UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
High Chances and Close Margins: How Different Forecast Formats Shape Beliefs

Permalink
https://escholarship.org/uc/item/64f2v89r

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 40(0)

Authors
Urminsky, Oleg
Shen, Lucy
Nero, Sondre

Publication Date
2018

eScholarship.org Powered by the California Digital Library
University of California

High Chances and Close Margins: How Different Forecast Formats Shape Beliefs

Oleg Urminsky (oleg.urminsky@chicagobooth.edu)
University of Chicago, 5807 S. Woodlawn Ave
Chicago, IL 60637 USA

Lucy Shen (luxi.shen@cuhk.edu.hk)
Chinese University of Hong Kong, 12 Chak Cheung Street
Hong Kong SAR, China

Sondre Nero (sondreskarsten@gmail.com)
University of Chicago, 5807 S. Woodlawn Ave
Chicago, IL 60637 USA

Abstract

While a large literature has studied how people make forecasts, less is known about how lay people process and interpret forecasts presented to them. We contrast two common ways of communicating an uncertain forecast, as either a chance (e.g., the probability of winning) or as an expected margin (e.g., the point spread). Across five studies, we find a robust chance-margin discrepancy: people tend to treat a chance forecast as conveying greater probability of the higher-likelihood outcome than the statistically equivalent margin forecast.

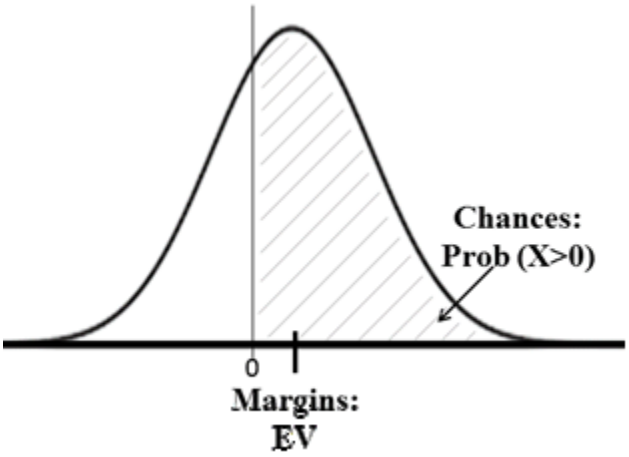
Keywords: Judgment; Decision Making; Linguistic Priming; Intertemporal Choice; Inference.

Forecasts are pervasive, with experts predicting outcomes ranging from election outcomes and economic circumstances to the weather and sports outcomes. Increasingly, forecasts are based on empirical data and statistical models, rather than expert opinion. For example, pundits used their judgment or exemplars ("bellweather" areas) to make election predictions in the past, but modern election forecasts primarily rely on political polling results, and more recently on model-based polling aggregation (Hilygus, 2011).

In statistical model-based forecasts, a probability distribution of outcomes is typically generated. Then, this distribution is used to generate the summary statistics that comprise the forecasts of interest. The two most commonly communicated summary statistics are the chances of a focal outcome (e.g., probability of winning an election) and the predict margin (e.g., the expected relative vote share of the candidates).

When the underlying form of the outcome distribution is understood, these summary statistics provide different but statistically equivalent information. For example, if the outcomes are normally distributed with a known variance (Figure 1), the margin (expected value of the outcome) and the probability of a positive outcome (mass above zero) are each sufficient to identify the mean of the distribution. In fact, the mean (expected margin) can be calculated from the tail mass for a given cutoff (chance of winning) and, conversely, the tail mass (chance of winning) can be calculated from the mean (expected margin).

Figure 1: Chance and Margin Forecasts



In this paper, we investigate how people process a chance or margin forecast, and the impact on their estimates, attitudes and behaviors. The key question we pose is whether people treat chance forecasts (e.g., the probability of a candidate winning an election and of a sports team winning a game) as if they conveyed the same information as the corresponding margin forecasts (e.g., predicted vote share or point-spread).

