Supplementary Materials for

Premature Predictions:

Accurate Forecasters Are Not Viewed as More Competent For Earlier Predictions

Data files, analysis code, and survey materials are available at https://researchbox.org/354

Table of Contents

Supplement 1: Study 3 Pretests

In Study 3 in the main manuscript, we manipulated whether or not the predicted outcome was reasonably knowable over a longer, one-year time frame. We conducted two pretests to help us develop stimuli for Study 3.

Pretest 1 – Open-ended

We recruited 183 participants on Mechanical Turk, of which 100 passed an attention check and finished the survey. Participants were asked to provide us with some examples of predictions that are reasonable to make at certain times. All participants read the following:

> *People can make predictions about many different things. For example, they could predict the winner of an election, how the economy will perform (e.g., whether the stock market will go up), whether a couple will eventually get married, what the weather will be on a certain day, and so on.*
>
> *However, it might not be reasonable to make these predictions at certain times. For example, it is reasonable to predict whether it will rain one day in advance or one week in advance, it is less reasonable to predict whether it will rain one year from today.*
>
> *Other predictions are more reasonable to make farther in advance. For example, it might be reasonable to predict who will win the Super Bowl or an election one year in advance but not ten years in advance.*
>
> ***We want to know some examples of predictions that you think are reasonable to make at certain times.***
>
> *There are no right or wrong answers, we are only interested in your opinions. You may write as much or as little as you'd like, but please provide an example for each question.*

We then gave participants open-ended questions asking them for examples of events that can be reasonably predicted one day, one week, one month, six months, one year, five years, and ten years in advance. Participants typed their answers in an open text box.

Participant responses from Pretest 1 generally clustered together. That is, participants gave similar responses for what events are predictable one day and one week in advance, 6 months and one year in advance, and so on. From participants' responses to Pretest 1, we developed the stimuli for Pretest 2 that we considered "short-term" (i.e., predictable less than one week in advance), "medium-term" (predictable 1 to 6 months in advance), and "long-term" (more than 1 year in advance), and we explain these stimuli in detail below. The complete data file including all participant responses from Pretest 1 is available at https://researchbox.org/354.

Pretest 2 – Closed Response

We recruited 110 participants on Mechanical Turk, of which 104 passed an attention check and completed the survey. Participants in this pilot were asked to indicate at what point in time certain predictions could reasonably be made.

Using participants' responses from Pretest 1, we developed nine predictions that we categorized into short-term, medium-term, and long-term events:

Short Term:
- *Local temperature* – "The high temperature in Washington, DC on October 16, 2020 will be between 65 and 70 degrees Fahrenheit."
- *Stock market* – "The NASDAQ will close between 11,500 and 12,000 on November 20, 2020."
- *Football game* – "The Denver Broncos will defeat the Miami Dolphins on November 22, 2020."

Medium-Term:
- *Super Bowl* – "The Kansas City Chiefs will win the Super Bowl in 2020."
- *Unemployment* – "The United States unemployment rate will be between 3 and 4 percent for the fourth quarter of 2019."
- *Yearly snow* – "Boston will have less snow than average in 2019."

Long-Term:
- *Global temperature* – "The Earth's average temperature will be between 60 and 62 degrees Fahrenheit in 2019."
- *Population trends* – "The population of Bangkok, Thailand will grow to over 10 million people in 2019."
- *National income* – "The annual household income of the United States (not adjusted for inflation) will pass $60,000 in 2018."

Participants were asked to indicate the earliest that they thought someone would have enough information to confidently make each of these predictions, "What is the earliest that you think someone would have had enough information to confidently make this prediction?" (answer choices: 1 day before, 1 week before, 1 month before, 6 months before, 1 year before, 5 years before). Table S1 shows the response frequencies for each possible time frame across the nine different stimuli.

Table S1. Results of Study 3 Pretest 2.

