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.

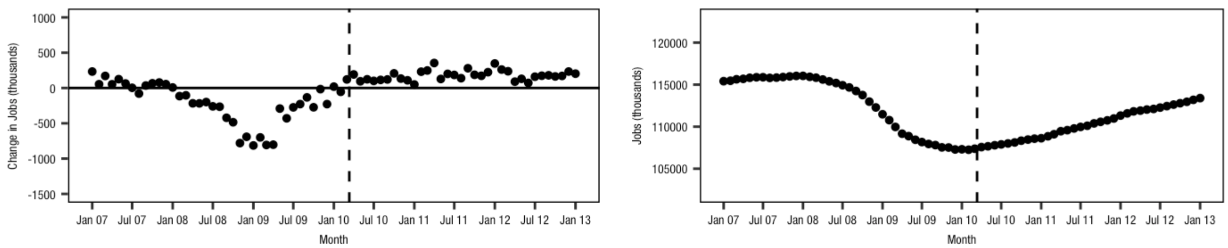


Fig. A1.
Job charts used in Study A1. The flow chart on the top shows the flow of jobs (jobs gained or lost). The stock chart on the bottom shows the same data presented as the stock (total number of jobs). The vertical dashed line indicates the passage of the Affordable Care Act.

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 (ps > .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; $t(297) = 0.70$, $p = .486$).

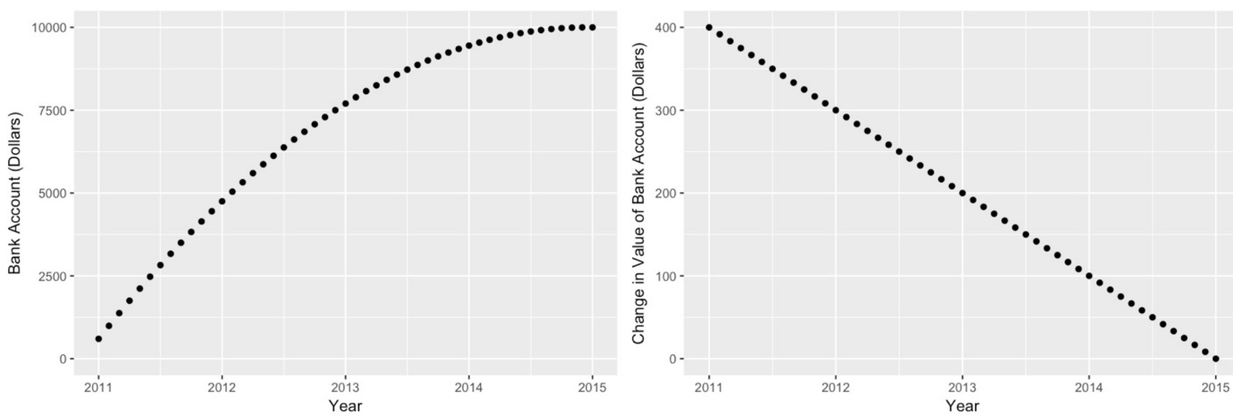
Subjective importance ratings did not significantly vary by condition (ps > .05). 73% of participants responded with top two responses ("somewhat" or "to a great extent") that major health care laws have the potential to impact the economy shortly after they become law.

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.