Supplementary Information for

A Strategy to Improve Expert Technology Forecasts

Tamara Savage, Alex Davis, Baruch Fischhoff, and M. Granger Morgan

Corresponding author: M. Granger Morgan
Email: granger.morgan@andrew.cmu.edu

This PDF file includes:

Supplementary text
Figures S1 to S13
Tables S1 to S20
Appendices
SI References

Table of Contents

I. Methods

As a first step in developing the improved forecasting method, a survey was developed in which respondents made forecasts about electric vehicles (EVs) and autonomous vehicles (AVs). Both EVs and AVs are emerging technologies, as defined by Rotolo et al. (2015): novel, relevant, and impactful, with the potential for change and disruption. They are also domains with many people involved as workers or enthusiasts, providing a large pool of potential survey participants. The entire survey can be found in Appendix A. The participants' task was to assess 80% interval estimates for the current value and the forecasted value in two years for ten questions related to EVs and AVs. The full list of questions and current values is provided in Table S1 and an example of the user interface is shown in Figure S2. Respondents were randomly divided into three groups: a control and two treatment groups. The *one-briefing treatment group* was given a briefing on the general concept of overconfidence (a copy of this briefing is provided in Appendix B and at https://youtu.be/0XuZ8Q6pxZE). The *two-briefing treatment group* was given the same briefing on overconfidence and then was given a second briefing on the poor performance of past forecasts and advice on how to consider policy, economic, and social factors that might affect their forecasts (a copy of this briefing is provided in Appendix C and at https://youtu.be/mQN1jt1dklU). Both briefings were provided as short recorded PowerPoint presentations. The *control group* did not see either of the briefings. The level of overconfidence in each of the groups was assessed. Figure S1 below shows the sequence of the survey elements. Data were collected in November and December 2019.

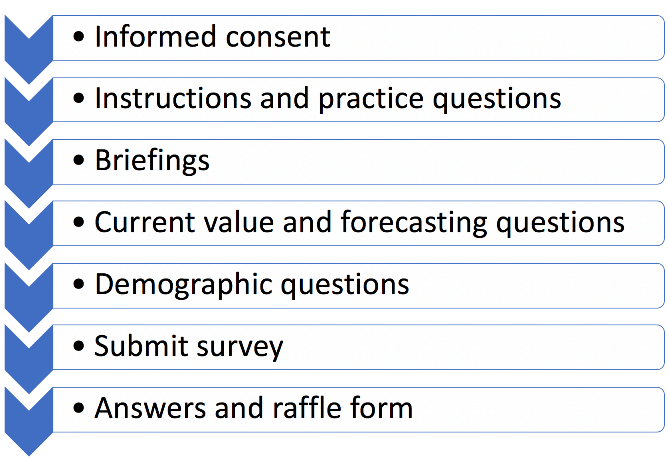


Figure S1: The sequence of survey elements.

1a. **Currently, what is the longest EPA-rated range (in miles) for a battery electric vehicle?** Only consider passenger vehicles that are offered for retail sale in the U.S. The EPA-rated range is the official measure of how far the vehicle can drive on one full battery charge.

Lower estimate for current value (only a 10% chance the value is lower)

Higher estimate for current value (only a 10% chance the value is higher)

1b. **At the end of 2021, what will be the longest EPA-rated range (in miles) for a battery electric vehicle?** Only consider passenger vehicles that are offered for retail sale in the U.S. The EPA-rated range is the official measure of how far the vehicle can drive on one full battery charge.

Your range estimate for the current value is to miles.

Lower estimate for 2021 value (only a 10% chance the value is lower)

Higher estimate for 2021 value (only a 10% chance the value is higher)

Figure S2: Example motor vehicle technology question, eliciting both the current value and the forecasted value.

Table S1: Motor vehicle technology survey questions and current value answers.

Motor Vehicle Technology Question	Current Answer
1. **Currently, what is the longest EPA-rated range (in miles) for a battery electric vehicle?** Only consider passenger vehicles that are offered for retail sale in the U.S. The EPA-rated range is the official measure of how far the vehicle can drive on one full battery charge.	370 miles (Tesla Model S) Wikipedia: https://en.wikipedia.org/wiki/Tesla_Model_S Kelley Blue Book: https://www.kbb.com/car-reviews-and-news/top-10/longest-range-electric-cars/2100006708/?slide=1
2. **Currently, what is the average cost for a lithium-ion battery pack (in $/kWh)?** Estimate the average industry-wide, volume-weighted cost for passenger electric vehicles that are sold in the U.S.	$180.5/kWh (average of Bloomberg New Energy Finance and Department of Energy) BNEF ($176/kWh): https://about.bnef.com/blog/behind

	-scenes-take-lithium-ion-battery-prices/ DOE ($185/kWh): https://www.energy.gov/eere/articles/new-lab-investment-incubator-program-helps-jumpstart-manufacturing-innovation
3. **Currently, how many companies have a permit issued by the California DMV for testing of autonomous vehicles <u>with a driver</u> on public roads in California?** By law (enacted in 2014), a permit is required before a company may test AVs on public roads. A company is granted one permit and can list multiple vehicles on the permit.	64 companies (California DMV) https://www.dmv.ca.gov/portal/dmv/detail/vr/autonomous/permit https://www.dmv.ca.gov/portal/dmv/detail/vr/autonomous/testing
4. **Currently, how many companies have a permit issued by the California DMV for <u>driverless</u> testing of autonomous vehicles on public roads in California?** Driverless testing means the AV is capable of operating without the presence of a driver inside the vehicle. By law (enacted in 2018), a permit is required before a company may test driverless AVs on public roads. A company is granted one permit and can list multiple vehicles on the permit.	1 company (California DMV) https://www.dmv.ca.gov/portal/dmv/detail/vr/autonomous/driverlesstestingpermits https://www.dmv.ca.gov/portal/dmv/detail/vr/autonomous/auto
5. The California DMV requires that a report be filed for every collision resulting in any damage of property or in bodily injury or death when testing an autonomous vehicle under a company's testing permit (with a driver or driverless). **In 2018, how many reports of collisions involving autonomous vehicles did the California DMV receive?**	75 reports (California DMV) https://www.dmv.ca.gov/portal/dmv/detail/vr/autonomous/autonomousveh_ol316
6. **Currently, how many public DC fast charging <u>outlets</u> are available in the U.S.?** Include CCS, CHAdeMO, and Tesla fast charging outlets.	11,778 outlets (DOE Alternative Fuels Data Center) https://afdc.energy.gov/stations/#/analyze?country=US&fuel=ELEC&ev_levels=dc_fast
7. **In 2018, how many hydrogen fuel cell passenger vehicles were sold in the U.S.?**	2,332 vehicles (Inside EVs) https://insideevs.com/news/342378/hydrogen-fuel-cell-car-sales-in-us-just-2300-in-2018/
8. **Currently, how many battery electric public buses are there in the U.S.?** Do not include buses/trolleys with overhead power cables.	1,100 buses (Electric Power Research Institute) https://www.epri.com/#/?lang=en-US
9. **By the end of 2018, how many U.S. states had authorized some form of automated vehicle platooning on public roads?** Platooning refers to a	17 states (Competitive Enterprise Institute) https://cei.org/content/authorizing-

group of vehicles (with or without a driver) that can communicate and therefore can drive very closely together and accelerate and brake simultaneously. There is currently a patchwork of state laws regarding platooning; states can authorize full use of platooning on public roads, only testing, or not allow platooning.	automated-vehicle-platooning-2019
10. **In 2018, what percentage of new passenger vehicles sold in the U.S. were electric vehicles?** Please enter a number between 0 and 100. Include both plug-in hybrid electric vehicles and battery electric vehicles.	2.1% (EV-Volumes) http://www.ev-volumes.com/country/usa/

Recruitment & Incentive

Participants were recruited from several places. The survey was posted on Reddit (electric vehicle and autonomous vehicle related sub-Reddits) and Twitter. Emails were sent to university research centers (such as Carnegie Mellon's Traffic21 lab) and to members of the National Academies' Transportation Research Board. The survey link was included in one issue of the Future of Transportation e-newsletter. Paid advertisements were posted on AutomotiveNews.com and in SAE International's eSource newsletter. Participants were also recruited through snowball sampling, in which people were asked to forward the survey to others who might be interested in taking it. As an incentive to participate, respondents who completed the survey could fill out a raffle form for a chance to win one of ten $50 Amazon gift cards.

Instructions

All participants first provided informed consent, which included affirming that they were at least 18 years old and at least somewhat knowledgeable about motor vehicle technology in the U.S. Those who passed were shown a 1-minute instruction video on what an 80% interval estimate is, along with an example question (see Appendix D or https://youtu.be/o4EPyeKFZnw for the instructions). The video was followed by an attention check question. Participants were asked to not look up answers online or ask anyone for advice while filling out the survey and had to click on a button agreeing to those conditions. They were then given two practice questions (without receiving answers) to familiarize themselves with the task and the user interface. Respondents were then randomly assigned to one of three groups using the randomizer in Qualtrics: control, one-briefing, or two-briefing. The control group was routed directly to the first motor vehicle technology question.

Treatment Groups & Briefing Materials

The first briefing (shown to both the one-briefing group and the two-briefing group) was two minutes long and provided information on what the behavioral bias of overconfidence is. This briefing is provided in Appendix B and at https://youtu.be/0XuZ8Q6pxZE. They were also shown a few illustrative examples. The first was Soll and Klayman's study (2004), where they elicited 80% interval estimates for general knowledge questions; only about 50% of those intervals contained the correct answer. The second example was the summary of Surprise Index results from Morgan (2017), which showed how pervasive overconfidence is in many studies of interval estimates. Participants were then instructed to be careful to avoid being overconfident when making their interval estimates and to consider widening their intervals to account for being overconfident. After the briefing, respondents were asked an attention check question. The one-briefing group was then routed to the motor vehicle technology questions.

The two-briefing group watched a second briefing that was six minutes long and showed experts' poor performance in past technology forecasting and provided advice on how to think more systematically about policy, economic, and social developments that could affect the outcome being forecasted. This briefing is provided in Appendix C and at https://youtu.be/mQN1jt1dklU. The examples included predictions of the price of solar electricity (Baker et al., 2015; Verdolini et al. 2015; Lazard, 2018), technology development (Fye et al., 2013), and investment returns (Ben-David et al., 2013). The briefing also showed examples of how policy, economic, and social factors could affect the forecasted events, such as changes in government regulations, gasoline prices, availability of EV charging stations, AV crashes, and ridesharing. Respondents were then encouraged to think about how these factors would affect their higher and lower estimates. For example, if gas prices increase, more people may be interested in buying EVs (increasing the higher estimate of the proportion of new cars sold that are EVs), but if gas prices decrease, people will probably buy more gas cars instead (decreasing the lower estimate of the proportion of new cars sold that are EVs). They were also encouraged to think of factors they could add to this list. After the briefing, respondents were asked an attention check question. This group was then routed to the motor vehicle technology questions.

Motor Vehicle Technology Questions

When answering the ten motor vehicle technology questions, respondents made an 80% interval estimate for the current value and then the forecasted value at the end of 2021, two years in the future. All respondents answered the questions in the same order. Of the ten questions, six were about EVs and four were about AVs. For the final two questions, respondents were asked to write down what they had thought about when answering the forecasting questions, to elicit broader policy, economic, and social factors that might have influenced their forecasts. The survey questions were formulated with the help of experts at Carnegie Mellon and the Department of Energy. The questions were designed to have publicly available answers, so we could see which intervals contained the answers and which did not.

Demographic Questions & Survey Submission

After completing the ten motor vehicle technology questions, respondents answered questions designed to help characterize the sample's expertise. These included self-assessments of respondents' expertise in EVs and AVs, a basic knowledge check question about each technology, and indicated whether they were employed working on motor vehicle technology or were an enthusiast or hobbyist. They then answered standard demographic questions about their age, gender, income, and level of education. After they submitted their survey, they were shown the current value answers along with their current value interval estimates, so they could see how well they did. A link on that page directed them to a separate Google Form where they could enter the raffle if they wished; this was done to preserve the anonymity of the survey.

II. Data Analysis & Results

We analyzed the effects of the briefings on three measures. The primary measure was the average hit rate (proportion correct) of each group. The two secondary measures we analyzed were interval widths and the number of policy, economic, and social factors keywords used by respondents. A proposed data analysis plan was pre-registered on Open Science Framework (see Appendix E or osf.io/vxdmb). The multilevel model analysis was not included in the pre-registered analysis plan. The data were analyzed using R and Excel. Two-sided statistical tests were used as they are more conservative and provided consistent tests for cases where there were and were not directional hypotheses.

Survey Participants & Demographics

A total of 133 participants completed the survey and were included in the data analysis, with 45 in the control group, 45 in the one-briefing treatment group, and 43 in the two-briefing treatment group. As seen in Table S2, most were recruited through Reddit, emails to their university research center, or a friend or colleague forwarding the survey to them. There were 106 men and 26 women. Reported ages ranged from 18 to 76 (median 34); 89% had a Bachelor's degree or higher. About 73% self-assessed their level of expertise about electric vehicles as "medium" or "high" and for autonomous vehicles this number was 58% (see Table S3 below for the breakdown). The demographics were also checked to verify that the randomization in Qualtrics worked.

Table S2: Respondents recruited from each source (sum is greater than 133 because some respondents heard about the survey from multiple sources).

Recruitment Source	**Number of Respondents**
Reddit	46 (35%)
Forwarded by a friend or colleague	45 (34%)
University research center or lab	27 (20%)
Future of Transportation weekly email	6 (5%)
Twitter	5 (4%)
SAE International advertisement	5 (4%)
AutomotiveNews.com advertisement	2 (2%)

Table S3: Respondents' self-assessed level of expertise about electric vehicles and autonomous vehicles.

		Self-Assessed Level of Expertise			
		None	Low	Medium	High
Electric vehicles	Technical and engineering	2%	27%	47%	24%
	Policy, economic, and social factors	1%	26%	52%	22%
Autonomous vehicles	Technical and engineering	3%	40%	38%	20%
	Policy, economic, and social factors	4%	40%	43%	14%

Attention & Knowledge Checks

Depending on their group, respondents were given three, four, or five attention and knowledge checks during the survey. All respondents answered the attention check on the instructions video and the two basic knowledge check questions (one each on EVs and AVs). Respondents in both treatment groups answered an attention check on the first briefing on general overconfidence. The two-briefing group answered an attention check for the second briefing. 77% of participants passed all of the attention and knowledge checks. Table S4 shows the respondents' performance on these questions. All respondents who completed all tasks were included in the statistical analyses. Their performance on the checks serves as a measure of their attention and knowledge, as well as the clarity of the materials. Table S5 below shows the amount of time it took the respondents to complete the survey, both including and excluding the time it took to watch the briefings. Most participants spent at least 14 minutes on the survey (a reasonable amount of time). The full timing distributions are in Appendix F.

Table S4: Percentage of respondents who passed each attention or knowledge check.

Check	**Group of Respondents**	**Passed**	**Failed**
Instructions on 80% interval estimates	All respondents	94%	6%
General overconfidence briefing	One-briefing and two-briefing	94%	6%
Briefing on forecasting and broader factors	Two-briefing	84%	16%
Electric vehicles knowledge	All respondents	96%	4%

Autonomous vehicles knowledge	All respondents	92%	8%

Table S5: The median amount of time (in minutes) respondents spent taking the survey, including and excluding the time spent on the briefings.

		Timing Data		
		Control	**One-Briefing**	**Two-Briefing**
Total time [min]	**1st quartile**	14	18	30
	Median	19	20	45
	3rd quartile	32	30	98
Total time minus time spent on briefings [min]	**1st quartile**	14	16	23
	Median	19	19	36
	3rd quartile	32	27	88

Hit Rates Analysis

For appropriately confident respondents, 8 out of 10 answers would fall inside their 80% interval estimates. A hit rate (the proportion correct) was calculated for each respondent. Answers that fell on an endpoint were treated as correct.

Hypotheses:

For the current values, we hypothesized that both treatment groups would have higher hit rates than the control group. We had no *a priori* hypothesis regarding the hit rates for the two treatment groups. For the forecasted values, we hypothesized that the two-briefing group would have hit rates higher than the one-briefing group, and both treatment groups would have higher hit rates than the control group. We plan to test this hypothesis in two years once the forecasted values have been realized.

The hit rates for each group are shown in Figure S3 below, with summary statistics in Table S6. The bottom row of Table S6 shows the results of one-sample t-tests (with $\mu = 0.8$), showing that the hit rates for all three groups were significantly lower than the ideal hit rate of 0.8, indicating that all three groups were overconfident. For checking the assumptions of the t-test, see Appendix F.

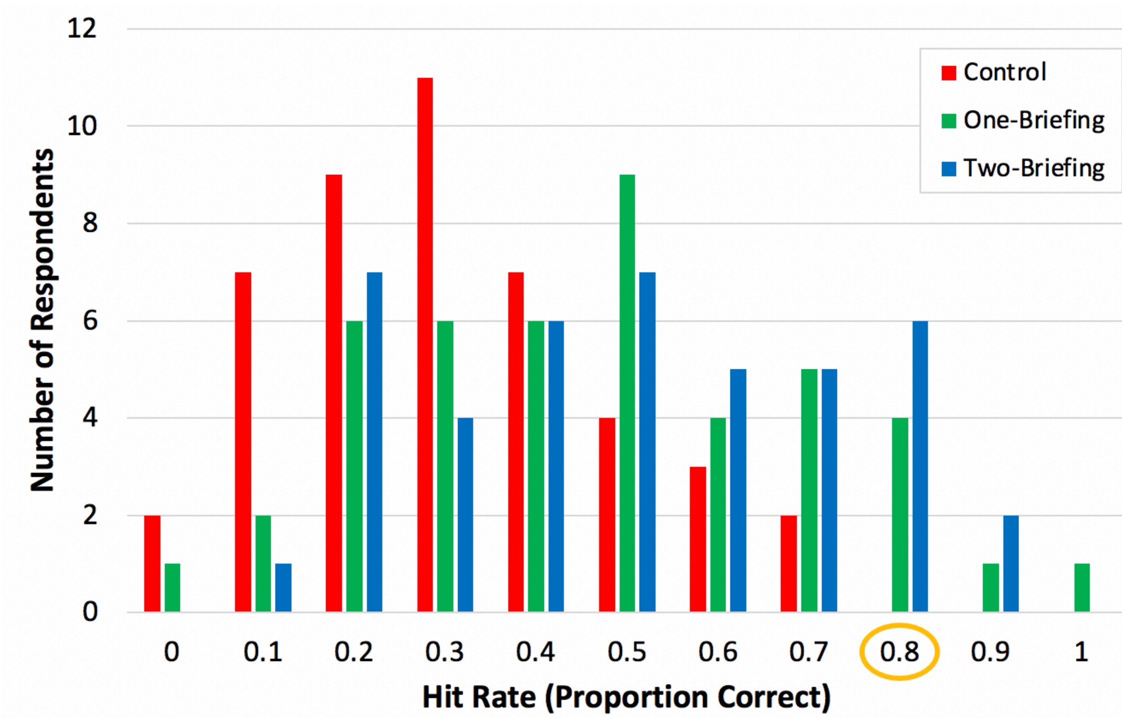


Figure S3: A histogram showing the distribution of hit rates in each group. The appropriately confident hit rate is 0.8 (circled in yellow). Respondents with hit rates below 0.8 are overconfident.

Table S6: Summary statistics for the distribution of hit rates for each group. Statistically significant one-sample t-tests are highlighted in green.

	Hit Rate Summary Statistics		
	Control	**One-Briefing**	**Two-Briefing**
Median	0.30	0.50	0.50
Mean	0.31	0.47	0.50
Standard deviation	0.18	0.23	0.22
Variance	0.031	0.054	0.050
One-sample t-test p-value	$p < 0.0001$	$p < 0.0001$	$p < 0.0001$
One-sample Wilcoxon test p-value	$p < 0.0001$	$p < 0.0001$	$p < 0.0001$

After verifying that the assumptions needed for an ANOVA were met (see Appendix F for assumptions and alternative tests), we ran a one-way ANOVA[1], which identified a significant difference ($p < 0.0004$). To check which means were different, we then ran a Tukey HSD (honest

[1] Note that this was a departure from the pre-registered data analysis plan. See Appendix E and Appendix F for details.

significant difference) test, and the results of this test are in Table S7. Both treatment groups had significantly higher hit rates than the control group ($p = 0.0012$ for one-briefing and $p = 0.00007$ for two-briefing). The hit rates of the two treatment groups were not significantly different. Thus, the treatment groups were less overconfident than the control group, but were still overconfident.

Table S7: P-values for the Tukey HSD test for pairwise comparisons of each group's hit rates ($p < 0.05$ highlighted in green). There was a significant difference in the hit rates for the control and one-briefing group and for the control and two-briefing group.

	Comparison		
	Control and One-Briefing	**Control and Two-Briefing**	**One-Briefing and Two-Briefing**
P-value	0.0012	0.00007	0.71

Table S8 shows the hit rates per question for each group. As shown, the two treatment groups did better than the control group on all questions. The two-briefing group did better than the one-briefing group on 6 out of 10 questions, although the average hit rates for the two treatment groups were not significantly different.

Table S8: Hit rates per question for each group (i.e., the proportion of respondents who got that question correct in each group).

	Hit Rate			
	Control	**One-Briefing**	**Two-Briefing**	**All**
Q1: EV range	0.56	0.73	0.72	0.67
Q2: Battery cost	0.33	0.47	0.44	0.41
Q3: Permits with driver	0.09	0.13	0.21	0.14
Q4: Permits without driver	0.58	0.60	0.70	0.62
Q5: Collision reports	0.29	0.42	0.47	0.39
Q6: DCFC outlets	0.24	0.29	0.44	0.32
Q7: Hydrogen cars	0.29	0.44	0.53	0.42
Q8: Electric buses	0.11	0.58	0.47	0.38
Q9: Platooning states	0.13	0.36	0.40	0.29
Q10: Percentage EVs	0.44	0.67	0.67	0.59

Interval Width Analysis

Each interval width was calculated by subtracting the respondent's lower estimate from their higher estimate. This was done for each current value question and each forecasting question.

Hypotheses:

My hypotheses for the interval widths were the same as for the hit rates. Wider intervals should include more correct answers, hence higher hit rates.

Table S9 shows the median interval width for each current value and forecasting question. Figure S4 (on the following page) shows a side-by-side comparison of the groups' intervals for two questions.

Figure S4: Example visualization of interval widths for Question 9: number of states that have authorized automated platooning (top) and Question 1: maximum range of an EV (bottom). Each line represents one respondent's interval. The actual answers are shown with black horizontal lines (17 states and 370 miles).

