Are Talking Heads Blowing Hot Air?

An Analysis of the Accuracy
of Forecasts in the Political Media

Public Policy 501

Holly Donaldson
Russ Doubleday
Scott Hefferman
Evan Klondar
Kate Tummarello

Table of Contents

Abstract

We evaluated the predictions of politicians, journalists, and pundits (collectively, "prognosticators") over a 16-month period surrounding the 2008 United States federal elections. We sought to test the accuracy of these predictions and understand what makes individuals unusually good or bad at prognostication. We evaluated a random sample of Meet the Press, This Week, *and* Face the Nation *transcripts and printed columns from the most prominent American prognosticators. Ultimately, we determined that some prognosticators are significantly better than others, and there are characteristics that significantly impact a prognosticator's predictive power.*

After finding which characteristics make an accurate prognosticator, we ranked prognosticators based on outcomes. We found that a number of individuals in our sample, including Paul Krugman, Maureen Dowd, Ed Rendell, Chuck Schumer, Nancy Pelosi, and Kathleen Parker were better than a coin flip (sometimes, substantially so.) A number of prognosticators were frequently inaccurate, including Cal Thomas, Lindsey Graham, and Carl Levin. Our paper explores the reasons for these differences and attempts to evaluate the purpose of prognosticators in the media in light of their varying degrees of accuracy.

Introduction

Parents frequently say that kids should not believe everything they hear on TV. Is this true? And does it apply to what people read in newspapers, too? We sought to test the powers of prediction possessed by political pundits and politicians, evaluate their accuracy, and test the qualities that made for good predictors. TV may (or may not) harm kids' brains, but it turns out that parents probably should not believe most of what they hear on TV, either.

Millions of Americans tune in for Sunday morning talk shows and read weekly op-ed columns to gain insight into current events and global politics and to anticipate what is coming next. They watch politicians, pundits, and journalists pontificate around the "round table" and read the opinions of newspaper columnists each week, but what do we stand to gain from this? It is possible that prognosticators offer deeper analysis of current issues and can more accurately predict future outcomes. It is possible that prognosticators serve solely as entertainers. Alternatively, prognosticators may present opposing viewpoints that are representative of the wide range of ideologies in our political spectrum. Regardless, one has to wonder whether certain prognosticators are more valuable than others.

We explored these questions through a large-scale analysis of the predictions made by politicians, pundits, and journalists in both TV and print. Taken together, we called predictors from all groups prognosticators. We evaluated the predictions of prognosticators over a 16-month period, from the beginning of September 2007 to the end of December 2008. This was designed to encompass the 2008 election season and capture the largest possible number of predictions. By analyzing the prognosticators' statements in transcripts and printed columns, we sought to find and test their predictions. Our data yields results with implications that begin to answer some of our initial questions.

It turns out that we, as consumers, should be watching and reading certain prognosticators over others based on their predictive ability. In addition, political topics tend to elicit certain predictive qualities, such as being phrased using positive/negative or extreme language. When we created an absolute, holistic scale on which we could place individual prognosticators, certain individuals did markedly better than others.

To our knowledge ours is the first attempt at creating a "consumer report" for the average American who wants to know when to tune in and who to take most seriously. Our hope is that with enhanced accountability, prognosticators will become better--and with enhanced visibility, citizens will be more discriminating, listening to more accurate predictors.

Literature Review

Existing literature relevant to political prognostication focuses, with one major exception, on critical skepticism rooted in theoretical opinion rather than statistical evidence. Much has been written about the "sin" of punditry, a trade accused of sacrificing factual integrity for the pageantry of content-free entertainment (Perlstein, 11). Authors label pundits "frauds" and "fools" who "blind us to the ineluctable reality that no one knows what the future holds" (Perlstein, 12). These accusations carry little weight, however, when accompanied by merely anecdotal evidence. The exception to the problem of unsupported criticism of prognosticators as terrible predictors comes from Philip Tetlock's book *Expert Political Judgment* about his twenty year study of the accuracy of political forecasting. Tetlock's book delves into "the psychology of judgment under uncertainty" and "the limits of human knowledge," drawing conclusions about both the accuracy of "expert" predictions and the reasoning behind them.

Tetlock's experiment had 284 experts make 82,361 predictions about political and economic trends. Over time those predictions were checked for accuracy, and Tetlock examined his data for trends. Among the trends Tetlock uncovered in his study:

* simple extrapolation using mathematical models typically does better than human predictors
* education and popularity increase the predictors' confidence but not their accuracy
* prognosticators overpredict change and underpredict the status quo
* extremists predict worse than moderates
* some people predict better than others, and their superiority will not be confined to their area of expertise (Tetlock, 23-163).

Tetlock's findings raise questions about the role of experts in society, the degree of faith the public should have in experts, and the underlying reason some people predict better than others.

While initially investigating whether higher levels of education and experience correspond to higher predictive accuracy, Tetlock ultimately concluded that cognitive style was the most important influence on prediction accuracy. Using the framework derived from Isaiah Berlin's essay *The Hedgehog and the Fox* that "hedgehogs know one big thing; foxes know many things, (Berlin, 3)" Tetlock separated experts into two groups with competing cognitive approaches to prediction and found "the hedgehog-fox dimension did what none of the other traits did: distinguish more accurate forecasters from less accurate ones in both economics and politics" (Begley, 45).

According to Tetlock, there are clear differences between hedgehogs and foxes. Hedgehogs "know one big thing" and "apply that one thing everywhere," express "supreme confidence in their forecasts, dismiss opposing views and are drawn to top-down arguments deduced from that Big Idea"; they "seek certainty and closure, dismiss information that undercuts their preconceptions and embrace evidence that reinforces them" (Begley, 45). Foxes "consider competing views, make bottom-up inductive arguments from an array of facts, doubt the power of Big Ideas" and "are cognitively flexible, modest and open to self-criticism" (Begley, 45). Ultimately, "what experts think matters far less than how they think: their cognitive style" (Begley, 45). Tetlock found that foxes outperform hedgehogs in prediction accuracy in virtually all fields, across all time periods, and across the various levels of expertise.

Expert Political Judgment also considers two types of general skepticism found in theoretical literature about prognostication. Tetlock mentions both radical skepticism, which is the belief that nobody knows anything, and ontological skepticism, the idea that the nature of the world is unpredictable. Both are ideas well illustrated by Rick Perlstein, a contributor to *The Nation*. Perlstein's disbelief in and distaste for prognosticators stems from a blend of radical and ontological skepticism. Perlstein's article "Pundits Who Predict the Future are Always Wrong" goes so far as to "call punditry a sin" (Perlstein, 12). Perlstein dismisses forecasting because of ontological skepticism, alleging "history does not repeat itself, nor does it unfold in cycles" (Perlstein, 13). Not only does Perlstein claim "there's nothing you can really know about the future at all," he warns that "to pretend otherwise is an insult to democracy" (Perlstein). Appealing to radical skepticism and criticizing conventional wisdom, Perlstein concludes that political prognostication "blinds us to the only actual, ineluctable reality--that no one knows what the future holds" (Perlstein, 11).

Another important comment on political prognosticating comes from Glenn Greenwald's editorial in *The National Interest*. His article "The Perilous Punditocracy" comments on political prognosticators in the news media. Greenwald criticizes prognosticators not for their inaccurate predictions, but for the lack of accountability in their profession. Greenwald despises prognosticating as "the ultimate accountability-free profession" in which inaccurate prognosticators "never learn their lesson, are never held accountable and virtually never acknowledge their errors" (Greenwald, 25). To Greenwald, the "distinction between reporters and opinionists has been eroded almost completely," and most predictions are based on personal bias and a "self-centered, self-referential" method that projects too much personal opinion into predictions rather than focusing on facts (Greenwald, 25).

Greenwald's article calls into question the ambiguous classification of prognosticators and whether they should

be considered academics or entertainers. Greenwald criticizes prognosticators for "never having to say you are sorry, or even admit that you are wrong" (Greenwald, 27). Instead, they "just move seamlessly to the next story line, ignoring their past record of fundamental mistakes as though it never happened" (Greenwald, 28). If accuracy is unimportant and professional accountability is nil, political prognosticators simply publish inaccurate, opinionated personal takes on the future.

While prognosticator accuracy is clearly a popular topic of criticism, so too is criticizing the experts themselves. Susan Douglas' article "You too can be a Pundit" in *The Progressive* offers a biting sarcastic how-to guide for someone considering a career in prognostication. According to Douglas, prognostication is about pageantry, not knowledge. To be a prognosticator "you don't need to know anything about anything" (Douglas, 17). Douglas' list of prognosticator necessities: "you have to be real good at vague, content-free assessments and far-fetched, irrelevant predictions; you need to focus on the imagery of events, not their substance; you must sound emphatic and exude total certitude" (Douglas, 17). Douglas assures that as a prognosticator you will never have to back up your predictions with facts, and stresses that "[prognosticators] can and should ignore all facts that don't fit in with your point of view" (Douglas, 17).

Authors such as Douglas, Perlstein, and Greenwald may personify the educated skeptics' perception of prognosticators, but they all still agree on the general public's dependence on prognosticators. While statistical models may be more accurate, Sharon Begley points out that "booking statistical models on talk shows probably wouldn't help their ratings" (Begley, 45). Even Tetlock was aware of this fact, and he concluded the public believes in prognosticators because of their need to believe that events are under our control, and their lack of understanding of the laws of probability. According to literature it seems that, to the average audience member, prognosticators are not about accurately predicting the future. Instead, prognosticators are entertainers fulfilling our human need to feel in control and to provide comfort that events are predictable. In this sense, being right about the future appears less important than speaking confidently about an exciting, changing future.

Methodology

Although our study is similar to Tetlock's in that ours also examines the accuracy of experts, our study differs both in its methodology and in the type of experts it samples. Tetlock had his prognosticators predict on certain subjects as a part of his study, whereas we examined the predictions our prognosticators made of their own accord. In addition, our sample was limited exclusively to professional political prognosticators who appear on TV or in print on a regular basis. Rather than consciously participating in our study, our prognosticators became subjects by way of random selection. We examined the predictions made by these prognosticators during a given time period surrounding the 2008 election. We looked for factors that could influence the accuracy of prognosticators generally. Our results may provide insights into the role of punditry in the American political sphere. We are the first to evaluate individuals who make their predictions exclusively in the media.

Obtaining Prognosticator Sample

Print Media:

To obtain a representative sample of print columnists, we researched the top 10 most widely read liberal and top 10 most widely read conservative columnists, along with the top two independent columnists (by total average number of readers) in the United States (Media Matters). We determined that a columnist would need to produce at least 20 columns in the sample period to be considered for analysis. Since all of these columnists write one to two columns a week, their total columns in our time period (September 2007-December 31, 2008) easily exceed our minimum necessary "appearances" in order to include each of the columnists in our study. As a result, all 22 columnists were included in our print media sample and in our final sample of prognosticators.

Television Media:

To obtain a representative sample of network television prognosticators, we looked at the Sunday morning talk shows on each of the three major broadcast networks: NBC's *Meet the Press*, ABC's *This Week*, and CBS' *Face the Nation*. For each show, we tallied all appearances by each guest during the sample period. From the aggregate list (i.e. the sum of all tallies from all three of the shows during our time period), we cut all guests that appeared less than five times, leaving a sample of 36 guests (pundits, politicians, and administration members).

Final Prognosticator Sample:

To obtain the final prognosticator (defined as the group of pundits, politicians, and administration officials) sample, we combined our print media and television media samples to form a list of 56 total prognosticators. The total was only 56 because two people appeared on both the print and television media lists.

From this list, we randomly selected five prognosticators for each analyst. We did this by numbering each prognosticator and then using a random number generator to select a prognosticator and assigning it to one of our five analysts in order. These five prognosticators were a starting point for each analyst. If, in the course of evaluation, it is discovered that a prognosticator did not make five testable predictions, the analyst picked a different prognosticator from the list of unassigned prognosticators (at random) to fill the vacancy.

Preparing to Code

Selecting Columns & Transcripts

Analysts randomly selected the prognosticators' appearances (columns or transcripts from TV appearances) to read in search of predictions. This process is described below:

TV Appearances

Our initial research showed that there are fewer predictions made during TV appearances. As a result, analysts took a census of the available appearances within our time frame, meaning they looked at every TV appearance between September 1, 2007 and December 31, 2008. The only exception was if a prognosticator appeared more than 20 times during the time frame, in which case the analyst generated a random sample of 20, as illustrated by the column selection method below.

Columns

To create a sample of columns, analysts first located all columns written within the sample time frame. They then chose 20 randomly generated numbers, arranged the columns in chronological order, and then assigned columns numbers (1 was oldest, counting up to the most recent.) Analysts included columns whose numbers corresponded to the 20 randomly generated numbers.

Coding System

Prediction Variables

Variable Code	Full Name	Explanation
PName	Prognosticator Name	A three-letter code representing the name of the prognosticator
PredAcc	Prediction Accuracy	A 1 (cannot happen) to 5 (will happen) rating of each statement by a prognosticator.
PredTrue	Prediction Truth	A 1 (event didn't happen) to 5 (did happen) rating of each event as made in a prediction.
PredTrue Ref	Prediction Truth Reference	Reference (hyperlink) for a prediction truth assessment
Conditional	Conditional Prediction	Whether or not a prediction is conditional (if/then.) 1 if conditional, 0 otherwise.
CondTrue	Conditional Prediction Truth	A 1 (didn't happen) to 5 (did happen) rating of the conditional ("if X") portion of a conditional prediction.
Timeframe	Chronological scope of the prediction	Measured in months, round up
PredMed	Prediction Medium	1 if TV, 0 otherwise.
RepeatPred	Repeat Prediction	1 if a person already made the same prediction, 0 otherwise.
DemPrimary	Democratic Primary Prediction	1 if a prediction about the Democratic Presidential primary, 0 otherwise.
GOPPrimary	GOP Primary Prediction	1 if a prediction about the GOP Presidential primary, 0 otherwise.
PresPred	Presidential Prediction	1 if a prediction about who will win the Presidency in 2008; 0 otherwise.
VPPred	Vice Presidential Prediction	1 if a prediction about who will be a candidate's Vice President; 0 otherwise.
PropoQ	Proposition/ballot initiative prediction	1 if a prediction about a state or local proposition or ballot initiative; 0 otherwise.

SenElect	Senate Election Prediction	For an individual Senate race. 1 if a prediction about a single race's outcome, 0 otherwise.
HouseElect	House Election Prediction	For an individual House race. 1 if a prediction about a single race's outcome, 0 otherwise.
SenComp	Senate composition prediction	A prediction about the party composition of the Senate. 1 if about that, 0 otherwise.
HouseComp	House composition prediction	A prediction about the party composition of the House. 1 if about that, 0 otherwise.
PredEcon	Economic prediction	1 if a prediction about the economy (incl. stock market, jobs, laws, etc.) 0 otherwise.
PredEnviro	Environmental prediction	1 if a prediction about the environment or environmental law, 0 otherwise.
PredFP	Foreign policy prediction	1 if a prediction about foreign policy, 0 otherwise. Note: not for dealing with national security/war issues.
PredHC	Healthcare prediction	1 if a prediction about healthcare laws/issues, including Medicare and Medicaid. 0 otherwise.
PredImm	Immigration prediction	1 if a prediction about the enforcement, creation, or application of laws relating to immigration and immigrants, 0 otherwise.
PredNatSec	National security prediction	1 when discussing "War on Terror (Afghanistan/Iraq)," terrorism, national security etc. 0 otherwise.
PredSoc	Social issues prediction	1 when prediction is made about gay rights, abortion, or other ostensibly "moral" issues; 0 otherwise.
PredOther	Other predictions	For an issue not captured by any of the other variables. 1 if all other prediction variables are zero; 0 otherwise.

