measures that assess whether each participant could correctly read and interpret each dataset 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 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 605 participants reported descriptions and forecasts for each of the four patterns, for 2,420 total 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% vs. 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% vs. 30%, though see caveat regarding magnitude of effect in Online Appendix F) and constant flow trends were more likely to be accurately described than varying flow trends (53% vs. 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. In 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) are given in Table 3 for both the accurate subsample as well as the whole sample.

**Qualitatively Different Forecasts from Stocks versus Flows.** As in Study 4, we analyze the signed outcome (decrease, no change, 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; quadratic contrast coded stock = -1, combined = 2, flow = -1. The linear contrast represents the comparison of stock vs. flow; the quadratic contrast represents the comparison of combined vs. 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 dataset are generally stronger than those for the accurate

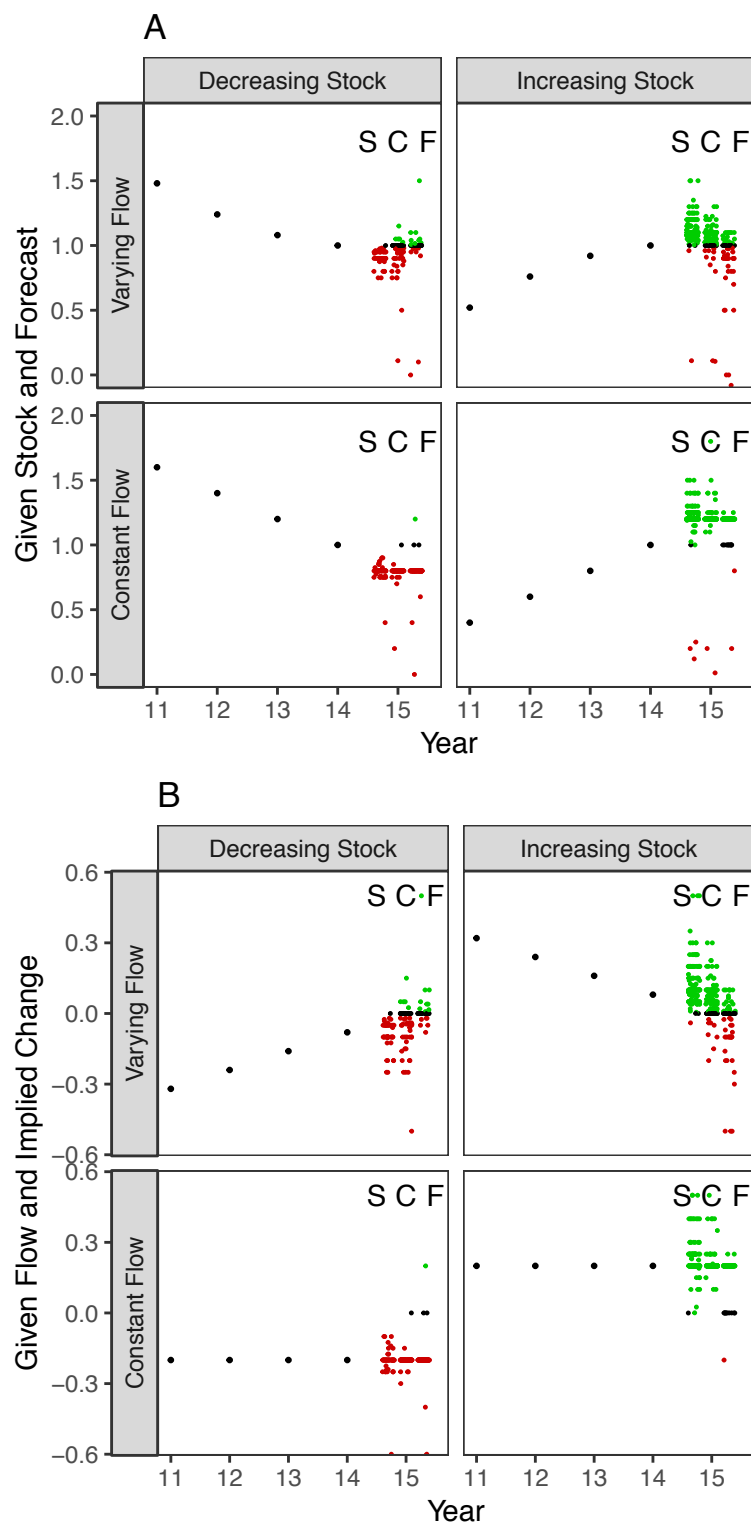


Fig. 8.