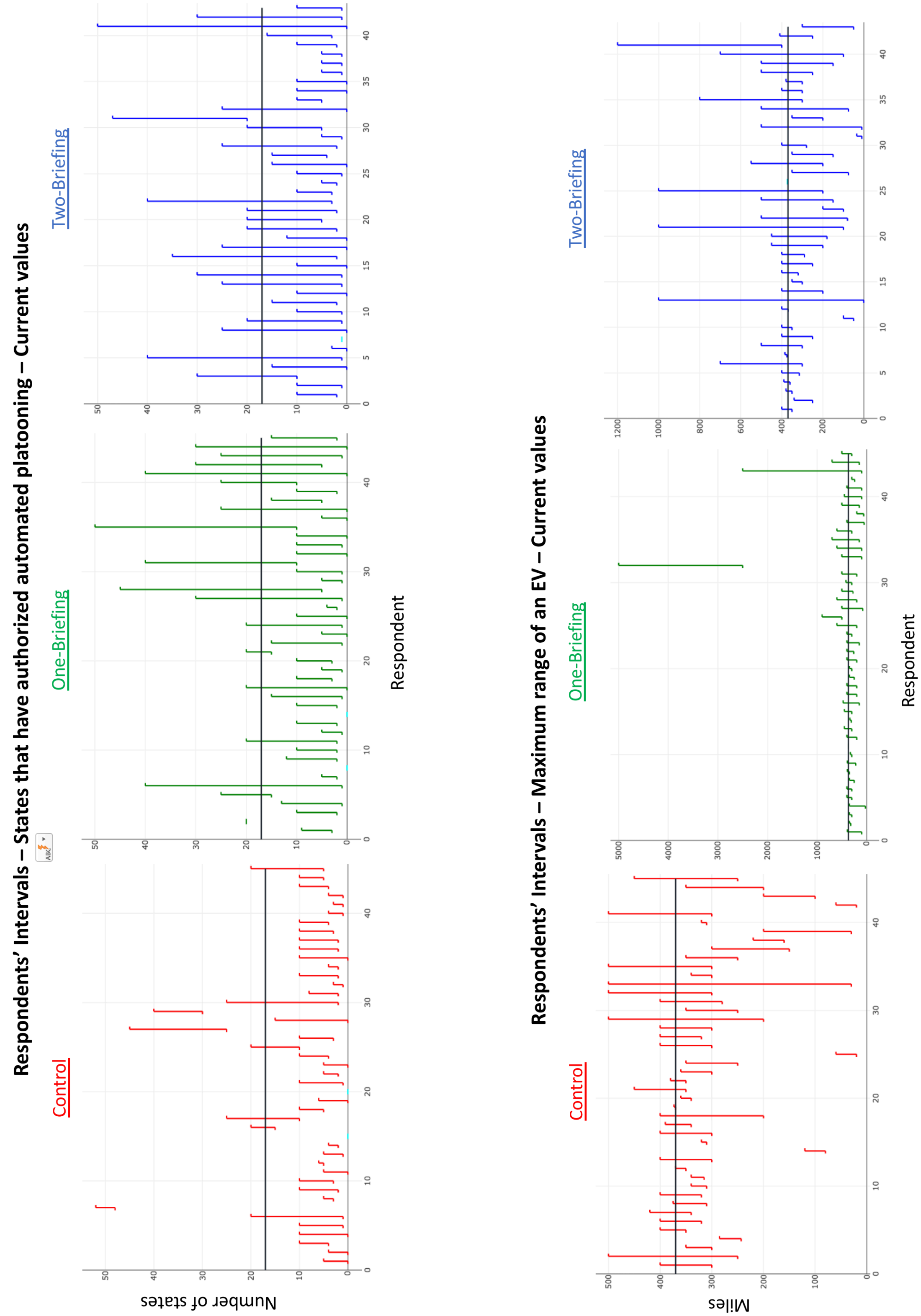


Table S9: The median interval width for each question for the three groups.

		Interval Width		
		Control	**One-Briefing**	**Two-Briefing**
Q1: EV range [miles]	Current	80	200	160
	Forecast	100	200	275
Q2: Battery cost [$/kWh]	Current	100	180	200
	Forecast	100	130	150
Q3: Permits with driver	Current	10	15	20
	Forecast	15	23	35
Q4: Permits without driver	Current	3	6	9
	Forecast	6	14	19
Q5: Collision reports	Current	35	45	60
	Forecast	70	140	390
Q6: DCFC outlets	Current	4,000	4,700	9,000
	Forecast	8,000	10,000	24,999
Q7: Hydrogen cars	Current	990	2,000	3,500
	Forecast	1,400	4,000	9,500
Q8: Electric buses	Current	600	2,650	4,000
	Forecast	1,800	8,500	8,000
Q9: Platooning states	Current	6	10	13
	Forecast	9	20	24
Q10: Percentage EVs [%]	Current	4	4.5	7
	Forecast	5	8	14

We ran pairwise comparisons on the interval widths for each question, to see if there were statistically significant differences between the three groups. Because these distributions were not normal (as you can see in Figure S4 and the further analysis in Appendix F), we compared medians with the non-parametric permutation test. The permutation test has 100% power (Siegel and Castellan, 1988) and is distribution-free. For more on the distributions and the permutation test, see Appendix F. The results from this analysis are shown below in Table S10, using the Bonferroni correction on the standard p-value of 0.05 (0.05 divided by the 60 tests is 0.00083).

Table S10: P-values for pairwise comparisons of the interval widths between groups. $P < 0.00083$ is highlighted in green (this is with the Bonferroni correction) and $p < 0.05$ is marked with *.

	P-Values		
	Control and One-Briefing	**Control and Two-Briefing**	**One-Briefing and Two-Briefing**

Q1: EV range [miles]	Current	0.00050	0.0030*	0.45
	Forecast	0.046*	0.00040	0.62
Q2: Battery cost [$/kWh]	Current	0.20	0.21	0.79
	Forecast	0.19	0.14	0.68
Q3: Permits with driver	Current	0.40	0.014*	0.40
	Forecast	0.13	0.014*	0.56
Q4: Permits without driver	Current	0.0018*	0.00010	0.39
	Forecast	0.011*	0.00010	0.23
Q5: Collision reports	Current	0.40	0.15	0.84
	Forecast	0.22	0.0010*	0.17
Q6: DCFC outlets	Current	0.66	0.13	0.38
	Forecast	0.64	0.056	0.068
Q7: Hydrogen cars	Current	0.28	0.14	0.50
	Forecast	0.14	0.0011*	0.052
Q8: Electric buses	Current	0.029*	0.039*	0.42
	Forecast	0.00050	0.0026*	1.0
Q9: Platooning states	Current	0.012*	0.00010	0.30
	Forecast	0.0017*	0.00010	0.75
Q10: Percentage EVs [%]	Current	0.82	0.054	0.32
	Forecast	0.19	0.00030	0.089

Table S11: Summary of the number of statistically significant differences in interval widths between the three groups.

	Control and One-Briefing	**Control and Two-Briefing**	**One-Briefing and Two-Briefing**
Current ($p<0.05$)	4 out of 10	5 out of 10	0 out of 10
Current (Bonferroni)	1 out of 10	2 out of 10	0 out of 10
Forecast ($p<0.05$)	4 out of 10	8 out of 10	0 out of 10
Forecast (Bonferroni)	1 out of 10	4 out of 10	0 out of 10

In comparison to the control group (at the Bonferroni level), the one-briefing group had significantly greater interval widths for one current value question and one forecasting question. In comparison to the control group (at the Bonferroni level), the two-briefing group had significantly greater interval widths for two current value questions and four forecasting questions (see Table S11). There were no significant differences in interval widths between the one-briefing and two-briefing groups, for the current value questions or the forecasting questions. These results indicate that at least sometimes, respondents in the treatment groups provided wider intervals than respondents in the control group. This corresponds to the same pattern that was seen with the hit rates for the current values.

Exploratory Post Hoc Analysis: Multilevel Model

To improve the statistical power, we conducted an exploratory post hoc multilevel model with per-person random effects to account for the repeated questions answered by each participant. This analysis was not pre-registered. See Appendix F for assumptions and additional details on this analysis.

Because this model assumes normal data, we applied a $\log_{10}(x+1)$ transform to the forecasted interval widths, then standardized the data for each question separately (across the three groups) by subtracting the mean and dividing by the standard deviation of the transformed intervals.

As seen in Table S12, the intervals were significantly wider for both briefing groups than for the control group. They were also wider for the two-briefing group than for the one-briefing group. That pattern appeared both with and without outliers (observations more than 3 standard deviations from the log-standardized mean). These results indicate that the briefings had the desired effect of widening the respondents' intervals and reducing overconfidence.

Table S12: Difference between the three treatment groups when performing the post hoc analysis using a multi-level model on the forecast interval widths. P-values that are significant are highlighted in green.

	Control vs. One-Briefing	**Control vs. Two-Briefing**	**One-Briefing vs. Two-Briefing**
Including outliers	$p = 0.002$ (df = 88, $t = 3.159$)	$p < 0.001$ (df = 86, $t = 5.157$)	$p = 0.045$ (df = 86, $t = 2.034$)
Outliers removed	$p < 0.001$ (df = 88, $t = 3.405$)	$p < 0.001$ (df = 86, $t = 5.566$)	$p = 0.030$ (df = 86, $t = 2.201$)

Keyword Counting Analysis

We also analyzed the number of policy, economic, and social factors keywords used by each respondent when they were asked, "What did you think about as you made your forecast?"

on the last two motor vehicle technology questions. This was intended to get the respondents to describe what they thought about as they came up with their estimates, to see if the second briefing was effective at getting respondents to think about policy, economic, and social factors as they made their forecasts. Keywords were pooled across policy, economic, and social factors, as the briefing treated these factors together. The keyword counting was done in Excel with the SEARCH function, using a list of keywords compiled as the survey was designed and after an initial reading of the survey responses (see Appendix G for details and the full list of keywords). Table 12 shows example keywords. A few example responses from participants are given in Table 13. We ran pairwise comparisons of the median number of keywords used by each respondent for each group for the two questions using the permutation test.

Hypotheses:

We predicted that respondents in the two-briefing group would use more keywords than the control and one-briefing group, because this group had been encouraged to think about policy, economic, and social factors when making their forecasting interval estimates. We did not think there would be a difference between the control and the one-briefing group.

Table S13: Examples of policy, economic, and social factors keywords.

Policy	**Economic**	**Social**
Regulations	Price	Acceptance
Election	Consumer	Environment
Lobbying	Demand	Behavior
Politics	Subsidy	Ethics
Progressive	Expensive	Perception
Infrastructure	Marketing	Collision
Government	Jobs	Privacy

Table S14: Example participant responses to the question, "What did you think about as you made your forecast?" Keywords are shown in orange.

Question 9: By the end of 2021, how many U.S. states will have authorized some form of automated vehicle platooning on public roads?	**Question 10: In 2021, what percentage of new passenger vehicles sold in the U.S. will be electric vehicles?**
New legislation necessitated by rapidly improving technology and adoption.	I think the adoption rate of EVs will increase significantly over the next few years as the price comes down and consumers see the

	benefits that electric cars provide (lower maintenance, no gas costs, cleaner)
Public perception of platooning and news/media coverage. Mostly negative and fear based thoughts by the public.	Cost reduction, increasing acceptance, word of mouth, pollution, lower TCO.
Politics, safety, accidents, IoT security, cost, need.	Cheaper cost, public policy towards pollution control. Automakers investing heavily.
I assume there is a growing demand for smart transportation which will cause an increase in adoption of automated platooning on public roads.	Growth in popularity, incentives as well as battery capability.
The likelihood of technological advances weighed against political action (or inaction).	Demand for EVs is significant, but the lithium supply chain limits how many cars the major automakers can produce. Adoption will continue to ramp up slowly.

The average of the number of keywords used by each respondent for each group is shown below in Table S15. The full distribution for Question 10 is shown below in Figure S5. As seen in Table S16, pairwise permutation tests using the medians showed that none of the differences were statistically significant for Question 9, indicating that all respondents across the three groups used approximately the same number of keywords in their responses. For Question 10, the two-briefing group used significantly more keywords than either of the other two groups. There was no difference between the control and one-briefing group for either question.

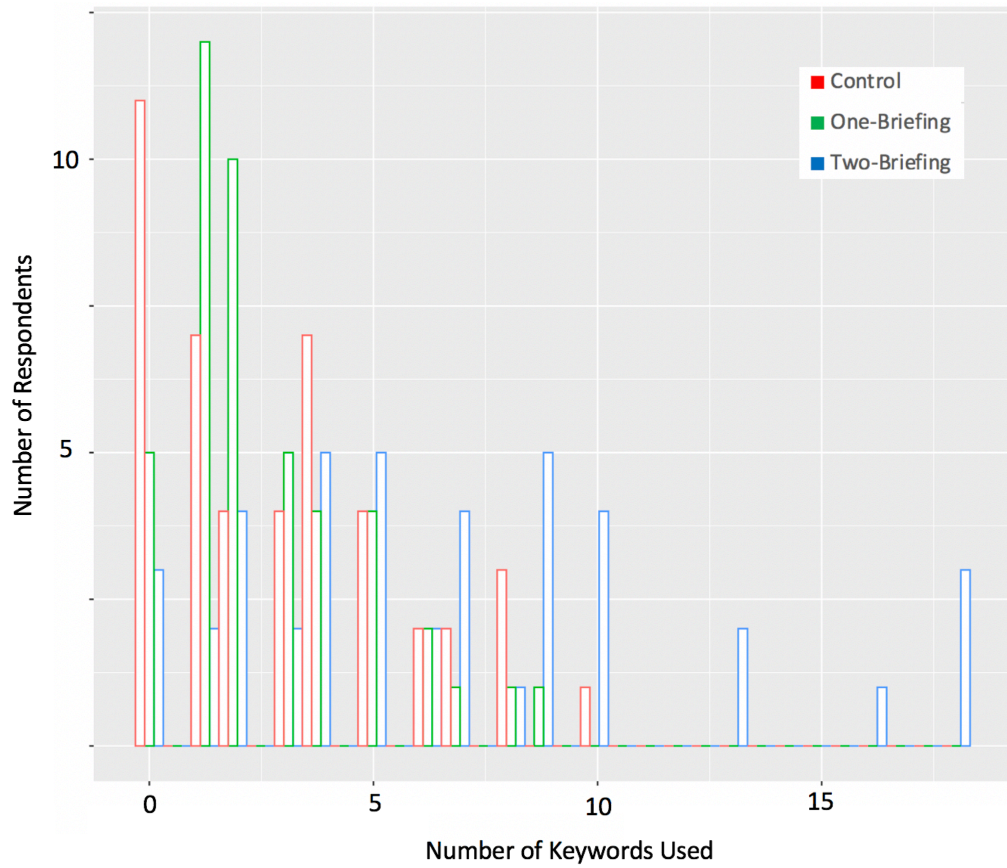


Figure S5: Distribution of the number of keywords used by each respondent in their response to Question 10.

Table S16: Average number of policy, economic, and social factors keywords used by each respondent for each group.

		Number of Keywords Used by Each Respondent		
		Control	**One-Briefing**	**Two-Briefing**
Question 9 (Platooning states)	Median	2	2	4
	Mean (SE)	2.96 (2.48)	2.56 (2.04)	4.67 (3.88)
Question 10 (Percentage EVs)	Median	3	2	6
	Mean (SE)	3.0 (2.73)	2.64 (2.20)	6.74 (4.85)

Table S17: P-values for pairwise comparisons of the keyword counts for the two questions. P-values that are significant with the Bonferroni correction are highlighted in green (0.005/6 tests = 0.0083).

	P-Values		
	Control and One-Briefing	**Control and Two-Briefing**	**One-Briefing and Two-Briefing**
Question 9 (Platooning states)	1.0	0.10	0.16
Question 10 (Percentage EVs)	0.40	0.0038	0.00010

Appendix A: Entire Survey

A Qualtrics PDF printout of the entire survey follows.

Welcome page

Hello, and thank you for your interest in this survey about the future of motor vehicle technology! The purpose of this research is to better understand how technology forecasts are made and how they can be improved.

If you decide to participate, you will provide informed consent, watch at least one short video, provide current and forecasted values for 10 questions about emerging motor vehicle technology, and answer some demographic questions. At the end you will see how well you did on the current value questions and can enter the gift card raffle. The entire survey will take approximately 15-30 minutes.

To begin, please click on the Next button below.

#EditSection, TimingExplanation#
First Click: *0 #EditSection, TimingSeconds#*
Last Click: *0 #EditSection, TimingSeconds#*
Page Submit: *0 #EditSection, TimingSeconds#*
Click Count: *0 #EditSection, TimingClicks#*

Informed Consent

Informed Consent: Motor Vehicle Technology Forecasts Survey

This survey is part of a research study conducted by PhD student Tamara Savage under the supervision of Prof. Granger Morgan at Carnegie Mellon University (CMU). It is funded by the Alfred P. Sloan Foundation and CMU's Center for Climate and Energy Decision Making/the National Science Foundation. The purpose of this research is to help improve technology forecasts. In this survey, you will be asked to make forecasts about motor vehicle technology, mostly relating to autonomous vehicles and electric vehicles. The survey will take approximately 15 to 30 minutes to complete.

Participant Requirements/Voluntary Participation

Participation in this study is limited to individuals age 18 and older. You must live in and currently be located in the United States. Your participation in this research is voluntary and you may stop at any time. You may print a copy of this consent form for your records.

We ask that those who take the survey are at least somewhat knowledgeable about motor vehicle technology in the U.S.

Risks

The risks and discomfort associated with participation in this study are no greater than those ordinarily encountered in daily life or during other online activities. During the survey, we will not be collecting any personally identifying information and all responses will be anonymous. Your IP address will not be captured.

If you complete the survey, you will be directed to a separate online form where you can provide your personal contact information to be entered into a random drawing for one of ten $50 Amazon gift cards. This form will not be linked in any way to your survey responses. You will be asked to enter your first and last name, email address, phone number, and preferred contact method. Winners will be informed by their preferred method of communication and the gift cards will be emailed to the winners. While we will treat the information you provide as confidential, there is a possibility for a breach of confidentiality of your participation in this study and personal information if you choose to enter the raffle and provide personal contact information.

Benefits/Compensation & Costs

There are no direct benefits to respondents for completing the survey. There will be no cost to you if you participate in this study. There is no guaranteed compensation for participation in this study. If you complete the survey, you can choose to enter the raffle, where you may win one of ten $50 Amazon gift cards, with approximate odds of winning of 1 in 30.

Future Use of Information

In the future, once we have removed all identifiable information from your data, we may use the data for our future research studies, or we may distribute the data to other researchers for their research studies. We would do this without getting additional informed consent from you (or your legally authorized representative). Sharing of data with other researchers will only be done in such a manner that you will not be identified.

Confidentiality

By participating in this research, you understand and agree that Carnegie Mellon may be required to disclose your consent form, data and other personally identifiable information as required by law, regulation, subpoena or court order. Otherwise, your confidentiality will be maintained in the following manner:

By participating, you understand and agree that the data and information gathered during this study may be used by Carnegie Mellon and published and/or disclosed by Carnegie Mellon to others outside of Carnegie Mellon. However, your name, address, contact information and other direct personal identifiers will not be mentioned in any such publication or dissemination of the research data and/or results by Carnegie Mellon. Note that per regulation all research data must be kept for a minimum of 3 years.

Right to Ask Questions & Contact Information

If you have any questions about this study, you should feel free to ask them by contacting the Principal Investigator now: Tamara Savage, PhD Student, Department of Engineering and Public Policy, thsavage@andrew.cmu.edu. If you have questions later, desire additional information, or wish to withdraw your participation, please contact the Principal Investigator by email in accordance with the contact information listed above.

If you have questions pertaining to your rights as a research participant; or to report concerns to this study, you should contact the Office of Research Integrity and Compliance at Carnegie Mellon University. Email: irb-review@andrew.cmu.edu. Phone: 412-268-1901 or 412-268-5460.

I am age 18 or older.

Yes

No

I have read and understand the information above.

Yes

No

I want to participate in this research and continue with the survey.

Yes

No

I live in and am currently located in the United States.

Yes

No

I am at least somewhat knowledgeable about motor vehicle technology in the United States.

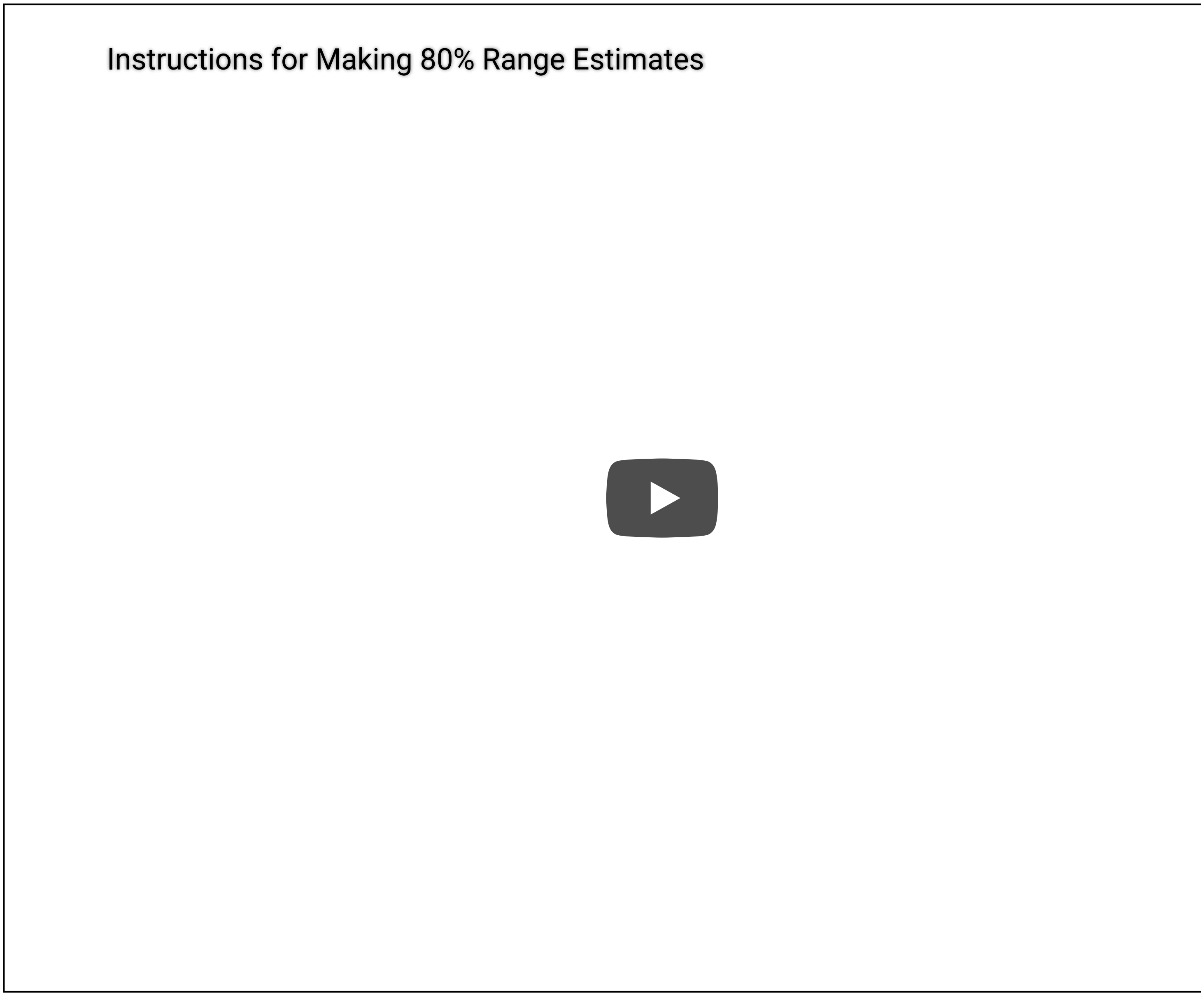
Yes

No

#EditSection, TimingExplanation#
First Click: *0 #EditSection, TimingSeconds#*
Last Click: *0 #EditSection, TimingSeconds#*
Page Submit: *0 #EditSection, TimingSeconds#*
Click Count: *0 #EditSection, TimingClicks#*

Instructions & No Lookups

Please watch this 1-minute instruction video and answer the question on the following page. The video is captioned. If the video is too small, use the full-screen button. If you are having trouble viewing the video, please click here to watch it directly on YouTube.