Prognosticator Variables

Variable Code	**Explanation**
Age	As of January 1, 2008
Gender	1 if male, 0 otherwise
Black	1 if Black or African-America, 0 otherwise.
Politician	1 if a current or former elected or appointed officeholder; 0 otherwise. Does not include executive branch support staff (Chief of Staff, Press Secretary, etc.) Reserved for Cabinet members, Congresspeople, Presidents, etc.
Current Adviser	1 if a current adviser to a President or member of Congress; 0 otherwise. Includes Chief of Staff, political workers (Karl Rove, etc.) press secretaries, etc.
Former Adviser	1 if a former adviser to a President or member of Congress; 0 otherwise.
Party	1 if a Democrat, 0 if a Republican (or other)
Journalist	1 if currently employed by a news organization, 0 otherwise
Advanced degree	1 if awarded a degree over a B.A. or B.S. (exluding an L.L.M. or J.D. and honorary degrees.) 0 otherwise.
Law degree	1 if awarded a J.D. or L.L.M., 0 otherwise.
Journalistic experience	Listed in years, round up.

Prediction Probability & the Dictionary

Analysts read transcripts and columns looking for predictive language, as outlined in our coding dictionary (see Appendix A). We designed this dictionary to rate the predictive language on a scale from one (prediction: no chance of occurring) to five (prediction: will absolutely occur,) with two through four in-between. Three served as our "hedge" indicator when people were noncommittal to an event occurring. For each number, one through five inclusive, we started with a list of words that we considered typical for that level. In our coding template, this variable was called the "Prediction Probability," or PredProb for short.

Additional words were added to the Dictionary as we encountered them during coding, and each analyst assigned the new words a number on the predictive language scale. We then placed each new word along the scale based on its average number assignment, and coded all future predictions using such language according to the updated dictionary. The overwhelming majority of words were added by the time the first two prognosticators were coded. On the inter-coder swaps, we checked for these new words, ensuring their inclusion into the sample.

In addition to accumulating new words, we also ended up deleting original words from the Dictionary as well. Specifically, this occurred in the case of the word "would," since this word rarely ended up being used in predictions. It was used to explain what a prognosticator, him/herself, or a political figure might do given a certain set of circumstances. In addition, the word "would" was often used when discussing conditionals that had already been proven false (for example, stating "If Hilary had won the primaries, Obama would have been her pick for Vice President," after the Democratic primaries). As a result, the word "would" was omitted from the Dictionary. No testable predictions were lost as a result.

Finally, we decided to omit predictions for which the PredTrue variable value was 0, or in other words, the conditional part of the prediction statement did not end up happening. When the conditional portion of the prediction (the "if" portion) is not accurate, then the rest of the statement (the "then" part) is by default inaccurate as well.

Prediction Accuracy

We also tested each prediction's accuracy (which we coded as the PredAcc variable) on a one-to-five scale. Based on the language of the prediction, an analyst asked: "Did (predicted event) occur?" If the event did occur, the PredAcc variable was scored a "five." If the event did not occur, the PredAcc variable was scored a "one."

Although PredAcc is a binary variable, we used the values 1 and 5 rather than 1 and 0 so that we could easily take the absolute value of the difference between the PredProb score and the PredAcc score of a prediction to keep the same scale when evaluating an individual prediction. The absolute value of the difference between the assigned probability and the accuracy of the prediction became the PredABS variable. Based on this data, we derived the Prognosticator Value Score (PVS), which we will discuss in the next section.

We, in conjunction with our research assistants, performed an extensive fact checking process to determine the prediction accuracy (PredAcc value) for each prediction. As noted, we performed fact checking both in groups and as individuals (both coders themselves, and our research assistants). PredAcc values were assigned to each prediction, and the coders and research assistants provided citations for each PredAcc value, as well.

Data Collected

We collected data on both predictions and prognosticators and evaluated them with the variables listed in our "evaluated variables" section.

In order to measure our partisanship variable, we surveyed the Hamilton College Government Department faculty and students, asking them to place each prognosticator on an ideological spectrum. The spectrum ranged from 1 (most conservative) to 9 (most liberal). We then averaged the responses for each prognosticator to generate a final placement along our spectrum in which every prognosticator received a partisanship value between one and nine.

Inter-Analyst Reliability

To ensure a standardized coding process for each prognosticator and to remove bias we set up a system for ensuring inter-analyst reliability. In this process, two analysts separately coded each prognosticator. After both analysts individually coded the prognosticator, they then discussed their findings and produced a final list of predictions for each prognosticator. This process also decided whether or not any predictions were untestable, in which case they were omitted from our sample.

The Good, the Bad, and the Ugly

Measuring Prognosticator Accuracy

In analyzing a prognosticator's predictive power, we created a "scoring system" to determine his or her overall accuracy. With large quantities of data about a prognosticator's prediction and the outcome of the event in question, we devised a simple scoring system. If a prognosticator "leaned" in the correct direction (predicted that an event would happen when it did, or correctly expected it would not happen), he or she is assigned a point. If a prognosticator hedged, he or she received no points. If a prognosticator leaned incorrectly, he or she lost a point. The points were tallied and divided by the total number of predictions made by that prognosticator. We then multiplied this number by ten to get a final Prognosticator Value Score. This left us with a final score in a range between -10 and 10. -10 represented the worst possible prognosticator, while 10 would be a perfect prognosticators.

This score was then broken down into three broad categories: the "good," the "bad," and the "ugly." The "good" were those prognosticators who received scores above five. The "bad" received scores between zero and five. The "ugly" received scores less than zero.

Example of Prognosticator Value Score System

Here, we use Paul Krugman, a *New York Times* columnist, as an example to show how our PVS system works. From a random sample of his columns and television appearances, we tallied 17 testable predictions. Krugman was awarded one point if he either predicted an event would happen and that event took place, or if he predicted an event not happening and that event ultimately did not take place. When Krugman made a prediction and the opposite outcome occurred, we subtracted one point from his total. Hedged predictions did not receive any points.

Of the 17 predictions, there were 15 instances where Krugman made a correct prediction and only one instance where he made an incorrect prediction. With the single hedged prediction included, Krugman's final tally was +14. We took the final tally (14) and divided that by the total number of testable predictions made (17), which we then multiplied by ten. Krugman's final score was 8.235.

(15-1)/17 = .8235
.8235 * 10 = 8.235

This placed Krugman in the "good" category.

Possible Criticisms of the Prognosticator Value Scoring System

We recognize there are unavoidable imperfections with the rating system we have devised and described above. One possible objection to such a system would be that it seems to fail to account for any predictions with outcomes that are not binary (for example, the prediction "Hillary Clinton will win the Democratic primary election" has an binary outcome – Hillary Clinton either did or did not win the primary and we would award this prediction with negative one points because the prediction did not come true). By non-binary predictions, we mean predictions that do not have only two possible outcomes, but rather a multitude of possible outcomes. For example, a binary prediction might be whether or not Obama will win the presidential election, whereas a non-binary prediction might be an estimated number of seats picked up by the GOP in a House election. However, there were relatively few non-binary predictions compared to the overwhelming majority of binary predictions, probably due to the overly complex and difficult nature of making such non-binary predictions.

Another possible criticism is that this method for assigning Prognosticator Value Scores does not account for coding system scores of two or four. While these ratings were used in later calculations, we feel that it was accurate enough to measure whether or not a prognosticator leans in a positive or negative direction. Also, by using a scale of negative one to positive one, we could more easily assign zero points to prognosticators who consistently hedge on predictions, thereby avoiding making definitive and truly useful predictions, while more thoroughly punishing prognosticators who spout incorrect predictions.

The Good, The Bad, and The Ugly

This section includes a description of each prognosticator's predictions. An information box lists their total number of predictions, the number they had correct, wrong, and hedged, their PVS, and the p-value of a binomial formula calculation; that is, the odds that the number that particular pundit answered correctly was obtained by chance, based on a p of .5. In other words, this tests whether a prognosticator is better than a coin flip; if the p-value is less than .05, that prognosticator has a statistically significant result. If the result has an asterisk, that prognosticator is statistically worse than a coin flip.

Our p-values were calculated as the probability of correctly guessing exactly k times out of the total N predictions. Traditionally, p-values are expressed as the probability of being equal to or greater than some observed value if the null possibility (in our case, that prognosticators are as good as a coin flip) is true. Using this method, the p-value for Paul Krugman would be his probability of guessing at least 15 right (the probability of getting 15, 16 and 17 out of 17 right). Calculating p-values as "great than or equal to" raises the p-values relative to our way, which would make even fewer prognosticators statistically better than a coin flip. Therefore, we chose to use the more conservative way to calculate our p-values.

Our p-values were calculated by comparing the number of correct predictions against the total number of predictions for each prognosticator (right, wrong, and hedged).

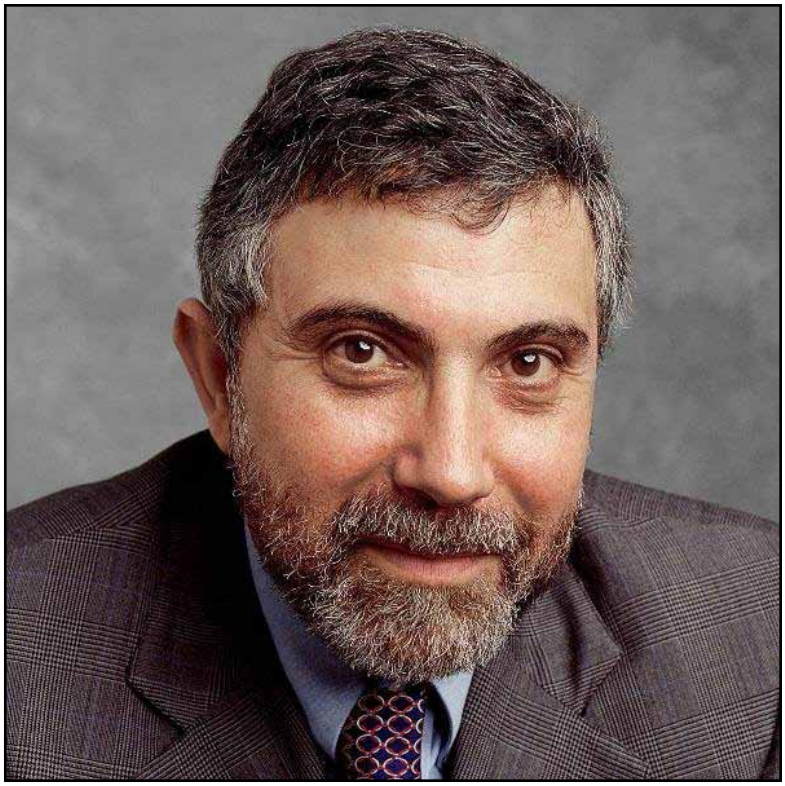
The Good

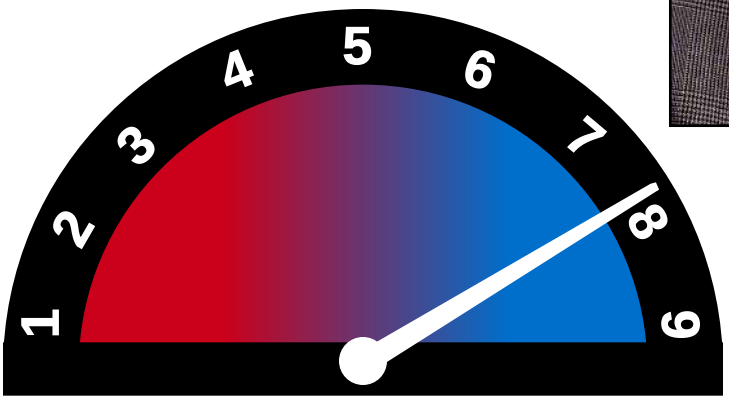
The Good comprised the best of our prognosticators, each of them with a PVS over 5. There are few trends among the good—some are male, some are female; some are politicians, others are journalists, and there is an Administration official included, too. They discuss all the "hot topics" of the 2008 election season, including the economy, the Democratic primaries, the Presidential contest, and politics generally. These individuals rarely hedge their predictions.

A number of analyses noted that predictors who made it to this list made a number of "easy" predictions, that were "obvious" and did not seem to be especially good tests of predictive capacity. A larger sample can subject these prognosticators to a more rigorous analysis to test their true predictive capacity.

Paul Krugman

Paul Krugman, an economist and *New York Times* columnist, made 17 predictions in his sampled columns from the examined period. He primarily discussed economics, predicting often about the impending economic crises. Housing and unemployment issues were frequent topics. He also talked about politics on occasion, especially as the election grew closer. Many of his predictions were fairly far into the future—a number of them discussed the crisis in a year or more from the date of the prediction. Krugman was also uncommonly accurate, only missing one prediction and hedging on one other. His powers of prognostication were impressive, but primarily confined to his field of expertise—he is, after all, a Nobel-winning economist.



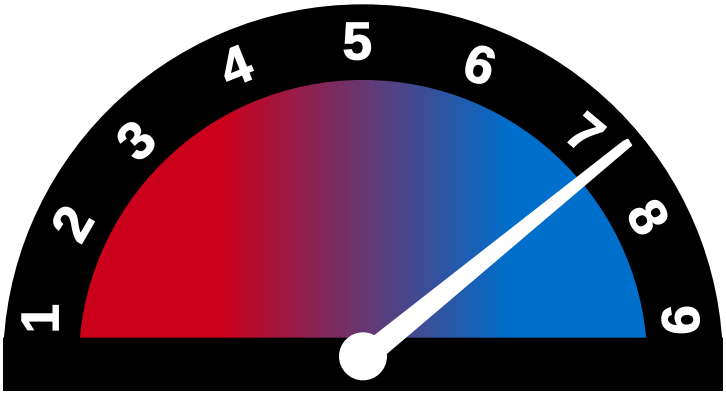


On our partisanship scale, where 1 is most conservative and 9 is most liberal, Paul Krugman scored a 7.90.

Total # of Predictions	17
Correct	15
Wrong	1
Hedged	1
Prognosticator Value Score	8.2
P-Value	0.0010

Maureen Dowd

Of the randomly selected 30 articles written by Maureen Dowd, she made 11 clear and testable predictions. Dowd would seem to be at a clear disadvantage in making predictions during the lead up to the 2008 elections because she picked Hillary Clinton to win the Democratic primaries. Obviously, this did not happen, although her PVS could remain high because she continued to focus on Clinton after the primaries, discussing things like the role she and her husband could and would play in the Obama administration if Obama were to win the general election. When discussing issues other than the binary "Will Hillary Clinton be the next President?" Dowd tended to make ambiguous predictions about the political atmosphere surrounding the 2008 elections with wording that rendered them immeasurable (for example, "Obama will never be at his best around Hillary"). She tended to focus on personalities rather than hard facts. On the rare occasions she stepped outside of this subject matter, she made clear cut but seemingly obvious predictions on subjects like foreign policy (as in, "We're going to have a huge number of troops in Iraq for a long time").