Observed stock (Panel A) and flow (Panel B) trends with participant forecasts by condition (S = stock, C = combined, F = flow). Black points indicate no change (1 in A, 0 in B), shaded points indicate a change (positive change > 1 in A, > 0 in B; negative change < 1 in A, < 0 in B). Panels contain the same data on different scales. Figure only includes responses associated with accurate descriptions of the data.

Table 3.

Proportion in each condition (stock, combined, flow) who report a forecast greater than, equal to, or less than the last observed data point for each of four combinations of stocks (increasing, decreasing) and flows (varying, constant). Top panel summarizes responses associated with accurate descriptions of the data. Bottom panel summarizes all responses. When the flow varied, its trend was opposite the stock trend.

		Accurate Responses					
		Decreasing Stock			Increasing Stock		
		Stock	Combined	Flow	Stock	Combined	Flow
Varying Flow	Increase	.04	.08	.21	.97	.68	.25
	No Change	.02	.34	.60	.01	.24	.34
	Decrease	.94	.58	.19	.01	.08	.41
Constant Flow	Increase	.00	.01	.03	.97	.99	.90
	No Change	.00	.01	.03	.01	.00	.09
	Decrease	1.00	.98	.95	.02	.01	.02
		All Responses					
		Decreasing Stock			Increasing Stock		
		Stock	Combined	Flow	Stock	Combined	Flow
Varying Flow	Increase	.03	.08	.53	.96	.72	.15
	No Change	.05	.22	.30	.02	.20	.24
	Decrease	.92	.70	.16	.02	.08	.61
Constant Flow	Increase	.02	.02	.05	.98	.98	.63
	No Change	.00	.00	.26	.00	.00	.32
	Decrease	.98	.98	.70	.02	.01	.05

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 ( $z = 8.33, p < .001$ ), and the combined condition did not differ from the midpoint ( $z = -1.31, p = .19$ ). For an increasing stock with a constant flow, nearly all forecasts indicated an increase, though there was a significant difference between the stock and flow conditions ( $z = 2.37, p = .018$ ), with the combined condition lying marginally significantly above the average of stock and flow ( $z = 1.91, p = .056$ ) such that it differed from the flow condition ( $z = -2.81, p = .005$ ) but not the stock condition (stock  $z = -0.89, p = .38$ ). 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 ( $z = -5.67, p < .001$ ) and again the combined condition did not differ from the midpoint ( $z = 1.04, p = .30$ ). 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 ( $ps < .001$ ), 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 ( $F(2, 602) = 34.54, p < .001$ ) such that it was higher for stock ( $M = +2.55, SD = 1.65$ ) than for combined ( $M = +2.12, SD = 1.60; t(602) = 2.316, p = .021$ ) and higher for combined than for flow ( $M = +1.05, SD = 2.26; t(602) = 5.798, p < .001$ ). 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.

## 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 assessing stocks from flows, especially when that flow varies over time or is negative. Second, we find that while 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 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 upon visual extrapolation or difficulty in reading precise values from the graph axes.

### Method

Four hundred and one participants (194 women, 206 men; median age = 32) were recruited from AMT and completed Study 6.<sup>22</sup> 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

---

<sup>22</sup> An additional 52 partial responses were recorded.

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 vs. flow), presentation type (contrast coded table vs. 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 ( $z = 6.87, p < .001$ ). We also find that the odds of an accurate description are greater when data are presented as a table rather than a graph ( $z = 7.45, p < .001$ ); 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:  $z = -0.91, p = .36$ ). Cell proportions are given in Table 4.

Table 4.  
Proportion of responses providing accurate descriptions of the data.

	Stock	Flow
Graph	.65	.20
Table	.91	.68

**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 (vs. flows) induced more positive forecasts ( $z = 6.50, p < .001$ ), and this did not vary depending on whether the data were presented as graphs ( $z = 4.24, p < .001$ ) or text ( $z = 5.80, p < .001$ ; interaction  $z = -0.89, p = .37$ ). There was no main effect of presentation format ( $z = -1.58, p = .11$ ). Cell proportions are given in Table 5.<sup>23</sup>

<sup>23</sup> When using all of the data, including inaccurate responses, there was an interaction ( $z = 3.16, p = .002$ ), but the simple effect of stock vs. flow was large and significant whether the data were presented as a graph ( $z = 8.15, p < .001$ ) or as a table ( $z = 6.51, p < .001$ ). This interaction can largely be attributed to the