One possibility is that people are skilled intuitive statisticians, particularly in domains in which they have experience, such as politics and sports. Recent research in cognitive psychology has argued that many decision-making phenomena that might be assumed to be errors can be represented by Bayesian models that account for uncertainty or computational cost of complex inferences (Lieder, Griffiths & Goodman, 2012; Pouget, Beck, Ma & Latham, 2013). In this view, people are able to near-optimally represent and update probability distribution information (Griffiths & Tenenbaum, 2006), perhaps spontaneously at the neural level (Knill & Pouget, 2004; Ma, Beck, Latham, & Pouget, 2006). If this is the case, people will recognize corresponding chance and margin forecasts from the same underlying outcome distribution as providing equivalent information.

Another possibility is that people are systematically biased in their statistical reasoning. A large literature on intui-

tive statistics suggests that this is the case, particularly in making inferences from samples (Kahneman & Tversky, 1972; Peterson, DuCharme and Edwards 1968; Wheeler and Beach 1968), largely due to using simplifying or simply incorrect heuristics (e.g., Tversky & Kahneman, 1974). Urminsky (2014) shows that people fail to accurately generalize from individual event probabilities to the overall outcome distribution. Instead, people systematically overestimate tail probabilities, even when their estimates of the central tendency of the distribution (e.g., mean, median and mode) are quite accurate. More generally, recent papers (Jones & Love, 2011; Griffiths, Chater, Norris, & Pouget, 2012; Bowers & Davis, 2012) have revisited such findings in light of the success of Bayesian models of reasoning and debated the degree to which Bayesian processes can plausibly explain performance in complex decision tasks.

Next, we present five studies that investigate the degree to which people treat chance and margin forecasts as equivalent (e.g., correctly estimating the forecast in one format when given the forecast in the other format), and the implications for attitudes (election optimism, Study 2) and choices (sports betting, Study 4). We begin with tests using actual public forecasts of election and sport outcomes, and we conclude (Study 5) with a statistical scenario in which correct inferences can be definitively identified.

Study 1: Presidential Election Forecasts of Winning Chances vs. Vote Margin

Method

We collected data in five waves, between August and November 2016, prior to the U.S. Presidential election, with a combined 1244 US Amazon Mechanical Turk (AMT) participants. We used actual election forecasts from the website *fivethirtyeight.com*, based on aggregated political opinion polls, of (1) the chances (probability) of each candidate winning and of (2) the predicted vote-margin forecasts for each candidate. Both forecasts were based on the same polling data, and can therefore be seen as summary statistics for approximately the same distribution of outcomes (ignoring the role of the Electoral College). The forecasts varied over time, as the polling results shifted.

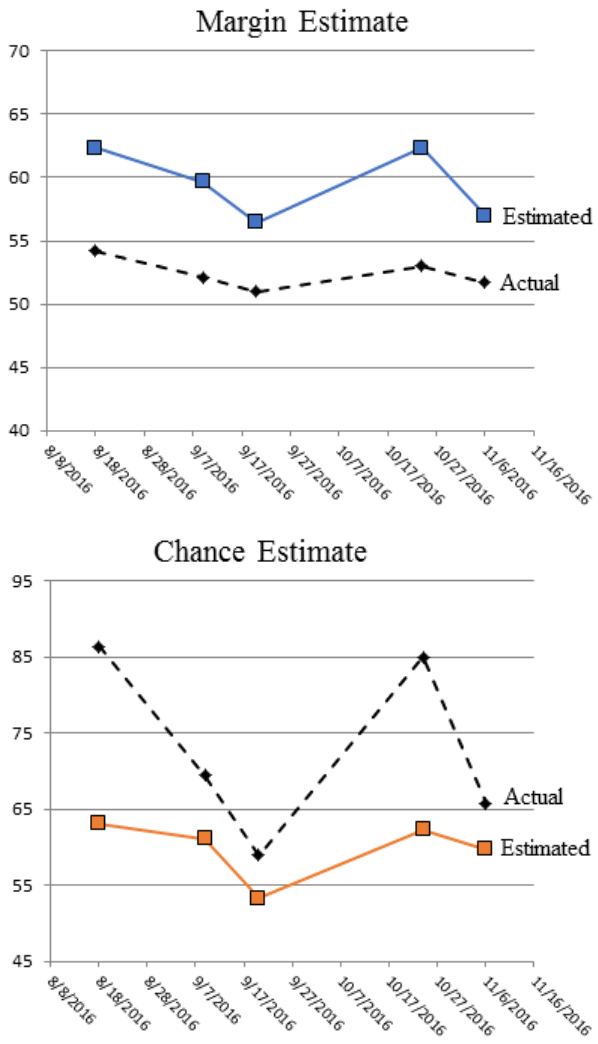
Participants were randomly assigned to one of two conditions: they either saw the chance forecast and were asked to estimate the margin forecast, or saw the margin forecast and were asked to estimate the chance forecast. Given that voters' own preferences end to distort their outcome estimates (Orhun & Urminsky 2012), participants were not asked to make their own personal predictions, but were instead asked to estimate the predictions made by the website.