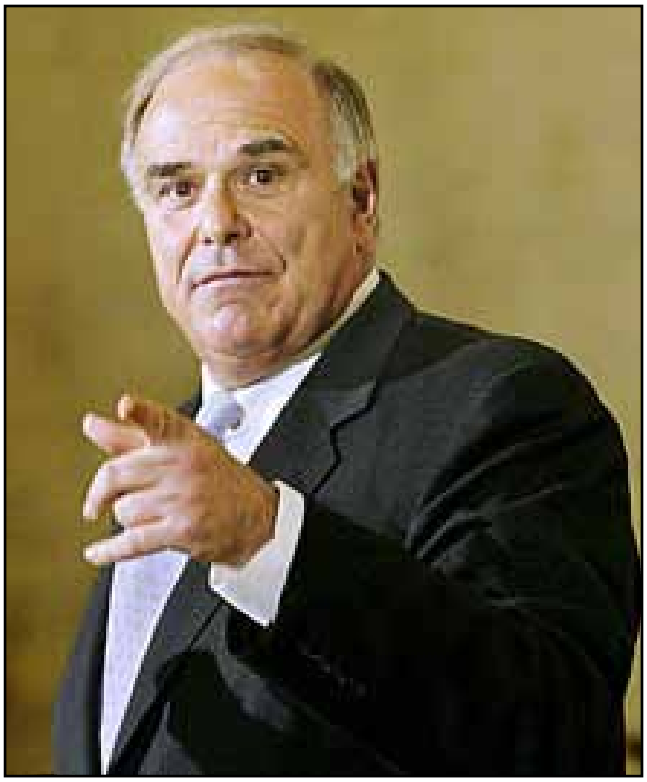


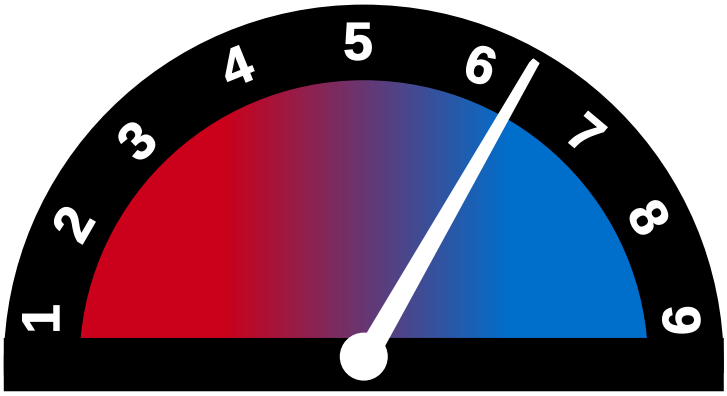
On our partisanship scale, where 1 is most conservative and 9 is most liberal, Maureen Dowd scored a 7.20.

Total # of Predictions	11
Correct	9
Wrong	1
Hedged	1
Prognosticator Value Score	7.3
P-Value	0.0269

Ed Rendell

Rendell's predictions were almost exclusively about the upcoming elections. He only hedged one prediction, and with exception of three predictions used language scored at either a 1 or 5 on the PredProb scoring scale. His predictions seemed like he was reciting partisan rhetoric, rather than making genuine personal predictions; however this can be excused since he was probably compelled to support the Democrats in the upcoming elections. He was relatively accurate, due to his partisan identification and his party's ultimate victories.



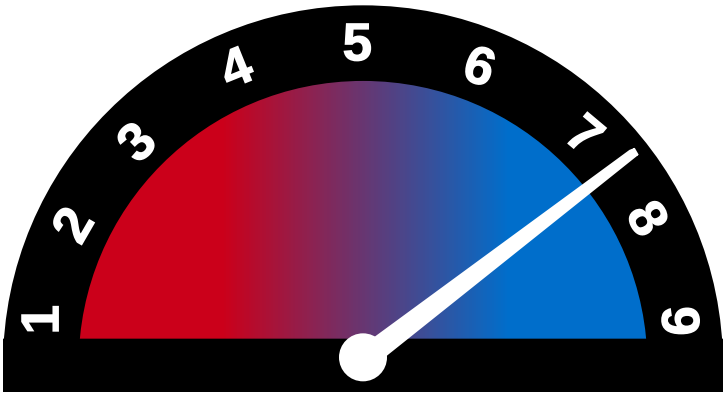


On our partisanship scale, where 1 is most conservative and 9 is most liberal, Ed Rendell scored a 6.20.

Total # of Predictions	20
Correct	15
Wrong	1
Hedged	4
Prognosticator Value Score	7.0
P-Value	0.0148

Chuck Schumer

Schumer was a supporter of Hillary Clinton in the primaries and most of his early appearances were dedicated to boosting her candidacy. The remainder of his ten predictions were primarily made in the full swing of the presidential election and supported Barack Obama. He was correct 80 percent of the time and never hedged, an impressive total for a prognosticator. A significant portion of his accuracy was aided by the fact that he was on the "right side" of the election; predicting Obama's rise and McCain's fall, common for a Democrat, put Schumer on the correct side of events. He also used carefully constructed language when talking about the Democratic primary, not always asserting that Hillary would win, but saying that a clear winner would be determined by a set date. This allowed Schumer to lower the risk level of his predictions. The use of such careful language overall boosted his score and made him seem a better pundit—and while he scored well, he is a lesson that the specific language of the prediction is important for a viewer to consider.

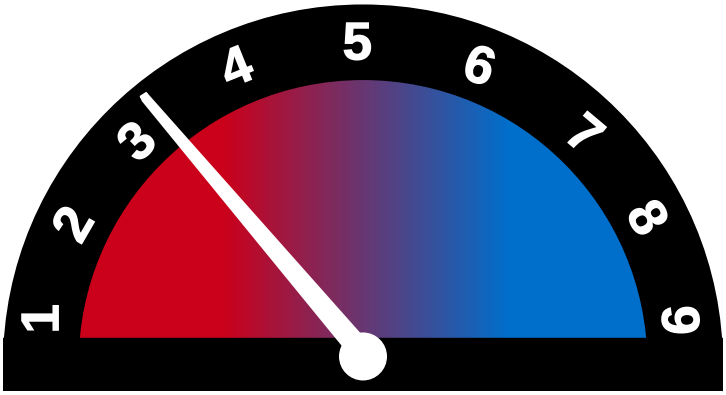


On our partisanship scale, where 1 is most conservative and 9 is most liberal, Chuck Schumer scored a 7.30.

Total # of Predictions	13
Correct	11
Wrong	2
Hedged	0
Prognosticator Value Score	6.9
P-Value	0.0095

Kathleen Parker

Of the 20 articles that were randomly selected, which covered both her writing at TownHall.com and The Washington Post, writer Kathleen Parker made only six testable predictions, barely qualifying her to be included in our regression. The majority of her five predictions focused on what was happening in Republican politics in the lead up to the 2008 election. She was often critical of those criticizing Palin, and Parker predicted that McCain and Palin would win the general election. Prior to the Republican primaries, she also discussed other Republican candidates, including Mitt Romney. Parker did make one prediction that relied on the conditional statement "If Palin were to become President," rendering that prediction untestable. Overall, her predictions were rare, but when she made them they were focused on Republican politics during the 2008 election cycle, and therefore they fell within our time frame and were concise and measurable.

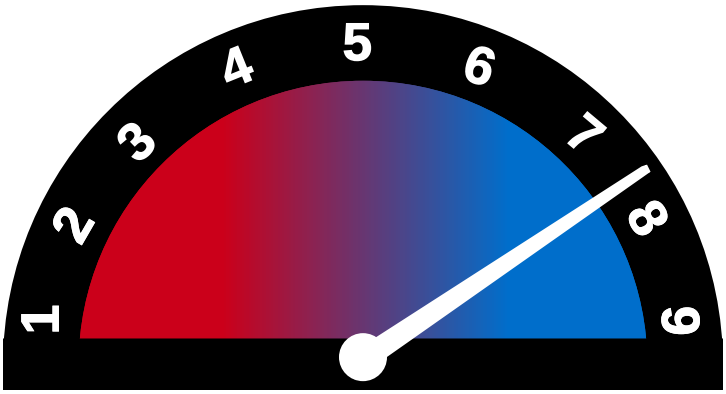


On our partisanship scale, where 1 is most conservative and 9 is most liberal, Kathleen Parker scored a 3.30.

Total # of Predictions	6
Correct	4
Wrong	0
Hedged	2
Prognosticator Value Score	6.7
P-Value	0.2343

Nancy Pelosi

During the time of our sample, Nancy Pelosi was one of the most important Democrats in Washington DC. She was a general spokeswoman both for the Democratic agenda and for Democratic candidates running in the 2008 election. She obviously had firm talking points which she wanted to address on each Sunday morning talk show (she often repeated phrases and assertions verbatim on the different shows), but this comes with the leadership position she held at the time. Her predictions scanned many different policy and political arenas, yet she could dodge a question with ease, and she avoided finding herself making uncomfortable predictions. For example, she stayed away from the Obama/Clinton controversy and predicting a winner before the convention. Her predictions were both smart and, perhaps, a little bit "lucky;" as a Democrat, she would not pick against her Democratic allies in the election. To her benefit, he 2008 election saw sweeping gains for the Democratic Party. That aside, she generally stayed with what she was familiar with: politics. Her final score placed her comfortably in the "Good" category.

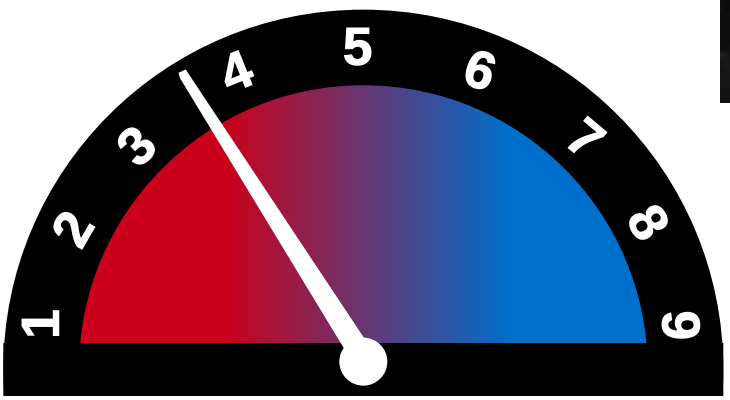


On our partisanship scale, where 1 is most conservative and 9 is most liberal, Nancy Pelosi scored a 7.70.

Total # of Predictions	27
Correct	20
Wrong	3
Hedged	4
Prognosticator Value Score	6.3
P-Value	0.0066

David Brooks

David Brooks is a prominent pundit for the *New York Times*, and a prolific prognosticator. He made 29 predictions, many about politics or how issues like the economic crisis related to the 2008 election. His most productive set of predictions occurred in one article where he discussed the future of financial regulation; he made 7 predictions, all of them correct. Some of these were, admittedly, low-hanging fruit; saying that "oversight [of the economy] will be put in place," for instance, is not an especially bold or farsighted prediction. But others showed vision, including the fact that the country will "turn to heads of the investment banks [after the economic crisis]," something that happened when the Obama administration came to power and appointed numerous former investment bank officials. While Brooks made a number of easy predictions, he also was willing to take on challenging issues.

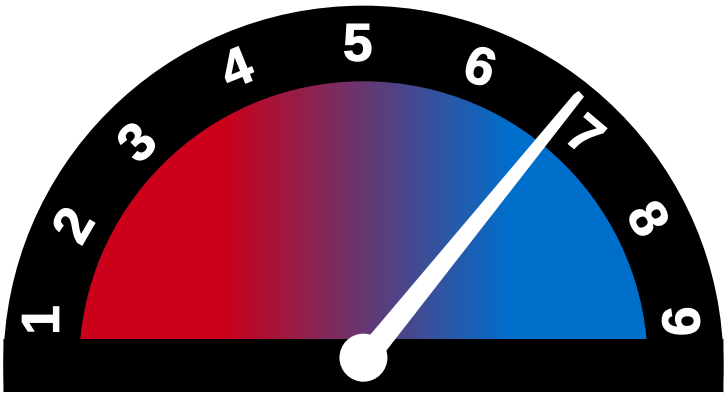


On our partisanship scale, where 1 is most conservative and 9 is most liberal, David Brooks scored a 3.80.

Total # of Predictions	19
Correct	13
Wrong	2
Hedged	4
Prognosticator Value Score	5.8
P-Value	0.0518

Eugene Robinson

Eugene Robinson cannot be classified strictly as a political columnist, even though he writes for the Washington Post, a newspaper with unrivaled political coverage and analysis. Rather, he seems to be a cultural commentator who will write about almost any topic or issue that he deems relevant or interesting. Often, this touches on politics. His writing style is fluid yet accessible, but the random sample of columns we examined did not often lend themselves to straight political predictions. He wrote about Oprah, the Beijing Olympics and the embargo on Cuba, just to give a sample of his breadth of work. The one issue Robinson was most interested in during the 2008 election cycle was race, and he often wrote about the election through this lens. His style and subject matter did not churn out many testable predictions, but when he did, he made smart and simple predictions, which weren't earth-shattering, but were correct the majority of the time.

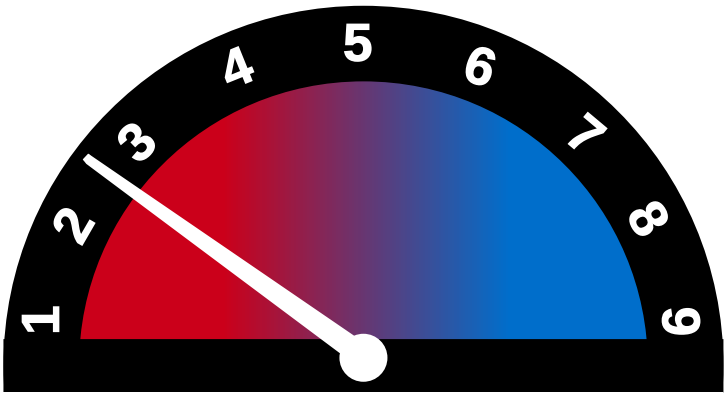


On our partisanship scale, where 1 is most conservative and 9 is most liberal, Eugene Robinson scored a 6.90.

Total # of Predictions	11
Correct	7
Wrong	1
Hedged	3
Prognosticator Value Score	5.5
P-Value	0.1611

Hank Paulson

Secretary of the Treasury Hank Paulson appeared on our Sunday morning talk shows six times during our designated time frame. In those six appearances, he made 22 testable predictions. As can be predicted by his title, the overwhelming majority of his predictions were related to the economy and the ongoing recession. His predictions often focused on the duration of the recession that he was anticipating. Some of his predictions were optimistic (for example, "We will work through this"), and some of his predictions earnestly recognized the obstacles facing the American economy (for instance, "We're going to be in a period of slow growth for awhile"). Aside from general feelings towards the future of the economy, Paulson predicted how policymakers would respond to the policies proposed to combat the recession; again, he was often optimistic. Many of his predictions were identical in sentiment if not wording. Regardless of their repetitiveness, almost all of Paulson's predictions were measureable within our designated time frame.



On our partisanship scale, where 1 is most conservative and 9 is most liberal, Hank Paulson scored a 2.56.