	What is the ***earliest*** that you think someone would have had enough information to confidently make this prediction?					
	1 day before	1 week before	1 month before	6 months before	1 year before	5 years before
Short-Term Events						
Local Temperature	18 (17.5%)	**51 (49.5%)**	15 (14.6%)	8 (7.8%)	9 (8.7%)	2 (1.9%)
Stock Market	27 (26.2%)	**35 (34.0%)**	29 (28.2%)	7 (6.8%)	4 (3.9%)	1 (1.0%)
Football Game	17 (16.5%)	**38 (36.9%)**	29 (28.2%)	16 (15.5%)	1 (1.0%)	2 (1.9%)
Medium-Term Events						
Super Bowl	14 (13.6%)	27 (26.2%)	**28 (27.2%)**	23 (22.3%)	9 (8.7%)	2 (1.9%)
Unemployment	10 (9.7%)	7 (6.8%)	36 (35.0%)	**37 (35.9%)**	12 (11.7%)	1 (1.0%)
Yearly Snow	12 (11.7%)	15 (14.6%)	**32 (31.1%)**	23 (22.3%)	19 (18.4%)	2 (1.9%)
Long-Term Events						
Global Temperature	10 (9.7%)	8 (7.8%)	10 (9.7%)	17 (16.5%)	**32 (31.1%)**	26 (25.2%)
Population Trends	3 (2.9%)	3 (2.9%)	7 (6.8%)	17 (16.5%)	24 (23.3%)	**49 (47.6%)**
National Income	3 (2.9%)	5 (4.9%)	11 (10.7%)	27 (26.2%)	**36 (35.0%)**	21 (20.4%)

Note: Table shows frequency of responses and row percentages in parentheses. Modal responses for each prediction are bolded.

Confirming the results from Pretest 1, majorities of participants in Pretest 2 indicated that the "short-term" predictions were knowable one week or less in advance, "medium-term" predictions were knowable one to six months in advance, and "long-term" predictions were knowable at least one year in advance. We used the "short-term" and "medium-term" predictions derived from this pretest as stimuli in Study 3.

Supplement 2: Additional Preregistered Analyses for Studies in Manuscript

Tables S2-S9 show the additional preregistered analyses for Studies 1a-5.

Study 1a

Table S2. Pairwise T-Tests Comparing Means for Each Time Condition in Study 1a.

	1 month $M = 4.47$	**1 year** $M = 4.78$	**5 years** $M = 5.05$	**10 years** $M = 5.09$
1 month $M = 4.47$		$t(201) = 3.10$, $p = .012$	$t(201) = 3.75$, $p = .001$	$t(201) = 3.14$, $p = .012$
1 year $M = 4.78$			$t(201) = 3.16$, $p = .011$	$t(201) = 2.38$, $p = .108$
5 years $M = 5.05$				$t(201) = .55$, $p = 1$
10 years $M = 5.09$				

Note: P-values are adjusted for multiple comparisons using Bonferroni correction.

Study 1b

Table S3. Study 1b Results By Prediction Domain.

Domain	**1 month**		**1 year**		**5 years**		**10 years**		**Effect of Time (in Months) on Forecaster Evaluation**
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	
Politics	4.48	1.04	5.01	1.12	5.53	0.77	5.08	1.49	$b = .004$, $SE = .002$, $p = .048$
Sports	4.84	1.16	5.21	1.04	5.13	1.21	4.74	1.81	$b = -.002$, $SE = .002$, $p = .425$
Investments	4.78	0.88	5.39	0.83	5.78	0.86	5.61	1.17	$b = .006$, $SE = .002$, $p = .001$
Oscars	5.06	0.79	5.24	1.07	5.01	1.06	5.03	1.25	$b = -.001$, $SE = .002$, $p = .578$
Brexit	4.92	1.15	5.01	1.09	5.55	0.89	5.55	1.25	$b = .006$, $SE = .002$, $p = .004$
Business	5.18	1.01	5.15	1.18	5.15	1.21	5.26	1.27	$b = .001$, $SE = .002$, $p = .733$
Overall	4.88	1.03	5.16	1.06	5.36	1.04	5.22	1.41	$b = .002$, $SE = .001$, $p = .007$

Study 2

Table S4. Study 2 Results by Prediction Domain.