Results

Participants' estimates were significantly biased (Figure 2), suggesting that they did not interpret the two types of forecasts as representing the same underlying political situation. Overall, when participants were shown a chance forecast (an average 75.35% chance of Clinton winning), they overestimated the forecast margin (60.56% estimated vs. 52.53% actual, $t(588) = 14.09$, $p < .001$). Conversely, participants who were shown a margin forecast (an average share for Clinton of 52.7%) under-estimated the chance forecast (59.27% estimated vs. 74.09% actual, $t(591) = 23.51$, $p < .001$). These finding replicated in each of the five tests, as the actual forecasts varied, when analyzed separately (all *p*s < .001).

These findings suggest not only that people's intuitions about election outcome chances and margins are inconsistent, but that they diverge in a systematic way: margins are overestimated and chances are under-estimated, relative to each other.

Figure 2: Voting Forecasts



Study 2: Forecast Format Impacts Subjective Assessments

Study 1 demonstrates a consistent discrepancy between estimates based on having seen chance projections and estimates based on having seen margin projections. These results suggest that the two kinds of projections will lead to

different assessments of the state of the election, which we test in the next study.

Method

In Study 2, 226 US AMT participants saw the election forecast, either presented in terms of the chances of Clinton winning or in terms of Clinton's predicted margin of victory. This study was conducted four days before the election, when Clinton was leading 51.5% to 48.5% for Trump, and was projected to have a 64% chance of winning, according to *fivethirtyeight.com*.

Participants were asked to assess their opinions of the current predictions, on a scale from 1 ("very good news") to 7 ("very bad news"), and were asked several questions about election-related behavioral intentions. Since a wider perceived lead for Clinton would be perceived as more good news by her supporters but more bad news by Trump supporters, we coded the extremity of attitudes as the absolute difference from the center of the scale (extremity =|rating-4|).

Results

Participants who were shown the chance forecasts gave more extreme assessments of the information than those shown the share forecasts, on the same attitudinal scale (Ms = 1.75 vs. 1.38, t(224)=2.31, p=.022). However, the difference in forecast format did not significantly impact measures of behavioral intentions (e.g., intent to vote).

Study 3: Sports Forecasts of Winning Chances vs. Point-Spreads

One concern about the election forecasts is that our analysis relies on the assumption that the two summary statistics (chance and share) are in fact equivalent. Different forecasters during the election disagreed about the relationship between the predicted share (which was less controversial, and more directly based on polling data) and the predicted chances of winning. In particular, factors such a distribution of votes across states and the role of the Electoral College complicate chance predictions. We selected *fivethirtyeight.com* because their chance predictions were the most conservative (e.g., compared to the Princeton Election Consortium).