Total # of Predictions	22
Correct	16
Wrong	4
Hedged	2
Prognosticator Value Score	5.5
P-Value	0.0178

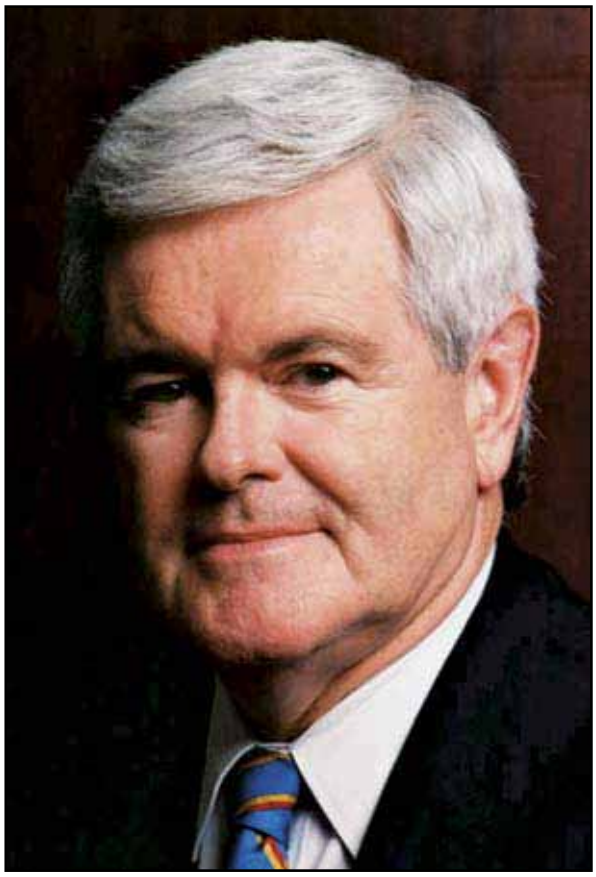
The Bad

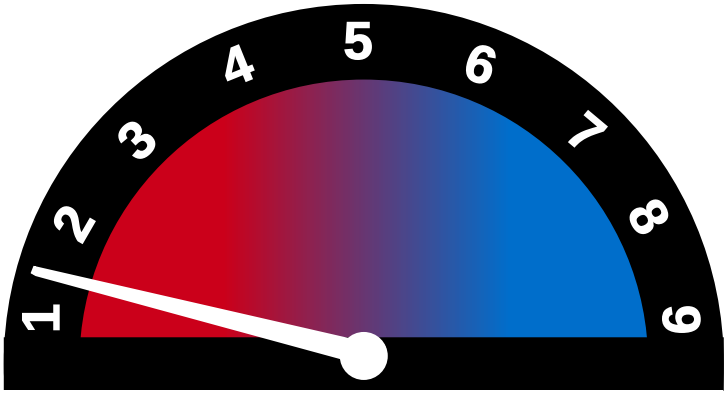
"The Bad" is comprised of those pundits whose Prognosticator Value Score is at or above a 0 but below a 5. In other words, they are more reliable than a coin toss, but less reliable than those who are considered "The Good."

"The Bad" category has a substantial number of politicians, all of whom were heavily invested in the 2008 presidential election. On one hand, there are Hillary Clinton and her communications director Howard Wolfson. On the other hand, we have McCain supporters Joe Lieberman and Newt Gingrich. "The Bad" group also presents us with a group of varied columnists. Writers like Thomas Friedman and Nicholas Kristof tend to tackle broader issues that just the political outcomes of the 2008 election cycle. These issues often included foreign policy. However, there are also writers in this group that predict mainly on politics, such as Clarence Page and David Broder.

Newt Gingrich

Although Gingrich only appeared a handful of times on *This Week* and *Face the Nation* during the given time period, he showed little fear in making confident predictions. He made 23 predictions, the vast majority of which used language corresponding to a 1 or 5 on our PredProb scoring scale. His confidence mainly extended to the upcoming elections; all but a few of his predictions dealt directly with primary and the presidential elections. This trend makes sense, since these shows likely invited him to speak on these topics, and then asked him questions accordingly. Columnists have more freedom to speak about that which interests them, whether it be upcoming elections, the economy, foreign relations, or other topics. It should not come as a surprise that Newt is rather opinionated and confident in his conjectures – this trend aligns with his past role as Speaker of the House.



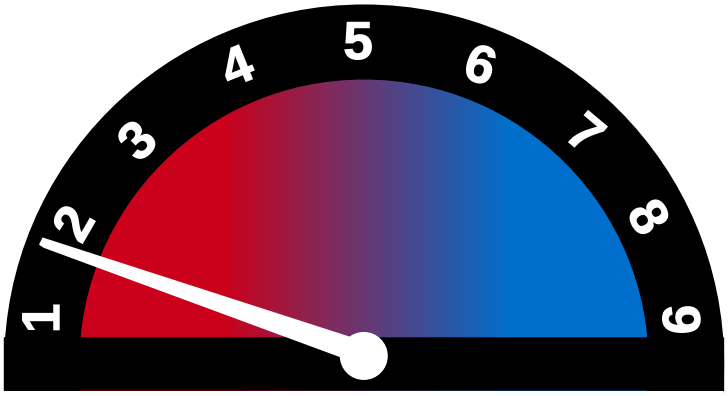


On our partisanship scale, where 1 is most conservative and 9 is most liberal, Newt Gingrich scored a 1.50.

Total # of Predictions	23
Correct	13
Wrong	5
Hedged	5
Prognosticator Value Score	3.5
P-Value	0.1364

Mike Huckabee

Huckabee talked prolifically, but made very few testable predictions. He was almost always speaking about himself, not surprisingly as someone would if they were a Presidential candidate. His few predictions discussed a wide variety of topics—terrorism, the upcoming Presidential election, his prospects in the Michigan and Iowa contests, and the potential for an Obama-Clinton ticket. He only made one incorrect prediction, about the location of the next terrorist attack. In such a small sample, however, it is difficult to see Huckabee's true prognosticating power—or even discern his areas of predictive expertise.

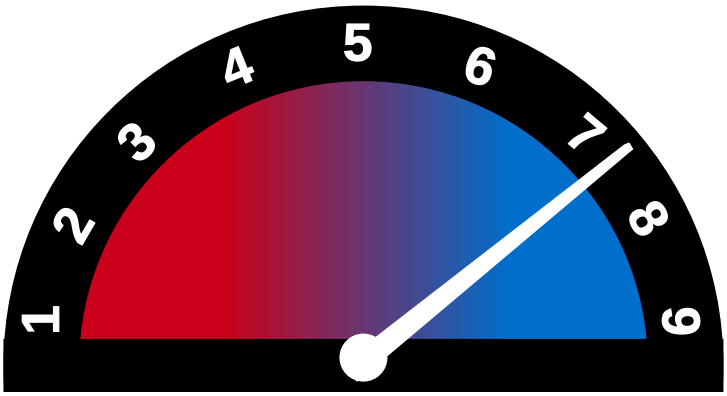


On our partisanship scale, where 1 is most conservative and 9 is most liberal, Mike Huckabee scored a 1.90.

Total # of Predictions	10
Correct	6
Wrong	3
Hedged	1
Prognosticator Value Score	3.0
P-Value	0.2051

Howard Wolfson

Howard Wolfson, as the Communications Director for Hillary Clinton's 2008 presidential campaign, was at a clear disadvantage in making predictions regarding the 2008 elections. Until the bitter end, Wolfson continued to predict that Hillary would beat Barack Obama and win the primaries. This obviously did not happen, rendering many of his predictions inaccurate. Of the six appearances he made on Sunday morning talk shows, we recorded 18 testable predictions, the majority of which were statements that Hillary Clinton would win the primary and go on to win the presidential election. After Clinton had clearly lost the primaries, Wolfson's predictions became discussions of how Clinton and her supporters would dedicate their support to Barack Obama. Although the majority of his predictions were measurable, the ones that weren't typically fell within this group of predictions that he made after Clinton lost the primaries.

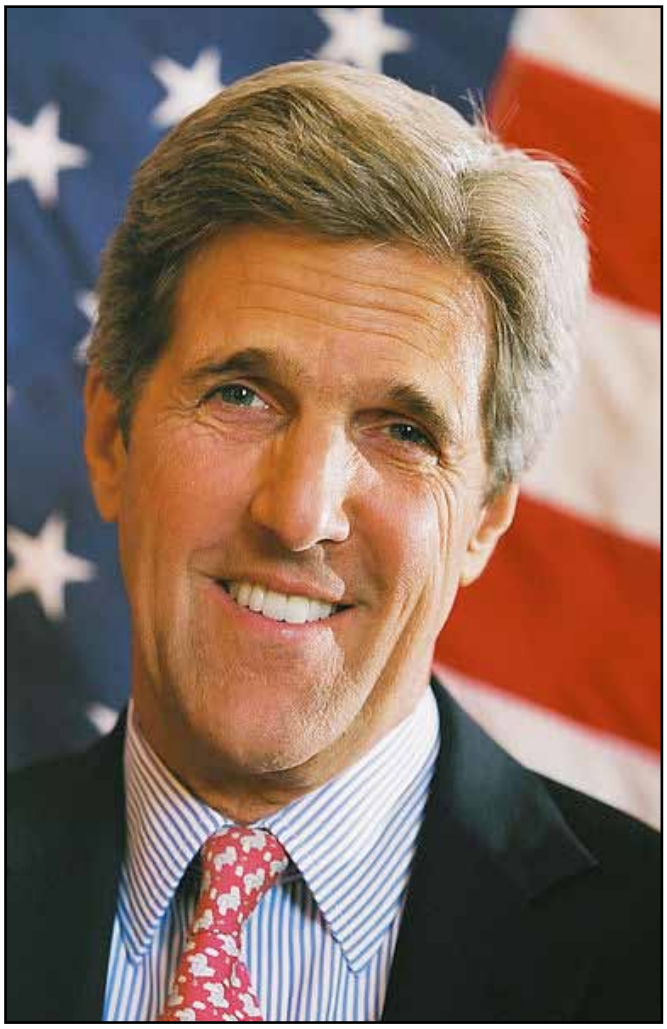


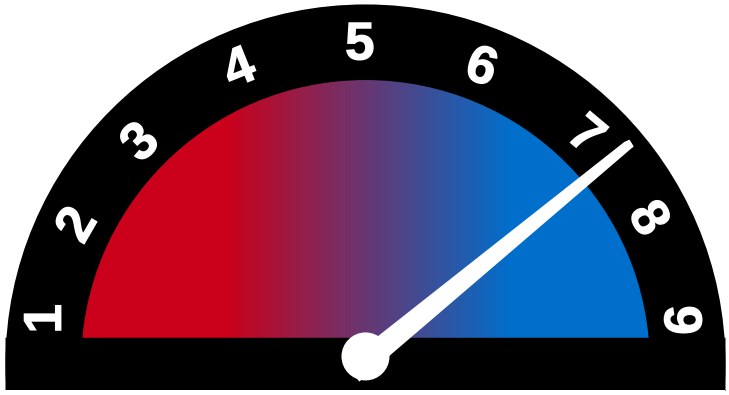
On our partisanship scale, where 1 is most conservative and 9 is most liberal, Howard Wolfson scored a 7.30.

Total # of Predictions	18
Correct	11
Wrong	6
Hedged	1
Prognosticator Value Score	2.8
P-Value	0.1214

John Kerry

While Kerry is extremely involved in foreign relations, he made primarily normative judgments, not predictions in the foreign policy sphere. He made predictions about domestic policy and politics regularly, prognosticating on the Florida/Michigan primary controversy, offshore drilling, and Obama's risk of getting "Swift Boated." He was generally accurate, getting more predictions right than he got wrong. He rarely hedged, and made a fairly significant number of predictions that were untestable because their conditional did not come true. He was a mediocre pundit, falling just barely in the upper half of the group.



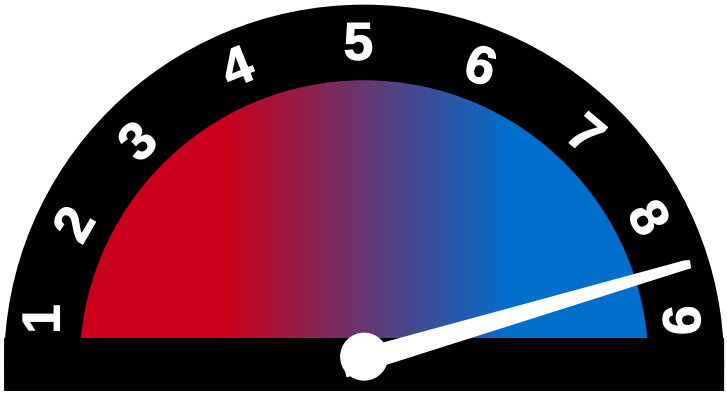


On our partisanship scale, where 1 is most conservative and 9 is most liberal, John Kerry scored a 7.20.

Total # of Predictions	12
Correct	7
Wrong	4
Hedged	1
Prognosticator Value Score	2.5
P-Value	0.1934

Bob Herbert

Bob Herbert of the *New York Times* wrote frequently on the Iraq war and the economic collapse. Another major topic of Herbert's writing is racism, which may be an underlying theme to his coverage of the Democratic presidential primaries (in which he favored Obama but feared a Hillary comeback until late in the primary season). Hedging is somewhat of an issue with Herbert, who hedged 3 out of 9 predictions. Herbert fell in the "bad" category, scoring a 2.2 on our scale. A final observation on Herbert is his tendency to discuss multiple possible scenarios in the future, such as predicting the policies of both a President Clinton and a President Obama during the primary season. As such, several predictions were unscored due to unmet conditionals.

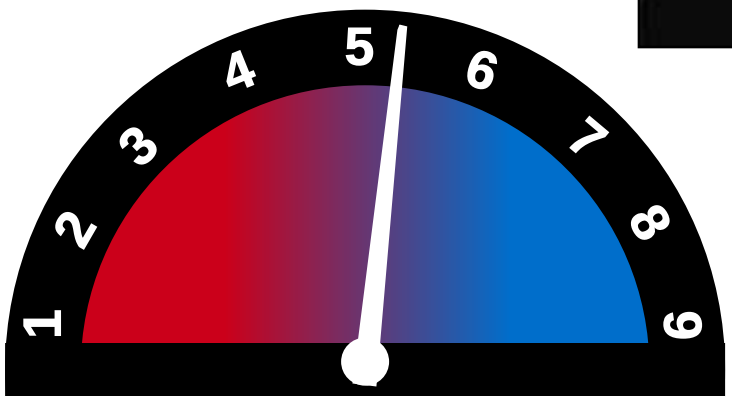


On our partisanship scale, where 1 is most conservative and 9 is most liberal, Bob Hebert scored a 8.50.

Total # of Predictions	9
Correct	4
Wrong	2
Hedged	3
Prognosticator Value Score	2.2
P-Value	0.2461

Andrea Mitchell

Andrea Mitchell is the epitome of a Sunday morning roundtable political pundit. Every single coded prediction of hers in this sample related to politics in some way. Compared with some of her journalist peers who examined either the left or the right, her focus and analysis splits evenly between the Democratic and Republican players for the presidency. She came across as remarkably neutral in other areas as well – other female or minority prognosticators tended to focus on Hillary Clinton and Barack Obama, respectively. Mitchell gave every candidate and issue a fair examination, which can be attributed as much to the roundtable discussion as much as her expertise in these various arenas. Her otherwise poor score may be attributed to a lack of in-depth focus on a particular aspect or candidate, yet many other Sunday morning regulars also did not seem to score all that well either.

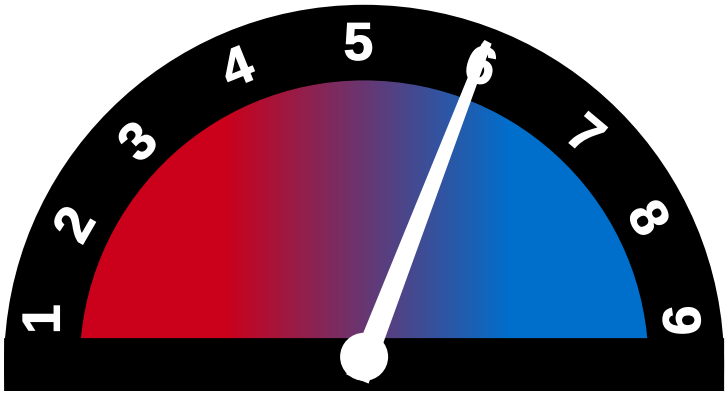


On our partisanship scale, where 1 is most conservative and 9 is most liberal, Andrea Mitchell scored a 5.20.