Domain	1 month		1 year		5 years		10 years		Effect of Time (in Months)	95% CI of Indirect Effect of EARS on DV
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*		
DV: Forecaster Evaluation										
Politics	4.46	1.13	4.46	0.96	4.24	1.41	4.54	1.33	b = .0004, SE = .002, p = .837	
Sports	4.48	1.28	4.49	1.00	4.42	1.25	4.54	1.21	b = .0003, SE = .002, p = .849	
Oscars	5.02	0.97	5.14	0.90	4.56	1.10	4.76	1.08	b = -.003, SE = .002, p = .042	
Brexit	5.19	0.84	5.38	0.84	5.15	1.12	5.06	0.89	b = -.002, SE = .001, p = .177	
Business	5.06	0.90	4.93	1.26	4.82	1.04	4.61	1.02	b = -.003, SE = .002, p = .034	
Overall	4.84	1.08	4.88	1.06	4.64	1.22	4.70	1.13	b = -.002., SE = .001, p = .035	
Mediator: EARS										
Politics	3.88	1.10	3.70	1.08	3.34	1.54	3.40	1.21	b = -.004, SE = .002, p = .039	(-.0018, -.00002)
Sports	3.70	1.18	3.52	1.20	2.48	1.29	2.91	1.41	b = -.007, SE = .002, p < .001	(-.0036, -.0009)
Oscars	4.43	1.04	3.83	1.21	3.02	1.22	3.12	1.28	b = -.010, SE = .002, p < .001	(-.0036, -.0006)
Brexit	4.86	1.24	4.63	1.05	3.76	1.17	3.33	1.32	b = -.013, SE = .002, p < .001	(-.0023, .0011)
Business	4.52	1.08	4.27	1.19	3.43	1.14	3.30	1.15	b = -.010, SE = .002, p < .001	(-.0037, -.0005)
Overall	4.27	1.20	3.99	1.21	3.20	1.34	3.21	1.28	b = -.009, SE = .001, p < .001	(-.003, -.001)

Study 3

Table S5. Results for Study 3 for All Participants.

	Short-term Prediction		Long-term Prediction		Effect of Long-term (vs. Short-term) Prediction	Effect of Knowable Long-term (vs. Short-term) Event	Interaction
	M	*SD*	*M*	*SD*			
DV: Forecaster Evaluation							
All Data	5.22	1.05	5.14	1.12	b = .034, SE = .072, p = .640	b = .134, SE = .102, p = .191	b = -.241, SE = .102, p = .019
Knowable Short-Term	5.13	1.10	4.92	1.20	b = -.207, SE = .076, p = .007		
Knowable Long-Term	5.32	1.00	5.35	0.99	b = .034, SE = 0.68, p = .619		
Mediator: EARS							
All Data	4.49	1.21	4.36	1.35	b = .041, SE = .078, p < .001	b = -.926, SE = .110, p < .001	b = -.363, SE = .110, p = .001
Knowable Short-Term	4.08	1.16	3.76	1.35	b = -.322, SE =.081, p < .001		
Knowable Long-Term	4.90	1.13	4.94	1.07	b = .041, SE = .075, p = .584		

Table S6. Results for Study 3 for Only Those Participants Who Passed The Manipulation Check.

	Short-term Prediction		Long-term Prediction		Effect of Long-term (vs. Short-term) Prediction	Effect of Knowable Long-term (vs. Short-term) Event	Interaction
	M	*SD*	*M*	*SD*			
DV: Forecaster Evaluation							
All Data	5.11	1.08	5.16	1.13	b = .190, SE = .145, p = .188	b = .372, SE = .176, p = .034	b = -.448, SE = .167, p = .008
Knowable Short-Term	5.09	1.09	4.85	1.25	b = -.257, SE = .088, p = .004		
Knowable Long-Term	5.21	0.96	5.37	0.99	b = .190, SE = .133, p = .154		
Mediator: EARS							
All Data	4.21	1.22	4.39	1.39	b = .120, SE = .154, p = .438	b = -1.145, SE = .187, p < .001	b = -.445, SE = .178, p = .013
Knowable Short-Term	4.08	1.17	3.53	1.37	b = -.564, SE = .092, p < .001		
Knowable Long-Term	5.06	1.13	4.97	1.07	b = -.120, SE = .147, p = .415		

Study 4

Table S7. Study 4 Results by Prediction Domain for Only Those Participants Who Passed the Manipulation Check.

Domain	Short-term Prediction		Long-term Prediction		Effect of Long-term (vs. Short-term) Prediction	95% CI of Indirect Effect of EARS on DV
DV: Forecaster Evaluation						
	M	*SD*	*M*	*SD*		
Politics	3.75	1.27	3.21	1.28	b = -.533, SE = .152, p < .001	
Sports	3.81	1.20	3.29	1.44	b = -.523, SE = .153, p < .001	
Entertainment	3.87	1.28	3.19	1.19	b = -.677, SE = .144, p < .001	
Overall	3.81	1.25	3.23	1.31	b = -.579, SE = .086, p < .001	
Mediator: EARS						
	M	*SD*	*M*	*SD*		
Politics	3.88	1.23	3.42	1.09	b = -.454, SE = .140, p = .001	(.064, .329)
Sports	3.05	1.09	2.69	1.15	b = -.357, SE = .130, p = .007	(.052, .360)
Entertainment	3.73	1.23	3.23	1.27	b = -.498, SE = .144, p < .001	(.105, .397)
Overall	3.56	1.24	3.10	1.21	b = -.436, SE = .080, p < .001	(.131, .296)

Table S8. Study 4 Results for All Participants.