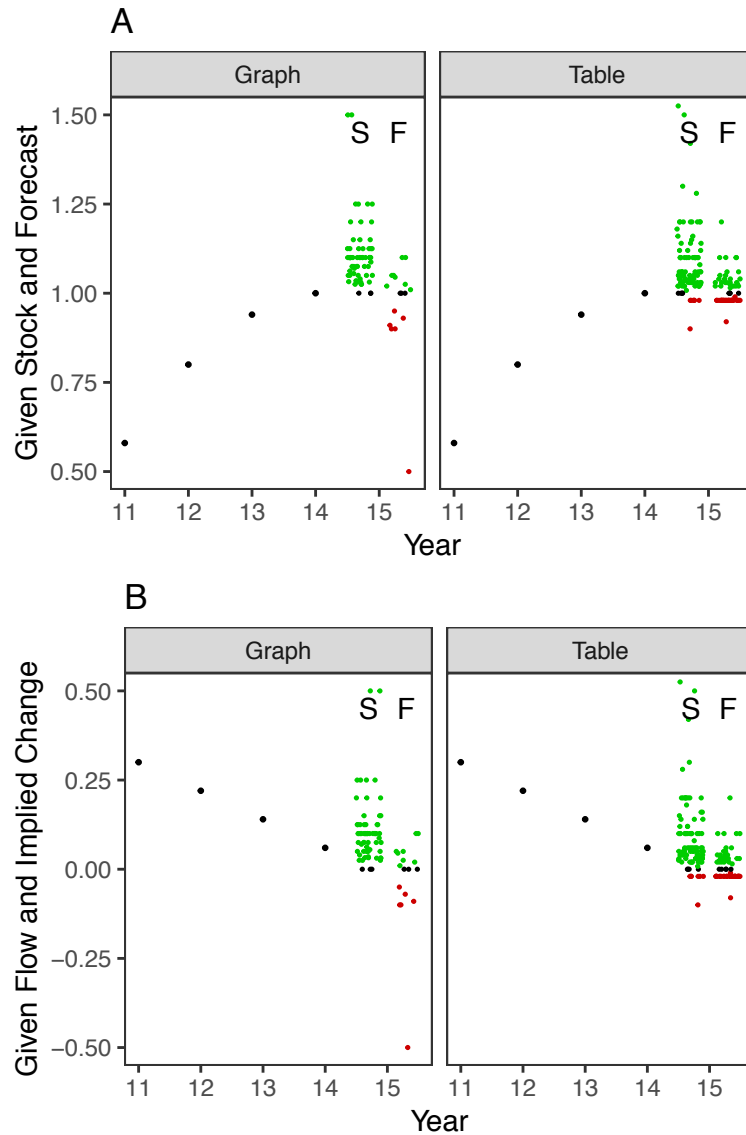


Fig. 9.

Observed stock (Panel A) and flow (Panel B) trends with participant forecasts by condition (S = stock, F = flow) among accurate responses. Black points indicate no change (1 in A, 0 in B), shaded points indicate a change. Panels contain the same data on different scales. Participants in the tabular condition did not see the past data depicted graphically.

**Subjective Understanding.** This study also included a report of subjective understanding. In the full sample, subjective understanding was higher for table than graph ( $t(397) = 8.16, p < .001$ ), higher for

---

fact that the effect is larger among those who are inaccurate, and participants were more inaccurate with the graph than the table.

stock than flow ( $t(397) = 8.27, p < .001$ ), and the difference between stock and flow was greater for graph ( $M_{\text{stock}} = +2.41, M_{\text{flow}} = +0.04, t(397) = 8.89, p < .001$ ) than for table ( $M_{\text{stock}} = +3.09, M_{\text{flow}} = +2.39, t(397) = 2.71, p = .007$ ; interaction  $t(397) = 4.50, p < .001$ ). Even among the subset of participants who accurately described the data, subjective understanding was higher for table than for graph ( $t(246) = 8.68, p < .001$ ), higher for stock than flow ( $t(246) = 7.58, p < .001$ ), and the difference between stock and flow was not only greater for graph ( $M_{\text{stock}} = +2.67, M_{\text{flow}} = -0.61, t(246) = 8.14, p < .001$ ), it was small and non-significant for table ( $M_{\text{stock}} = +3.20, M_{\text{flow}} = +2.92, t(246) = 1.15, p = .25$ ; interaction  $t(246) = 6.41, p < .001$ ). Among those who accurately described the data, subjective understanding did not moderate the effects of stock versus flow presentation format, table versus graph format, or their interaction ( $ps > .1$ ). The estimated simple effect at the highest level of subjective understanding remained significant for a table format ( $z = 4.94, p < .001$ ), though it was not significant for a graph format ( $z = 0.81, p = .42$ ), in large part because the smaller cell size and lower subjective understanding led to larger standard errors (.61 for graph, .23 for table).