Total # of Predictions	18
Correct	9
Wrong	5
Hedged	4
Prognosticator Value Score	2.2
P-Value	0.1855

Thomas Friedman

Thomas Friedman's columns span a wide variety of subjects. He often writes about the environment, foreign relations, and domestic policy. He predicts some election outcomes, but his subject varies far more than that of any politician prognosticators. Friedman makes more complex predictions that require a larger breadth of knowledge of subject matter than just a simple election prediction. His predictions are genuine and tied to his research and experience, not just partisan rhetoric. His PredProb scores fluctuate a little more – he doesn't hedge often, but he uses language scored as a 2 or 4 more often than politicians appear to use it. This may be because his predictions are based on his own contemplation, not just regurgitated from a party sound-byte.

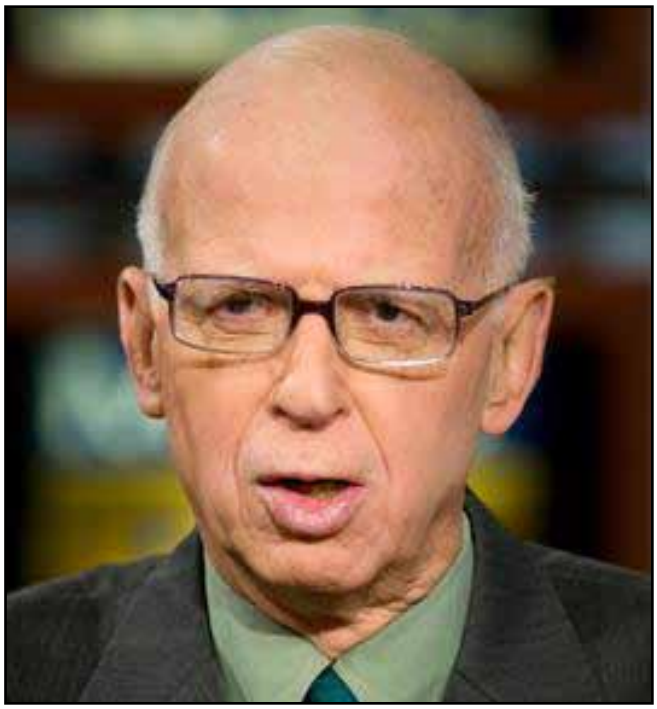


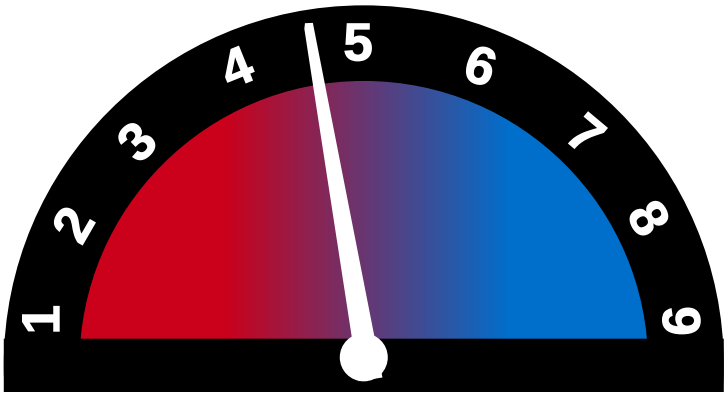
On our partisanship scale, where 1 is most conservative and 9 is most liberal, Thomas Friedman scored a 6.00.

Total # of Predictions	10
Correct	5
Wrong	3
Hedged	2
Prognosticator Value Score	2.0
P-Value	0.2461

David Broder

David Broder of the Washington Post prognosticated almost exclusively on primary politics and the policies of the presidential nominees. Making 11 predictions in total, six were about primary decisions, including running mate selections and party success, and the other five were about policy decisions of the upcoming president. David Broder correctly predicted a President Obama and some of the President's policy decisions; however most of those predictions came late in 2008, just weeks before the election. David Broder did not have much of a problem with hedging, but perhaps his overconfidence hurt him. His final score was a 1.8, resulting from just over a 1:1 right-to-wrong prediction rate coupled with the penalty of his few hedges. Broder was most often wrong when predicting how Obama would handle the financial crisis.



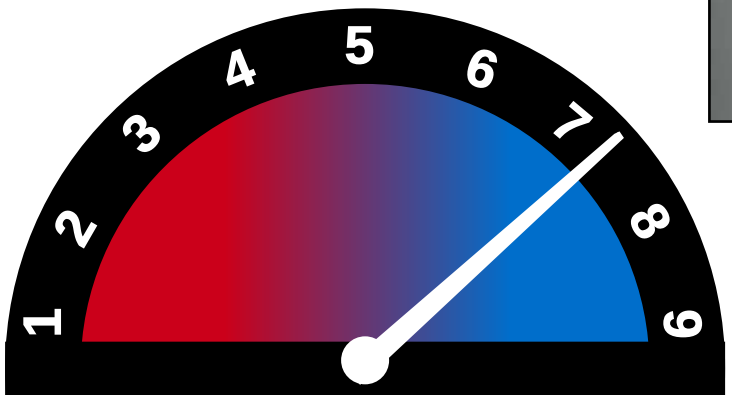


On our partisanship scale, where 1 is most conservative and 9 is most liberal, David Broder scored a 4.70.

Total # of Predictions	11
Correct	5
Wrong	3
Hedged	3
Prognosticator Value Score	1.8
P-Value	0.2256

Clarence Page

Clarence Page is a regular contributor to *The Chicago Tribune*. The most obvious element of Page's prognostication is rarity: Page recorded only six predictions in his sample; barely enough to even make it into the data. His "bad list" score of 1.7 is a combination of the weight of one wrong prediction in a small sample size and the small penalty for hedging. Although hedging is punished less than a wrong answer, Page hedged three of his six predictions, enough to bring down his score significantly. Page predicted in a short time-frame, making all his predictions within a year of the event. Page focused on politics and policy as the subjects of his predictions.

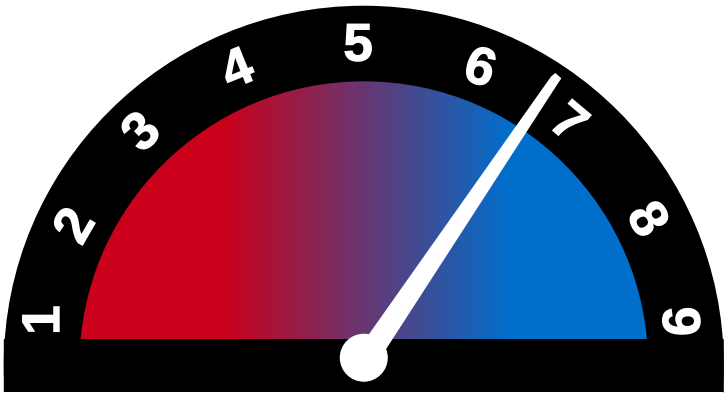


On our partisanship scale, where 1 is most conservative and 9 is most liberal, Clarence Page scored a 7.30.

Total # of Predictions	6
Correct	2
Wrong	1
Hedged	3
Prognosticator Value Score	1.7
P-Value	0.2344

Nicholas Kristof

Nicholas Kristof, similarly to Friedman, writes about a wide breadth of material. His subject matter spans from U.S. domestic policy to international policy, but he also dabbles in political campaigns. He made relatively few predictions overall, and he was less likely to use confident language. He leaned in either direction more often than he committed to either side, but he also avoided making predictions in a lot of cases altogether. He was far more likely to speak hypothetically, and then give hypothetical solutions, rather than committing to a particular outcome. His columns seemed more entertaining and probing, rather than focusing on short term issues and their possible outcomes.

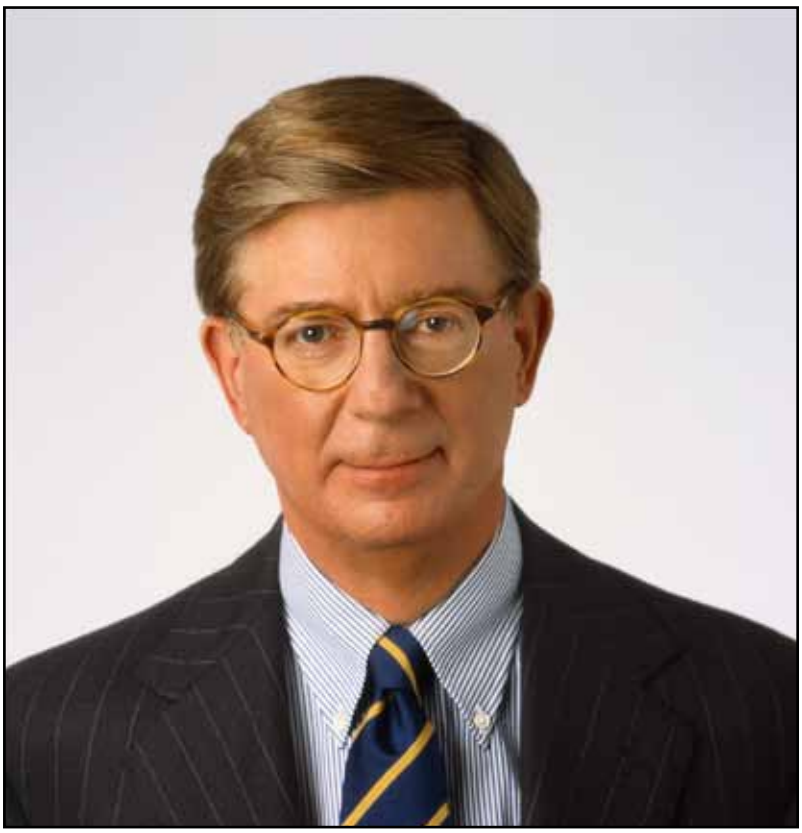


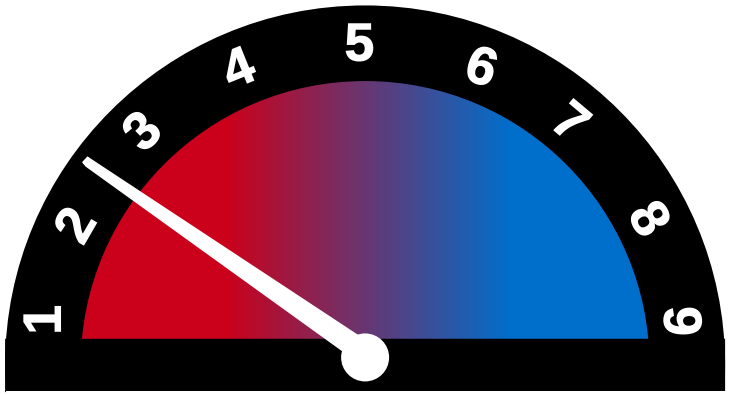
On our partisanship scale, where 1 is most conservative and 9 is most liberal, Nicholas Kristof scored a 6.80.

Total # of Predictions	12
Correct	6
Wrong	4
Hedged	2
Prognosticator Value Score	1.7
P-Value	0.2256

George Will

George Will is the most prolific of journalists in our sample – appearing weekly on ABC's *This Week* roundtable since 1981 as well as writing columns for *The Washington Post*. Will's predictions stayed mainly in the political sphere, and on the whole, he scored slightly above zero. A few of his predictions that we pulled from a random sample were spot on – his prediction about the Democratic primary in Iowa and Hillary's weakness there was certainly true. But this is one anecdote among a larger sample of predictions which were not often correct. He did talk about a few other things – most his love for baseball, and he accurately predicted that Major League Baseball would adopt instant replay under a small set of circumstances (such as reviewing home run calls). His language is fluid and sophisticated, and he is the most widely read journalist in the country. But when it comes to accurate predictions, there are better much journalists in our sample.



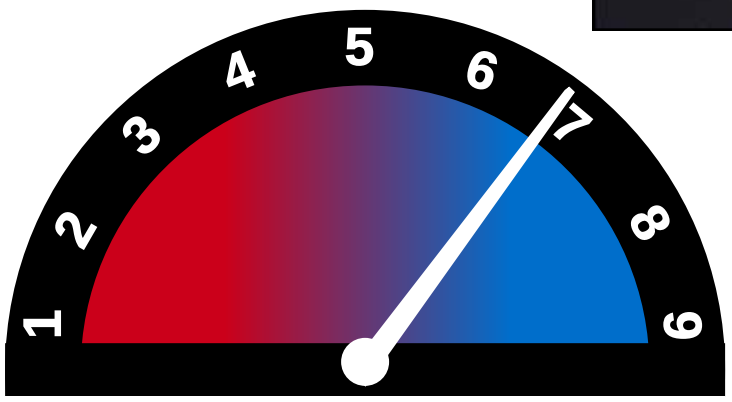


On our partisanship scale, where 1 is most conservative and 9 is most liberal, George Will scored a 2.50.

Total # of Predictions	48
Correct	18
Wrong	14
Hedged	16
Prognosticator Value Score	0.8
P-Value	0.0260*

Hillary Clinton

Hillary Clinton's prognostication serves to illustrate a key finding of Philip Tetlock's study: people tend to over-predict change and under-predict the status quo. Of her six predictions, all were rated as 5s, showing both total confidence and the expectation that events will occur. Hillary was half right, correctly prognosticating 3 out of 6 events. She focused her predictions on the war in Iraq and fallout from the economic struggle with equal accuracy. Hillary over-predicted the negative effects the recession would have on US families, and over-predicted a death toll rate increase in Iraq. These both probably result from a political agenda to emphasize the negative effects of a Republican administration. Clinton's final score of 0 places her at the bottom of the "bad" category, the result of getting exactly half right and half wrong.



On our partisanship scale, where 1 is most conservative and 9 is most liberal, Hillary Clinton scored a 6.90.

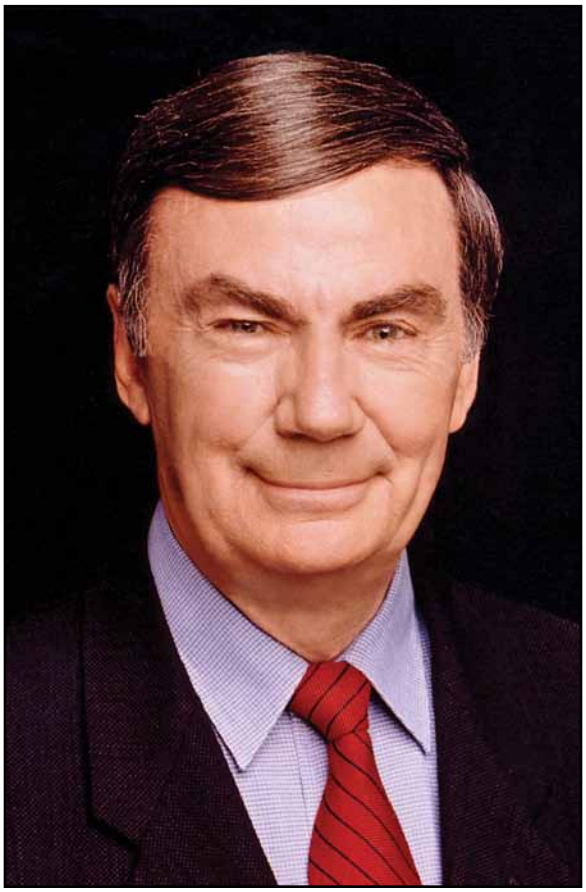
Total # of Predictions	6
Correct	3
Wrong	3
Hedged	0
Prognosticator Value Score	0
P-Value	0.3125

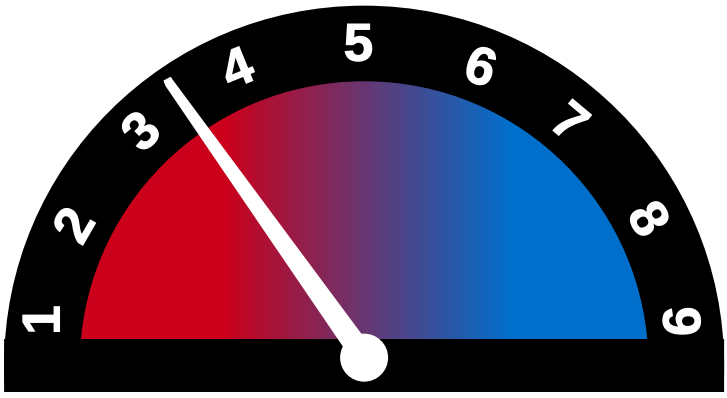
The Ugly

Five prognosticators had negative PVS, meaning each person in the following group was wrong more often than they were right. This group is not large enough to make any sort of overall claim that can explain their poor scores, but none of these prognosticators seem to have much in common anyway. Three senators, one Democrat and two Republican, found themselves in the "Ugly" category, along with a newspaper columnist and a Sunday morning network television pundit.