#EditSection, TimingExplanation#
First Click: *0 #EditSection, TimingSeconds#*
Last Click: *0 #EditSection, TimingSeconds#*
Page Submit: *0 #EditSection, TimingSeconds#*
Click Count: *0 #EditSection, TimingClicks#*

If you are asked to provide 80% range estimates for several questions, roughly what percentage of the actual answers should fall **outside** of your ranges?

10%

20%

80%

90%

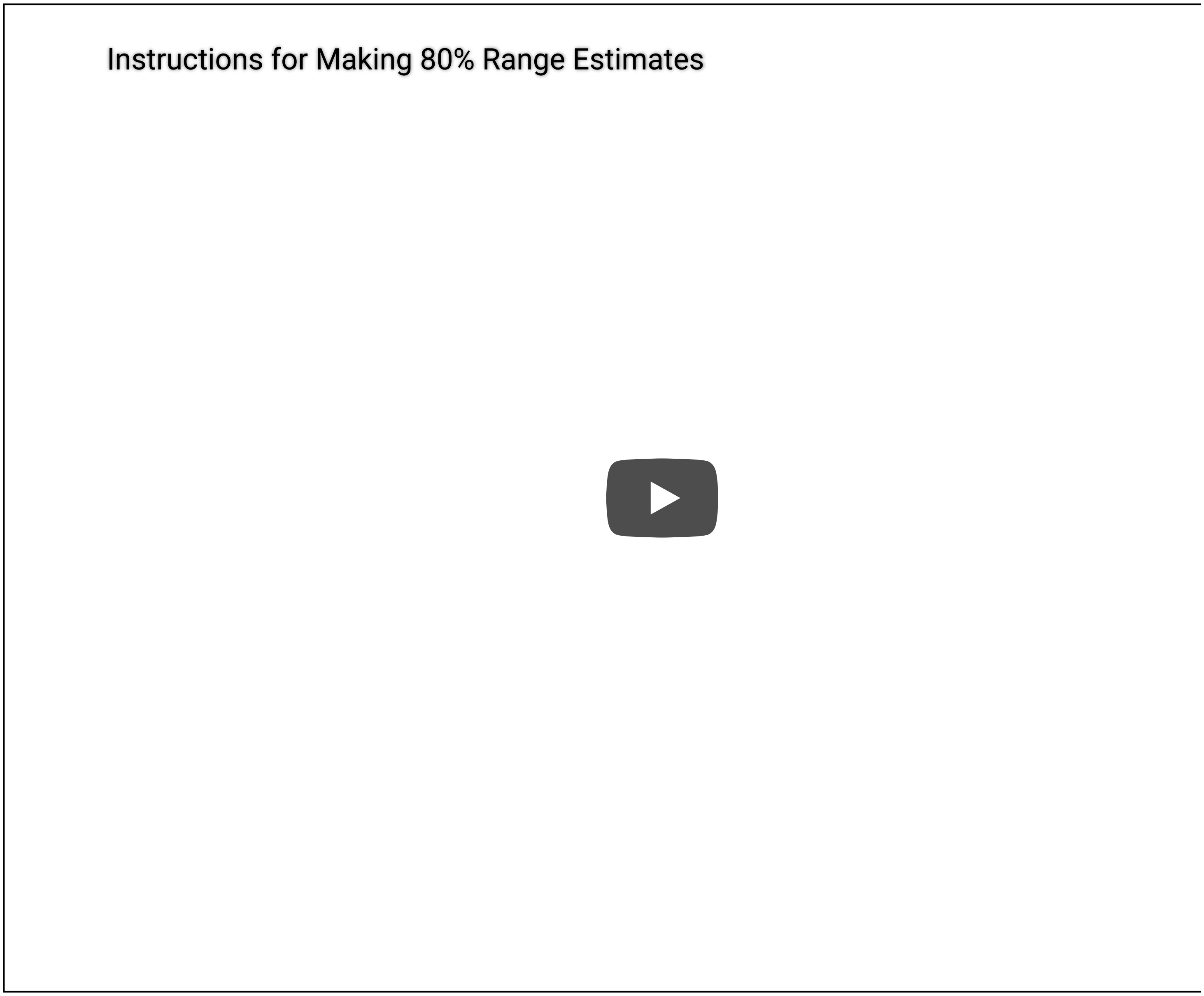
#EditSection, TimingExplanation#
First Click: *0 #EditSection, TimingSeconds#*
Last Click: *0 #EditSection, TimingSeconds#*
Page Submit: *0 #EditSection, TimingSeconds#*
Click Count: *0 #EditSection, TimingClicks#*

Correct! Please click the Next button to proceed.

#EditSection, TimingExplanation#
First Click: *0 #EditSection, TimingSeconds#*
Last Click: *0 #EditSection, TimingSeconds#*
Page Submit: *0 #EditSection, TimingSeconds#*
Click Count: *0 #EditSection, TimingClicks#*

If you are asked to provide 80% range estimates for several questions, roughly 20% of the actual answers should fall outside of your ranges.

Please consider watching the instruction video again before proceeding. The video is captioned. If the video is too small, use the full-screen button. If you are having trouble viewing the video, please click here to watch it directly on YouTube.



#EditSection, TimingExplanation#
First Click: *0 #EditSection, TimingSeconds#*
Last Click: *0 #EditSection, TimingSeconds#*
Page Submit: *0 #EditSection, TimingSeconds#*
Click Count: *0 #EditSection, TimingClicks#*

Please do not look up any answers or ask anyone for advice when you answer the survey questions. It is very important for our research that you do not look anything up. Please just answer to the best of your ability--you will not be penalized for wrong answers and you will not be rewarded for right answers. You will see the answers to the

10 current value questions at the end of the survey. Thank you very much for your cooperation.

I understand that it is important for the purposes of this research that I do not look up answers or ask anyone for advice on the answers I give:

Yes

No

#EditSection, TimingExplanation#

First Click: *0 #EditSection, TimingSeconds#*

Last Click: *0 #EditSection, TimingSeconds#*

Page Submit: *0 #EditSection, TimingSeconds#*

Click Count: *0 #EditSection, TimingClicks#*

Practice Questions

Practice Question 1: In what year did Tesla launch its initial public offering (IPO) on the NASDAQ stock exchange?

Lower estimate (only a 10% chance it happened before this year)

Higher estimate (only a 10% chance it happened after this year)

#EditSection, TimingExplanation#

First Click: *0 #EditSection, TimingSeconds#*

Last Click: *0 #EditSection, TimingSeconds#*

Page Submit: *0 #EditSection, TimingSeconds#*

Click Count: *0 #EditSection, TimingClicks#*

Practice Question 2: **In 1980, how many new passenger cars were sold in the U.S.?** New sales include sales to individuals and corporate fleets and leased cars.

Lower estimate (only a 10% chance the value is lower)

Higher estimate (only a 10% chance the value is higher)

#EditSection, TimingExplanation#
First Click: *0 #EditSection, TimingSeconds#*
Last Click: *0 #EditSection, TimingSeconds#*
Page Submit: *0 #EditSection, TimingSeconds#*
Click Count: *0 #EditSection, TimingClicks#*

Control Block

Treatment 1 Block

Please watch this 2-minute video on overconfidence bias and answer the question on the following page. The video is captioned. If the video is too small, use the full-screen button. If you are having trouble viewing the video, please click here to watch it directly on YouTube.

#EditSection, TimingExplanation#
First Click: *0 #EditSection, TimingSeconds#*
Last Click: *0 #EditSection, TimingSeconds#*
Page Submit: *0 #EditSection, TimingSeconds#*
Click Count: *0 #EditSection, TimingClicks#*

Which of the following did **not** appear in this video? Please select all that apply.

Hot air balloon trivia question

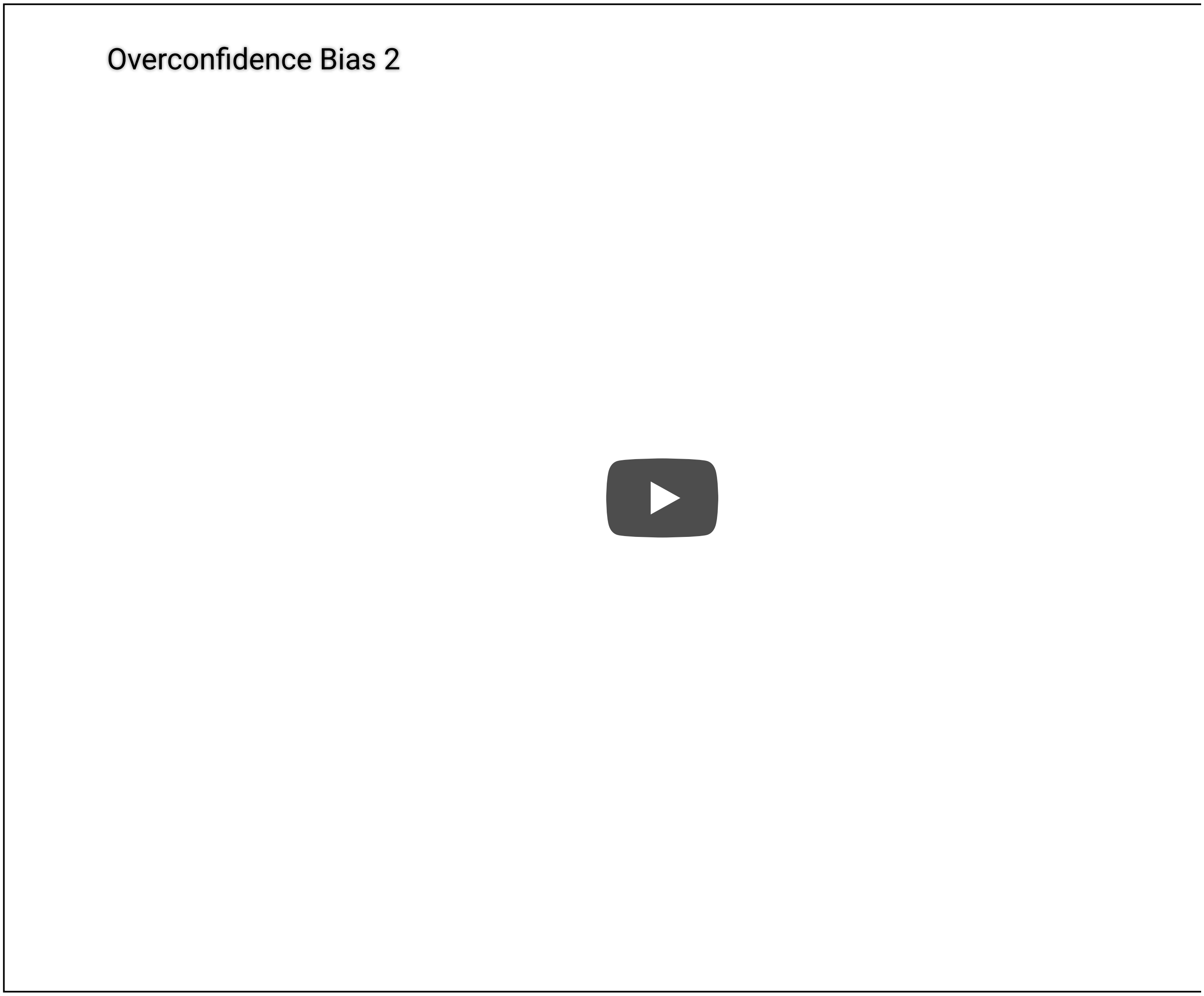
Florida and sea level rise

The Surprise Index

#EditSection, TimingExplanation#
First Click: *0 #EditSection, TimingSeconds#*
Last Click: *0 #EditSection, TimingSeconds#*
Page Submit: *0 #EditSection, TimingSeconds#*
Click Count: *0 #EditSection, TimingClicks#*

Treatment 2 Block

Please watch this 2-minute video on overconfidence bias and answer the question on the following page. The video is captioned. If the video is too small, use the full-screen button. If you are having trouble viewing the video, please click here to watch it directly on YouTube.



#EditSection, TimingExplanation#
First Click: *0 #EditSection, TimingSeconds#*
Last Click: *0 #EditSection, TimingSeconds#*
Page Submit: *0 #EditSection, TimingSeconds#*
Click Count: *0 #EditSection, TimingClicks#*

Which of the following did **not** appear in this video? Please select all that apply.

Hot air balloon trivia question

Florida and sea level rise

The Surprise Index

#EditSection, TimingExplanation#

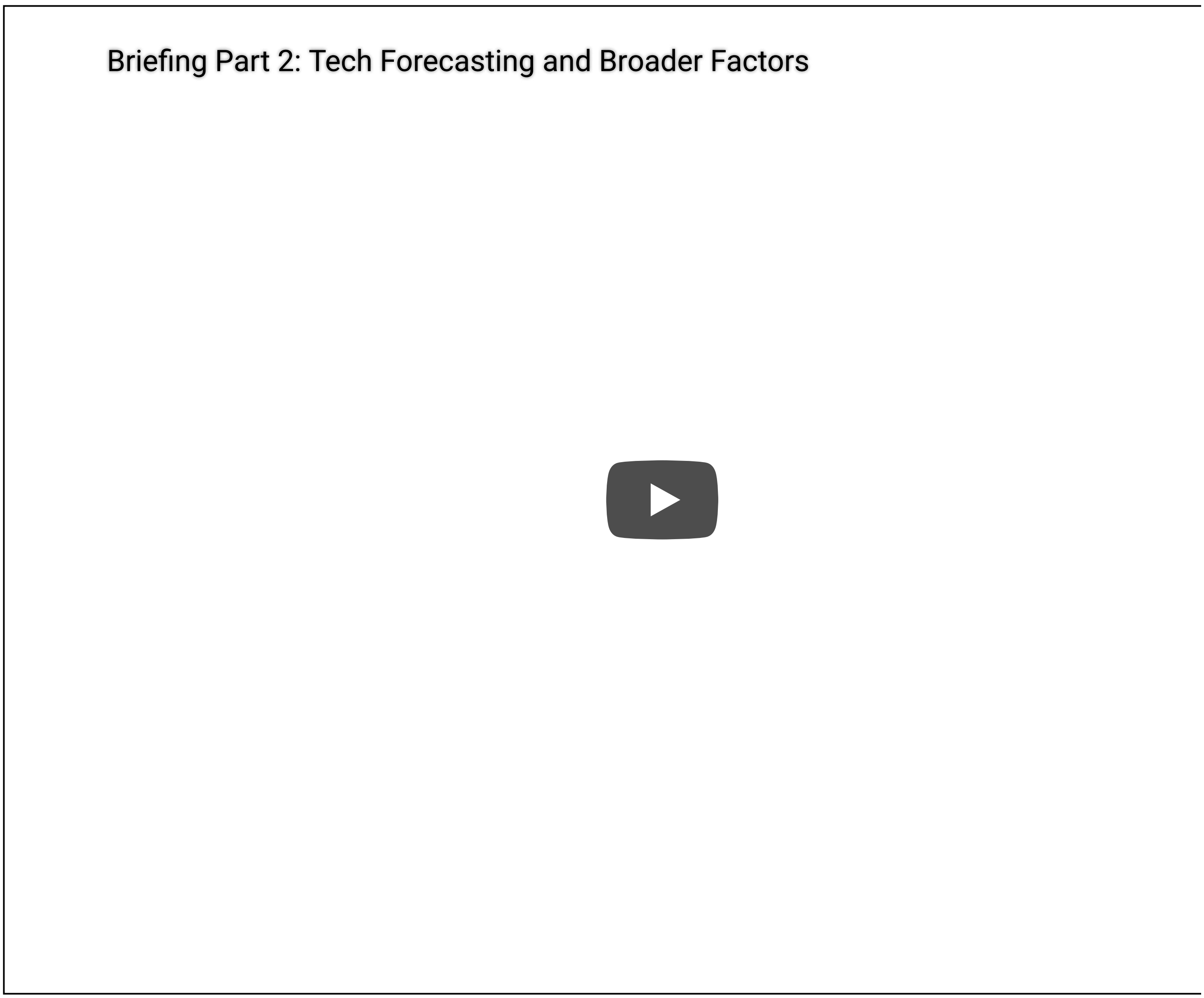
First Click: *0 #EditSection, TimingSeconds#*

Last Click: *0 #EditSection, TimingSeconds#*

Page Submit: *0 #EditSection, TimingSeconds#*

Click Count: *0 #EditSection, TimingClicks#*

Please watch this less than 6-minute video on past technology forecasting performance and broader policy, economic, and social factors and answer the question on the following page. The video is captioned. If the video is too small, use the full-screen button. If you are having trouble viewing the video, please click here to watch it directly on YouTube.



#EditSection, TimingExplanation#
First Click: *0 #EditSection, TimingSeconds#*
Last Click: *0 #EditSection, TimingSeconds#*
Page Submit: *0 #EditSection, TimingSeconds#*
Click Count: *0 #EditSection, TimingClicks#*

Which of the following did **not** appear in this video? Please select all that apply.

Solar electricity price forecasting

Cost forecasts of roads, bridges, and airplanes

The Eiffel Tower

A picture of taxis

#EditSection, TimingExplanation#
First Click: *0 #EditSection, TimingSeconds#*
Last Click: *0 #EditSection, TimingSeconds#*
Page Submit: *0 #EditSection, TimingSeconds#*
Click Count: *0 #EditSection, TimingClicks#*

Forecasting Questions

1a. **Currently, what is the longest EPA-rated range (in miles) for a battery electric vehicle?** Only consider passenger vehicles that are offered for retail sale in the U.S. The EPA-rated range is the official measure of how far the vehicle can drive on one full battery charge.

Lower estimate for current value (only a 10% chance the value is lower)

Higher estimate for current value (only a 10% chance the value is higher)

#EditSection, TimingExplanation#
First Click: *0 #EditSection, TimingSeconds#*
Last Click: *0 #EditSection, TimingSeconds#*
Page Submit: *0 #EditSection, TimingSeconds#*
Click Count: *0 #EditSection, TimingClicks#*

1b. **At the end of 2021, what will be the longest EPA-rated range (in miles) for a battery electric vehicle?** Only consider passenger vehicles that are offered for retail sale in the U.S. The EPA-rated range is the official measure of how far the vehicle can drive on one full battery charge.

Your range estimate for the current value is ${q://QID9/ChoiceTextEntryValue/1} to ${q://QID9/ChoiceTextEntryValue/2} miles.

Lower estimate for 2021 value (only a 10% chance the value is lower)

Higher estimate for 2021 value (only a 10% chance the value is higher)

#EditSection, TimingExplanation#

First Click: *0 #EditSection, TimingSeconds#*

Last Click: *0 #EditSection, TimingSeconds#*

Page Submit: *0 #EditSection, TimingSeconds#*

Click Count: *0 #EditSection, TimingClicks#*

2a. **Currently, what is the average cost for a lithium-ion battery pack (in $/kWh)?** Estimate the average industry-wide, volume-weighted cost for passenger electric vehicles that are sold in the U.S.

Lower estimate for current value (only a 10% chance the value is lower)

Higher estimate for current value (only a 10% chance the value is higher)

#EditSection, TimingExplanation#

First Click: *0 #EditSection, TimingSeconds#*

Last Click: *0 #EditSection, TimingSeconds#*

Page Submit: *0 #EditSection, TimingSeconds#*

Click Count: *0 #EditSection, TimingClicks#*

2b. **At the end of 2021, what will be the average cost for a lithium-ion battery pack (in $/kWh)?** Estimate the average industry-wide, volume-weighted cost for passenger electric vehicles that are sold in the U.S.

Your range estimate for the current value is ${q://QID41/ChoiceTextEntryValue/1} to ${q://QID41/ChoiceTextEntryValue/2} $/kWh.

Lower estimate for 2021 value (only a 10% chance the value is lower)

Higher estimate for 2021 value (only a 10% chance the value is higher)

#EditSection, TimingExplanation#
First Click: *0 #EditSection, TimingSeconds#*
Last Click: *0 #EditSection, TimingSeconds#*
Page Submit: *0 #EditSection, TimingSeconds#*
Click Count: *0 #EditSection, TimingClicks#*

3a. **Currently, how many companies have a permit issued by the California DMV for testing of autonomous vehicles with a driver on public roads in California?** By law (enacted in 2014), a permit is required before a company may test AVs on public roads. A company is granted one permit and can list multiple vehicles on the permit.

Lower estimate for current value (only a 10% chance the value is lower)

Higher estimate for current value (only a 10% chance the value is higher)

#EditSection, TimingExplanation#
First Click: *0 #EditSection, TimingSeconds#*
Last Click: *0 #EditSection, TimingSeconds#*
Page Submit: *0 #EditSection, TimingSeconds#*
Click Count: *0 #EditSection, TimingClicks#*

3b. **At the end of 2021, how many companies will have a permit issued by the California DMV for testing of autonomous vehicles with a driver on public roads in California?** By law (enacted in 2014), a permit is required before a company may test AVs on public roads. A company is granted one permit and can list multiple vehicles on the permit.

Your range estimate for the current value is ${q://QID51/ChoiceTextEntryValue/1} to ${q://QID51/ChoiceTextEntryValue/2} companies.

Lower estimate for 2021 value (only a 10% chance the value is lower)	
Higher estimate for 2021 value (only a 10% chance the value is higher)	

#EditSection, TimingExplanation#
First Click: *0 #EditSection, TimingSeconds#*
Last Click: *0 #EditSection, TimingSeconds#*
Page Submit: *0 #EditSection, TimingSeconds#*
Click Count: *0 #EditSection, TimingClicks#*

4a. **Currently, how many companies have a permit issued by the California DMV for driverless testing of autonomous vehicles on public roads in California?** Driverless testing means the AV is capable of operating without the presence of a driver inside the vehicle. By law (enacted in 2018), a permit is required before a company may test driverless AVs on public roads. A company is granted one permit and can list multiple vehicles on the permit.