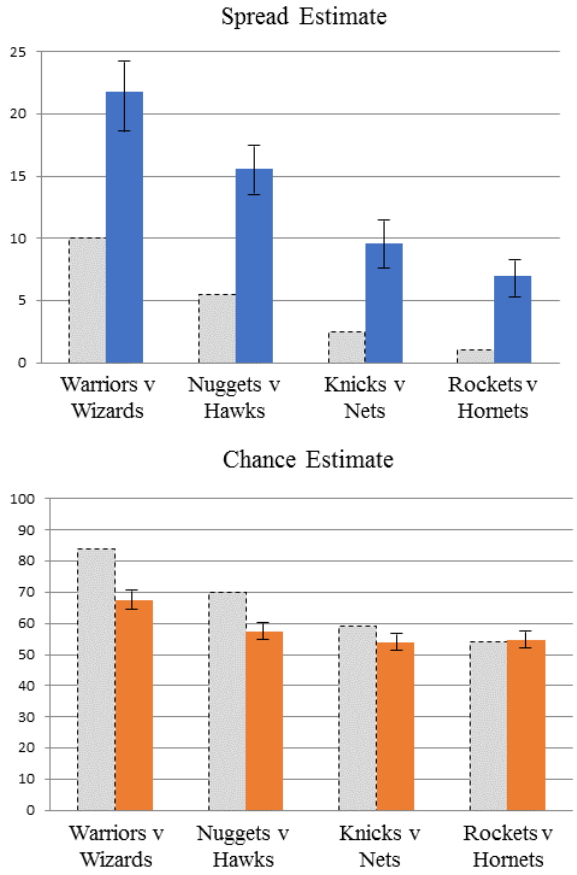
However, particularly in light of the fact that the 2016 US Presidential election outcome was not well-predicted, we cannot rule out the possibility that, when averaged, lay people's projections were in fact more accurate than the forecasts. Therefore, it is important to test the generality of the findings across contexts. Accordingly, we tested people's inferences in the context of NBA basketball games for the predicted chances of winning and the predicted margin of winning (point-spread).

Method

In Study 3 (N = 221 US AMT, pre-registered AsPredicted #6391), we used chance and margin of winning (e.g., point spread) forecasts from *fivethirtyeight.com* for four basketball games that took place in October of 2017, shortly before the games were played. The games ranged between distant predicted outcomes (Warriors vs. Hornets, 84% vs. 16%, 10 points spread) to narrower predictions (Rockets vs. Hornets, 54% vs. 46%, 1 point) with two other games in between (Nuggets vs. Hawks, 71% vs. 29%, 5.5 points; Knicks vs. Nets, 59% vs. 41%, 2.5 points).

Each participant either saw the predicted chances of winning for both teams in each game and estimated the point-spreads, or saw the point-spread, and estimated the predicted chances of winning. As in Study 1, participants were instructed to estimate what the website predicted (rather than their own beliefs).

Figure 3: NBA Game Forecasts



Legend: Gray bars represent actual projections, error bars=95% CI

Results

Participants who saw the chances of winning over-estimated the point-spread ($p < .001$ for all four games, top pane of Fig. 2). In contrast, participants who saw the point-spread significantly under-estimated the forecasted chance of winning for three of the four games ($p<.001$ for three games, $p = .54$ for the lowest win probability game, Rockets vs. Hornets; bottom panel of Figure 3).

Study 4: Impact of Forecast Format on Betting

In Study 4 (N = 418 US AMT, pre-registered AsPredicted #6762), we test whether the impact of forecast format on people's beliefs about NBA games, demonstrated in Study 3, will impact their willingness to bet on an upcoming game. In particular, the results of Study 3 indicated that the effect of presentation format on beliefs was weaker for games predicted to be close. Thus, we compare the impact of forecasts on betting choices for games forecasted to have close vs. distant scores.

Method

We showed participants one of two actual upcoming NBA games. In the close game, one team was weakly favored to win (Heat 56% chance vs. 44% chance Wizards, point-spread of 1.5). In the distant game, one team was strongly favored to win (Thunder 91% chance vs. 9% chance Bulls, point-spread of 14). Participants saw the team names and either the point-spread or the chance prediction for both teams.

Participants were told that five lottery winners would receive a $5 bonus each. They could choose to bet as much of the $5 as they wished on whichever team they preferred, and keep the remainder, should they win. Their bets would pay off $3 for each $1 bet on the winning team, but they would lose the money bet if the team they selected lost. Participants then allocated the $5 among three options: bet on one team, bet on the other team or keep and not bet.