Domain	Short-term Prediction		Long-term Prediction		Effect of Long-term (vs. Short-term) Prediction	95% CI of Indirect Effect of EARS on DV
DV: Forecaster Evaluation						
	M	*SD*	*M*	*SD*		
Politics	3.74	1.27	3.37	1.29	b = -.367, SE = .138, p = .008	
Sports	3.83	1.20	3.40	1.37	b = -.429, SE = .140, p = .002	
Entertainment	3.85	1.27	3.34	1.24	b = -.516, SE = .135, p < .001	
Overall	3.81	1.25	3.37	1.30	b = -.437, SE = .079, p < .001	
Mediator: EARS						
	M	*SD*	*M*	*SD*		
Politics	3.88	1.23	3.47	1.13	b = -.402, SE = .128, p = .002	(.058, .292)
Sports	3.04	1.10	2.77	1.15	b = -.265, SE = .122, p = .030	(.016, .293)
Entertainment	3.75	1.23	3.33	1.25	b = -.419, SE = .133, p = .002	(.081, .337)
Overall	3.56	1.24	3.19	1.21	b = -.363, SE = .074, p < .001	(.101, .252)

Study 5

Table S9. Study 5 Results by Prediction Domain.

Domain	1 month		1 year		5 years		10 years		Effect of Time (in Months)	95% CI of Indirect Effect of EARS on DV
DV: Forecaster Evaluation										
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*		
Politics	5.05	0.89	4.95	0.98	5.03	0.90	4.80	1.18	b = -.002, SE = .001, p = .118	
Sports	5.06	1.07	5.04	1.14	5.03	1.13	4.82	1.27	b = -.002, SE = .001, p = .110	
Investments	4.96	1.20	5.01	1.21	4.83	1.20	4.89	1.17	b = -.001, SE = .001, p = .458	
Economics	5.10	1.12	5.31	1.03	5.17	1.09	5.14	1.11	b = -.0004, SE = .001, p = .712	
Overall	5.04	1.07	5.08	1.10	5.02	1.09	4.91	1.19	b = -.001, SE = .0006, p = .035	
Mediator: EARS										
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*		
Politics	4.62	0.90	4.08	0.99	3.68	0.99	3.81	1.18	b = -.006, SE = .001, p < .001	(-.0015, -.0001)
Sports	3.95	1.00	3.57	0.98	3.09	1.10	3.14	1.09	b = -.006, SE = .001, p < .001	(-.003, -.001)
Investments	4.00	1.14	3.95	1.23	3.40	1.09	3.25	1.06	b = -.007, SE = .001, p < .001	(-.004, -.001)
Economics	4.68	1.09	4.32	0.92	3.86	1.03	3.81	1.13	b = -.007, SE = .001, p < . 001	(-.003, -.001)
Overall	4.31	1.09	3.98	1.07	3.51	1.09	3.50	1.16	b = -.006, SE = .0006, p < .001	(-.002, -.001)

Supplement 3: Supplementary Studies Not Presented in the Manuscript

In this supplement, we report five studies not included in the main manuscript. Studies S1 and S2 provide replications of Study 2 (with and without the mediator measure), using different prediction stimuli. Study S3 examines whether viewing accurate forecasts in contrast with less-accurate forecasts and how forecasters perform over multiple predictions influences how forecasters are perceived. Study S4 tests evaluations of forecasters who provide predictions that turn out to be incorrect. Study S5 extends our finding to predictions about future events (like Study 4) and tests whether participants view premature predictions as norm violations, causing them to rate those forecasters less positively.

Studies S1 and S2

Studies S1 and S2 were run prior to Study 2 using different prediction stimuli. We do not include them in our main manuscript because we realized that most target events in these studies were largely influenced by the COVID-19 pandemic, making them particularly difficult to predict in advance. Therefore, the results from these studies may not generalize. However, we include them here for completeness. Study S1 includes one item on perceived luck vs. knowledge that was the precursor to our epistemicness measure that we used in the subsequent studies as a mediator. Study S2 was the first study that we conducted in which we included epistemicness as a mediator. Confirming our results from Study 2, participants in these studies again evaluated forecasters less positively the earlier the prediction was made, and this was mediated by the perceived epistemicness of the events (Study S2).