Lower estimate for current value (only a 10% chance the value is lower)	
Higher estimate for current value (only a 10% chance the value is higher)	

#EditSection, TimingExplanation#
First Click: *0 #EditSection, TimingSeconds#*
Last Click: *0 #EditSection, TimingSeconds#*
Page Submit: *0 #EditSection, TimingSeconds#*
Click Count: *0 #EditSection, TimingClicks#*

4b. **At the end of 2021, how many companies will have a permit issued by the California DMV for driverless testing of autonomous vehicles on public roads in California?** Driverless testing means the AV is capable of operating without the presence of a driver inside the vehicle. By law (enacted in 2018), a permit is required before a company may test driverless AVs on public roads. A company is granted one permit and can list multiple vehicles on the permit.

Your range estimate for the current value is ${q://QID53/ChoiceTextEntryValue/1} to ${q://QID53/ChoiceTextEntryValue/2} companies.

Lower estimate for 2021 value (only a 10% chance the value is lower)

Higher estimate for 2021 value (only a 10% chance the value is higher)

#EditSection, TimingExplanation#
First Click: *0 #EditSection, TimingSeconds#*
Last Click: *0 #EditSection, TimingSeconds#*
Page Submit: *0 #EditSection, TimingSeconds#*
Click Count: *0 #EditSection, TimingClicks#*

5a. The California DMV requires that a report be filed for every collision resulting in any damage of property or in bodily injury or death when testing an autonomous vehicle under a company's testing permit (with a driver or driverless). **In 2018, how many reports of collisions involving autonomous vehicles did the California DMV receive?**

Lower estimate for 2018 value (only a 10% chance the value is lower)

Higher estimate for 2018 value (only a 10% chance the value is higher)

#EditSection, TimingExplanation#
First Click: *0 #EditSection, TimingSeconds#*
Last Click: *0 #EditSection, TimingSeconds#*
Page Submit: *0 #EditSection, TimingSeconds#*
Click Count: *0 #EditSection, TimingClicks#*

5b. The California DMV requires that a report be filed for every collision resulting in any damage of property or in bodily injury or death when testing an autonomous vehicle under a company's testing permit (with a driver or driverless). **In 2021, how many**

reports of collisions involving autonomous vehicles will the California DMV receive?

Your range estimate for the 2018 value is ${q://QID44/ChoiceTextEntryValue/1} to ${q://QID44/ChoiceTextEntryValue/2} reports.

Lower estimate for 2021 value (only a 10% chance the value is lower)

Higher estimate for 2021 value (only a 10% chance the value is higher)

#EditSection, TimingExplanation#

First Click: *0 #EditSection, TimingSeconds#*

Last Click: *0 #EditSection, TimingSeconds#*

Page Submit: *0 #EditSection, TimingSeconds#*

Click Count: *0 #EditSection, TimingClicks#*

How old is former U.S. President Barack Obama?

Lower estimate (only a 10% chance he is younger than this)

Higher estimate (only a 10% chance he is older than this)

#EditSection, TimingExplanation#

First Click: *0 #EditSection, TimingSeconds#*

Last Click: *0 #EditSection, TimingSeconds#*

Page Submit: *0 #EditSection, TimingSeconds#*

Click Count: *0 #EditSection, TimingClicks#*

6a. **Currently, how many public DC fast charging outlets are available in the U.S.?** Include CCS, CHAdeMO, and Tesla fast charging outlets.

Lower estimate for current value (only a 10% chance the value is lower)

Higher estimate for current value (only a 10% chance the value is higher)

#EditSection, TimingExplanation#

First Click: *0 #EditSection, TimingSeconds#*
Last Click: *0 #EditSection, TimingSeconds#*
Page Submit: *0 #EditSection, TimingSeconds#*
Click Count: *0 #EditSection, TimingClicks#*

6b. **At the end of 2021, how many public DC fast charging outlets will be available in the U.S.?** Include CCS, CHAdeMO, and Tesla fast charging outlets.

Your range estimate for the current value is ${q://QID42/ChoiceTextEntryValue/1} to ${q://QID42/ChoiceTextEntryValue/2} outlets.

Lower estimate for 2021 value (only a 10% chance the value is lower)

Higher estimate for 2021 value (only a 10% chance the value is higher)

#EditSection, TimingExplanation#
First Click: *0 #EditSection, TimingSeconds#*
Last Click: *0 #EditSection, TimingSeconds#*
Page Submit: *0 #EditSection, TimingSeconds#*
Click Count: *0 #EditSection, TimingClicks#*

7a. **In 2018, how many hydrogen fuel cell passenger vehicles were sold in the U.S.?**

Lower estimate for 2018 value (only a 10% chance the value is lower)

Higher estimate for 2018 value (only a 10% chance the value is higher)

#EditSection, TimingExplanation#
First Click: *0 #EditSection, TimingSeconds#*
Last Click: *0 #EditSection, TimingSeconds#*
Page Submit: *0 #EditSection, TimingSeconds#*
Click Count: *0 #EditSection, TimingClicks#*

7b. **In 2021, how many hydrogen fuel cell passenger vehicles will be sold in the U.S.?**

Your range estimate for the 2018 value is ${q://QID64/ChoiceTextEntryValue/1} to ${q://QID64/ChoiceTextEntryValue/2} vehicles.

Lower estimate for 2021 value (only a 10% chance the value is lower)

Higher estimate for 2021 value (only a 10% chance the value is higher)

#EditSection, TimingExplanation#

First Click: *0 #EditSection, TimingSeconds#*

Last Click: *0 #EditSection, TimingSeconds#*

Page Submit: *0 #EditSection, TimingSeconds#*

Click Count: *0 #EditSection, TimingClicks#*

8a. **Currently, how many battery electric public buses are there in the U.S.?** Do not include buses/trolleys with overhead power cables.

Lower estimate for current value (only a 10% chance the value is lower)

Higher estimate for current value (only a 10% chance the value is higher)

#EditSection, TimingExplanation#

First Click: *0 #EditSection, TimingSeconds#*

Last Click: *0 #EditSection, TimingSeconds#*

Page Submit: *0 #EditSection, TimingSeconds#*

Click Count: *0 #EditSection, TimingClicks#*

8b. **At the end of 2021, how many battery electric public buses will there be in the U.S.?** Do not include buses/trolleys with overhead power cables.

Your range estimate for the current value is ${q://QID66/ChoiceTextEntryValue/1} to ${q://QID66/ChoiceTextEntryValue/2} buses.

Lower estimate for 2021 value (only a 10% chance the value is lower)

Higher estimate for 2021 value (only a 10% chance the value is higher)

#EditSection, TimingExplanation#
First Click: *0 #EditSection, TimingSeconds#*
Last Click: *0 #EditSection, TimingSeconds#*
Page Submit: *0 #EditSection, TimingSeconds#*
Click Count: *0 #EditSection, TimingClicks#*

9a. **By the end of 2018, how many U.S. states had authorized some form of automated vehicle platooning on public roads?** Platooning refers to a group of vehicles (with or without a driver) that can communicate and therefore can drive very closely together and accelerate and brake simultaneously. There is currently a patchwork of state laws regarding platooning; states can authorize full use of platooning on public roads, only testing, or not allow platooning.

Lower estimate for 2018 value (only a 10% chance the value is lower) [ ]

Higher estimate for 2018 value (only a 10% chance the value is higher) [ ]

#EditSection, TimingExplanation#
First Click: *0 #EditSection, TimingSeconds#*
Last Click: *0 #EditSection, TimingSeconds#*
Page Submit: *0 #EditSection, TimingSeconds#*
Click Count: *0 #EditSection, TimingClicks#*

9b. **By the end of 2021, how many U.S. states will have authorized some form of automated vehicle platooning on public roads?** Platooning refers to a group of vehicles (with or without a driver) that can communicate and therefore can drive very closely together and accelerate and brake simultaneously. There is currently a patchwork of state laws regarding platooning; states can authorize full use of platooning on public roads, only testing, or not allow platooning.

Your range estimate for the 2018 value is ${q://QID82/ChoiceTextEntryValue/1} to ${q://QID82/ChoiceTextEntryValue/2} states.

Lower estimate for 2021 value (only a 10% chance the value is lower) [ ]

Higher estimate for 2021 value (only a 10% chance the value is higher) [ ]

#EditSection, TimingExplanation#
First Click: *0 #EditSection, TimingSeconds#*
Last Click: *0 #EditSection, TimingSeconds#*
Page Submit: *0 #EditSection, TimingSeconds#*
Click Count: *0 #EditSection, TimingClicks#*

9c. **For the previous question (reproduced below), what did you think about as you made your forecast?** Please do not include any private or personally identifiable information about yourself or others in your response.

Previous question: 9b. By the end of 2021, how many U.S. states will have authorized some form of automated vehicle platooning on public roads?

Your range estimate for the 2021 value is ${q://QID85/ChoiceTextEntryValue/1} to ${q://QID85/ChoiceTextEntryValue/2} states.

#EditSection, TimingExplanation#
First Click: *0 #EditSection, TimingSeconds#*
Last Click: *0 #EditSection, TimingSeconds#*
Page Submit: *0 #EditSection, TimingSeconds#*
Click Count: *0 #EditSection, TimingClicks#*

10a. **In 2018, what percentage of new passenger vehicles sold in the U.S. were electric vehicles?** Please enter a number between 0 and 100. Include both plug-in hybrid electric vehicles and battery electric vehicles.

Lower estimate for 2018 value (only a 10%

chance the value is lower)

Higher estimate for 2018 value (only a 10% chance the value is higher)

#EditSection, TimingExplanation#
First Click: *0 #EditSection, TimingSeconds#*
Last Click: *0 #EditSection, TimingSeconds#*
Page Submit: *0 #EditSection, TimingSeconds#*
Click Count: *0 #EditSection, TimingClicks#*

10b. **In 2021, what percentage of new passenger vehicles sold in the U.S. will be electric vehicles?** Please enter a number between 0 and 100. Include both plug-in hybrid electric vehicles and battery electric vehicles.

Your range estimate for the 2018 value is ${q://QID45/ChoiceTextEntryValue/1}% to ${q://QID45/ChoiceTextEntryValue/2}%.

Lower estimate for 2021 value (only a 10% chance the value is lower)

Higher estimate for 2021 value (only a 10% chance the value is higher)

#EditSection, TimingExplanation#
First Click: *0 #EditSection, TimingSeconds#*
Last Click: *0 #EditSection, TimingSeconds#*
Page Submit: *0 #EditSection, TimingSeconds#*
Click Count: *0 #EditSection, TimingClicks#*

10c. **For the previous question (reproduced below), what did you think about as you made your forecast?** Please do not include any private or personally identifiable information about yourself or others in your response.

Previous question: 10b. In 2021, what percentage of new passenger vehicles sold in the U.S. will be electric vehicles?

Your range estimate for the 2021 value is ${q://QID63/ChoiceTextEntryValue/1}% to ${q://QID63/ChoiceTextEntryValue/2}%.

#EditSection, TimingExplanation#
First Click: *0 #EditSection, TimingSeconds#*
Last Click: *0 #EditSection, TimingSeconds#*
Page Submit: *0 #EditSection, TimingSeconds#*
Click Count: *0 #EditSection, TimingClicks#*

Bonus Question: **Do you think a Level 5 autonomous vehicle will be ever be available in the U.S.?** At Level 5, the vehicle is capable of performing all driving functions under all conditions without a driver. The vehicle can be offered for retail sale or as a ridesharing or subscription service.

Yes

No

Bonus Question: **In what year do you think a Level 5 autonomous vehicle will be available in the U.S.?** At Level 5, the vehicle is capable of performing all driving functions under all conditions without a driver. The vehicle can be offered for retail sale or as a ridesharing or subscription service.

Lower estimate (only a 10% chance it will happen before this year)

Higher estimate (only a 10% chance it will happen after this year)

#EditSection, TimingExplanation#
First Click: *0 #EditSection, TimingSeconds#*
Last Click: *0 #EditSection, TimingSeconds#*
Page Submit: *0 #EditSection, TimingSeconds#*

Click Count: *0 #EditSection, TimingClicks#*

Filtering Questions

How would you rate your knowledge of the technical and engineering aspects of **electric** vehicles?

None	Low	Medium	High

How would you rate your knowledge of the policy, economic, and social aspects of electric vehicles?

None	Low	Medium	High

How would you rate your knowledge of the technical and engineering aspects of **autonomous** vehicles?

None	Low	Medium	High

How would you rate your knowledge of the policy, economic, and social aspects of autonomous vehicles?

None	Low	Medium	High

Which of the following **cannot** use gasoline as fuel? Please select all that apply.

Internal combustion engine-electric hybrid

Plug-in hybrid electric vehicle

Battery electric vehicle

Don't know

Has a vehicle with some autonomous capabilities ever been involved in a fatal collision?

Yes

No

Don't know

Have you taken a probability and statistics course? Please select all that apply.

No

Yes, in high school

Yes, in college (undergraduate)

Yes, in graduate school

Would you describe your interest in motor vehicle technology as part of your job and/or as an enthusiast or hobbyist? Please select all that apply.

Job

Hobbyist/Enthusiast

Neither

Do you own an electric vehicle or other alternative-fuel vehicle? Please select all that apply.

Yes - PHEV

Yes - BEV

Yes - Other

No

Unsure

Do you own a vehicle with any autonomous capabilities (adaptive cruise control, lane-staying, self-parking, etc.)?

Yes

No

Unsure

#EditSection, TimingExplanation#
First Click: *0 #EditSection, TimingSeconds#*
Last Click: *0 #EditSection, TimingSeconds#*
Page Submit: *0 #EditSection, TimingSeconds#*
Click Count: *0 #EditSection, TimingClicks#*

Demographic Questions

In which state do you currently reside?

What is your age?

Age:

What is your gender identity?

Female

Male

Non-binary/Third gender

Prefer to self-describe:

Decline to state

Are you of Hispanic, Latino, or Spanish origin?

Yes

No

Decline to state

How would you describe yourself? Please select all that apply.

American Indian or Alaska Native

Asian

Black or African American

Native Hawaiian or other Pacific Islander

White

Other:

Decline to state

What is your annual income?

$0 - $50,000

$50,001 - $100,000

$100,001 - $150,000

$150,001 - $200,000

$200,001 and above

Decline to state

Which best describes your highest completed level of education?

Less than a high school diploma

High school diploma or equivalent

Some college, no degree

Associate's degree

Bachelor's degree

Master's degree

Professional degree

Doctorate

Decline to state

Which of the following areas do you currently work in or have you previously worked in? Please select all that apply.

Academia/Education

Government

Industry

Non-profit

Decline to state

How did you hear about this survey? Please select all that apply.

Email through university/lab

Forwarded by a friend or colleague

Reddit

National Academies TRB

SAE International eSource

Automotive News website

Other (please specify)

Do you have any comments, questions, or suggestions about this survey? Anything you care to share is appreciated.

#EditSection, TimingExplanation#
First Click: *0 #EditSection, TimingSeconds#*
Last Click: *0 #EditSection, TimingSeconds#*
Page Submit: *0 #EditSection, TimingSeconds#*
Click Count: *0 #EditSection, TimingClicks#*

Contact information: Tamara Savage, thsavage@andrew.cmu.edu | Carnegie Mellon University's Office of Research Integrity and Compliance: irb-review@andrew.cmu.edu

Appendix B: First Briefing

The links below will take you to the briefing videos on YouTube. PDFs of the PowerPoints follow.

Briefing on overconfidence bias (for one-briefing group):
https://youtu.be/0XuZ8Q6pxZE

Briefing on overconfidence bias (for two-briefing group, only differs by the last slide):
https://youtu.be/GQNWWsadOUo

Carnegie Mellon University

Overconfidence Bias

Tamara Savage & Prof. Granger Morgan
Department of Engineering & Public Policy

People are generally *overconfident*

Cartoon by David G. Kline in the *New York Times,* Aug. 21, 2010. https://www.nytimes.com/2010/08/22/business/economy/22view.html

The two practice questions you have just answered are to get you to think about ***overconfidence***.

Experimental studies show that most of us are systematically overconfident when we provide range estimates like the ones you just did.

When asked to provide a range such that they are 80% sure that the answer is inside that range, the answer actually falls inside the range much less than 80% of the time. We think that our answers are much more accurate than they really are and that more of our ranges will contain the answers than they really do.

We'll now show you some examples where this was the case.

An example of overconfidence bias

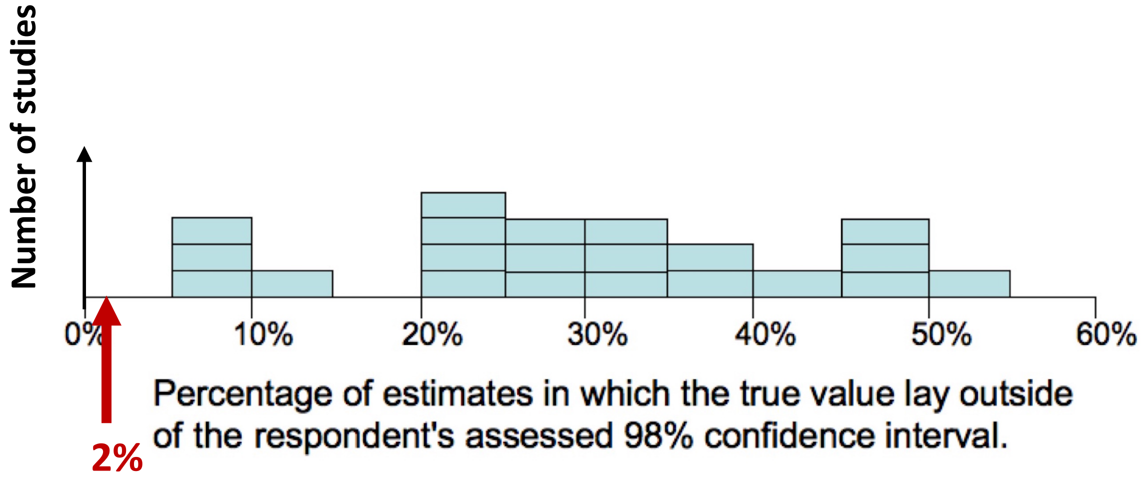
In a psychology research study, people were asked to make range estimates about trivia questions, such as: In what year was the first flight of a hot air balloon?

The researchers asked for 80% range estimates, which should have contained the answers 80% of the time.

However, the ranges people gave ***only contained the answers 48% of the time.***

Soll and Klayman, "Overconfidence in Interval Estimates," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 2004, Vol. 30, No. 2, 299-314.

The Surprise Index



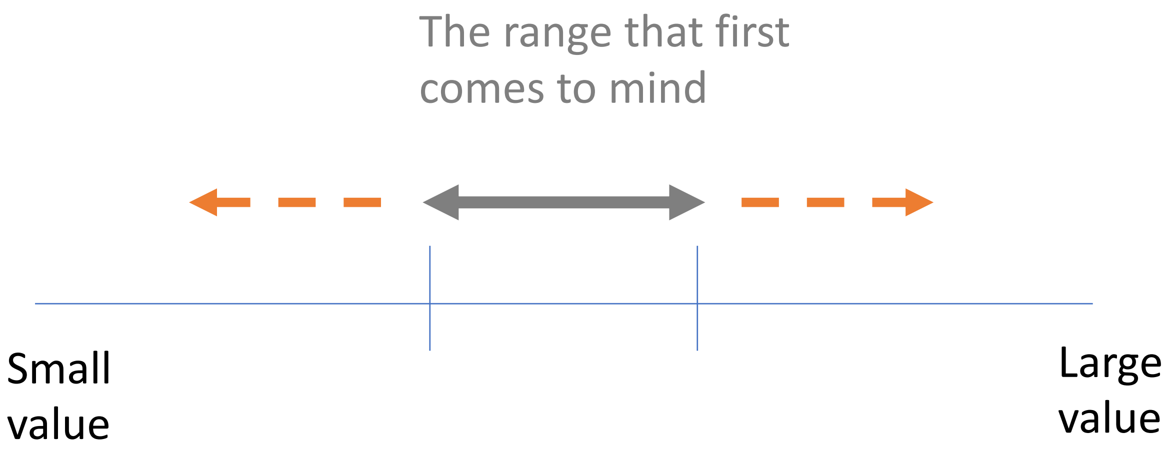
Here are the results from several studies in which researchers collected more than 10,000 range estimates like the ones you just provided.

In this case, respondents were asked to give a range for which there was only a 1% chance that the answer lay outside each end of the range. That means that only 2% of the answers should have fallen outside their ranges. This number is called the Surprise Index.

But, as you can see, the answers fell outside of their ranges 40%, 50%, or even 55% of the time!

This is what we mean by ***overconfidence***—people think their range estimates are better than they actually are and contain the answers more often than they really do, even when asked to give very wide ranges.

Morgan, M. G., *Theory and Practice in Policy Analysis: Including Applications in Science and Technology*. Cambridge, United Kingdom; New York, NY: Cambridge University Press, 2017. (Figure also from this source, slightly modified)



Consider making your range wider to take into account possibly being overconfident, as the range that first comes to mind may not be wide enough to contain the answer.

Please keep this in mind as you answer the forecasting questions in this survey.

The last slide differs slightly for the two treatment groups. The following slide is for the **one-briefing** group.

We went through this short briefing because we are going to ask you to make range estimates about the future of automotive technology, and we wanted to caution you against being overconfident.

We are now going to move on to the main part of our study. Please click the Next button to advance to the survey questions.

The last slide differs slightly for the two treatment groups. The following slide is for the **two-briefing** group.

Before we move on to the main part of our study, we also want to show you how poorly experts making forecasts have performed in the past.

We will also give you some advice on how to consider factors other than the technology itself—that is, policy, economic, and social factors that may affect how the technology develops—that could shape the results of a forecast.

Please click the Next button to advance to the second part of the briefing.

Appendix C: Second Briefing

The link below will take you to the briefing video on YouTube. A PDF of the PowerPoint follows.

Briefing on past technology forecasting performance and broader policy, economic, and social factors:
https://youtu.be/mQN1jt1dklU

Carnegie Mellon University

Past Technology Forecasting Performance and Broader Policy, Economic, and Social Factors

Tamara Savage & Prof. Granger Morgan
Department of Engineering & Public Policy

When we look back at how well experts have performed in the past when they have made technology forecasts, we find that they have often been ***overconfident.***

Let’s quickly look at some examples…

Technology Forecasting Success Rate

Forecast Time Horizon	Success Rate
Short term (1-5 years)	38%
Medium term (6-10 years)	39%
Long term (11-25 years)	14%

One study looked back at about 300 forecasts for a variety of technologies, such as computers, robotics, materials, energy, and sensors. The forecasts came from market research firms, industry organizations, government reports, and academic publications.

The study found that forecasts over the short and medium terms were equally successful, at about 38%, but long term forecasts were much less successful, at about 14%.

Fye, Charbonneau, Hay, and Mullins, "An examination of factors affecting accuracy in technology forecasts," *Technological Forecasting and Social Change*, vol. 80, no. 6, pp. 1222–1231, Jul. 2013.