Results

In the weakly favored case (Heat vs. Wizards), when both prediction formats conveyed the closeness of the game, the information format made no difference for how much more they bet on the team predicted to win than the team predicted to lose (Ms = +$0.28 after seeing the chances vs. +$0.11 after seeing the point-spread, $t(211)=.60$, $p = .55$). However, in the other condition, where one team was strongly favored (Thunder vs. Bulls), participants bet more on the favored than unfavored team moreso when shown the chance information rather than the margin information (Ms = +$2.39 after seeing the chances vs. +$1.62 after seeing the point-spread, $t(203)=2.37$, $p = .02$). In a regression model, the difference in impact of forecast type on betting for close vs. distant games was significant (interaction between team pair and presentation format, $\beta=.95$, $t(414)=2.19$, $p = .03$).

Study 5: Accuracy of Estimates Based on Forecast Format in a Statistical Scenario

The results of Studied 3 and 4 generalize the effect of forecast format to the domain of sports outcomes. Combined with the election findings in Studies 1 and 2, these results suggest that people treat ostensibly equivalent forecasts very differently, responding to chance forecasts as if they represent a reality with more extreme projected differences, compared to margin forecasts.

However, this interpretation is still contingent on the assumption that the full outcome distribution from which the chance and margin forecasts are generated is correct in form. If this is not the case, then the forecasts may not actually be equivalent. Furthermore, in both the voting and sports contexts, the shape of the full distribution may not be knowable to participants, conditional on their knowledge about the electorate and about NBA games.

In the next study, we test the difference between chance and margin forecasts on people's judgments in a statistical scenario. In this scenario, we provide people with sufficient objective information to enable a Bayesian to accurately generate the exact margin forecast when presented with a chance forecast, and vice versa.

Method

In Study 5, participants (N = 197 US AMT, pre-registered AsPredicted #6686) read a statistical scenario involving a jar containing 99 marbles. The marbles were an unknown mix of red and green marbles only, with the number of red marbles chosen at random from a uniform distribution. Four samples were generated, by drawing 20 marbles from the jar with replacement, and the number of red marbles were recorded for each sample. The samples were not observed by the participants. However, a statistician accurately determined, based on each sample, both (1) the chance that there were more red than green marbles in the jar and (2) the expected margin (i.e. how many more red marbles than green marbles were expected).

Participants were assigned to one of two conditions. In one condition, they were shown the chance prediction (probability of more red than green marbles) for each of the four samples, and estimated the corresponding margin predictions. In the other condition, participants were shown the margin prediction (how many more red than green marbles) for each of the four samples, and estimated the corresponding chance predictions. Participants were paid for the accuracy of their estimates, with a maximum accuracy bonus of $1 per person.

Derivation of Chance and Margin Forecasts

To calculate the forecasts, we apply Bayes Theorem. Let $n \sim U[0,N]$ be the unknown number of red marbles in a jar (N-n marbles are green, for a total of N). A sample of S marbles are drawn with replacement, and there are s red marbles in the sample. We can think of the expected number of red marbles as the expected value of the posterior distribution. We can think of the probability of a majority of marbles being red, conditional on the sample outcome s, as the mass of the posterior distribution at or above N/2, after accounting for the information in the sample.

Thus, we want to compute $E(n \mid s = i)$ and $P(n \geq \frac{N}{2} \mid s = i)$, where i is a given outcome from the sample. We use Bayes' Rule:

$$P(n = j \mid s = i) = \frac{P(s = i \mid n = j)P(n = j)}{P(s = i)}$$

Since n is drawn from a uniform distribution, $P(n=j)=\frac{1}{N+1}$ for all integer values of j in [0,N]. $P(s=i \mid n=j)$ is defined by the binomial distribution: $P(s=i \mid n=j)=\binom{S}{i}\left(\frac{j}{N}\right)^{i}\left(1-\frac{j}{N}\right)^{S-i}$. Lastly, $P(s=i)=\sum_{k=0}^{N} P(s=i \mid j=k)P(j=k)$.