Study S1

Sample. We recruited 1,097 participants on MTurk, of which 1,009 (45.6% female, M_{age} = 41.0 years) passed an attention check that was embedded at the beginning of the study. Participants were paid $0.40 to complete the survey.

Design. As in Study 2, participants were randomly assigned to one of 20 conditions in a 5 (domain: health vs. economics vs. politics vs. sports vs. business) x 4 (prediction timing: 1 vs. 12 vs. 60 vs. 120 months) between-subjects design.

Participants evaluated an expert who made a (correct) prediction about a specific event in advance of the event. Participants were asked to imagine that they were reading an article where an expert had offered that prediction. To manipulate the prediction domain, we randomly assigned participants to read about one of the following five predictions:

1. *Health* – “There will be a major global pandemic in 2020.”
2. *Economics* - “The United States economy will enter a recession in 2020.”
3. *Politics* - “The President of the United States will be impeached in 2020.”
4. *Sports* - “The 2020 Summer Olympics will be postponed.”
5. *Business* - “Tesla will become the world's most valuable car manufacturer in 2020.”

To manipulate prediction timing, participants learned that the prediction was made either 1 month, 1 year, 5 years, or 10 years in advance of the event happening. We manipulated prediction timing by changing the date that the article was ostensibly published. For example, in the health condition where the prediction was made one year in advance, participants saw the following:

> Imagine that you are reading an article **published in March 2019**. An expert is quoted in the article as saying:
> **“There will be a major global pandemic in 2020.”**
> This prediction was correct. The World Health Organization declared that COVID-19 was a global pandemic in March 2020.

Participants in all conditions always learned that the prediction the expert had made was correct. After seeing the prediction and the outcome, participants then answered the same five questions about the person making the prediction that we used in Study 2 (see Table 1 in the main text). That is, they were asked to indicate how knowledgeable, credible, and competent they perceived this person to be, how much they trusted this person (7-point scales from 1 = not at all; 7 = extremely), and to indicate whether they would seek additional information or advice from this person in the future (1 = definitely not; 7 = definitely). In this study, we also included an additional item that asked participants to indicate whether they thought that this person made a correct prediction because they were lucky or because they were knowledgeable (1 = definitely lucky; 7 = definitely knowledgeable). This was a precursor of the epistemicness measure that we used in the subsequent studies. As pre-registered, we collapse all six items to create a single measure of forecaster evaluation ($\alpha = .93$).

Results. For our analysis, we collapsed across prediction domains and used OLS regression to regress evaluation of forecasters on how many months in advance the prediction was made, with fixed effects for prediction domain.

Participants evaluated forecasters less positively the earlier the prediction was made. Specifically, participants' average evaluations decreased from 4.96 out of 7 (SD = 1.13) when the prediction was made one month in advance to 4.53 (SD = 1.31) when the prediction was made 10 years in advance, b = -.003, t(999) = -4.30, p < .001.

Study S2

Sample. We recruited 1,334 participants on MTurk, of which 1,033 (54.5% female, M_{age} = 39.6 years) passed an attention check that was embedded at the beginning of the study. Participants were paid $0.50 to complete the survey.

Design. The design of this study was almost identical to that of Study S1. As in Study S1, participants were asked to judge an expert making a (correct) prediction and were randomly assigned to one of 20 conditions in a 5 (domain: health[1] vs. economics vs. politics vs. sports vs. business) x 4 (prediction timing: 1 vs. 12 vs. 60 vs. 120 months) between-subjects design. However, this study also included our epistemicness mediator (EARS items; see Table 1 in the main text). Because we included this new set of questions, we also removed the question about whether the forecaster was lucky or knowledgeable that was part of our dependent measure in Study S1. We presented the dependent variables and EARS items to participants on separate pages in counterbalanced order (i.e., half of participants answered the DVs first and half answered the EARS first). The order of presentation has no impact on our dependent variable, b = -.043, t(962) = -.506, p = .613.

Results. For our analysis, we collapsed across prediction domains and used OLS regression to regress the respective dependent measure (forecaster evaluation or epistemicness measure) on months, using fixed effects for prediction domain and for whether participants answered the DVs or EARS first.

Participants evaluated forecasters less positively the farther in advance they made their prediction, decreasing from 4.97 (SD = 1.27) when the prediction was made one month in advance to 4.46 (SD = 1.38) when the prediction was made 10 years in advance, b = -.003, t(961) = -3.12, p = .002.

[1] Due to a coding error, one of the conditions (health/120 months) displayed the wrong prediction date to participants. We have removed that condition from our analysis.