Tesla's Model 3 Production Forecasts

Bloomberg the Company & Its Products ▾ | Bloomberg Anywhere Remote Login | Bloomberg Terminal Demo Request

Bloomberg

Technology

Tesla Falters With Model 3 as Initial Output Trails Forecast

By Dana Hull
October 2, 2017, 4:15 PM EDT *Updated on October 3, 2017, 9:39 AM EDT*

- Company blames 'bottlenecks' for production of just 260 units
- Deliveries for Model S and Model X climb to quarterly record

BUSINESS INSIDER DEUTSCHLAND INTERNATIONAL

Tesla misses its Model 3 deliveries by a mile

BI Joe Ciolli, Business Insider
3.01.2018, 22:37

An example in the automotive industry is Tesla's Model 3 production issues.

These issues have caused the company to badly miss their forecasted delivery targets to customers.

In the third quarter of 2017, Tesla forecasted it would produce 1,500 Model 3's, but only ended up producing 260.

Dana Hull, Bloomberg, Oct. 2, 2017. https://www.bloomberg.com/news/articles/2017-10-02/tesla-sales-climb-as-model-3-stokes-demand-in-pricier-offerings
Joe Ciolli, Business Insider International, Mar. 1, 2018. https://www.businessinsider.de/tesla-model-3-delivery-report-misses-fourth-quarter-2018-1?r=US&IR=T

Stock Market Returns Forecasting

Image from Wikimedia Commons: By Kevin Hutchinson - Flickr,
https://commons.wikimedia.org/w/index.php?curid=28923260

Another study asked companies' chief financial officers to forecast the average annual S&P 500 stock market returns for one year in the future.

The researchers evaluated more than 13,000 forecasts, and found that when they asked for 80% range estimates, they ***only contained the actual returns 36% of the time***.

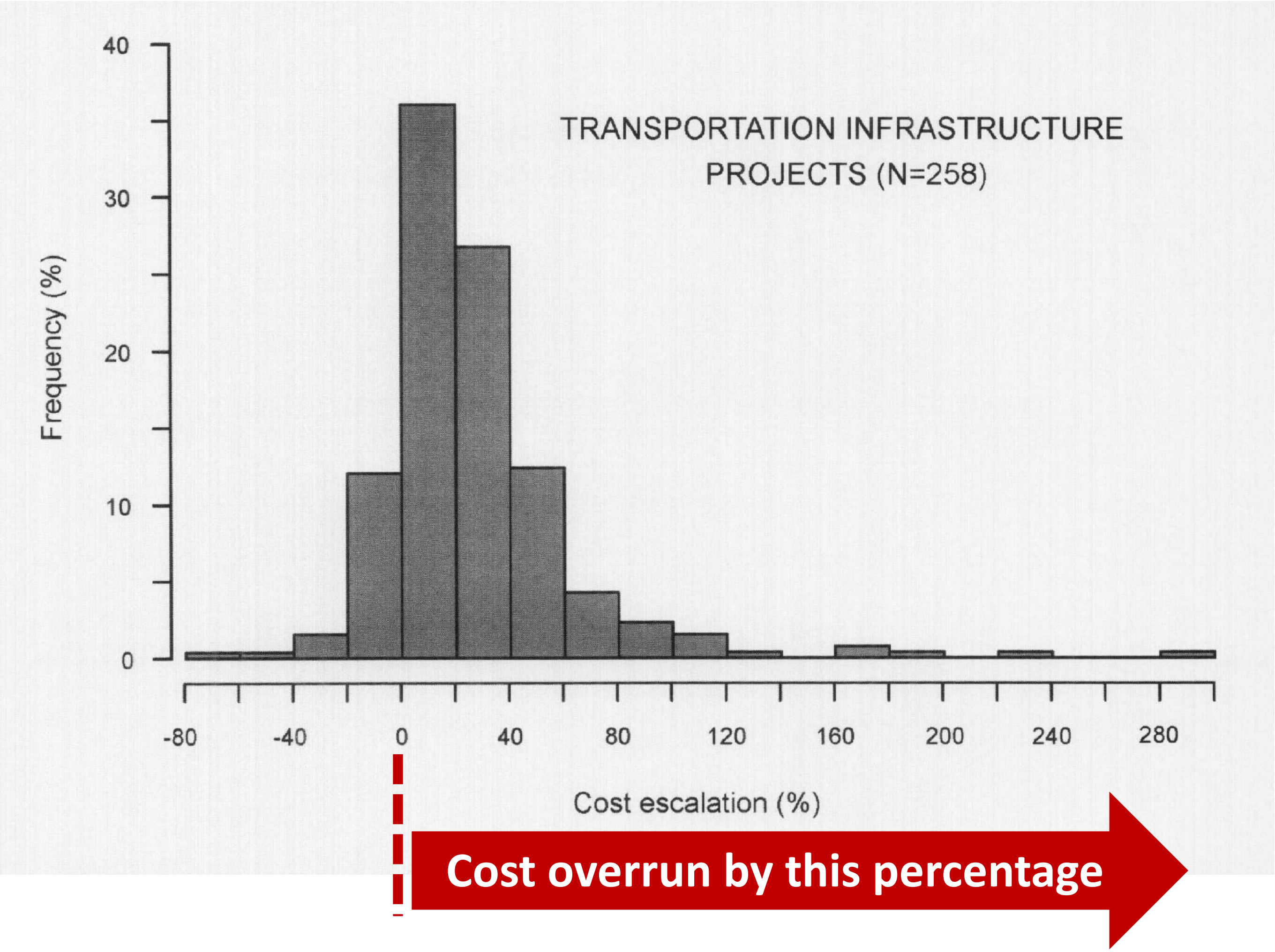
Ben-David, Graham, and Harvey, "Managerial Miscalibration," *The Quarterly Journal of Economics* (2013), 1547-1584.

Large Project Cost Forecasting

Sometimes the forecasts experts make are far too ***optimistic***.

This often happens when people make forecasts about how much large projects like new roads, bridges, nuclear power plants, or airplanes will cost.

Large Project Cost Forecasting

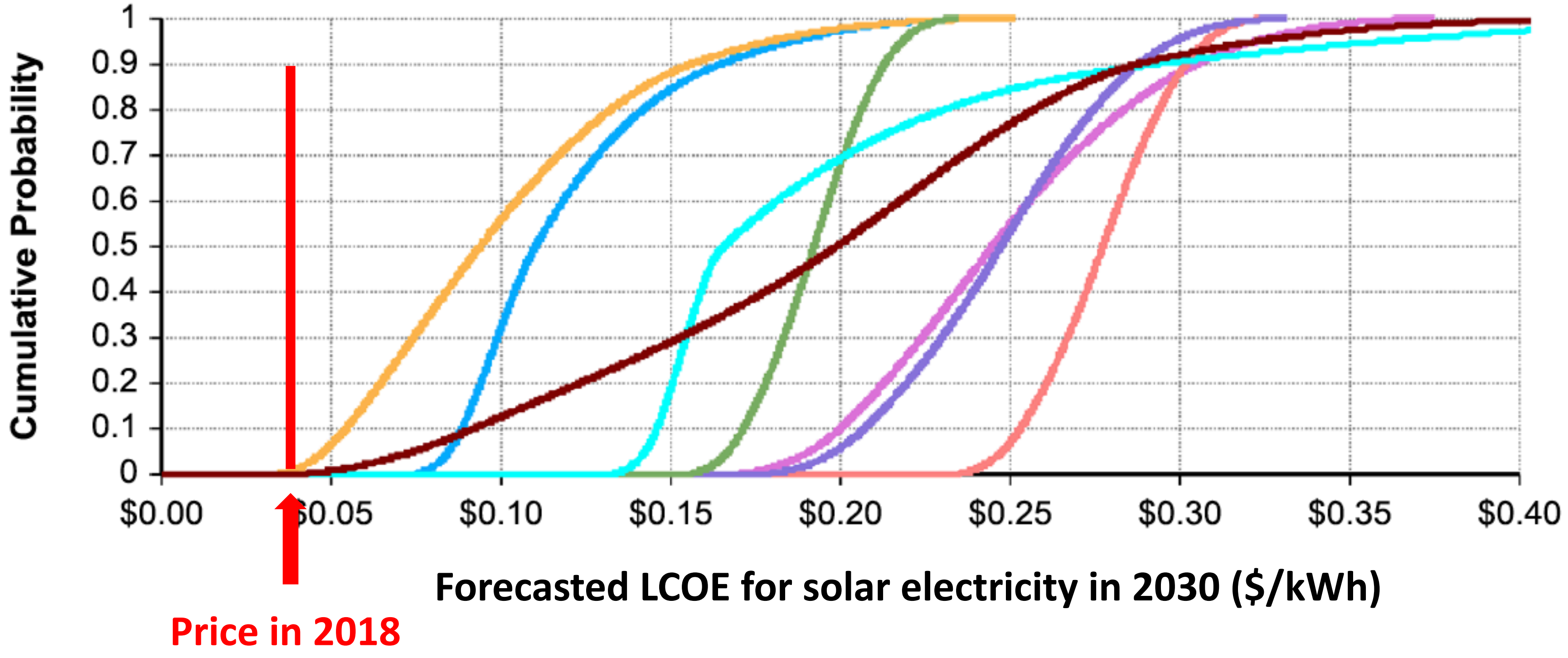


Here are the results from a study in which researchers compared the forecasted costs of 258 large transportation infrastructure projects to what those projects actually cost.

You can see that the experts were usually too optimistic. Some of the projects ended up costing twice what experts originally thought they would. Only a few projects ended up costing less than the original estimate.

Flyvbjerg, Holm, & Buhl, "Underestimating costs in public works projects: Error or lie?" *Journal of the American Planning Association;* Summer 2002; 68, 3, pp. 279-295. (Figure also from this source, slightly modified)

Solar Electricity Price Forecasting



From 2007 to 2011, researchers ran studies in which they asked experts to predict the price of solar electricity in 2030.

In contrast to the transportation project cost estimates, the ranges these experts gave were way too ***pessimistic***. The actual price today is already lower than the bottom of their ranges forecasted for 2030.

Baker, Bosetti, Anadon, Henrion, & Reis, "Future costs of key low-carbon energy technologies: Harmonization and aggregation of energy technology expert elicitation data," *Energy Policy* 80 (2015), 219-232.
Verdolini, Anadon, Lu, & Nemet, "The effects of expert selection, elicitation design, and R&D assumptions on experts' estimates of the future costs of photovoltaics," *Energy Policy* 80 (2015) 233-243.
Actual price from Lazard, "Levelized Cost of Energy and Levelized Cost of Storage 2018," Nov. 8, 2018. https://www.lazard.com/perspective/levelized-cost-of-energy-and-levelized-cost-of-storage-2018/

Solar Electricity Price Forecasting

Image from Wikimedia Commons: By U.S. Army Corps of Engineers Sacramento District - Solar panels at Presidio of Monterey, Public Domain, https://commons.wikimedia.org/w/index.php?curid=52215100

It is not entirely clear why these experts were so wrong, but here are a few possible reasons.

Decades ago, experts may not have imagined that:

- Many states would create renewable portfolio standards that would create a big market for solar cells
- State and federal governments would start subsidizing solar power
- China would start mass producing solar cells

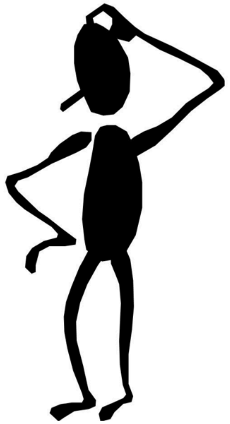
By now you get the basic idea: none of us, even the experts, are very good at accurately forecasting the future. This is likely due to the forecasters being overconfident.

Forecasters may be overconfident because they do not consider how broader policy, economic, and social factors may affect their forecasts. We would like you to consider these things as you make your range estimates.

What might affect your range estimates?

Things that could happen

1. ______________________
2. ______________________
3. ______________________
4. ______________________
5. ______________________
6. ______________________
7. ______________________
8. ______________________
9. ______________________
10. ______________________



Before you answer the forecasting questions, we are asking you to think up a list of policy, economic, and social factors that might affect your answers to questions about the future of motor vehicle technology.

To help get you started, we'll make a few suggestions for the kinds of things you might put on your list.

Government Regulations

There could be changes in government regulations…

For example, the U.S. federal government might:

- Dramatically tighten the Corporate Average Fuel Economy (CAFE) standards so auto companies must make a very fuel efficient mix of cars
- Or, it might eliminate the CAFE standards so that auto companies could make whatever mix of cars they wanted

Images from NHTSA.gov and EPA.gov

Fuel Prices

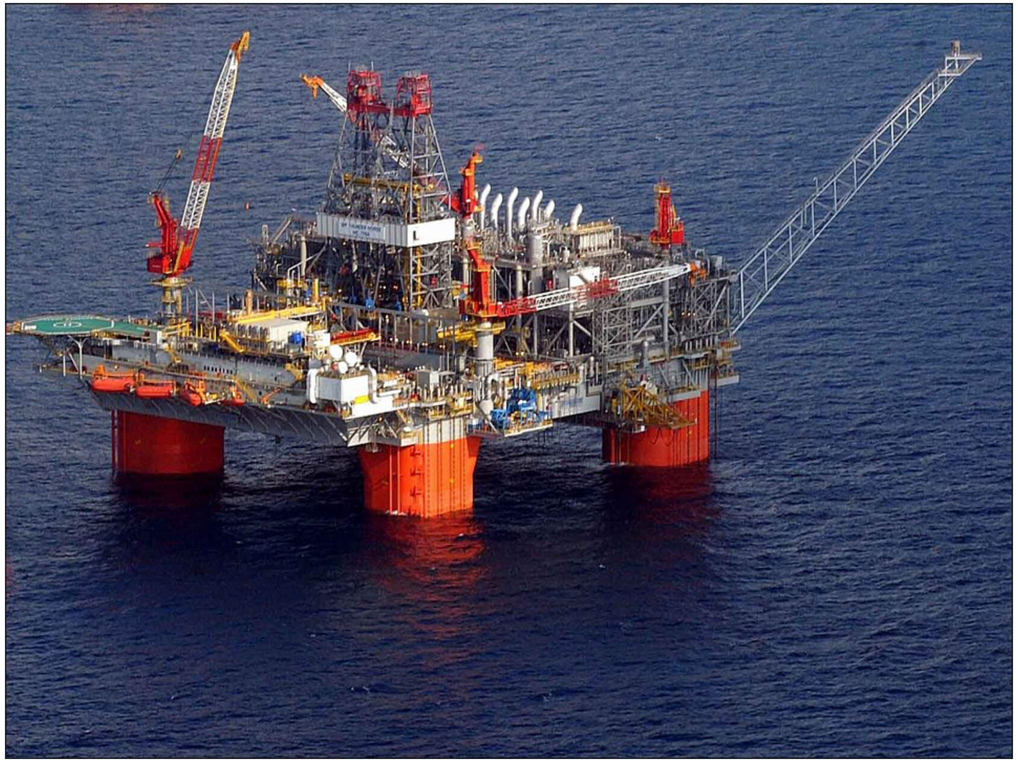


Image from Wikimedia Commons: By Andyminicooper - Own work, https://commons.wikimedia.org/w/index.php?curid=11701670

There could be changes in the market…

For example:

- OPEC might get stronger and gasoline prices might get very high
- Or, new global oil reserves might be found and gasoline prices might get very low

Infrastructure Development

Image from Wikimedia Commons: By Mariordo (Mario Roberto Durán Ortiz) - Own work
https://commons.wikimedia.org/w/index.php?curid=63358372

Needed infrastructure might or might not get built…

For example:

- Cities, towns, and major highways might install public fast charging stations everywhere
- Or, public fast charging stations might remain pretty scarce

Public Perception

Public attitudes might shift…

For example:

- Autonomous vehicles might be involved in many collisions that get the public very concerned
- Or, extensive road tests might show that autonomous vehicles are significantly safer than driver-operated cars, resulting in wide public support

Interest Group Advocacy

Image from Wikimedia Commons: By Ajay Suresh from New York, NY, USA - 20150627-P6270022
https://commons.wikimedia.org/w/index.php?curid=70532535

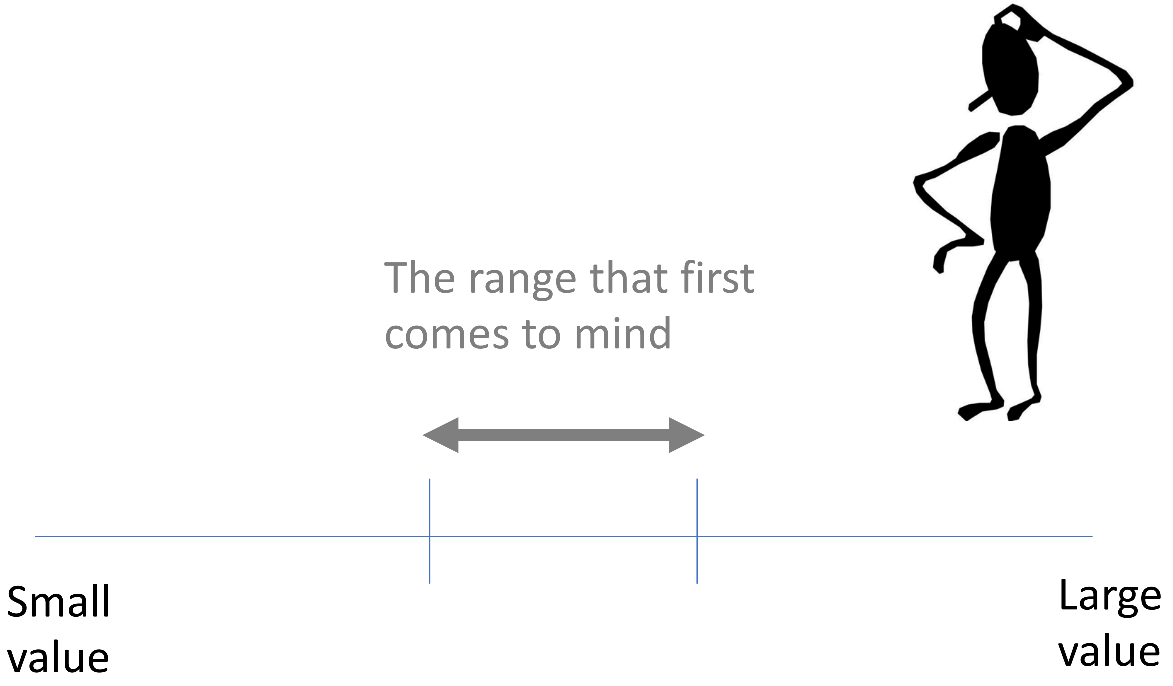
Interest groups may advocate for policies they would find beneficial…

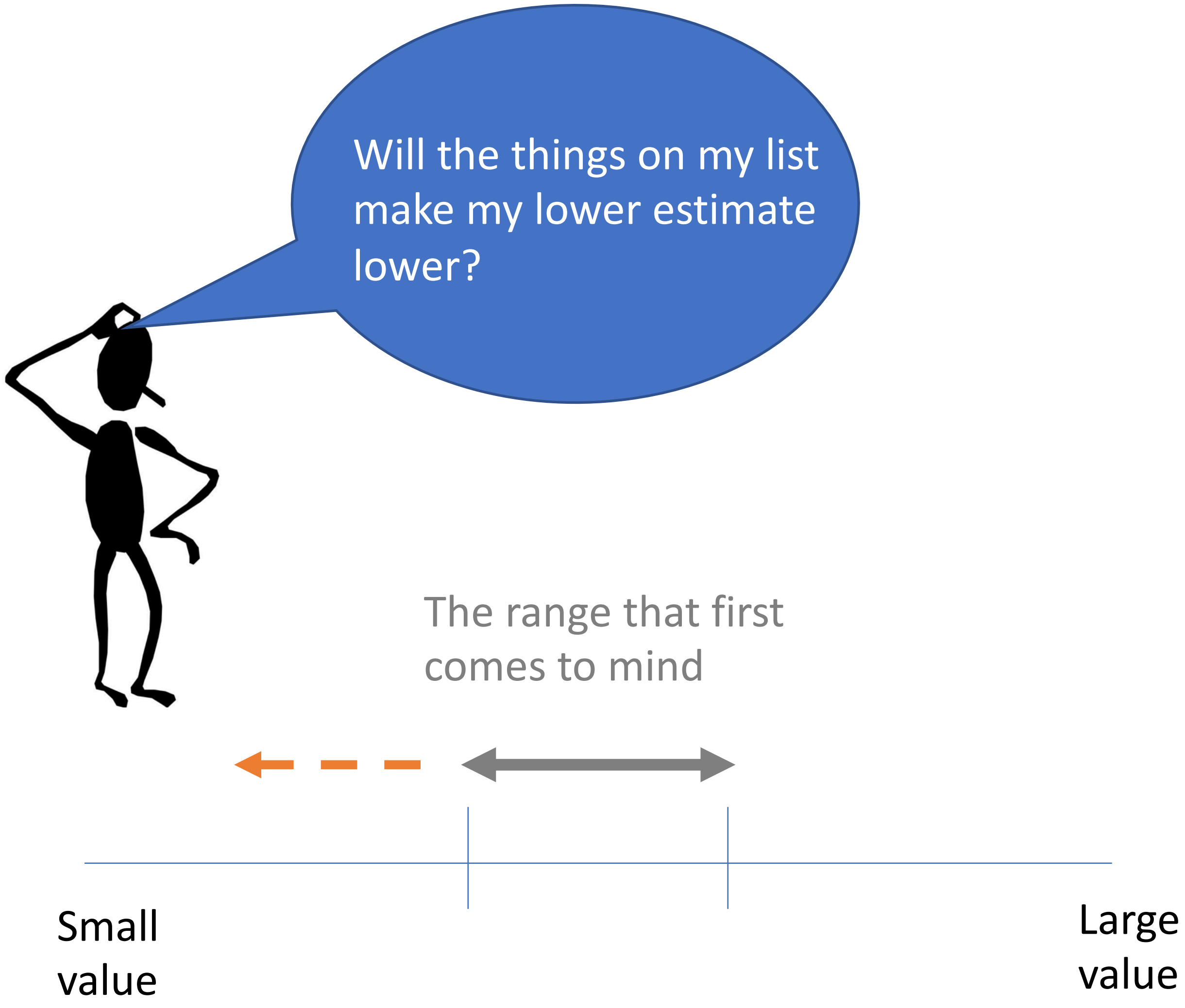
For example:

- Auto companies could lobby the government to eliminate regulations related to autonomous vehicles
- Or, there may be pushback from taxi drivers, who do not want more autonomous vehicles on roads

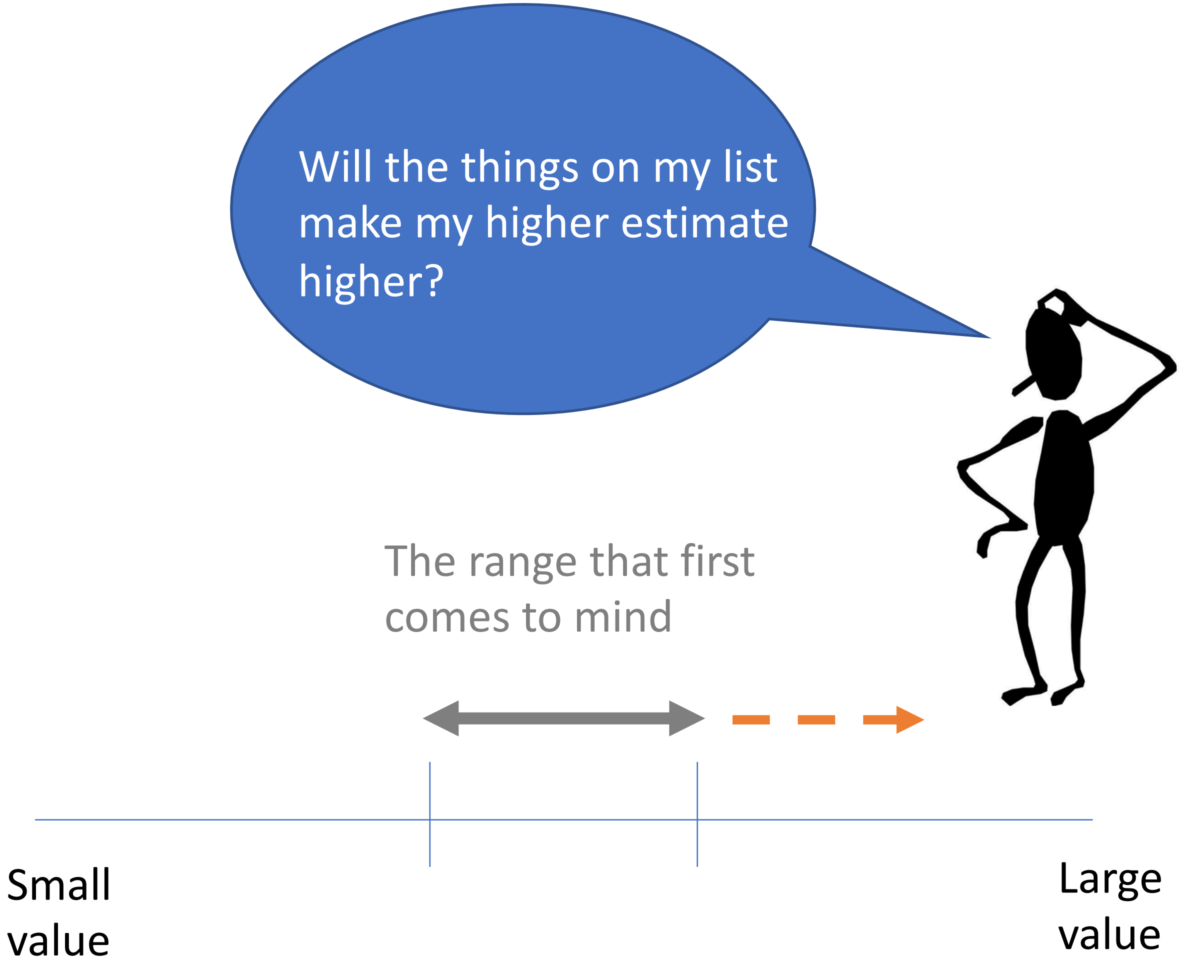
These are just some ideas to get you started on your list of policy, economic, and social factors to keep in mind as you make your forecasts.