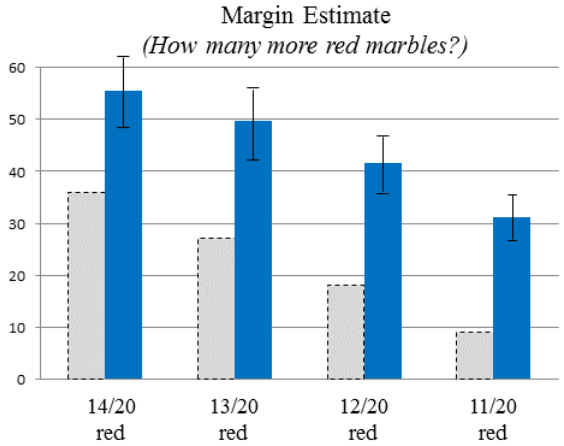
The expected number of red marbles is therefore given by $E(n \mid s=i)=\sum_{r=0}^{N} r * P(n=r \mid s=i)$. The probability that at least half of the marbles are red is $P(n \geq \frac{N}{2} \mid s=i)=\sum_{r=N/2}^{N} P(n=r \mid s=i)$.

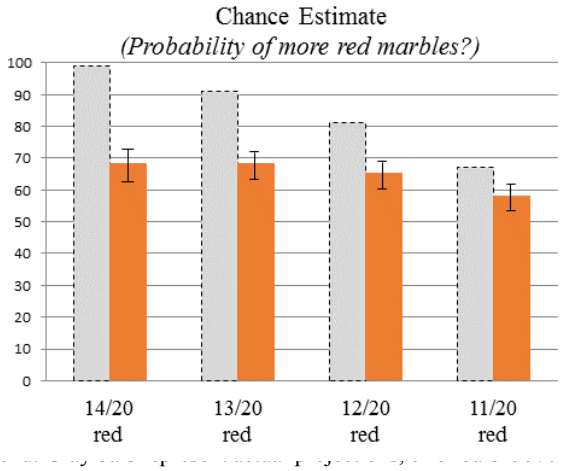
We computed the relevant quantities for N=99 and S=20 in R, and the stimuli shown to participants were based on sample outcomes of 11 through 14 out of 20 marbles being red (Table 1).

Table 1

Number of red marbles *i* in sample	11	12	13	14
Expected number of red marbles, E(n | s=i)	54.0	58.5	63.0	67.5
Expected margin	9	18	27	36
Probability that most marbles are red, P(n>N/2 | s=i)	.6682	.8084	.9055	.9609

Figure 4: Statistical Forecasts





Leg... I

Results

As shown in Figure 4, participants who were presented with the margin predictions significantly underestimated the chances of more marbles being red for all four scenarios (all ps < .001). Conversely, participants shown the chance predictions significantly overestimated the margin prediction of how many more marbles were red in all four scenarios (all *p*s < .001).

These results confirm that providing more precise information, which would allow a true Bayesian to make completely accurate judgments, does not eliminate people's incompatible estimates based on chance vs. margin forecast formats.

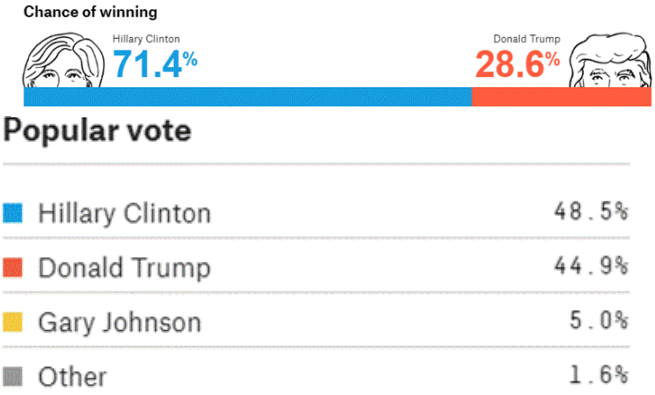
General Discussion

Overall, the results demonstrate that equivalent forecasts are seen as more extreme when framed as predicting chances rather than as predicting margins. These findings suggest that the supposedly irrelevant choice of format for a forecast can have a meaningful impact on people's interpretation of the forecast and can even shift attitudes and behaviors.

These findings are difficult to reconcile with the view of people as skilled intuitive statisticians who are adept at Bayesian reasoning. In Bayesian terms, the findings suggest a systematic discrepancy in beliefs about the expected value and tail-mass of the posterior distribution, particularly when considering the mass of the more extreme ends of the tails. It may be that the flaws in Bayesian reasoning about chances and margins identified in this research reflects an active reliance on a mis-specified distribution, conditional on which people's chance and margin beliefs are in fact consistent. Alternatively, it may be that people use different heuristics when evaluating and interpreting chances and margins. It would be useful for future research to investigate the underlying causes of the systematic discrepancy documented in this paper.

