A Better Way to Forecast

Uriel Haran

Ben-Gurion University of the Negev

Don A. Moore

University of California at Berkeley

Article Reference:
Haran, U., & Moore, D. A. (2014) A better way to forecast. *California Management Review, 57*(1), 5-15.
doi: 10.1525/cmr.2014.57.1.5.

The final, published version of the article can be accessed here:
http://journals.sagepub.com/doi/pdf/10.1525/cmr.2014.57.1.5

Abstract: Every business decision depends on making a forecast of the consequences of the decision. Although most organizations do forecasting, most do so badly. They ask either for a point prediction—a single "best guess" forecast, when everyone knows that this is an oversimplification of the truth, or for a simple range forecast, which is likely to result in biased predictions more often than not. In this article, we propose a better approach, one that takes seriously the uncertainty in forecasting and the most common errors in the way people think about this uncertainty.

A Better Way to Forecast

Every decision depends, to some degree, on a forecast. Big decisions, such as how many people to hire or how many factories to build, depend fundamentally on what the demand for your product will be in the future. Even small decisions, like what route to take to work, depend on forecasts, such as what traffic will be like. Yet companies routinely make basic forecasting mistakes that can have expensive consequences. The most elementary mistake most companies make is asking their people the wrong forecasting questions.

The most common approach to forecasting relies on a point prediction: an attempt to guess precisely what the future will hold. This is a bit like asking your doctor how long you will live. It is not worth trying to predict, down to the second, how long you will live. Just so, it is not worth trying to predict, down to the last vehicle, what demand for a new model of car will be next year. Yet companies typically ask their people to make forecasts as a point prediction.

A slightly better approach is to ask for a range of plausible estimates. A range forecast leaves some margin of error, by asking for the range within which the most probable outcomes lie. While this type of question produces more meaningful answers than does a point prediction, it sets the forecaster up for failure, by exploiting people's cognitive shortcomings. This often results in biased, overconfident estimates that are wrong more often than they are right. In this paper we confront the shortcomings of both point predictions and range forecasts. We then propose a new way of asking for forecasts—the SPIES approach—that offers a variety of advantages over these traditional methods.

The Point Prediction

The most common way organizations forecast is by asking their people for a point prediction: a single "best guess" of what will happen—of how many units they will sell, how high the price will go, or how long it will take to complete the project. This approach has two obvious shortcomings. First, research shows that asking this single question maximizes overconfidence, exacerbating the natural human tendency to believe too strongly that you have seen the truth and know what is going to happen. If you ask people how sure they are that the truth will fall close to their point prediction, they will give you wildly overconfident estimates.[i] In our studies, people routinely claim to be 80% sure the truth will fall within 5% of their best-guess forecast, but it actually happens less than 20% of the time. When companies commit to a future strategy based on an excessive faith that they know what that future looks like, they will fail to prepare for other possibilities. This can make it difficult for them to change course when they realize, belatedly, that they were heading in the wrong direction.

The other problem with just asking for a point prediction is that it ignores the fact that what we are forecasting is not a predetermined quantity.[ii] The outcome depends on future contingencies that are not yet known—including political events, the broader economic conditions, and even the weather. These things can be difficult enough to predict that it makes sense to think of them as random—and treat the future as if it will be drawn from a distribution of possible futures. Similarly, a roll of the dice will produce one specific outcome, but when predicting what will happen it is silly to claim with confidence that you know how the dice will come up. Instead, it makes much more sense to talk about a distribution of possible outcomes, think about their respective probabilities, and try to make wise bets given the full probability distribution.

Although there is uncertainty about any future event, some events are easier to predict than others. When we receive a forecast, we need to know how uncertain it is. A point prediction does not allow forecasters to express their level of uncertainty in the estimate. Estimating a volatile future value is a fundamentally different task than making a forecast in stable market conditions.[iii] However, a point prediction can look the same in both cases; the one "best guess" does not reflect the forecaster's uncertainty. Knowing this uncertainty is crucial for the decision maker, who can determine from it whether she should try to collect more information before choosing a course of action, or how much caution to exercise when making this choice. For example, when the CFO's best guess for the company's next quarter's growth will be 2%, does it mean the she thinks they can go as high as 3%? Can they go as low as 1%? Would it be a surprise if sales declined, relative to last year? Knowing the probabilities attached to these outcomes is crucial for any decision that depends on the forecast.