Sam Donaldson

Sam Donaldson works for ABC, and it is his job to appear on the roundtable for *This Week* from time to time and weigh in with political observations and predictions. Unlike some of the high profile politicians who also regularly appear on these shows, Donaldson is free to say what he wants since his "constituents" are paying him to do just that and do not have repeated chances to vote him out of office. Like anyone in his position, Donaldson takes his position seriously and puts his neck on the line by making lots of political predictions. He was bold in making predictions well before many others would have, like stating who he thought would win the Democratic nomination back in September of 2007, four months before the Iowa caucuses. However, Donaldson was often wrong (he thought Hillary Clinton would win the nomination) early on, and then he did not have new information to correct himself for a long time. Donaldson was bold and unflinching, but he was wrong much more often than he was correct.



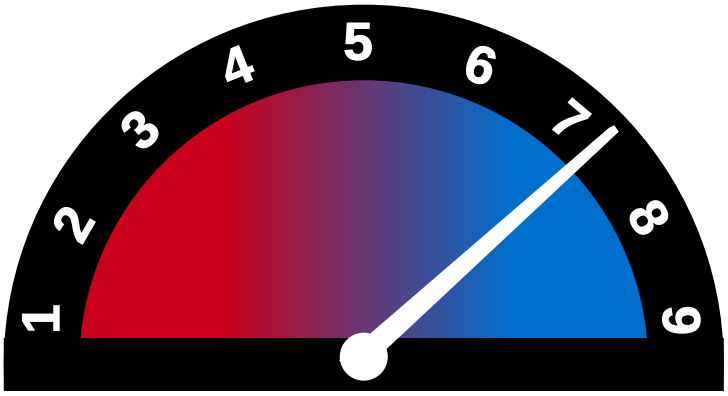


On our partisanship scale, where 1 is most conservative and 9 is most liberal, Sam Donaldson scored a 3.40.

Total # of Predictions	25
Correct	7
Wrong	9
Hedged	9
Prognosticator Value Score	-0.8
P-Value	0.0143*

Carl Levin

Senator Carl Levin focused a significant amount of his attention on the issues surrounding the Michigan and Florida primaries in the 2008 Presidential election. This makes sense, seeing as he is a Senator from Michigan, and his state's delegates were the ones in question. As the Chairman of the Senate Committee on Armed Services, he also discussed foreign policy, specifically withdrawal from Iraq, and how he anticipated such policies would be received in the Senate. Levin appeared on television less frequently than other prognosticators, and it is likely that this is one of the reasons we found only nine testable predictions. Although there was a smaller sample size, the majority of the predictions he made were clear, concise, and testable.

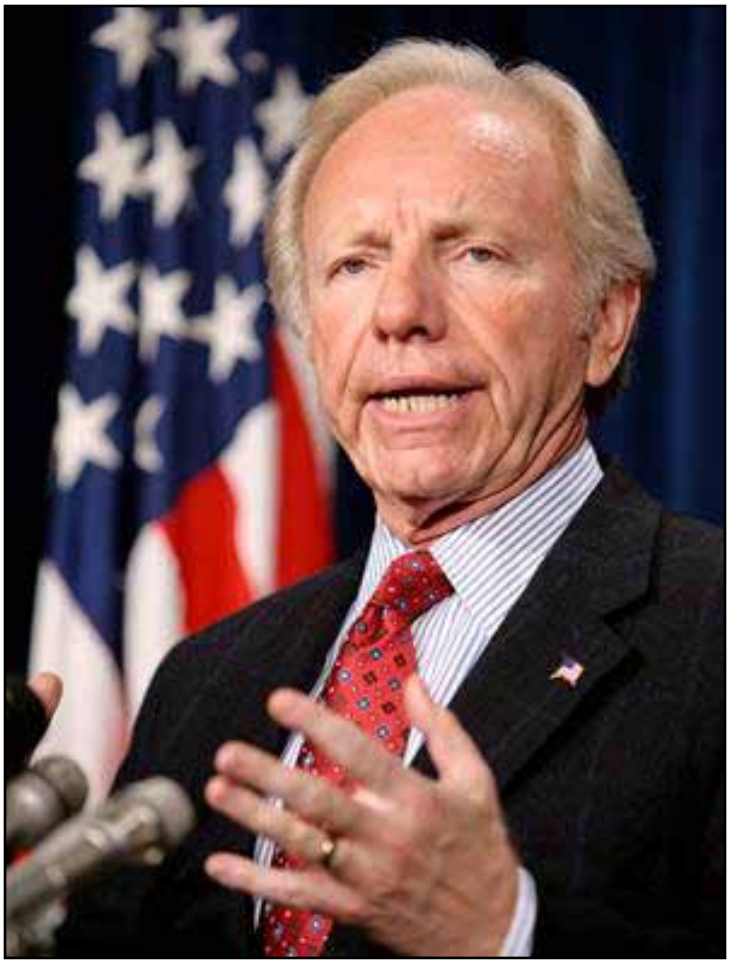


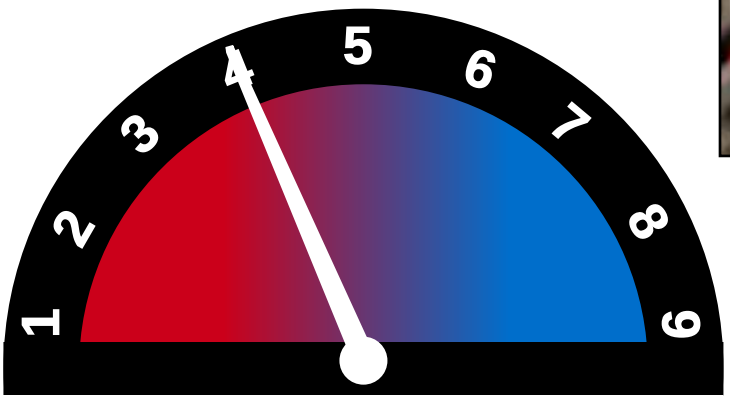
On our partisanship scale, where 1 is most conservative and 9 is most liberal, Carl Levin scored a 7.20.

Total # of Predictions	9
Correct	3
Wrong	4
Hedged	2
Prognosticator Value Score	-1.1
P-Value	0.1641

Joe Lieberman

Joe Lieberman is a good friend of John McCain, so, of course, his predictions regarding the 2008 Presidential election are somewhat biased. He makes many predictions about domestic politics – and almost exclusively about John McCain and his run for presidency. However, Lieberman also makes a few predictions about foreign policy. He hedges three predictions, but the vast majority of his predictions are said with confidence, scoring a 1 or 5 on the PredProb score scale.



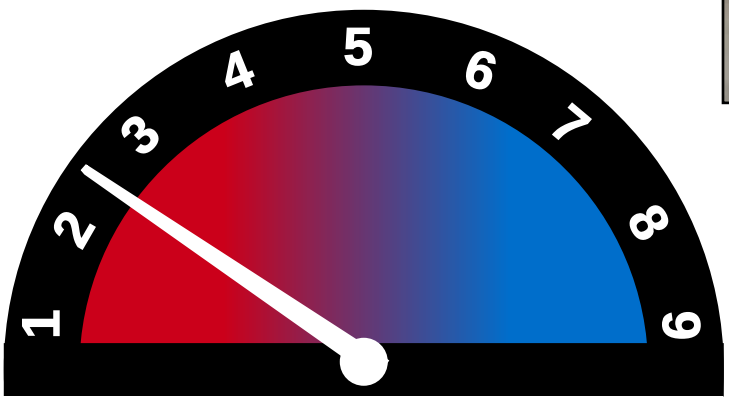


On our partisanship scale, where 1 is most conservative and 9 is most liberal, Joe Lieberman scored a 4.00.

Total # of Predictions	17
Correct	6
Wrong	8
Hedged	3
Prognosticator Value Score	-1.2
P-Value	0.0944

Lindsey Graham

Lindsey Graham found himself in the unfortunate position of being a close friend and adviser to the eventual loser in the 2008 presidential election. Many of his political predictions appeared to be skewed as a result of this alliance with Senator John McCain. Even in the face of clear polling data two days before the race, Graham was making bold predictions in support of McCain which did not go his way. He did more damage to his final score by finding new ways to express his belief that McCain would win certain swing states and the presidency. Graham was a constant presence on the Sunday morning TV shows both as an influential senator and spokesman for McCain, and he used his face time to make many, many predictions. These generally consisted of predictions only about politics and the wars we are currently fighting in Iraq and Afghanistan.

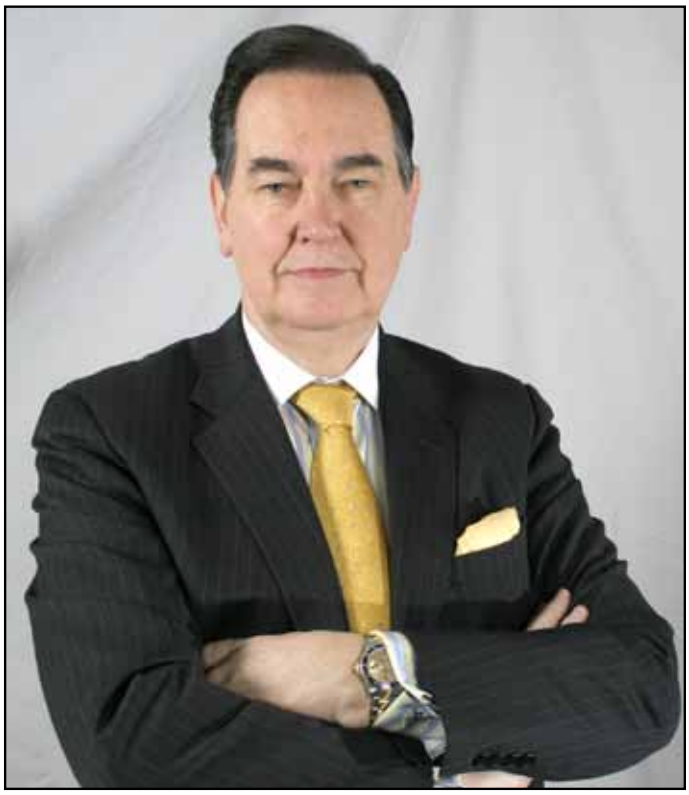


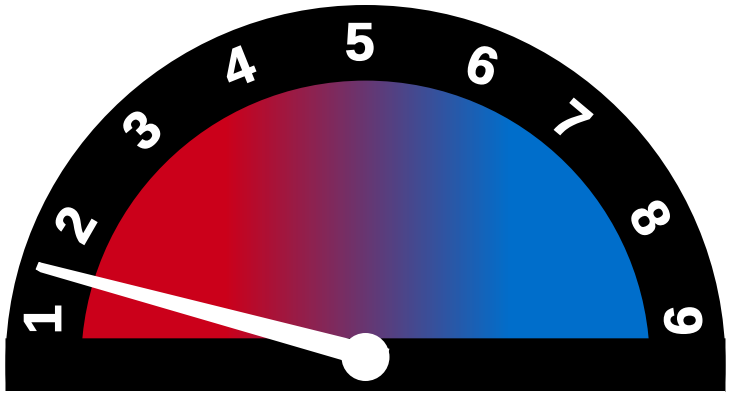
On our partisanship scale, where 1 is most conservative and 9 is most liberal, Lindsey Graham scored a 2.50.

Total # of Predictions	41
Correct	13
Wrong	27
Hedged	1
Prognosticator Value Score	-3.4
P-Value	0.0080*

Cal Thomas

Cal Thomas of the *Chicago Tribune* was the worst of all the prognosticators in our sample. Scoring an -8.7, readers could literally gain insight by believing the opposite of what they read in his weekly columns. Of his 15 predictions, 13 were wrong and only one was correct. Although occasionally Mr. Thomas was close (predicting the Nobel Peace Prize would go to Bill Clinton after Bush left office as a political statement when many would argue it went to Obama for the same reason), more often than not his predictions were overly supportive of the Republican party (predicting a Republican president, the end of immigration law enforcement under a liberal Congress, and Palin lifting her ticket to victory). Another Republican influence in Mr. Thomas' prognostication can be seen in his insistence that "the next terrorist attack" is "coming soon." Cal Thomas discussed at length this perceived threat, yet none actually occurred. Mr. Thomas focused on a short time frame, yet this did not aid his prognosticating accuracy as much as his Republican support hurt it.





On our partisanship scale, where 1 is most conservative and 9 is most liberal, Cal Thomas scored a 1.50.

Total # of Predictions	15
Correct	1
Wrong	14
Hedged	0
Prognosticator Value Score	-8.7
P-Value	0.0004*

Components of Predictive Accuracy

Deciding on which Variables to Regress

After we had a complete sample of the predictions of all twenty-six prognosticators, we had to decide how many, and on which combinations of variables we wanted to regress. We decided to run the following four regressions:

Regression	Dependent Variable	Independent Variables
1	PredABS	JournalisticExp, PredNatSec, PredHC, HouseComp, HouseElect, PropoQ, PresPred, Conditional, PredSoc, AdvancedDegree, PredEnviro, PredFP, VPPred, SenComp, GOPPrimary, CurrentAdviser, PredImm, Gender, PredEcon, RepeatPred, Race, DemPrimary, PredMed, LawDegree, Age, Partisanship, FormerAdviser, Journalist, SenElect, Politician
2	PredExtreme	JournalisticExp, PredNatSec, PredHC, HouseComp, HouseElect, PropoQ, PresPred, Conditional, PredSoc, AdvancedDegree, PredEnviro, PredFP, VPPred, SenComp, GOPPrimary, CurrentAdviser, PredImm, Gender, PredEcon, RepeatPred, Race, DemPrimary, PredMed, LawDegree, Age, Partisanship, FormerAdviser, Journalist, SenElect, Politician
3	PredProb	JournalisticExp, PredNatSec, PredHC, HouseComp, HouseElect, PropoQ, PresPred, Conditional, PredSoc, AdvancedDegree, PredEnviro, PredFP, VPPred, SenComp, GOPPrimary, CurrentAdviser, PredImm, Gender, PredEcon, RepeatPred, Race, DemPrimary, PredMed, LawDegree, Age, Partisanship, FormerAdviser, Journalist, SenElect, Politician
4	PVS	Total Number of Predictions, JournalisticExp, Race, LawDegree, Age, Partisanship, FormerAdviser, Journalist, Politician

Analysis

The first regression tested correlation with a correct prediction. This was done by generating a new variable, PredABS, that was the absolute value of the difference between the prediction's truth (PredTrue) and the predicted outcome (PredProb). This PredABS could be between 0 (guessed correctly) and 4 (guessed the absolute opposite of what really occurred). Because a 0 represents a correct guess, variables with negative coefficients improve a prediction's accuracy.