In answering the forecasting questions, you'll probably start with a range that first comes to mind…



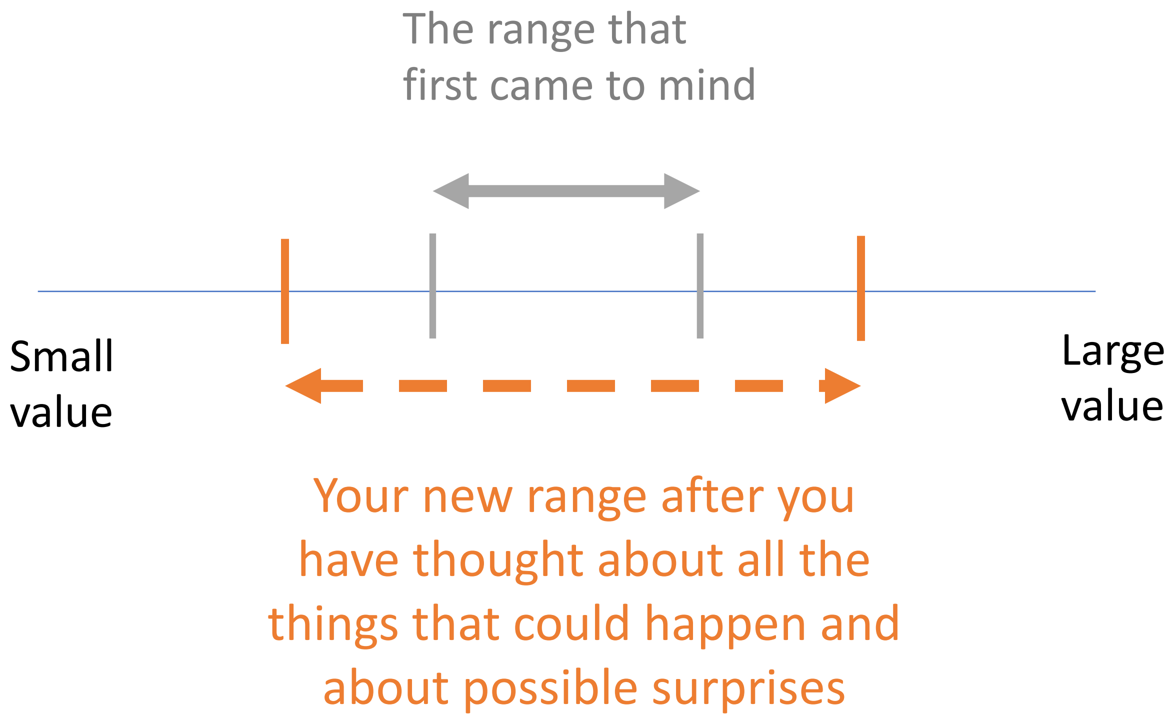


Once you have done that, consider if the things on your list might make your lower estimate lower…



And consider if the things on your list might make your higher estimate higher…

Also consider that there could be surprises that today you and nobody else imagine could occur but that could have a large impact on the forecast…



Then, once you have thought about policy, economic, and social factors that could affect the forecast, and realized that there may be more things that you can’t even imagine… consider widening your range to take all that into consideration.

We went through these short briefings because we are going to ask you to make range estimates about the future of automotive technology, and we wanted to caution you against being overconfident.

We are now going to move on to the main part of our study. Please click the Next button to advance to the survey questions.

Appendix D: Instructions

The link below will take you to the instructions video on YouTube. A PDF of the PowerPoint follows.

Instructions on making 80% interval estimates:
https://youtu.be/o4EPyeKFZnw

Carnegie Mellon University

Instructions for Making 80% Range Estimates

Tamara Savage & Prof. M. Granger Morgan
Department of Engineering & Public Policy

Thank you for participating in our research study about the likely future evolution of motor vehicle technology.

Before we get started, we'd like you to review these instructions and answer a couple of practice questions.

Instructions

In this survey, you will be asked to make 80% confidence range estimates. For example, consider this question: What year was the Ford Model T first produced?

Instead of providing just your best guess of a single year, we'd like you to provide a range: both a lower estimate and a higher estimate. When you are asked many questions like this, about 80% of the ranges you provide should contain the actual answer.

Instructions

In this survey, you will be asked to make 80% confidence range estimates. For example, consider this question: What year was the Ford Model T first produced?

Instead of providing just your best guess of a single year, we'd like you to provide a range: both a lower estimate and a higher estimate. When you are asked many questions like this, about 80% of the ranges you provide should contain the actual answer.

The survey you are about to take will look like this:

Lower estimate (You think there is only a 10% chance the Model T was first produced before this year)	
Higher estimate (You think there is only a 10% chance the Model T was first produced after this year)	

Instructions

In this survey, you will be asked to make 80% confidence range estimates. For example, consider this question: What year was the Ford Model T first produced?

Instead of providing just your best guess of a single year, we'd like you to provide a range: both a lower estimate and a higher estimate. When you are asked many questions like this, about 80% of the ranges you provide should contain the actual answer.

The survey you are about to take will look like this:

Lower estimate (You think there is only a 10% chance the Model T was first produced before this year)	1900
Higher estimate (You think there is only a 10% chance the Model T was first produced after this year)	

Instructions

In this survey, you will be asked to make 80% confidence range estimates. For example, consider this question: What year was the Ford Model T first produced?

Instead of providing just your best guess of a single year, we'd like you to provide a range: both a lower estimate and a higher estimate. When you are asked many questions like this, about 80% of the ranges you provide should contain the actual answer.

The survey you are about to take will look like this:

Lower estimate (You think there is only a 10% chance the Model T was first produced before this year)	1900
Higher estimate (You think there is only a 10% chance the Model T was first produced after this year)	1920

Instructions

In this survey, you will be asked to make 80% confidence range estimates. For example, consider this question: What year was the Ford Model T first produced?

Instead of providing just your best guess of a single year, we'd like you to provide a range: both a lower estimate and a higher estimate. When you are asked many questions like this, about 80% of the ranges you provide should contain the actual answer.

The survey you are about to take will look like this:

Lower estimate (You think there is only a 10% chance the Model T was first produced before this year)	1900
Higher estimate (You think there is only a 10% chance the Model T was first produced after this year)	1920

The actual answer to this question is 1908.

Practice Questions

Please answer the multiple choice question on the next page and then proceed to the two practice questions.

Appendix E: Pre-Registered Data Analysis Plan

I pre-registered my data analysis plan on Open Science Framework. All materials can be found here: osf.io/vxdmb. The data analysis plan follows.

Introduction

I have designed a survey to assess the level of overconfidence among survey respondents when making forecasts about motor vehicle technology. Respondents are asked to make 80% interval estimates in response to questions such as, "How many DC fast charging outlets are there in the US?" and "How many companies in California have permits to test autonomous vehicles on public roads?" They will be asked to provide interval estimates for both the current value and the forecasted value (for 2021) for 10 questions. See the survey PDF or the Qualtrics .qsf file for the complete survey.

There will be a control group and two treatment groups. The two treatment groups will watch one or two short briefing videos before responding to the questions. The first treatment group (I'm calling this group Treatment 1) will watch a briefing on overconfidence bias. The second treatment group (I'm calling this group Treatment 2) will watch the overconfidence bias briefing and an additional briefing that tells them about the poor past performance of technology forecasts and asks them to consider how policy, economic, and social factors might affect their forecasts and gives some examples. See Attachment C for the briefing videos or the attached PDFs. The survey is designed to measure the difference in overconfidence between the three groups.

Data Collection & Sample

I will collect data from November 6, 2019 until December 4, 2019. As more data will give my tests higher statistical power, I want to leave the survey open as long as is feasible to collect as many responses as possible. This was the maximum amount of time I could leave the survey open given other scheduling constraints.

Participants were recruited from several places. The survey was posted on Reddit (electric vehicle and autonomous vehicle related sub-Reddits) and Twitter, and emails were sent to university research centers/lab groups and to members of the National Academies' Transportation Research Board. Paid advertisements were posted on AutomotiveNews.com and in SAE International's eSource newsletter. Participants were also recruited through daisy-chaining/snowballing, were I emailed people I knew and asked them to forward the survey to people who might be interested in taking it. The incentive was a raffle with a chance to win one of ten $50 Amazon gift cards.

Filtering Out Respondents

If people do not pass any of the following checks, they will be routed to the end of the survey. They will not answer the rest of the questions and their responses will be filtered out.

- They do not pass any of the five informed consent questions
- They do not agree to not look up answers online or ask anyone for advice while taking the survey

Attention Checks

Respondents will get 3, 4, or 5 attention/comprehension/knowledge check questions as they do the survey (listed below). I am not going to exclude people from my official analysis based on whether they get these correct or not. I decided to include people who did not pass all of the checks because even during pre-testing, when I could verify that people had just watched the briefings or knew a lot about car technology, they still occasionally missed one of the questions. In addition, the people who respond to the survey are semi-experts who are interested in the topic, not a population like Mturkers who are just quickly trying to get through the survey.

- Attention/comprehension check on the instructions video on making 80% interval estimates
- Attention check on overconfidence bias briefing (only treatment group 1 and treatment group 2)
- Attention check on past forecasting performance and broader factors briefing (only treatment group 2)
- Basic knowledge check question on electric vehicles
- Basic knowledge check question on autonomous vehicles

Measures, Hypotheses, and Statistical Tests

I will look at three measures for comparison with statistical tests. The primary measure will be the hit rates of each group. I will also look at two secondary measures: the interval widths and the number of keywords used by respondents. I have included my hypotheses for each measure.

For all comparisons, I will use two-sample, two-sided permutation tests, comparing the medians. All tests will be pairwise comparisons between Control and Treatment 1, Control and Treatment 2, and Treatment 1 and Treatment 2. I chose two-sided tests as they are more conservative, and I do not always know the direction of the comparisons (for example, sometimes the control group and treatment group 1 may not differ). Also, journals seem to prefer reporting two-sided p-values (for example, see Jain et al., 2013). From pre-testing, the distributions of the interval widths will very likely be non-normal, as some intervals were quite small (when pre-testers had a good idea of what the answer was) and some intervals were quite wide, spanning multiple orders of magnitude (when pre-testers were very unsure of the answer). These distributions will likely have outliers, which would skew the results of a parametric test (such as a t-test). The permutation test has 100% power (Siegel and Castellan, 1988) and is distribution-free. I will use 10,000 repetitions for each permutation test (this seems to be a standard value). I will use the correction from Phipson and Smyth (2010) when calculating the p-value (this avoids p-values of zero).

I will use a significance level of p = 0.05 for all tests. If the p-value from the permutation test is close to 0.05 (roughly 0.045-0.055), I will run the permutation test 10 times to see if the values are above or below 0.05. If they are both above and below 0.05, I will consider that to not be a statistically significant result.

Hit Rates

The hit rate is the number of questions each respondent gets correct out of the 10 questions. With 80% interval estimates and 10 questions, the hit rate should be 0.8 if respondents are well-calibrated (not overconfident), but they will likely be overconfident. A question is considered correct if their interval contains the answer. Endpoints are included in the intervals (i.e., if the answer is 5 and the interval is [5, 10], that counts as correct). The answers to the current value questions (and their sources) are in Appendix A. There will be six pairwise permutation tests total for this measure (three for the current values and three for the forecasting values).

Hypotheses:

For the current values: I hypothesize that the magnitude of the median hit rates will be Control < Treatment 1 and Control < Treatment 2. Since the second briefing is focused on factors that affect forecasts specifically, there may or may not be a difference in the hit rates between Treatment 1 and Treatment 2.

For the forecasted values: I hypothesize that the magnitude of the median hit rates will be Control < Treatment 1, Control < Treatment 2, and Treatment 1 < Treatment 2. This is the effect that the briefings are intended to have.

In order to verify that each group is overconfident, I will look at the histogram of the hit rates and run a one-sample, two-sided t-test with mu = 0.8. This distribution will likely be either mostly normally distributed or possibly uniformly distributed. There will at least not be outliers, as the hit rates are bounded at 0 and 1. The one-sample permutation test is a test of symmetry, and this distribution may or may not be symmetric (and is unlikely to be symmetric around mu = 0.8). The one-sample Kolmogorov-Smirnov test is a test of distribution, which is not what I want to do. Thus, the t-test is imperfect but probably the best option in this case. I hypothesize that the control group will be overconfident, and the two treatment groups will also be overconfident, but less so than the control group. This will probably be the case for both the current values and the forecasting values.

Interval Widths

I will also compare the interval widths for all 20 of the current value and forecasting questions. The interval width is just the higher estimate minus the lower estimate. If anyone flips the two values (put the lower estimate in the higher estimate box and vice versa), I will just flip them back. In pre-testing, people occasionally did this, it just seemed to come more naturally to some people to make the higher estimate first, but they did not do it very often. There will be 60 pairwise permutation tests total for this measure (3 times the 20 questions).

Hypotheses:

My hypotheses for the interval widths are the same as for the hit rates. Wider intervals correspond to more answers captured inside the intervals, meaning higher hit rates.

Keyword Counting Analysis

On the last two forecasting questions, respondents are asked, "What did you think about as you made your forecast?" The third measure I will analyze is the number of policy, economic, and social factors keywords that respondents use in their written responses. This is intended to see if the second briefing was effective at getting respondents to think about these factors as they made their forecasts. There will be six pairwise permutation tests total for this measure (three for each of the two questions).

The keywords are listed in Appendix B. I came up with this list over a few months as I was developing the survey. As responses came in, I added words to the list (blinded to the group the respondent was assigned to). I decided to count specific country, state, city, and company names. To ensure I captured the specific names that were used by respondents, I read through the replies as they came in and added those words to my list. I will use Excel's SEARCH function to count the number of keywords used in each response.

Hypotheses:

Since these questions refer to forecasts, I expect that there may not be a difference between Control and Treatment 1, but I expect Treatment 2 to use more keywords than both Control and Treatment 1. So my hypotheses are: Control = Treatment 1, Control < Treatment 2, and Treatment 1 < Treatment 2.

Evaluating the Forecasts in Two Years

I am planning on returning to this study in two years (at the end of 2021) when the forecasts have been realized. I will look up the answers to the questions and assess if the briefings reduced overconfidence in the forecasted values and retrospectively assess how accurate the forecasts were. I will run the same permutation tests on the hit rates for the forecasted values. I hypothesize that the magnitude of the median hit rates will be Control < Treatment 1, Control < Treatment 2, and Treatment 1 < Treatment 2. I may also do some of the supplementary analyses (listed below).

Exploratory, Supplementary, and Explanatory Analyses

There are a variety of analyses I can do after I have the data and run my primary analyses. These analyses may provide additional insight into my results or provide other noteworthy observations.

- Normality tests (such as QQ plots) to see if the data are actually normal or not

- I will also run t-tests and Mann-Whitney U-tests and compare those to the permutation test p-values (as a sanity check)
- A binomial bootstrap test may be a better test than the one-sample t-test for assessing if each group is overconfident, so I can run one of those tests as a check on the t-test
- I may run regressions to see if variables such as self-assessed level of expertise correspond to more or less overconfidence
- Take a wisdom of the crowd approach and average the responses and see how accurate those are
- Run the same primary analyses, but excluding anyone who failed any of the attention checks
- Additional measures to run pairwise comparisons for:
    - Time spent on each question/all of the 20 current and forecasting questions
    - Total number of words used in the text fill-in responses on the last two forecasting questions
    - Total number of "uncertainty" keywords used (such as "uncertain," "overconfident," "unsure," etc.)
    - Use a proper scoring rule such as the interval rule in Jose and Winkler (2009), as used by Jain et al. (2013)
- Seeing if there was a hard-easy effect based on the number of people who got each question correct
- In two years, when I look retrospectively at the forecasts, I may filter by who got the current value questions correct and see if that was an indicator of who got the forecasting questions correct
- Comparison of the interval widths of respondents whose intervals contained the answers and those whose intervals did not contain the answers
- Compute Cohen's d to find the effect size of the treatments
- Sensitivity analysis on the answers to the questions

Appendix F: Additional Data Analysis

This Appendix covers the testing of the assumptions necessary for the statistical tests used in the data analysis and the results of alternative statistical tests.

COMMON ASSUMPTIONS FOR ALL TESTS:
Random Sampling, Independence of Respondents, Random Assignment

We assumed all respondents were independent (between subjects), as they presumably did not speak to one another as they were taking the survey. Respondents were randomly assigned to a group using the randomizer in Qualtrics. The randomization was checked by looking at the demographics. This sample (as with pretty much all experimental samples) was not a perfectly random sample. However, we did what we could to get the survey out to a large number of people from a variety of backgrounds and experiences. There may have been a selection bias toward people who are very interested in the topic and therefore might spend longer on the survey and give wider intervals.

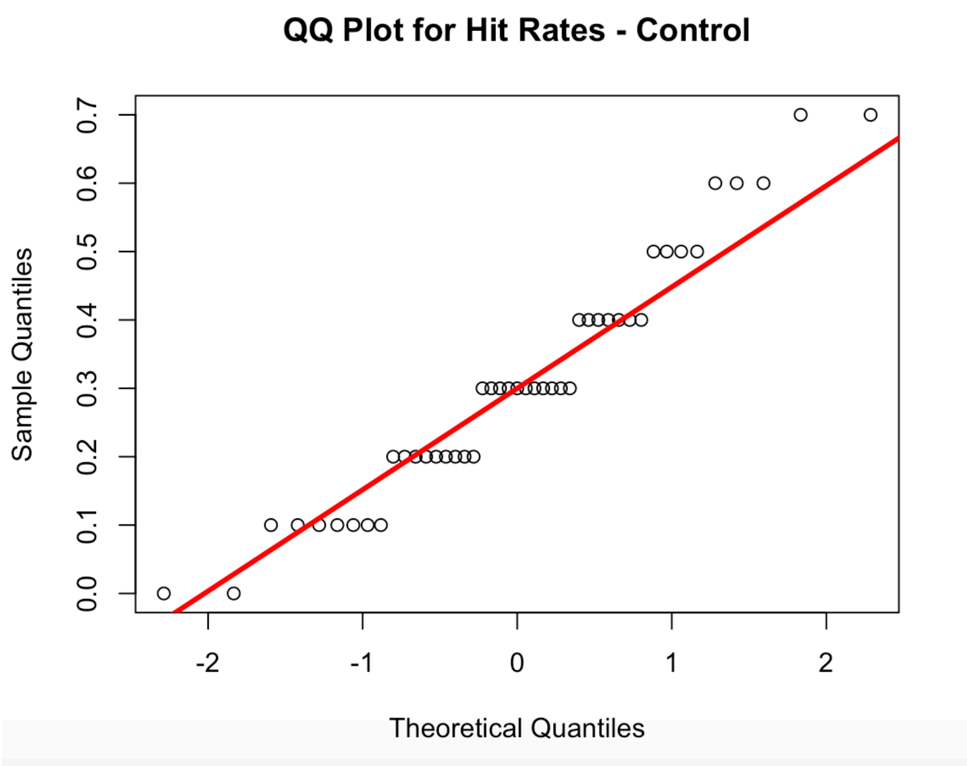
CHECKING NORMALITY OF HIT RATES FOR T-TEST/ANOVA

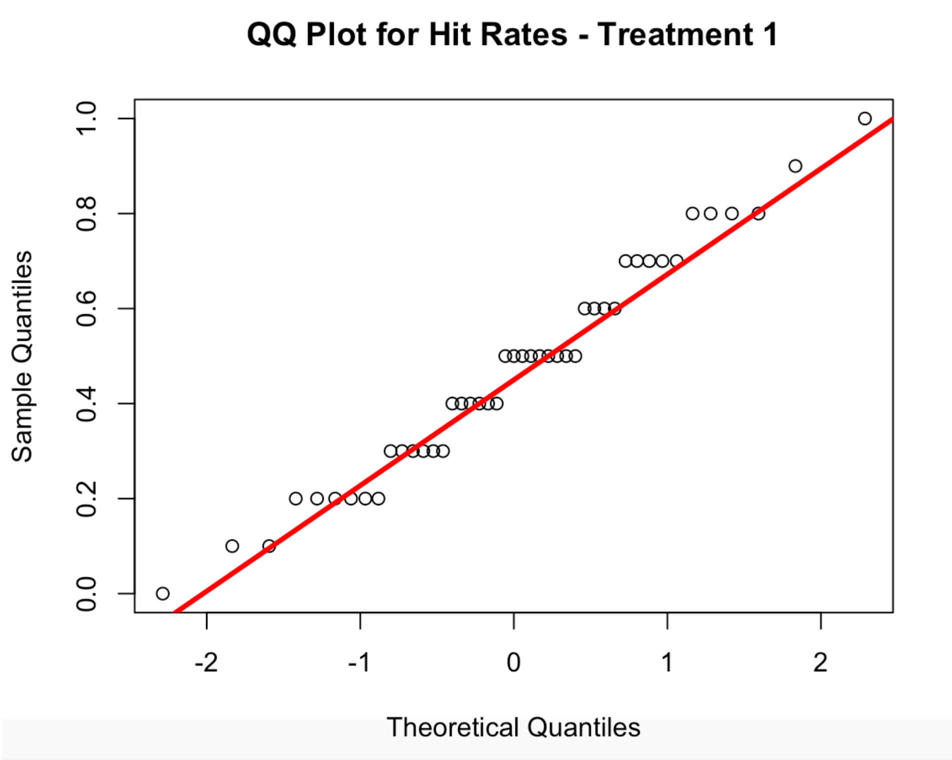
We were originally going to do permutation tests here also (that's what was pre-registered), but after seeing that the histogram of the hit rates was pretty normal, and considering that the ANOVA is a pretty standard test used in the overconfidence literature, we switched to just using the ANOVA as the test.