Participants also perceived the events as less epistemic (i.e., more determined by chance) the farther in advance the prediction was made. The average perceived epistemicness decreased from 4.73 when the prediction was made one month in advance to 3.12 when the prediction was made 10 years in advance, $b = -.012$, $t(960) = -11.87$, $p < .001$. When participants' responses to the EARS are included as a predictor of participants' evaluation of the forecaster, EARS mediates the main effect of months in advance that we obtained. Using a bootstrapped mediation model, the indirect effect of epistemicness excludes zero, 95% CI (-.005, -.003).

Study S3

Study S3 examines three additional aspects of a forecast that may be important when evaluating forecasters—forecasters' accuracy over multiple predictions, their accuracy compared to other forecasters, and whether the forecasters volunteer their predictions or are specifically asked for them (similar to Study 5 in the main manuscript).

Sample

We recruited 1,235 participants on Prolific (53.6% male, 44.4% female, 1.4% another option, $M_{age} = 39.5$ years) passed an attention check that was embedded at the beginning of the study. Participants were paid $0.80 to complete the survey.

Design

All participants were told to imagine that they were reading an article where three experts were making predictions about the 2022 NFL Draft. Participants were told that the article was written on March 28, 2022 (one month before the draft), October 28, 2022 (six months before the draft), April 28, 2021 (one year before the draft), or April 28, 2019 (four years before the draft).[2] The experts were forecasting who would be the first five players selected in the 2022 NFL Draft (providing their predictions in alphabetical order). One of the experts got all five selections correct,

[2] Unfortunately, there were two typos in this Qualtrics survey. In the "six months before" condition, the date given was actually 6 months *after* the draft (October 2022 instead of October 2021), and in the "4 years before" the date given was actually *three* years before (April 2019 instead of April 2018). However, all conditions had the correct timing with regard to our manipulation (i.e., "six months before the draft"). Our primary analysis includes all of the data, but we also provide results excluding these two conditions.

one got three out of five correct, and the third got none of the five correct (the accuracy of each expert was randomized within subjects, such that one-third of the time, Expert A got all five correct, one-third of the time Expert B got all five correct, and so on). The experts' predictions (and the actual selections) were:

Actual Selections (order drafted)	**All 5 Correct**	**3 out of 5 Correct**	**None Correct**
Travon Walker	Sauce Gardner	Charles Cross	Charles Cross
Aidan Hutchinson	Aidan Hutchinson	Derek Stingley, Jr.	Kenyon Green
Derek Stingley, Jr.	Derek Stingley, Jr.	Kayvon Thibodeaux	Kyle Hamilton
Sauce Gardner	Kayvon Thibodeaux	Travon Walker	Evan Neal
Kayvon Thibodeaux	Travon Walker	Garrett Wilson	Garrett Wilson

Participants were told who the first five selections actually were, and how each expert performed. After reading the scenario, participants answered the same dependent variables (α = .98) as in Studies 2-5 for each of the three experts.

Results

Analysis plan. As preregistered, we used OLS regression to regress the forecaster evaluations on how far in advance (in months) the prediction was made, how many predictions (out of five) the forecaster got correct, and the interaction of these two independent variables, including fixed effects for which expert participants were evaluating (i.e., Expert A, B, or C) and clustering standard errors by participant.

Including all data. Means and standard deviations for all conditions are available in Table S10. For the experts that got all of the selections correct, prediction timing had no effect on perceptions of forecaster competence, b = -.001, t(1,202) = -.50, p = .615. It is possible that this finding is due to a ceiling effect, potentially arising from the comparison to the less accurate forecasters, since the means across the prediction timing conditions for the expert who got all of the selections correct are all very close to the scale maximum, ranging from 6.16 (12 months in advance) to 6.22 (1 month in advance) out of 7. For experts that got three out of five correct, prediction timing had a positive but non-significant effect on perceptions of competence, ranging from 4.64 (1 month in advance) to 4.77 (4 years in advance) out of 7, b = .002, t(1,205) = 1.38, p = .168. Similarly, for forecasters that did not get *any* correct, there was a positive and significant effect of forecast timing, increasing from 2.56 (1 month in advance) to 2.83 (4 years in advance), b = .005, t(1,203) = 2.25, p = .025. That is, in this study, the less accurate forecasters were actually evaluated more *negatively* the closer to the NFL draft they made their prediction.