An R-squared value of .156 was obtained for the regression. This means that about 16 percent of the total variance in prediction outcomes is explained by the variables in our equation. This may be due to the fact that there is a substantial amount of inherent randomness in predictions, since it seems unlikely that we missed variables.

A handful of the variables tested showed statistical significance. They are Conditional, Partisanship, and having a Law Degree. None of the prediction category variables (for instance, economic or presidential prediction) were statistically significant.

If a prediction was conditional, the coefficient of .108 suggests that it is moderately more likely to be a wrong prediction. When making a conditional, the prediction is really "double-barreled;" it relies on the conditional being true before evaluating the main prediction. This may strain the predictive powers of even the most seasoned prognosticators. An important addendum is that only predictions where the conditional came true were included in the sample.

Partisanship had a significant role in a prognosticator's overall accuracy. Our scale measured it from 1 (most conservative) to 9 (most liberal) and as partisanship "went up" one level (a person was rated more liberal) there was a moderate increase in their predictive capacity. The difference between the most conservative individual in our sample and the most liberal, according to this model, is nearly 2 points in the overall PredABS score. The standardized coefficient of -.264 was the largest of any variable.

People with a law degree were somewhat more likely to get predictions wrong; the coefficient was fairly large, positive, and statistically significant (.239). This could be an artifact of the legal training process, which places great emphasis on argumentation with less on analysis and evaluation of data. All three variables were significant at the $p = .05$ level, with partisanship significant at the $p = .001$ level.

Our second regression examined the factors that influence extreme predictions, where the prognosticator used

predictive language coded in our dictionary as either a 1 or a 5. To create this dependent variable, we took all the prediction probability scores and took the absolute value of this number after subtracting by three. These prediction probability scores could now be examined in a new way – the most extreme predictions were now scored as 2 (the absolute value of 5 or 1 minus 3), the more moderate predictions were now scored as 1 (the absolute value of 2 or 4 minus 3) and hedged predictions were scored as 0. This calculation clumped all the similarly extreme types of predictions together under one number, making it easy to measure their influence together.

Of the 30 independent variables we had in the PredEXTREME regression, a HouseComp and Conditional were found to be statistically significant at the $p < .1$ level. At the $p < 0.05$ level, Age, VPPred, Politician, and JournalisticExp were statistically significant. GOPPrimary was significant at the $p<0.01$ level.

Most of these results are not surprising. We should expect prognosticators with political agendas to use stronger language than those who work in the media. Media figures may need to stay more neutral with their points of view and use less extreme language. That said, a journalist's experience is significant – the longer a journalist has been working, the more likely he/she will use extreme language. But being a journalist is nowhere close to being significant. Interestingly, age correlates negatively while journalist experience positively.

We ran the third regression with PredProb as the dependent variable to see what, if any, of our independent variables affect the framing of predictions. That is, we sought to test if anything had an impact on someone saying an event "won't" or "will" happen. It is important to differentiate between the tone of the prediction, which is really what PredProb records, versus a change/status quo dichotomy. Change or remaining in a status quo can both be predicted in either a positive or negative manner (i.e. change will happen vs change will not occur).

Our regression found 8 variables to be significant. Our r-squared is a rather low .18. This r-squared value tells us that our independent variables account for only 18 percent of the change in our dependent variable, PredProb.

The variables found to be significant were Conditional, House Election, Senate Election, Senate Composition, House Composition, Immigration Predictions, Age, Current Adviser, and Journalist. Senate Election was significant at the $p < .1$ level; all others were significant at the $p < .05$ level. The other three significant variables were significant at the .05 level. Predictions about Senate Composition after the election and the Senatorial Elections were more likely to be made using negative language than "other" predictions, our omitted condition. Predictions about the House of Representatives' composition and elections, however, were more likely to be made using positive language than other predictions. Predictions about immigration were more likely to be made

using negative language than "other" predictions. Predictions made by current political advisers are more likely to be made using positive language. Age tended to make people more likely to make predictions using negative languge. Lastly, predictions made by journalists are more likely to be made using negative language.

We also decided to run a regression that measured what influenced the overall Prognosticator Value Score (PVS) that each prognosticator had been scored with. We originally ran 10 independent variables to determine what most heavily influenced each person's final PVS score, but we had to drop the Political Party variable since we could only apply this data to a handful of our prognosticators.

Of the nine independent variables we did end up using, only two appeared to significantly influence the PVS scores: partisanship and having a law degree. The unstandardized b coefficient for partisanship is positive, which suggests that the more liberal prognosticators (as the Hamilton College government department faculty and politically-motivated student groups scored them) are more likely to have higher PVS scores. For the prognosticators who did have a law degree, the negative coefficient suggests that having this degree lowers one's PVS score. When we examined Tetlock's book (2006) on prognosticators, he classified everyone as either a fox or a hedgehog (for a more detailed discussion on Tetlock, please refer to our literature review). It appears that these eight prognosticators with a law degree were trained to examine issues like a hedgehog, using only one analytical technique in all circumstances.

The R-squared value for this regression comes in at .500. Our nine independent variables captures half of the variance in predictions. This R-squared value is reasonably good and we could not think of any additional variables to include in the regression.

Reinforcing Tetlock

Our findings seem to agree with Tetlock's research. Some studies suggest that conservatives have more rigid ideologies (Jost et al., 2003) In other words, they would be considered "hedgehogs." Similarly, lawyers are taught to argue one side with a single analytical method; they, too, would be "hedgehogs" under Tetlock's model. While not all liberals are foxes and not all conservatives are hedgehogs, these trends may be informative in explaining why our results are as they are. It may be that conservatives are inherently disadvantaged as prognosticators due to their ideological "hedgehoginess."

What Does Not Matter

We found that the majority of measured factors have no statistically significant influence on a prognosticator's predictive ability. Age, race, gender, previous employment as either a politician, adviser to the President, or journalist, and number of years as a journalist all had no statistically significant impact in the PredABS regression. This suggests that they have no bearing on a prognosticator's ability to be correct.

The implications of this are significant. Prognosticators should only be chosen on a few measurably important variables, namely liberalism and the lack of a law degree. Everything else is unimportant; women are no better than men, whites are no better than blacks, and experience has no advantage over youthful vitality. The converse, in all cases, is also true. So while it does not take much to be a good prognosticator, there also is not much that can disqualify someone from being a good prognosticator.

Perhaps most importantly, being a good prognosticator seems to be a product of choices, not birth. Anyone can be good; all they need to do is avoid law school and buy into liberalism as an overarching philosophy. There is no inferior ability associated with being born, say, black or female.

Implications

In this section, we address numerous conclusions we drew based on the available data. First, we address the question of whether or not individuals' predictions are, in aggregate, superior to flipping a coin. Next, we question the role prognosticators play in our society and discover that there are a few viable theories, including prognosticators as news reporters and as entertainment. We explore the implications of tone--that is, whether saying something "will" happen causes a different result than saying it "won't." Finally, we get "into the weeds" and address questions about specific variables. We explore whether an individual's career path or partisanship have an impact on their predictive accuracy.

Better than a Coin Flip

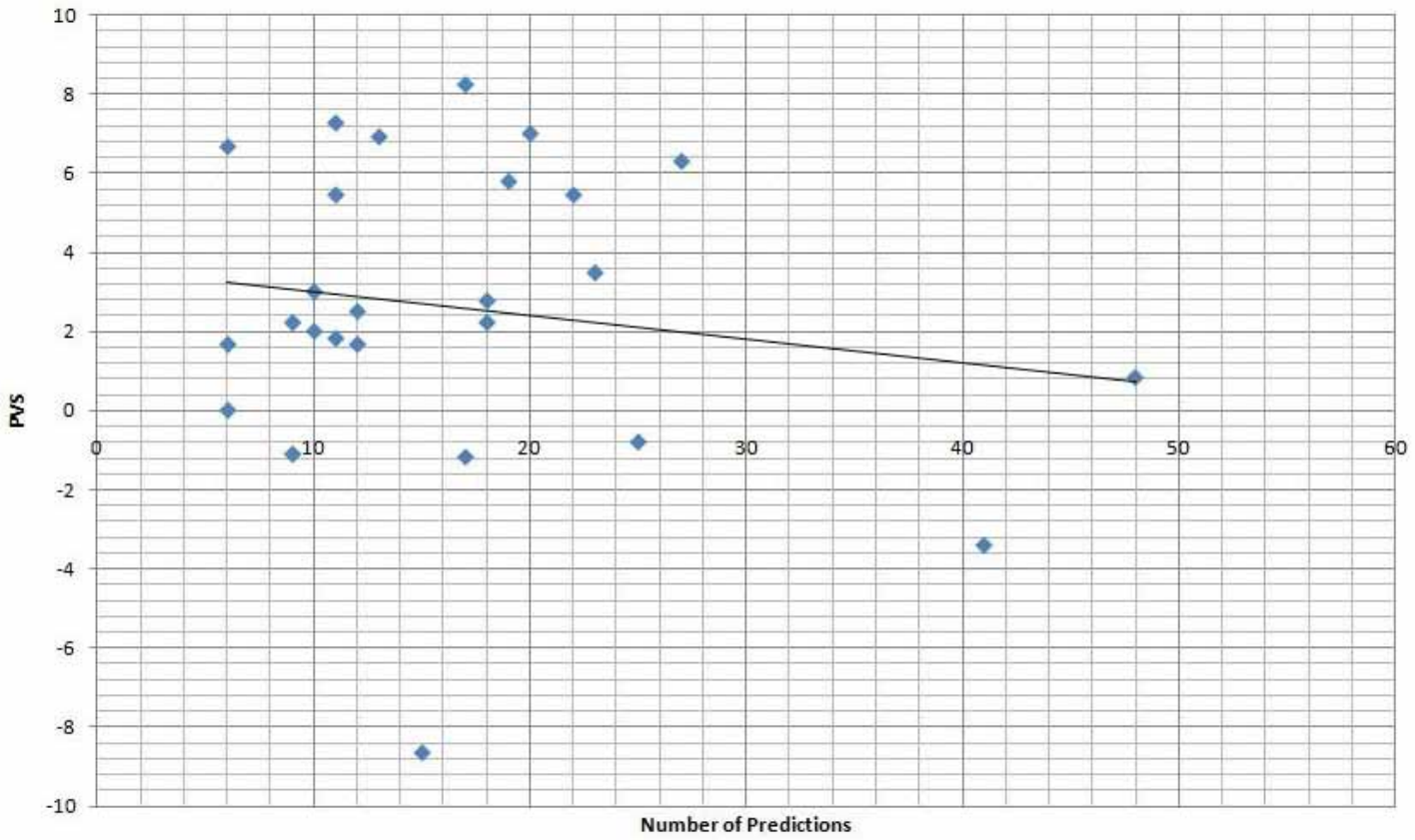
So, are prognosticators more accurate with their predictions than a 50-50 coin flip? Basic evidence from our descriptive statistics suggests that prognosticators, in aggregate, may be more effective than a simple coin flip at obtaining the correct results. On the whole, more prognosticators had correct predictions than incorrect predictions, as our Good, Bad, and Ugly section shows. However, many of these people also had very small sample sizes.

To parse out the coin flip question using more statistical measures, we decided to use the binomial formula to calculate the probability that these scores for each prognosticator were statistically significant. Six prognosticators scored significantly better than a coin flip and four scored significantly worse according to this metric. The remaining 16 were not significant either way, which shows that calculating p-values using the binomial formula has not produced conclusive results. It is clear that we need more predictions from many of the prognosticators, several of whom tallied 10 or fewer testable predictions. With so few predictions for several of these prognosticators, their scores were never going to be statistically significant. We simply lack the data to categorize 16 of our 26 prognosticators who were not significantly better or worse than a coin flip.

However, the prognosticators with much larger samples also tended to have Prognosticator Value Scores that edged toward zero. This could be proof that PVS fall closer to zero as the number of predictions rise. George Will had the highest number of predictions in our sample (48) and has a PVS of 0.83, which is fairly close to zero. Obviously, not all of our prognosticators fit this rule, but with only four people recording more than 25 or more predictions, it is hard to see this trend with such a small sample size for the majority of people.

From these trends in the data, our team formed a hypothesis stating that predictions become equally right and

wrong as the number of predictions increase. In essence, flipping a coin would be just as accurate. To test this, we ran a scatterplot measuring the number of predictions on the x-axis over versus PVS for each prognosticator. A fit line that had a negative slope would show that for our prognosticators on the right side of the scatterplot (those with a larger sample) would fall closer to zero and would statistically be no better than a coin flip. Indeed, the scatterplot below has a negative slope. The nature of the scatterplot that we ran did not lend itself to exact measurements, but the approximate slope is -0.028. This slope seems small, but the scales of the two axes are very different. The fact that the slope of the best fit line is negative helps show that prognosticators do not make better predictions when they increase to *x* amount of predictions.



Our reasonably small sample of 26 prognosticators was definitely representative of the universe of prognosticators. It included males and females, politicians and journalists, and people all over the ideological spectrum. Individuals ranged in age from 41 to 78—all individuals in the mature stages of their respective careers. Every category of prognosticator is represented. But on the other side of the coin, our sample window was very small relative to the careers of many of our prognosticators. The 16-month span of our sample is a tiny fraction of Clarence Page's 45 years of journalistic experience, Cal Thomas' 50, or David Broder's 48. But even though there are limits to the sample we constructed, the random sample of chosen prognosticators and their body of work helps alleviate many of these concerns.

Overall, we cannot definitively prove our original hypothesis to be true, but the data we do have seems to in-

dicate that prognosticators are indeed no better than a coin flip. Certainly a select few are statistically better by a wide margin, but these people seem to be the exception to the rule and not the norm. It is clear that we need more data to classify the remaining 16 prognosticators who are not statistically significant for being better or worse than a coin flip, but our scatterplot shows that their Prognosticator Value Scores will likely go down to when this data is obtained. Taken all together, it seems clear that prognosticators are no better than a coin flip.

Role of Prognosticators

What is the purpose of a prognosticator? With a few exceptions, the vast majority in our sample have proved to be less-than-excellent predictors of future events in everything from the upcoming election to the economy. If this is the case, why do these journalists get paid to write columns predicting future events and appear on Sunday morning shows? Why do politicians comment on events as predictions when everyone knows they are simply following their party line? And most importantly, why do citizens still consume what these people are saying?

This paper has found several people who are worth listening to and will provide a list of desirable qualities for prognosticators in general. However, prognosticators may fulfill some non-predictive purpose since their prognostication skills are on the whole lacking. For some members of society, TV and radio personalities and the politicians that appear on Sunday morning talk shows may fill an informational role by providing insight into current events through accessible media. However, these prognosticators may simply be entertainment. They are ubiquitous, appearing constantly on television and in newspapers. Both mediums can serve to entertain Americans at otherwise-boring parts of the day.

In the 2008 election season, liberal prognosticators were more accurate than conservative prognosticators. But we are reluctant to say that a certain ideology is incorrect on a larger timeframe because of its failures in our sample. The world is, simply put, too fluid for us to make this judgment. Conservatism may be an invalid philosophy in the 2008 election timeframe, but may do a good job of explaining phenomena from 2000 to 2007. From September 2007 to December 2008, though, liberalism was the ideology of the best prognosticators.