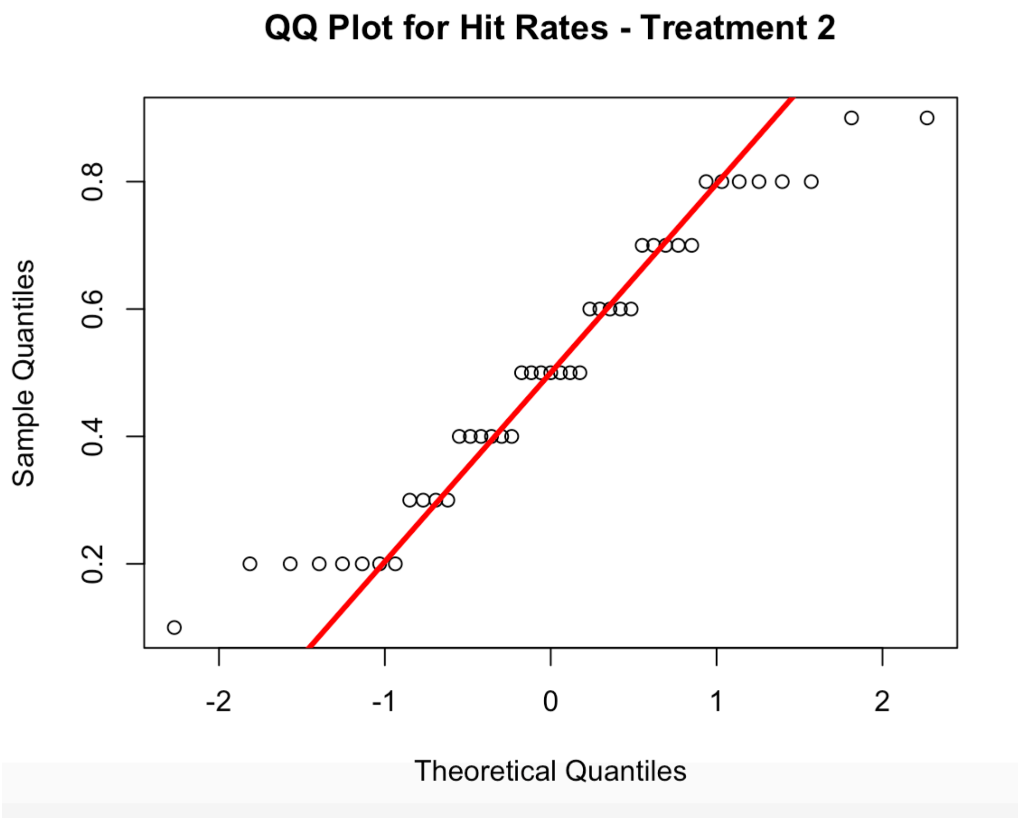
The QQ plots (Figure S6) show that the normality assumption is reasonable for the hit rates. They look sort of odd, but the horizontal clusters are likely due to the discrete values the hit rates can take (0, 0.1, 0.2…, 1.0). There are no outliers (this is mostly because the distribution is bounded at 0 and 1). The variances of the three distributions are similar (this is needed for the ANOVA).

If you don't think the normality assumption has been met, we also ran the non-parametric versions of the one-sample t-test and ANOVA test, and those results follow the QQ plots below. Those tests gave the same results as the results from the t-tests and ANOVA that are reported in the paper.

Figure S6: QQ plots for the three groups.







Non-parametric alternative to one-sample t-test: Wilcoxon Signed Rank Test

This test does not assume normality of the data. It determines whether the median of the sample is equal to mu, in this case 0.8. This gives the same results as the t-tests (see Table S17).

Table S17: P-values for Wilcoxon Signed Rank Test of each group's hit rates ($p < 0.05$ highlighted in green).

	Control	**One-Briefing**	**Two-Briefing**
P-value	4.698e-09	4.178e-08	1.403e-07

ANOVA Test Equivalents with Relaxing Assumptions:

The following tables show the results of the ANOVA-equivalents while relaxing various assumptions. They all give the same results as the regular ANOVA.

Relaxing the equal variances assumption:

The ANOVA equivalent when you don't assume equal variances is called the one-way test (oneway.test() in R). This gives a p-value of 1.359e-05. The Tukey HSD equivalents are just pairwise t-tests, using the pairwise.t.test() function in R that uses the BH correction, shown in Table S18. the table below.

Table S18: Results of the one-way test.

	Comparison		
	Control and One-Briefing	**Control and Two-Briefing**	**One-Briefing and Two-Briefing**
P-value	0.00053	5e-05	0.46492

Non-parametric equivalent: Kruskal-Wallis Test

This is the non-parametric version of the ANOVA test. The p-value is 8.677e-05. The pairwise comparisons are then done with the Wilcoxon rank sum test, which also uses the BH correction (shown in Table S19).

Table S19: Results of the Kruskal-Wallis Test.

	Comparison		
	Control and One-Briefing	**Control and Two-Briefing**	**One-Briefing and Two-Briefing**
P-value	0.00111	0.00019	0.46898

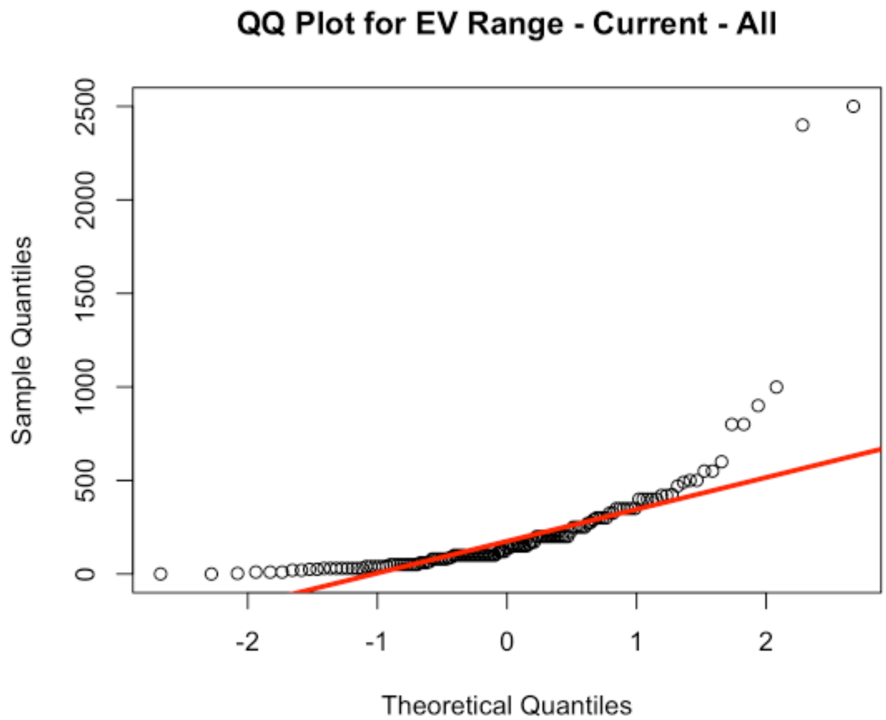
PERMUTATION TEST

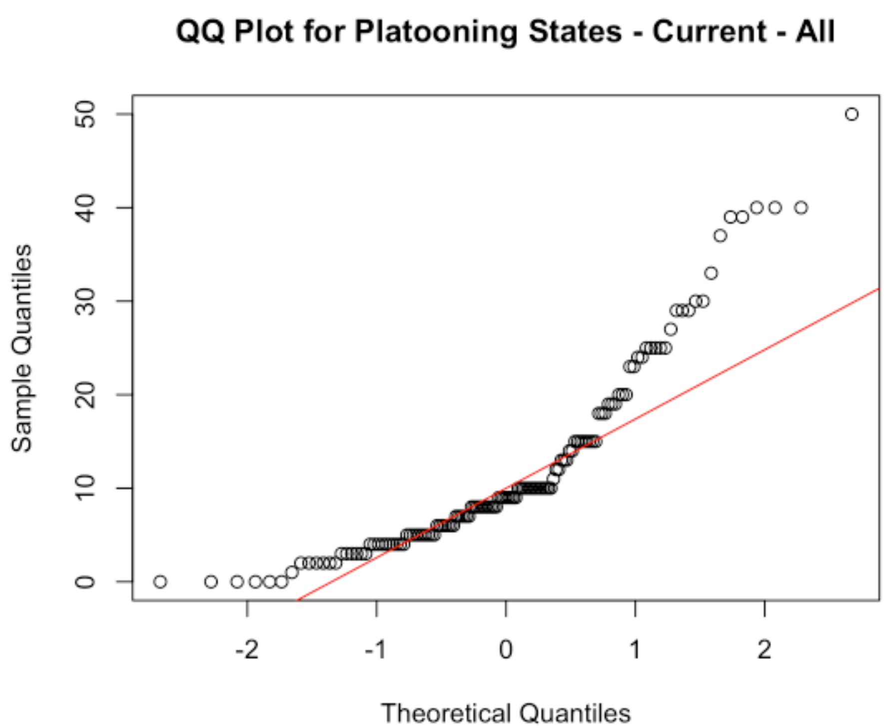
From pre-testing, we knew that the distributions of the interval widths would very likely be non-normal, as some intervals were quite small (when pre-testers had a good idea of what the answer was) and some intervals were quite wide, spanning multiple orders of magnitude (when pre-testers were very unsure of the answer). Given what we saw in the pre-testing data, we would likely have right-skewed data with outliers. This is what we saw in the actual data as well (see QQ plots below). The permutation test has 100% power (Siegel and Castellan, 1988) and is distribution-free. We compared the medians so the permutation tests would be resistant to outliers. We used the correction from Phipson and Smyth (2010) when calculating the p-value (this avoids p-values of zero in permutation tests).

Interval Widths:

Below are examples of QQ plots for the interval widths. There are outliers and (looking at the histograms in Figure S7) the distributions are right-skewed. The graphs are similar for most of the questions, both current and forecasting.

Figure S7: QQ plots for interval widths for two questions – not normal.





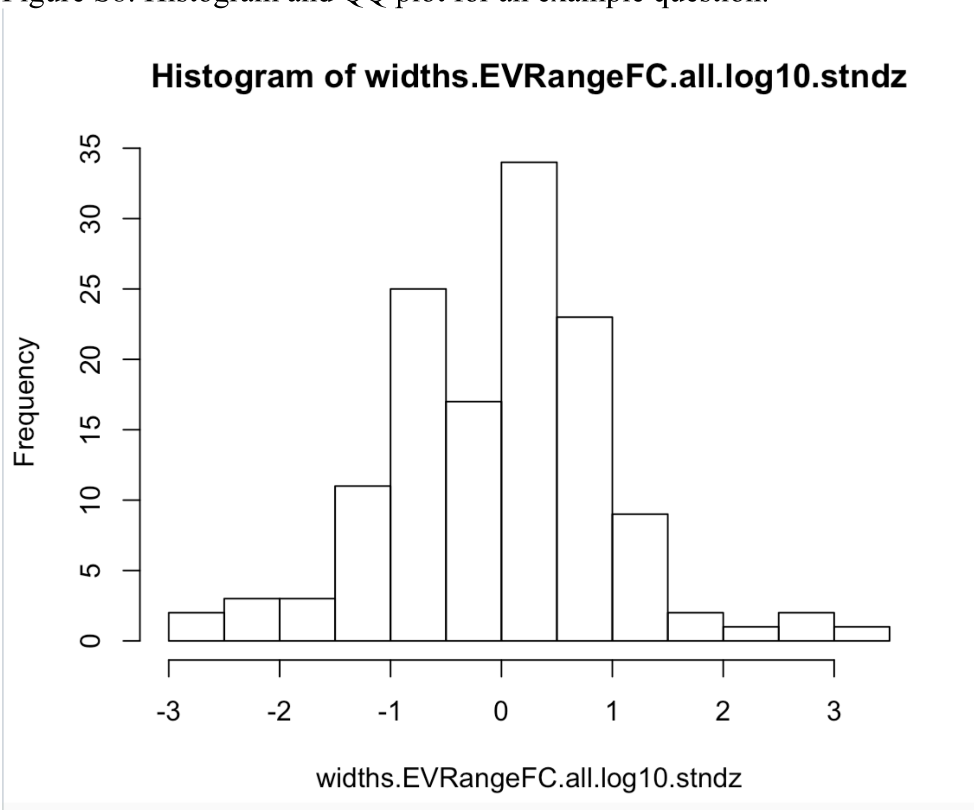
MULTILEVEL MODEL:

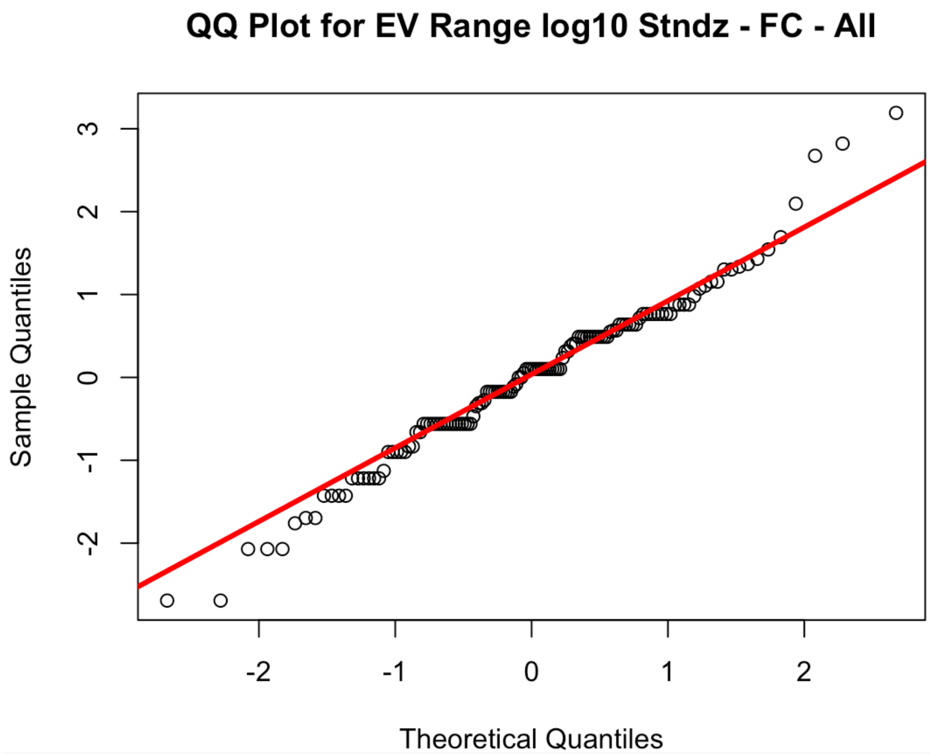
Multilevel models assume normal data for random effects and statistical inference. Since the raw data was not normal, it needed to be transformed in order to satisfy this assumption. $Log_{10}(x+1)$ was used to transform the data to make it normal (the x+1 used because some observations were 0). The log-transformed data was then standardized by subtracting the transformed question mean and dividing by the transformed question standard deviation. This standardization allowed us to compare across the questions, since their numerical values are on different scales.

Checking normality of the log-standardized data:

Figure S8 shows a representative histogram and QQ plot for one of the forecasting questions (about EV range). They look approximately normal.

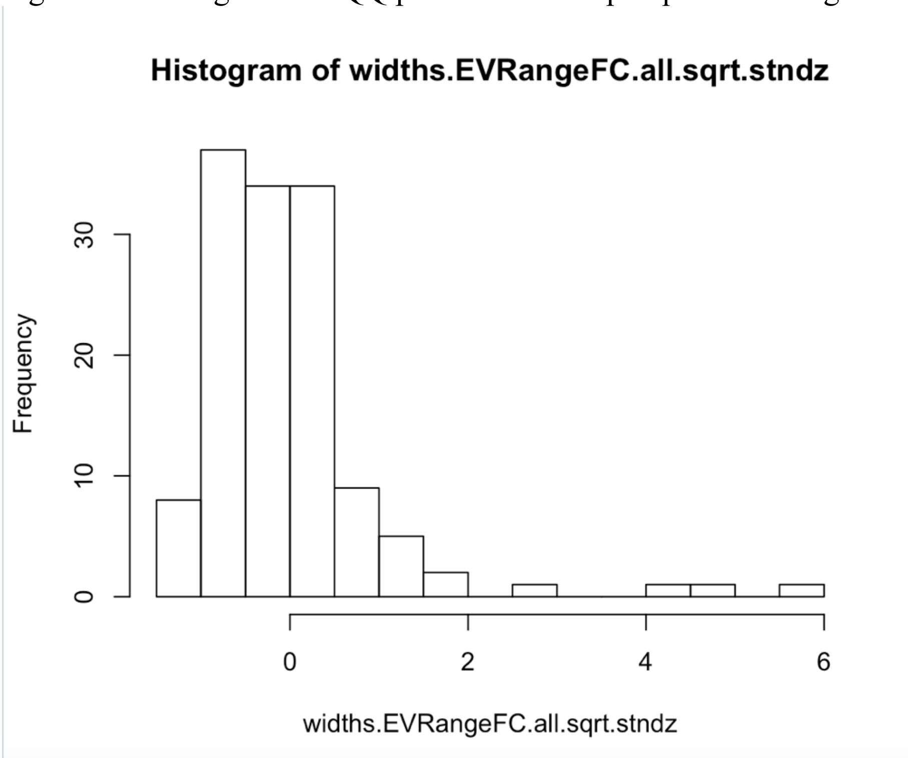
Figure S8: Histogram and QQ plot for an example question.

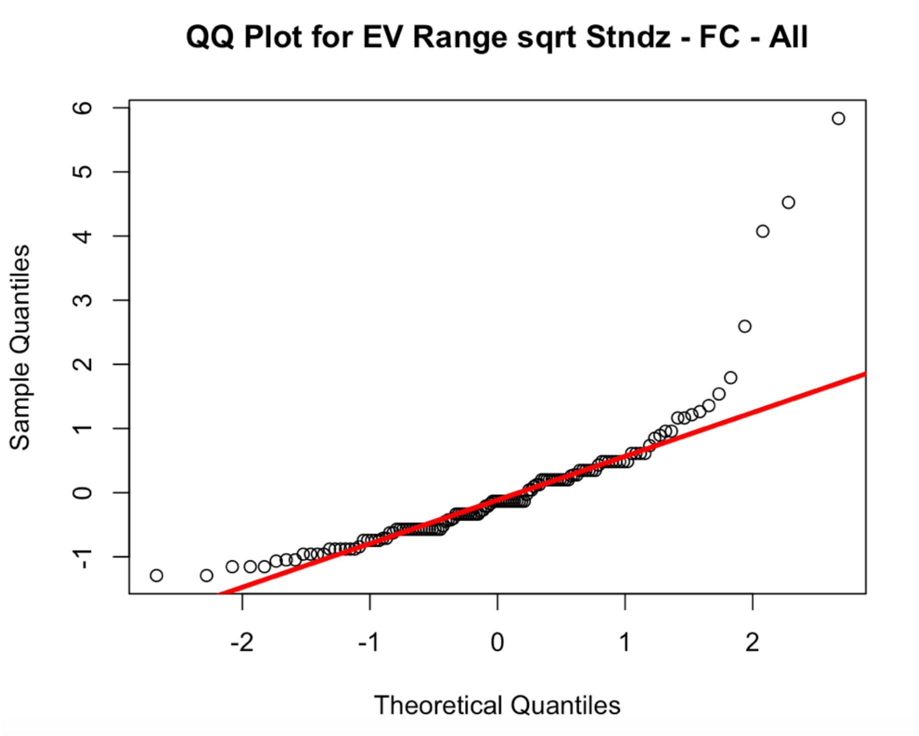




As a check, we also looked at doing a square root transformation. The square root transformed data did not look as good (see Figure S9), so we used the $\log_{10}(x+1)$ transformation.

Figure S9: Histogram and QQ plot for an example question using the square root transformation.





Removing outliers

In order to check that the results were robust to outliers, the same multilevel model was run with and without outliers in the data. Outliers were removed in three different ways: observations more than 3 standard deviations from the log-standardized mean, observations more than 2 standard deviations from the log-standardized mean, and respondents with a high Cook's distance. For the 3 SD cutoff, 9 observations were removed. For the 2 SD cutoff, 62 observations were removed. For the Cook's distance cutoff, 90 observations (9 respondents) were removed (see Figure S10). The results are summarized in Table S20.