Table 5.

Forecast trend by condition and graphical vs. tabular display. Top panel summarizes responses associated with accurate descriptions of the data. Bottom panel summarizes all responses.

Accurate Responses			
		Stock	Flow
Graph	Increase	.94	.44
	No Change	.05	.17
	Decrease	.02	.39
Table	Increase	.86	.42
	No Change	.04	.08
	Decrease	.10	.50
All Responses			
		Stock	Flow
Graph	Increase	.95	.23
	No Change	.04	.08
	Decrease	.01	.70
Table	Increase	.86	.42
	No Change	.04	.05
	Decrease	.10	.53

## 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. Further, as all participants described the data prior to making a forecast, 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 datasets, 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 ( $z = 2.72, p = .007$ ).

## 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., Paich & Sterman 1993; Sterman 1987, 1989). The present investigation underscores the fact that data are never presented in a vacuum, transmitting information in a purely abstracted, neutral form. Rather, in order 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 US 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 either a graph of stocks or flows. However, we still observed inconsistencies when the data were presented numerically (Study 6). This suggests that even non-visual 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 non-graphical means; Cronin et al. 2009; Maglio & Polman, 2016; *cf.* Larkin & Simon, 1987).

While 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 accurate judgments than another (*cf.*, de Langhe & Puntoni, 2016; Larrick & Soll, 2008). Nevertheless, we note that this is a variable that could be used by marketers or other choice architects to influence opinions and decisions (Thaler & 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 that 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.

### References

- Andreassen PB, Kraus SJ (1990) Judgmental extrapolation and the salience of change. *Journal of Forecasting* 9:347-372.
- Andreassen PB (1988) Explaining the price-volume relationship: The difference between price changes and changing prices. *Organizational Behavior and Human Decision Processes* 41:371-389.
- Booth Sweeney L, Sterman JD (2000) Bathtub dynamics: Initial results of a systems thinking inventory. *System Dynamics Review* 16:249-286.
- Bureau of Labor Statistics (2017) *Labor force statistics from the current population survey*. <https://www.bls.gov> (retrieved June 15, 2017).
- Burson KA, Larrick RP, Lynch JG Jr. (2009) Six of one, half dozen of the other: Expanding and contracting numerical dimensions produces preference reversals. *Psychological Science* 1074-1078.
- Camilleri AR, Larrick RP (2014) Metric and scale design as choice architecture tools. *Journal of Public Policy & Marketing* 33:108-125.
- Carlson KA, Shu SB (2007) The rule of three: How the third event signals the emergence of a streak. *Organizational Behavior and Human Decision Processes* 104:113-121.
- Cleveland WS, McGill R (1984) Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association* 79:531-554.
- Cronin MA, Gonzalez C, Sterman JD (2009) Why don't well-educated adults understand accumulation? A challenge to researchers, educators, and citizens. *Organization Behavior and Human Decision Processes* 108:116-130.
- de Langhe B, Puntoni S (2016) Productivity metrics and consumer's misunderstanding of time savings. *Journal of Marketing Research* 53:396-406.
- de Langhe B, Puntoni S, Larrick RP (2017) Linear thinking in a non-linear world. *Harvard Business Review* 95:130-139.