A widespread tendency to misinterpret chance and margin forecasts, as suggested by our results, might have important implications for decisions made based on forecasts, and ultimately for how to effectively communicate forecasts. It is notable that some sources, such as *fivethirtyeight.com*, provide both kinds of forecasts (Figure 5). Doing so may reduce the potential for bias, assuming that people attend to both sources of information and incorporate both in their subjective beliefs, rather than primarily focusing on and evaluating only one of the cues (Shen & Urminsky 2013).

Figure 5: Communicating Forecasts



However, in many settings, only one type of forecast may be widely available. For example, political gerrymandering (creating districts to optimize one party's electoral chances) is difficult to establish and define. Policy and media discussions of gerrymandering tend not to quantify the probabilities of a party winning a gerrymandered district. Instead both media discussions (Ingraham 2015) and proposed tests (McGhee 2014) of gerrymandering tend to focus on the more margin-like and easily quantified issue of relative vote shares.

If people are capable intuitive statisticians, this distinction should not matter, and people would be likely to draw the same conclusions about the consequences of gerrymandering whether presented with forecasts of the probability that a party wins a gerrymandered district or the predicted vote share in that district.

On the other hand, if people draw very different conclusions from the two types of forecasts, as our results suggest, their understanding of the implications of gerrymandered districts may be significantly biased by the way in which forecasts are presented. When newspaper readers learn that a gerrymandered district has a 60%-40% split between voters of two political parties, for example, they may see the district as more competitive than it is, underestimating the chances that the dominant party will win the election. As a result they may erroneously conclude that gerrymandering has less of an impact on election outcomes than it does, potentially with important consequences for their support of policies intended to curb gerrymandering.

References

Bowers, J. S., & Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin*, 138(3), 389.

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological Science*, 17(9), 767-773.

Griffiths, T. L., Chater, N., Norris, D., & Pouget, A. (2012). How the Bayesians got their beliefs (and what those beliefs actually are): Comment on Bowers and Davis (2012). *Psychological Bulletin*, 138(3), 415-422.

Hillygus, D. S. (2011). The evolution of election polling in the United States. *Public Opinion Quarterly*, 75(5), 962-981.

Ingraham, C. (2015). This is the best explanation of gerrymandering you will ever see. *The Washington Post*.

Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, 34(04), 169-188.

Kahneman, D., & Tversky, A. (1972) Subjective probability: A judgment of representativeness. *Cognitive Psychology* 3 (3), 430-454.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *TRENDS in Neurosciences*, 27(12), 712-719.

Lieder, F., Griffiths, T., & Goodman, N. (2012). Burn-in, bias, and the rationality of anchoring. In *Advances in Neural Information Processing Systems* (pp. 2690-2798).

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, 9(11), 1432-1438.

McGhee, E. (2014) Measuring partisan bias in single-member district electoral systems, 39 *Legislative Studies Quarterly*. 55, 68.

Orhun, Y. and O. Urminsky (2012) Conditional projection: How own evaluations impact beliefs about others whose choices are known," *Journal of Marketing Research*, 50(1), 111-124.

Peterson, C. R., DuCharme, W. M. & Edwards, W. (1968) Sampling distributions and probability revisions. *Journal of Experimental Psychology* 76 (2): 236.

Pouget, A., Beck, J. M., Ma, W. J., & Latham, P. E. (2013). Probabilistic brains: knowns and unknowns. *Nature Neuroscience*, 16(9), 1170-1178.

Shen, L., & Urminsky, O. (2013). Making sense of nonsense: The visual salience of units determines sensitivity to magnitude. *Psychological Science*, 24(3), 297-304.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124-1131.

Urminsky, O. (2014). Misestimating probability distributions of repeated events. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, 36 (36).

Wheeler, G., & Beach, L. B. (1968) Subjective sampling distributions and conservatism. *Organizational Behavior and Human Performance* 3 (1), 36-46.