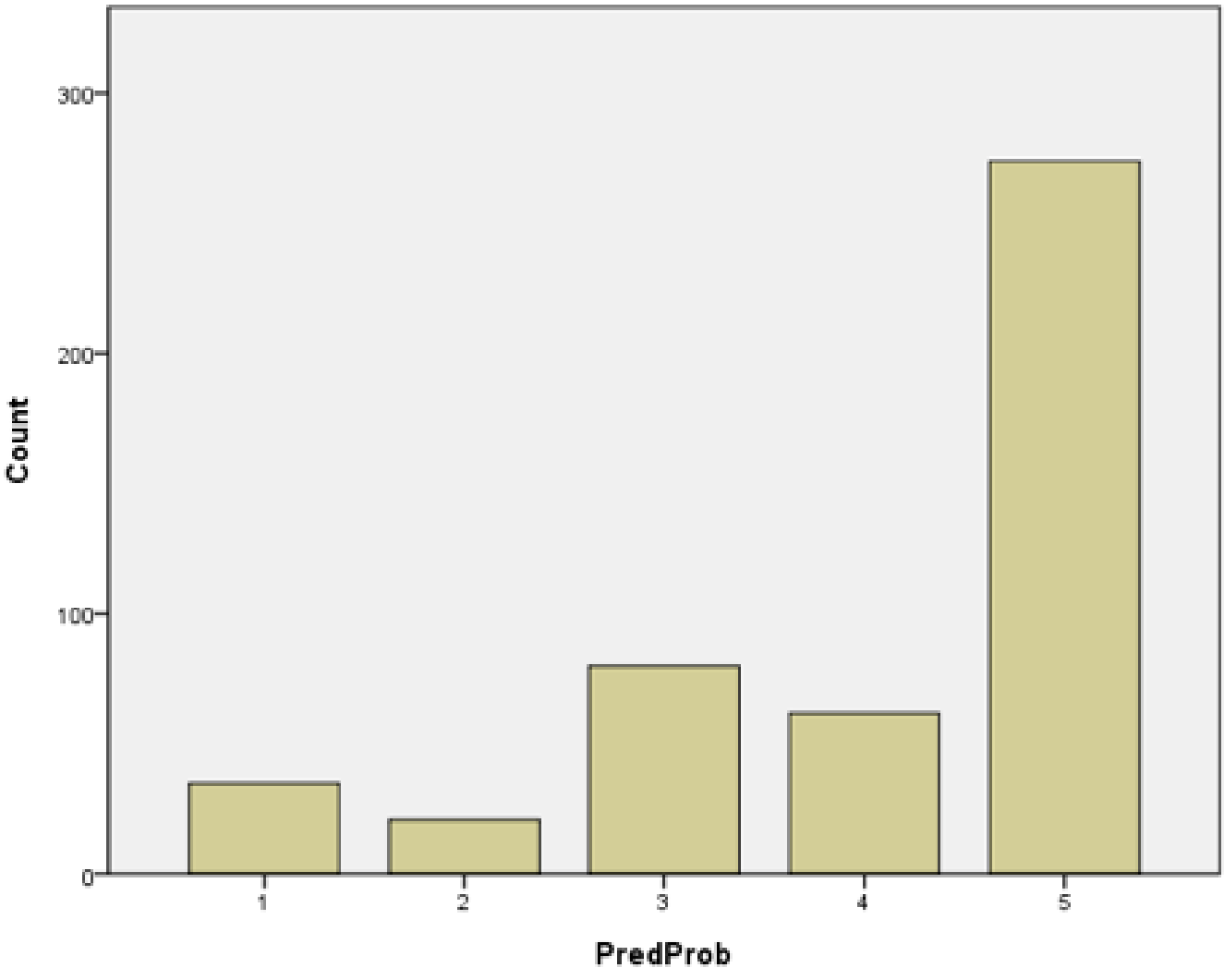
Is it worth listening to what prognosticators have to say? It depends on what each person values out of them. If someone is looking for quality predictions that have a better-than-average chance of coming true in the future, the general answer is no, with a few notable exceptions. Prognosticators on TV provide a unique and different style of delivering the news, in either a roundtable or one-on-one format absent from evening newscasts. As an information resource, prognosticators may deliver varying points in unique ways; as an entertainment resource,

they provide lively and active debate.

Implications of tone

The regression with PredProb as a dependent variable was fundamentally a test of the tone prognosticators took on certain issues. The low r-squared value of .18 could be the result of simply looking at the wrong independent variables. Maybe other factors have a much greater effect on the way predictions are phrased in a positive or negative manner. Many factors that could influence someone's tone are difficult or impossible to measure. This could include factors like a prognosticator's mood while writing the column, an impossible variable to measure. Another possibility is that PredProb is a rather randomly determined variable. Because both change and non-change can be predicted either in a positive or a negative manner, many times the decision as to which phrasing to use could be arbitrary. A largely random use of positive/negative phrasing would therefore result in a low r-squared no matter what variables were used in the regression.

Why are Predictions so Extreme?



Predictions that are more "extreme" (rated as 1 or 5 in Prediction Probability) were more likely to occur than predictions rated 2 or 4--that is, predictions that suggest an event will "probably" or "probably not" occur. A number of causes are apparent. First, prior to appearing on a show, journalists and politicians will prepare briefing books that outline nearly every possible topic that can be addressed on a talk show. This means that they have time to formulate opinions on recent events before being asked by the show's host. This, in turn, could lead them to say will or won't. If they do not want to answer the question because it is controversial or risky, a prognosticator can simply "punt" on the topic by saying "maybe." Alternatively, prognosticators may be set in their ideological ways; "hedgehogs," to use Tetlock's turn of phrase. If they are using one tool to analyze all information, that monochromatic approach could lack the nuance of "probably" or "doubtful." Finally, prognosticators may simply want to sound sure of themselves when making commentary about the future, which would explain the overall prevalence of "will" and "won't" compared to the dearth of "unlikely," "probable" and "maybe."

This tendency suggests that society at large should not look to prognosticators for nuance. While they may end up vindicated by a correct prediction or shamed by a misguided one, they will rarely reflect the uncertainty that exists in the real world. Very few events have a 100 or zero percent chance of happening; the underlying odds usually lie somewhere in between. The public at large should not believe the absolutes presented regularly in newspapers and on the Sunday morning talk shows.

Very few prognosticators were keen to use this nuanced language as our PredEXTREME regression shows. Politicians were much more likely to use extreme language, and journalists with more experience were also much more likely to as well. The R-squared value for this regression comes in at .151. Our long list of independent variables does not appear to capture much of the reason for variance in explaining this use of extreme language. Clearly, there are other factors are at work which determine when a prognosticator will use extreme predictive language outside of these two significant variables. Whether these other factors can be recorded and measured in an unbiased manner remains to be seen.

Partisanship

The variable for partisanship measured how liberal or conservative each prognosticator was relative to each other and on a scale of 1 (most conservative) to 9 (most liberal). As stated in the analysis of the predABS regression, as partisanship increased one level (a person was rated more liberal), so did predictive capacity as measured by the variable PredABS. In fact, the standardized coefficient of -.264 was the largest of any variable in the PredABS regression, so this variable had the largest individual effect on PredABS of all variables. This

finding was statistically significant, and so at face value it appears that Democrats make better prognosticators, at least within our sample. Similarly, the regression in which the independent variable was Prognosticator Value Score (PVS), the Partisanship variable was statistically significant and had a positive coefficient, indicating that prognosticators that are more liberal have higher PVS.

However, this large statistically significant coefficient signaled a red flag for us – it seems unlikely that Democrats would beat out Republicans by such a large margin. Our first instinct was that the election outcome (with Democrats and Barack Obama winning by large margins) could have had a confounding effect on the partisanship results. Democrats predicted optimistically that their candidates would win election, and the election swung in their favor, thus making their rhetorically-driven predictions correct. In order to decipher whether or not the election had skewed our results, we decided to run the regression again, this time excluding all the predictions having to do with the November election. Partisanship was once again statistically significant, showing that the liberal prognosticators had actually predicted more accurately than their Republican counterparts during our time period.

Democrats seem to be better at predicting than Republicans, but we are cautious in claiming that this is a generalizable conclusion outside of our time period. There may be underlying factors in the 2008 elections that might not occur in other time periods. It's also plausible that partisan strategies may lead to inaccurate predictions by Republicans--Republicans, unlike Democrats, may hope to face who they see as the weaker candidate from the opposing party in November. This may mean that some predictions are not really meant to be predictive--rather, they hope to shape the debate. Of course, it is certainly possible that Democrats really are better predictors.

Regardless, our findings on partisanship provide insight into the rhetorical and partisan-driven world of politics and the media that surrounds it. When making predictions, all prognosticators (both journalists and politicians) are aware of the fact that their predictions can impact voters. What's interesting, though, is that it's not purely candidates spewing confidence building rhetoric and optimistic promises – it's also less obviously affiliated prognosticators. According to our data, 67 percent of the predictions made by non-politicians were about the election, even more than the politician's corresponding 64 percent.

Future studies could look at the predictions of both Democrats and Republicans over a longer time period spanning multiple elections to see whether our finding of liberal success can be extrapolated.

Career Variables

The variable that was most frequently statistically significant was an individual's status as a current adviser. According to our regression where Prediction Probability (PredProb) was the dependent variable, current advisers have higher PredProb values, meaning that current advisers are more likely to predict using positive language. For example, a current adviser might say that Hillary Clinton will be elected, or that a policy has a real chance of working (These two predictions would be coded as a five PredProb and a four PredProb, respectively). This makes intuitive sense, as current advisers have a large incentive to tell the public that there is a higher chance that a policy is working, meaning that prediction would be coded as a four or a five.

However, it is important to note that only one of the prognosticators was coded as being a current adviser. Because our prognosticator sample was selected at random, Howard Wolfson, who served as Hillary Clinton's Communications Director during her 2008 presidential campaign, and Hank Paulson, Bush's Secretary of the Treasury, were the only current advisers included in the sample. Future studies could investigate whether this finding holds true if the selected sample of prognosticators were to include more current advisers.

According to our regressions, journalists are less likely to have higher Prediction Probabilities, meaning that journalists are more likely to predict using negative language; i.e.: Hillary Clinton will not be elected, this policy probably won't work (These two predictions would be coded as a one PredProb and a two PredProb, respectively). The journalists in our study include David Broder, David Brooks, Sam Donaldson, Maureen Dowd, Thomas Friedman, Bob Hebert, Mike Huckabee, Nicholas D. Kristof, Paul Krugman, Andrea Mitchell, Clarence Page, Eugene Robinson, Cal Thomas, George Will, and Howard Wolfson. The finding that journalists make negatively phrased predictions seems to run in opposition to our earlier finding that television has more negative predicting than print. However, it is important to keep in mind that journalists make predictions in both print and on television. Therefore these findings do not contradict one another, but perhaps show that journalists, like all the prognosticators on television, may be looking to entertain their audience through clashing viewpoints and refutations of the views of the opposition.

The final statistically significant career-related variable was whether or not an individual held a law degree. According to our final regression, in which we held Prognosticator Value Score (PVS) as the dependent variable, individuals holding law degrees are more likely to have lower PVS, meaning they are more likely to be classified by our PVS system as poor prognosticators. This is compatible with the fox/hedgehog model established by Tetlock and referenced in our literature review. If an individual holds a law degree and is taught to consistently approach problems and predictions attempting to identify and defend a certain outcome – as lawyers are

taught to do – they are less likely to view things from a broad range of perspectives.

Age, while not necessarily a "career" related variable, was also statistically significant in two of the regressions: PredExtreme and PredProb. According to these regressions, an older prognosticator is less likely to make extreme predictions than younger prognosticators. Additionally, older prognosticators are more likely to predict using negative language. For instance, an older prognosticator is more likely to say, "This bill might not not pass," as opposed to, "This bill will definitely pass."

Prediction Variables

There were a number of statistically significant variables in regards to the types of predictions being made by prognosticators. One such variable was HouseComp, or predictions made about the overall composition of the House after the 2008 election. In this case, it was a significant variable in both the PredExtreme and PredProb regressions, meaning predictions made about the House composition were more likely to made using both positive and extreme language. In other words, a prediction about the composition of the House is more likely to be, "the House will be overwhelmingly Republican," rather than "the Republicans might not win a large number of seats in the House." Interestingly, our PredProb regression also shows that predictions made about the composition of the Senate are more likely to be negative.

According to our PredProb regression, predictions made about individual elections to the House are more likely to be positive: "This candidate will win the election for a seat in the House," or, "This candidate may win the election for a seat in the House." On the other hand, predictions made about individual elections to the Senate are more likely to be negative: "This candidate will not win his election," or, "This candidate may not win his election."

Predictions made about the GOP primary are more likely to made using less extreme language. For instance, "McCain may not win the primary," or, "McCain may win the primary." Predictions made about the Vice Presidential choices were also more likely to be made using less extreme language. Predictions made about immigration policy were more likely to be made using negative language, such as, "This bill will not pass," or, "This bill may not pass."

Conclusion

Overall, our results indicate that most prognosticators are not very accurate predictors, but only four of 26 are worse than a coin flip (with statistical significance). Few offer reliably accurate predictions, but even fewer are wrong more than half of the time – most barely hover above the dreaded "ugly line." This should be startling, considering the number of Americans who rely on these prognosticators for their supposedly superior knowledge of the political environment.

Given this sad reality, who should you listen to? Good predictors tend to be liberal and are not lawyers. More rigorous study can confirm our findings, especially the question of whether partisanship has an impact on an individual's ability to make accurate predictions. There were nine prognosticators who were classified as "good" predictors. They were Paul Krugman, Maureen Dowd, Ed Rendell, Chuck Schumer, Kathleen Parker, Nancy Pelosi, David Brooks, Eugene Robinson, and Hank Paulson. Five of these were journalists at the time of our sample. Six are considered liberal, and the average partisanship score was 5.87 (with 5 being a perfect moderate.) Three are female and one is black. It is clear that there is a significant amount of diversity in the "good" category.

And who should you not listen to? Individuals who hold law degrees are less accurate when making predictions. Conservatives, according to our data, are also less accurate. But it is also important to keep in mind that the overwhelming majority of demographic factors have no bearing on a prognosticator's accuracy. Gender, race, and age are all irrelevant, as are most career path choices, such as becoming a journalist.

When listening to current political advisers, keep in mind that they are more likely to make predictions using positive language. Similarly, it is important to remember that professional journalists and male prognosticators are more likely to make predictions using negative language. Younger prognosticators, politicians, and journalists with more experience are more likely to predict using extreme language.

Future studies can expand upon our research in a number of ways. Increasing the number of prognosticators beyond 26 could lead to more significance among the variables within similar regressions. Different or longer time periods could be examined, such as non-election seasons or election seasons which favored Republicans rather than Democrats. Implications from such studies could uncover potential biases within our study tied to the nature of the prediction season. Taking these things into account, we hope that future studies can more fully examine the role of political prognosticators in the national dialogue. For now, however, our study is the first to create some sense of accountability for these prognosticators, while also examining their value in political media.

Works Cited

Begley, Sharon. "Why Pundits Get Things Wrong." Newsweek. February 23, 2009. Volume 153 Issue 8. P 45.

Berlin, Isaiah. "The Hedgehog and the Fox: an essay on Tolstoy's view of history." Ivan Dee, 1993.

Douglas, Susan. "You, too, can be a Pundit." Progressive. September 1995. Volume 59 Issue 9. P 17.

Greenwald, Glen. "The Perilous Punditocracy." National Interest. May/June 2008. Issue 95. Pp 25-29.

Jost, John T., Jack Glaser, Arie W. Kruglanski, and Frank J. Sulloway. "Political Conservatism as Motivated Social Cognition." Psychological Bulletin 129.3 (2003): 339-75. Online.

Media Matters for America, Black and White and Re(a)d All Over, http://mediamatters.org/reports/oped/

Perlstein, Rick. "Pundits who Predict the Future are Always Wrong." Nation. April 23, 2001. Volume 272 Issue 16. Pp 11-14.

Tetlock, Philip. Expert Political Judgment: How good is it? How can we know?. Princeton University Press, 2009.