Analyzing all of the conditions together, there is a positive effect of prediction timing on perceptions of competence, such that forecasters that make earlier predictions are perceived as more competent, b = .005, t(1,210) = 2.39, p = .017, although this is qualified by a marginally significant and negative interaction with the number of correct forecasts, b = -.001, t(1,210) = -1.79, p = .075, such that the effect of forecast timing diminishes for forecasters that get more predictions correct.

Table S10. Full Study S3 Results

DV: Forecaster Evaluation						
Prediction Made ____ Before Draft	**All 5 Correct**		**3 out of 5 Correct**		**None Correct**	
	M	*SD*	*M*	*SD*	*M*	*SD*
1 Month	6.22	0.95	4.64	1.14	2.56	1.35
6 Months	6.20	0.95	4.70	1.13	2.76	1.43
1 Year	6.16	0.99	4.66	1.02	2.63	1.35
4 Years	6.17	0.92	4.77	1.01	2.83	1.36
Overall	6.19	0.95	4.69	1.08	2.70	1.38

Including only 1-month and 12-month conditions. Due to a typo in two of the prediction timing condition, we re-run our analysis including only the 1-month and 12-month conditions. For the experts that got all of the selections correct, prediction timing had a negative, but non-significant, effect on perceptions of forecaster competence, b = -.006, t(596) = -.80, p = .426. For experts that got three out of five correct, prediction timing had a positive but non-significant effect on perceptions of competence, b = .002, t(598) = .249, p = .803. For forecasters that did not get *any* correct, there was a positive but non-significant effect of forecast timing, b = .009, t(597) = .87, p = .385.

Analyzing all of the conditions together, there is a positive but non-significant effect of prediction timing on perceptions of competence, such that forecasters that make earlier predictions are perceived as more competent, b = .007, t(602) = .68, p = .500, and there is a negative, but non-significant, interaction with the number of correct forecasts, $b = -.002$, $t(602) = -.79$, $p = .428$. We should note that, although none of the effects described here reach significance, the point estimates of the effects are larger than the effects when all data are included, suggesting that perhaps we lose too much statistical power by excluding these conditions.

Study S4

In Study S4 we test how forecasters who make incorrect predictions are evaluated.

Sample. We recruited 1,197 participants on MTurk, of which 1,007 (52.1% female, M_{age} = 39.7 years) passed an attention check that was embedded at the beginning of the study. Participants were paid $0.40 to complete the survey.

Design. Participants in this study were randomly assigned to one of 16 conditions in a 4 (domain: economics vs. politics vs. sports vs. business) x 4 (prediction timing: 1 vs. 12 vs. 60 vs. 120 months) between-subjects design.

Participants were asked to judge an expert who made a prediction about a specific event happening in the future, and we manipulated both the prediction domain and how much in advance of the event the prediction was made. However, in this study, we presented participants with incorrect rather than correct prediction. We randomly assigned participants to read about one of the following predictions:

1. *Economics* - "The United States unemployment rate will be above 5% at the end of 2019."
2. *Politics* - "Hillary Clinton will be elected President of the United States in 2016."
3. *Sports* - "The 2024 Summer Olympics will be held in Hamburg, Germany."
4. *Business* - "Amazon will be the largest company in the world by the end of 2019."

Participants learned that the prediction they read about was made either 1 month, 1 year, 5 years, or 10 years in advance of the event happening. As in Study 2, we manipulated the time frame via the date on which we told participants the article was published. For example, for the one-year-in-advance economics prediction, participants read that the article was published in December 2018. In all conditions, participants learned that the prediction was incorrect.

After seeing the prediction and the outcome, participants answered the same questions on forecaster evaluation as in Study S1 ($\alpha = .92$), except that we phrased the luck vs. knowledge item to reference an incorrect, rather than a correct, prediction and was reverse-coded: "Do you think this person made an incorrect prediction because they were unlucky or because they were not knowledgeable?" (1 = definitely unlucky; 7 = definitely not knowledgeable).

Results. For our analysis, we collapsed across prediction domains and used OLS regression to regress forecaster evaluations on months, with fixed effects for prediction domain.

Participants judged forecasters who made incorrect predictions more positively the further in advance the prediction was made. Specifically, participants' average rating of forecasters marginally increases from 3.44 for a prediction made one month in advance to 3.76 for a prediction made 10 years (120 months) in advance, $b = .002$, $t(1001) = 3.05$, $p = .002$. Importantly, however, this finding appears to be driven by the business (Amazon) stimulus in particular which has a particularly large effect ($M_{1\ month} = 3.82$ vs. $M_{120\ months} = 4.61$), $b = .007$, $t(249) = 4.33$, $p < .001$, although the other stimuli are directionally consistent. Excluding the business stimulus, the effect of months in advance on evaluation of forecasters is not significant ($M_{1\ month} = 3.32$ vs. $M_{120\ months} = 3.48$), $b = .001$, $t(751) = 1.093$, $p = .275$.