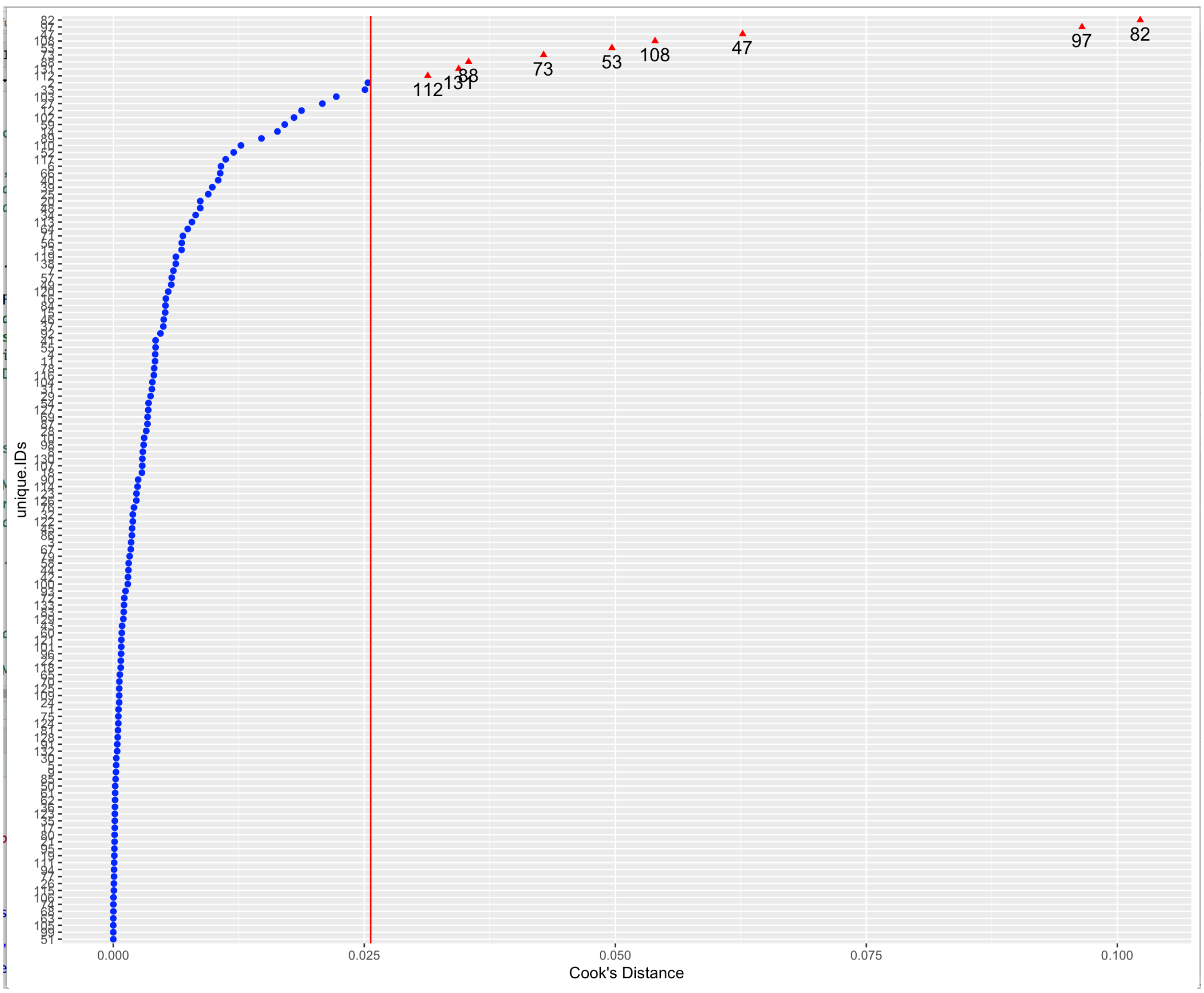


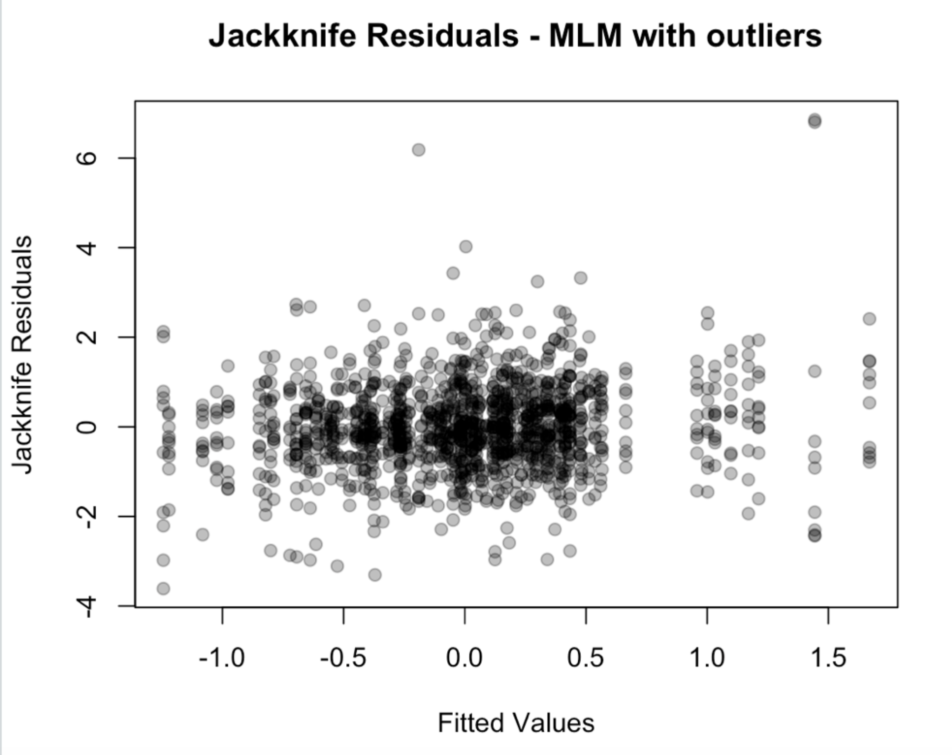
Figure S10: Cook's distance for each respondent.

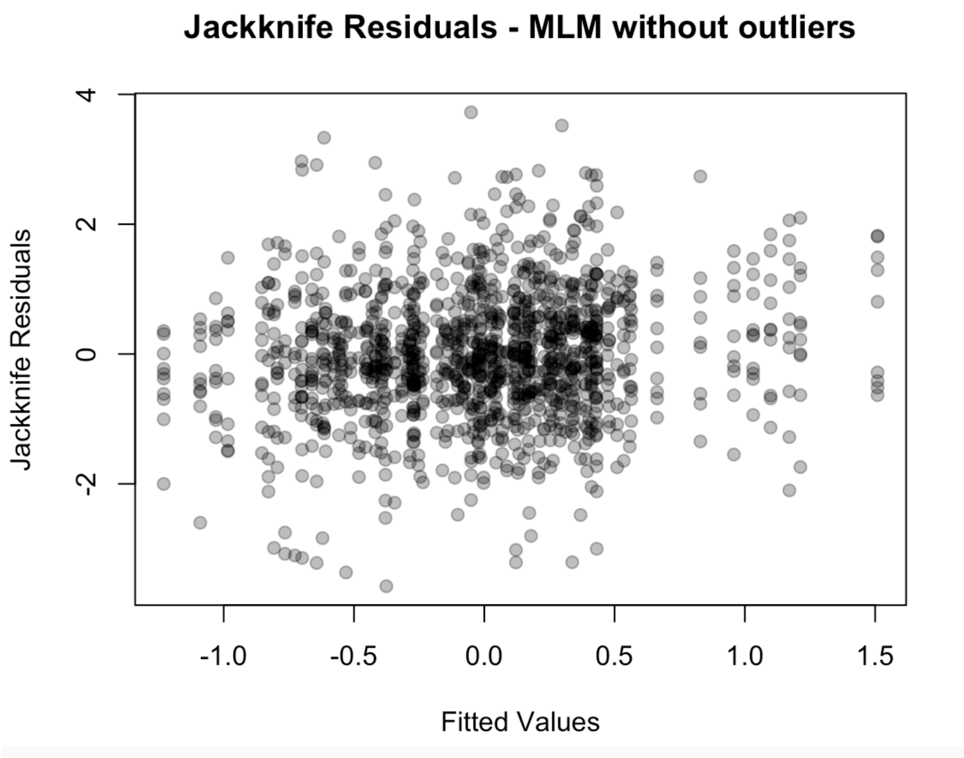
Table S20: P-values for comparisons with and without outliers.

	Control vs. One-Briefing	Control vs. Two-Briefing	One-Briefing vs. Two-Briefing
Including outliers	p = 0.002 (df = 88, t=3.159)	p < 0.001 (df = 86, t=5.157)	p = 0.045 (df = 86, t=2.034)
3 SD outliers removed	p < 0.001 (df = 88, t=3.405)	p < 0.001 (df = 86, t=5.566)	p = 0.030 (df = 86, t=2.201)
2 SD outliers removed	p < 0.001 (df = 88, t=3.595)	p < 0.001 (df = 86, t=5.988)	p = 0.017 (df = 86, t=2.428)
Cook's distance outliers removed	p < 0.001 (df = 83, t=3.746)	p < 0.001 (df = 82, t=6.348)	p = 0.013 (df = 77, t=2.554)

The plots in Figure S11 show the jackknife residuals for the multilevel model with and without outliers (the outliers that were removed were those more than three standard deviations from the mean). Both plots look reasonable.

Figure S11: Jackknife residual plots for the multilevel model with and without outliers.





TIMING DATA

To see how long each respondent spent taking the survey, we looked at the timing data. While survey timing data tends to be noisy (people stopping in the middle of the survey and then returning later, etc.), the graphs give an overview of approximately how long respondents in each group spent taking the survey. Figure S12 shows the timing data for the entire survey and Figure S13 shows the timing data for how long respondents took when the time spent on the briefings was removed.

Figure S12: The amount of time respondents in each group took to complete the survey (outliers have been omitted from the graph).

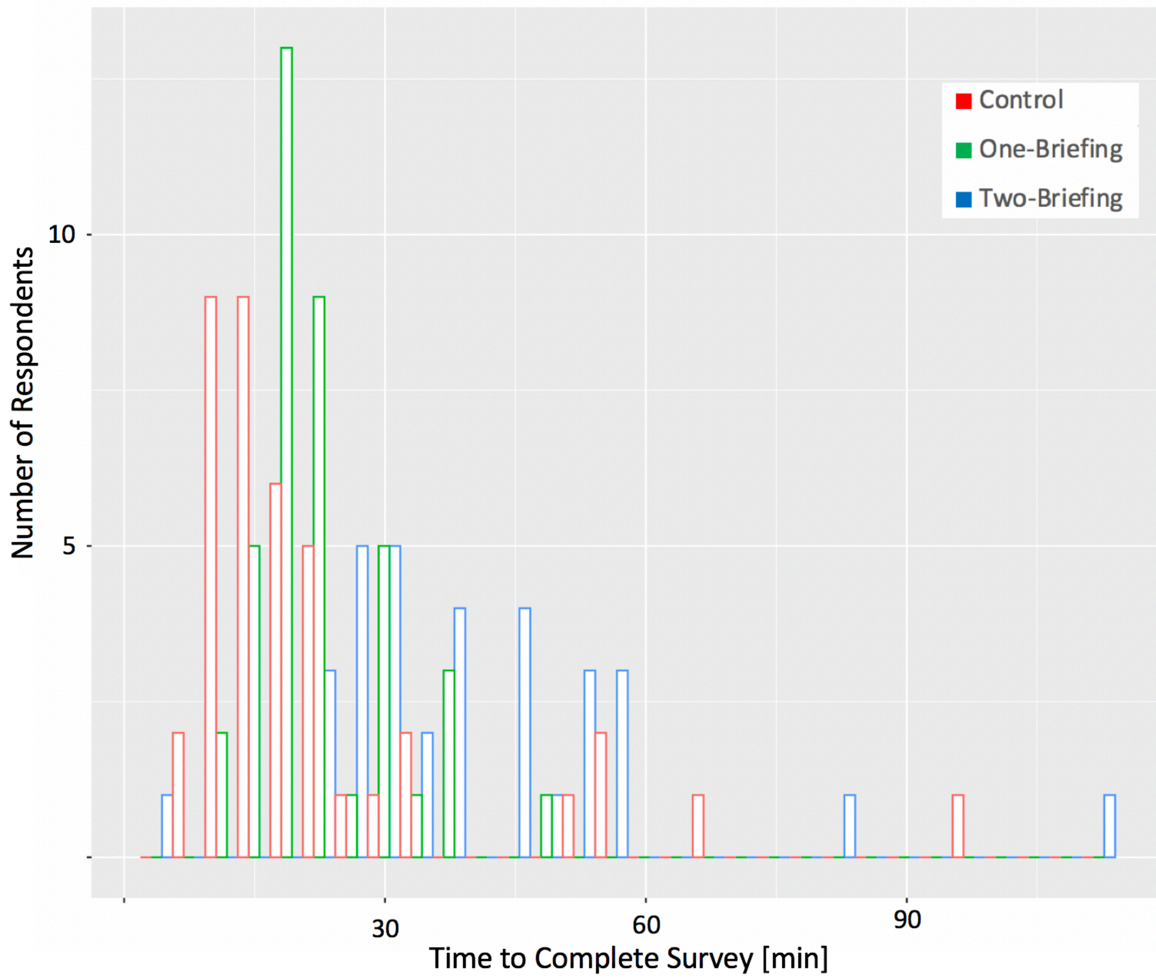
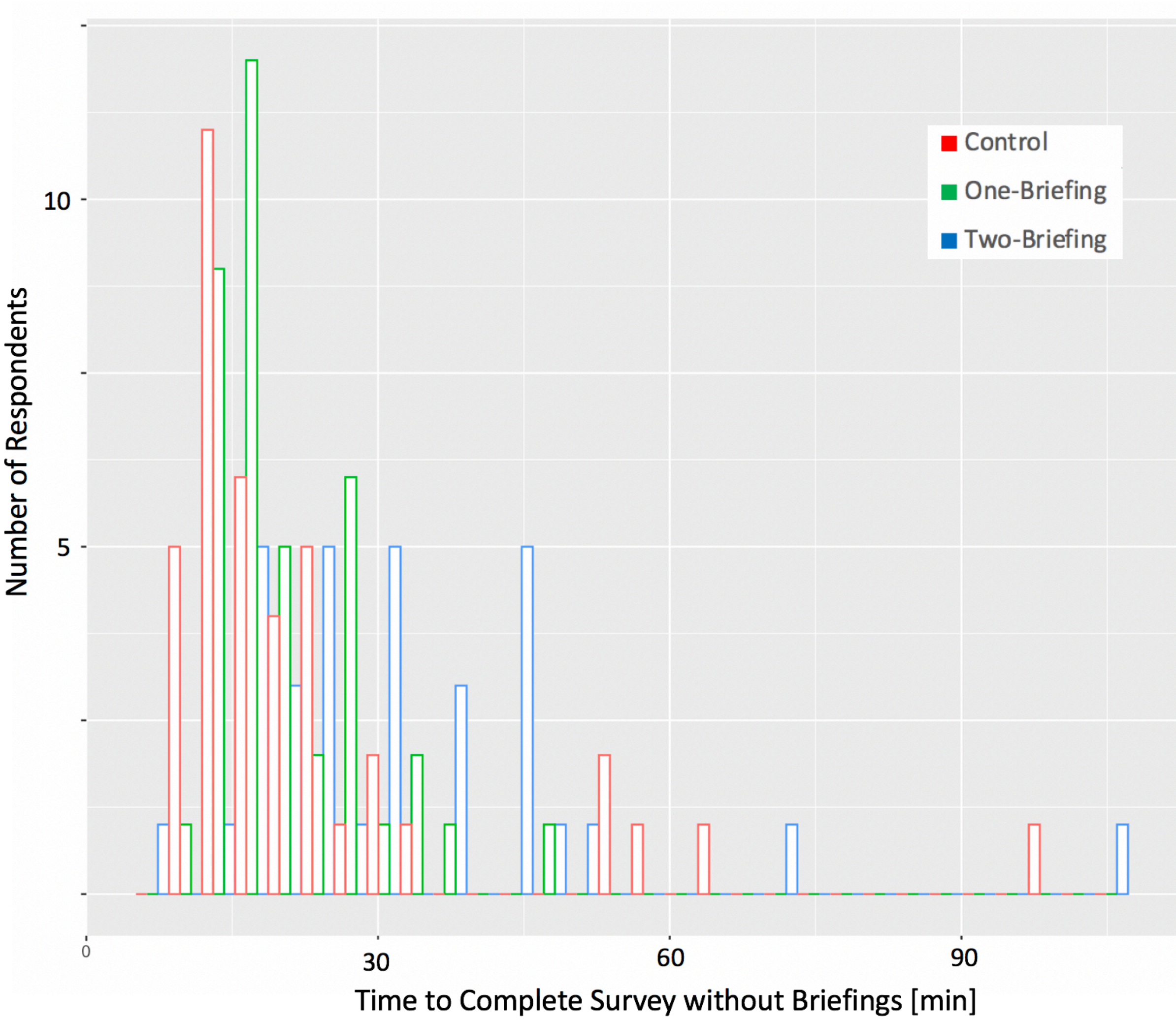


Figure S13: The amount of time respondents in each group took to complete the survey once the time spent on the briefings was subtracted (outliers have been omitted from the graph).



Control
One-Briefing
Two-Briefing
10
5
0
30
60
90
Number of Respondents
Time to Complete Survey without Briefings [min]

Appendix G: Full List of Keywords

We came up with this list over a few months as the survey was developed. As responses came in, words were added to the list (blinded to the group the respondent was assigned to). We decided to count specific country, state, city, and company names. To ensure we captured the specific names that were used by respondents, we read through the replies as they came in and added those words to the list. We will use Excel's SEARCH function to count the number of keywords used in each response.

The keywords are listed below, loosely sorted into policy, economic, and social keywords and in no specific order. The Excel SEARCH function accepts wildcards, so we can use "econom*" and the function will recognize "economic," "economics," "economy," and "economically" as the same keyword. It is not case sensitive. When we enter the keywords in Excel, we will use wildcards at the end of nearly all of them to account for differences in verb tenses and parts of speech. The SEARCH function does not double count keywords (i.e., if someone uses "policy" twice, the keyword count will only increase by 1).

Policy

- President
- Congress
- Local
- Legislation
- Federal
- (Inter)Nation(al)
- Domestic
- Foreign
- Global
- Administration
- Election
- Elected
- Trump
- Obama
- Candidate
- Democrat
- Republican
- Lobbying
- Politics
- Mayor
- Governor
- Supreme Court
- Liberal
- Conservative
- Progressive
- Blue
- Bureaucracy

- Rural
- (Sub)Urban
- Government
- Constitution
- Metro(politan)
- World
- Law(yer/suit)
- Legal
- Regulation
- Infrastructure
- Station
- NHTSA
- EPA
- CAFE
- EIA
- Standard
- Emission
- Support
- Incentive
- Priority
- Invest(ment)
- Clean
- Green
- Energy
- Policy
- Park(ing)
- ZEV
- Department of Transportation
- Mandate
- DMV
- Enact
- Risk
- Public
- State
    - California
    - Arizona
    - Texas
    - Michigan
    - Washington
    - New York
    - New Jersey
    - Pennsylvania
    - Virginia
    - Oregon
- City
    - Detroit

    - Silicon Valley
    - Bay Area
    - Spartanburg
    - Pittsburgh
    - Boston
    - Chicago
- Country
    - United States
    - China
    - Germany
    - United Kingdom
    - Europe
    - Asia
    - Norway
    - Netherlands
    - America

<u>Economic</u>

- Approve
- Authorize
- Allow
- Rule
- Follow
- Inertia
- Agency
- Price
- Cost
- Consumer
- Buy/bought
- Sell/sold
- Sale(s)
- Demand
- Producer/tion
- Adoption
- Mature
- Startup
- Supply
- Manufacturer/ing
- Tax
- Subsidy
- (In)Expensive
- Cheap
- Money
- Trade

- Economy/ic
- Market/ing
- Advertising/ment
- Business
- Competitor/ion
- Commerce/ial/ize
- Industry
- Efficient
- Workforce
- Labor
- Employ/ment/ees
- Job(s)
- Venture (capital)
- Revenue
- Profit
- Automaker
- Customer
- Pefer/ence
- Appeal
- Plant
- Purchase
- Available/ility
- Afford
- OEM
- Penetrate
- MSRP
- Company
    - Toyota
    - Uber
    - Waymo
    - Tesla
    - Amazon
    - BMW
    - VW
    - Rivian
    - Ford
    - Hyundai
    - Electrify (America)
    - Nissan
    - GM
    - Honda
    - Volvo
    - Lyft

<u>Social</u>

- Media
- News
- Taxi
- Rideshare/ing
- Ownership
- Crash
- Fatal/ity
- Kill
- Death
- Safe/ty
- Collision
- Accident
- Perception
- Perceive
- Accept/ance
- Privacy
- Hack
- Concern
- Anxiety
- Population
- Popular/ity
- Environment
- Society
- Pollute/ion
- Carbon
- Ethics/al
- Climate
- Behavior
- Dangerous
- (Mis)Trust
- Fear
- Nervous
- Security
- Traffic
- Permit
- Permission
- Reliable/ility
- Congestion
- Influence/tial
- Benefit/ial
- Resistant
- Feasible
- Viable
- Misconception
- Convenient
- Access/ible

- Reluctant
- Skeptical
- Culture/al
- Alternative
- Social

IV. References

E. Baker, V. Bosetti, L. D. Anadon, M. Henrion, and L. Aleluia Reis, "Future costs of key low-carbon energy technologies: Harmonization and aggregation of energy technology expert elicitation data," *Energy Policy* 80, 219–232 (2015).

I. Ben-David, J. R. Graham, and C. R. Harvey, "Managerial Miscalibration," *The Quarterly Journal of Economics* 128 (4), 1547–1584 (2013).

S. R. Fye, S. M. Charbonneau, J. W. Hay, and C. A. Mullins, "An examination of factors affecting accuracy in technology forecasts," *Technological Forecasting and Social Change* 80 (6), 1222–1231 (2013).

K. Jain, K. Mukherjee, J. N. Bearden, and A. Gaba, "Unpacking the Future: A Nudge Toward Wider Subjective Confidence Intervals," *Management Science* 59 (9), 1970–1987 (2013).

V. R. R. Jose and R. L. Winkler, "Evaluating Quantile Assessments," *Operations Research* 57 (5), 1287–1297 (2009).

Lazard, "Lazard's Levelized Cost of Energy Analysis – Version 12.0," 2018. Available at https://www.lazard.com/perspective/levelized-cost-of-energy-and-levelized-cost-of-storage-2018/

M. G. Morgan, *Theory and Practice in Policy Analysis: Including Applications in Science and Technology*, (Cambridge University Press, 2017).

B. Phipson and G. K. Smyth, "Permutation P-values Should Never Be Zero: Calculating Exact P-values When Permutations Are Randomly Drawn," *Statistical Applications in Genetics and Molecular Biology* 9 (1), (2010).

D. Rotolo, D. Hicks, and B. R. Martin, "What is an emerging technology?" *Research Policy* 44 (10), 1827–1843 (2015).

S. Siegel and N. J. Castellan, Jr., *Non-Parametric Statistics for the Behavioral Sciences,* second edition, (McGraw-Hill, 1988).

J. B. Soll and J. Klayman, "Overconfidence in Interval Estimates.," *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30 (2), 299–314 (2004).

E. Verdolini, L. D. Anadon, J. Lu, and G. F. Nemet, "The effects of expert selection, elicitation design, and R&D assumptions on experts' estimates of the future costs of photovoltaics," *Energy Policy* 80, 233–243 (2015).