- Duclos, R (2015). The psychology of investment behavior: (De)biasing financial decision-making one graph at a time. *Journal of Consumer Psychology*, 25:317-325.
- Freyd JJ, Finke RA (1984) Representational momentum. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 10:126–132.
- Grice HP (1975) Logic and conversation. *Syntax and Semantics* 3 eds. Cole P & Morgan JL (Academic Press, New York, NY).
- Hohle SM, Teigen KH (2015) Forecasting forecasts: The trend effect. *Judgment and Decision Making* 10:416–428.
- Hsee CK, Abelson RP (1991) Velocity relation: Satisfaction as a function of the first derivative of outcome over time. *Journal of Personality and Social Psychology* 60:341–347.
- Hsee CK, Abelson RP, Salovey P (1991) The relative weighting of position and velocity in satisfaction. *Psychological Science* 2:263–266.
- Huff D (1954) *How to Lie with Statistics* (Norton, New York, NY).
- IDC (2014) The digital universe of opportunities: Rich data and the increasing value of the internet of things. Retrieved 7/5/2017 from <https://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm>.
- Kahneman D (2011) *Thinking, Fast and Slow* (Farrar, Strause and Giroux, New York, NY).
- Kahneman D, Fredrickson BL, Schreiber CA, Redelmeier DA (1993) When more pain is preferred to less: Adding a better end. *Psychological Science* 4:401–405.
- Larkin, JH, Simon, HA (1987) When a diagram is (sometimes) worth ten thousand words. *Cognitive Science* 11:65–99.
- Larrick RP, Soll JB (2008) The MPG illusion. *Science* 320:1593–1594.
- Loewenstein GF, Prelec D (1993) Preferences for sequences of outcomes. *Psychological Review* 100:91–108.
- Lurie NH, Mason CH (2007) Visual representation: Implications for decision making. *Journal of Marketing* 71:160–177.

- Maglio SJ, Polman E (2014) Spatial orientation shrinks and expands psychological distance. *Psychological Science* 25:1345–1352.
- Maglio SJ, Polman E (2016) Revising probability estimates: Why increasing likelihood means increasing impact. *Journal of Personality and Social Psychology* 111:141–158.
- Maglio SJ, Trope Y (2011) Scale and construal: How larger measurement units shrink length estimates and expand mental horizons. *Psychonomic Bulletin and Review* 18:165–170.
- McKenzie CRM, Nelson JD (2003) What a speaker's choice of frame reveals: Reference points, frame selection, and framing effects. *Psychonomic Bulletin and Review* 10:596–602.
- McKenzie CRM, Liersch MJ (2011) Misunderstanding savings growth: Implications for retirement savings behavior. *Journal of Marketing Research* 48:1–13.
- Pacini R, Epstein S (1999) The relation of rational and experiential information processing styles to personality, basic beliefs, and ratio-bias phenomenon. *Journal of Personality and Social Psychology* 76:972–987.
- Paich M, Sterman, JD (1993) Boom, bust, and failures to learn in experimental markets. *Management Science*, 39:1439-1458.
- Parker JF, Paul I, Reinholtz N (2016) Perceived momentum influences responsibility judgments. *Proceedings of the 38<sup>th</sup> Annual Conference of the Cognitive Science Society* eds. Papafragou A, Grodner D, Mirman D, Trueswell JC (Cognitive Science Society, Austin, TX).
- Pelham BW, Sumarta TT, Myaskovsky L (1994) The easy path from many to much: The numerosity heuristic. *Cognitive Psychology* 26:103–133.
- Pettit NC, Sivanathan N, Gladstone E, Carson Marr J (2013) Rising stars and sinking ships: Consequences of status momentum. *Psychological Science* 24:1579–1584.
- Raghubir P, Srivastava J (2002) Effect of face value on product valuation in foreign currencies. *Journal of Consumer Research*, 29:335-347.
- Reyna VF, Brainerd CJ (2008) Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. *Learning and Individual Differences* 18:89–107.

- Slovic P (1972) *From Shakespeare to Simon: Speculations—and Some Evidence—about Man's Ability to Process Information* (Oregon Research Institute).
- Spiller SA, Fitzsimons GJ, Lynch Jr JG, McClelland GH (2013) Spotlights, floodlights, and the magic number zero: Simple effects tests in moderated regression. *Journal of Marketing Research*, 50: 277-288.
- Sterman, JD (1987) Testing behavioral simulation models by direct experiment. *Management Science* 33:1572-1592.
- Sterman, JD (1989) Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment. *Management Science* 35:321-339.
- Sterman JD (2002) All models are wrong: Reflections on becoming a systems scientist. *System Dynamics Review* 18:501–531.
- Sterman JD, Booth Sweeney (2002) Cloudy skies: Assessing public understanding of global warming. *System Dynamics Review* 18:207-240.
- Sterman JD, Booth Sweeney L (2007) Understanding public complacency about climate change: Adults' mental models of climate change violate conservation of matter. *Climatic Change* 80:213-238.
- 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. *International Conference on Thinking*.
- Tufte ER (2001) *The Visual Display of Quantitative Information* (Graphics Press LLC, Cheshire, CT).
- Wagenaar WA, Sagaria SD (1975) Misperception of exponential growth. *Perception and Psychophysics* 114:830–831.