Study S5

Study S5 extends our finding to predictions about future events (like Study 4 in the main manuscript) and shows that participants also rate forecasters less positively if they make predictions too far out in the future. We furthermore test whether participants view premature predictions as norm violations, causing them to rate those forecasters less positively.

Sample

We recruited 647 participants on MTurk, of which 536 (52.2% female, $M_{age} = 38.3$ years) passed an attention check that was embedded at the beginning of the study. Participants were paid $0.40 to complete the survey.

Design

The design of Study S5 is identical to that of Study 4 but includes new stimuli and tests an additional mediator. Participants were randomly assigned to one of eight conditions in a 4 (domain: sports vs. economics vs. entertainment vs. investments) x 2 (event timing: this year vs. next year) between-subjects design.

As in Study 4, participants learned that they are reading an article published on the day of the experiment (March 25, 2021). The event was either scheduled to take place within a month of the experiment (April 2021) or a year after the experiment (April 2022) and was in one of four domains (sports, economics, entertainment, investments). They were:

> *Sports* – "The Gonzaga Bulldogs will win the NCAA men's basketball tournament in April [2021/2022]."
> *Economics* - "The United States unemployment rate will be between 5% and 6% at the end of April [2021/2022]."
> *Entertainment* – "A film released by Netflix will win the Academy Award for Best Picture in [April 2021/February 2022[3]]."
> *Investments* –"The price of one Bitcoin will be above $80,000 at the end of April [2021/2022]."

After reading the scenario, participants were asked the same dependent variables ($\alpha = .93$) and EARS items ($\alpha = .71$) as in Studies 2-5. To capture whether or not participants felt that premature predictions violated norms, we also asked participants the following four questions ($\alpha = .82$):

1. How weird is it for someone to make this prediction? (1 = Not at all; 7 = Extremely)
2. How appropriate is it for someone to make this prediction? (1 = Extremely inappropriate; 7 = Extremely appropriate; reverse-coded)'
3. How arrogant is it for someone to make this prediction? (1 = Not at all; 7 = Extremely)
4. How unusual is it for someone to make this prediction? (1 = Not at all; 7 = Extremely)

The three groups of questions (DV, EARS, and prediction weirdness items) were presented in random order.

Finally, we again included a manipulation check question at the end of the survey to ensure that participants noticed that the events were scheduled for different dates in different conditions (86.0% answered correctly).

Results

Analysis plan. As preregistered, we excluded those participants who failed to answer the manipulation check correctly and collapsed across prediction domains for our analysis. We used OLS regression to regress the respective measure (forecaster evaluation, EARS, or prediction weirdness measure) on whether the prediction was a short-term or long-term prediction, including fixed effects for domain.

Main effect. Participants rated those who made predictions about events a year in the future directionally less positively ($M = 3.71$, $SD = 1.26$) than those who made predictions about events

[3] The Academy Awards are typically held in February but were delayed to April in 2021 due to COVID-19.

taking place within the next month (M = 3.91, SD = 1.34). However, this effect was only marginally significant, $b = -.210$, $t(436) = 1.69$, $p = .092$.

Mediation. Consistent with Studies 2-4, participants in this study also felt that events in the far future were less epistemic (M = 3.22, SD = 1.28) than events in the near future (M = 3.60, SD = 1.28), $b = .388$, $t(436) = 3.32$, $p < .001$. The indirect effect is significant, 95% CI (.08, .32), accounting for nearly all (93.8%) of the overall effect.

Perceived Norm Violations. Participants directionally, although not significantly, perceived predicting events in the far future to be more of a norm violation (M = 3.53, SD = 1.48) than predicting events in the near future (M = 3.33, SD = 1.46), $b = -.208$, $t(436) = -1.51$, $p = .132$. Although this did not reach significance, this may be due to heterogeneity across stimuli. As in Study 3, for instance, if an event is considered to be epistemic over a longer period of time, an early prediction may not be considered unusual and prediction weirdness may still be a viable mediator (Zhao et al., 2010). Even with this in mind, however, the indirect effect is not significant, 95% CI (-.02, .20). This suggests that it is unlikely that our effect is being driven by the perceived norm violation of making predictions far in advance of an event.