Asking for a point estimate creates obvious problems with some types of probability distributions. What's the best point-estimate of the number of Americans who will be infected with anthrax next year? The most likely number (the mode of the probability distribution) is low—probably zero. Does this mean the government should invest zero dollars on preparing vaccines? Probably not, as long as there is some chance that the number of people affected will be in the hundreds, or even thousands. But what is this chance? A point prediction does not tell us that.

Having a sense of what the probability distribution looks like is essential for making decisions like how many anthrax vaccines to produce. Even the most elementary production and inventory models require that companies understand the probability distribution. That allows them to better calculate the costs and benefits of producing too much versus too little, when combined with crucial considerations such as the carrying cost of inventory or the profitability of sales.

Some managers try to infer a possible range from a set of point predictions. Unfortunately, this technique can lead to false conclusions. A point prediction simply contains too little information. A decision maker cannot assume that uncertainty in the ultimate outcome will be reflected in differing forecasts from different people. If all forecasters hold the same probability distribution, they may all provide the same point prediction yet all have little confidence in it. For instance, if a pregnant woman asks ten different obstetricians when she will deliver, they will all provide the same due date, approximately 38 weeks from the date of conception. And while this may be the single most likely outcome, the probability that a particular baby is born on its due date is only about 5%.

A second challenge in building probability distributions around point predictions arises from asymmetric distributions, in which the forecaster's best guess is not centered between the confidence interval's low and high bounds. For example, when a music recording executive is trying to predict sales of a new song by a new artist, the best guess would be low, since big hits are rare. But the reason to bet on the artist in the first place is the possibility that the song will become a big hit. Plotting the distribution of music sales produces a large number of songs with low sales and a few hits. Therefore, it is likely that the forecaster's confidence interval will not be centered on the most likely value (the mode) of the distribution. For all these reasons and more, forecasts are substantially more useful when they contain more information than a single point prediction.

The Range Forecast

How many times were you advised to "expect the best and prepare for the worst"? The decision maker who uses forecasts as a major source of information upon which to base her decisions wants to be prepared for more than one exact outcome. Point predictions do not tell us what it is we should prepare for. Rather, a decision maker needs multiple estimates, representing a number of different scenarios that might happen; let's start with two – a reasonable "best case" and a reasonable "worst case." Note that we do not mean the absolute best and worst scenarios one could imagine. It would be nice to be dealt a royal flush in poker, but it would be overly optimistic to devise your betting strategy based on such an expectation. The most fantastic or catastrophic outcomes, while intriguing, may be exceedingly unlikely. It is more useful, then, to focus the decision maker's attention on the more probable outcomes for which she should prepare. This brings us to the other popular mode of forecasting – the confidence interval. A confidence interval is the range of probable outcomes that lie between the reasonable best and worst cases. Each morning we know when we should think about heading to work if we want to get in early, and when we should start rushing if we do not want to be late. Any time between these two is inside our confidence interval for getting to work at an appropriate time.

When forecasting a particular value, such as what time we will get to work, we can make that forecast more precise at a potential cost to accuracy. Forecasting with 100% certainty that you will be at work by 7:30, and scheduling an important meeting with your boss for that time, might be reckless if there is a chance you won't leave the house when you need to. Therefore, this range is more useful when accompanied by a confidence level, a percentage that indicates how confident the forecaster is that the true outcome will fall within the bounds of the interval. The more precise estimates include a narrower range of values and lower confidence, whereas the more conservative ones display a higher confidence level, but include more possibilities.

The tradeoff inherent in making a range forecast

Confidence intervals do provide more useful information for the decision maker than do point predictions, but they still leave much to be desired. Confidence intervals require a trade-off between accuracy and informativeness. If the only goal were accuracy, meaning that the interval will include the actual value, the best way to achieve it is a forecast of infinite size. What will profits be next quarter? Somewhere between infinitely negative and infinitely positive. This range would certainly include the true value, but it would not be the least bit helpful for deciding whether to invest. To be more useful, a range must be focused. It should help the decision maker understand which outcomes are likely to happen and which are not. The narrower the range of the estimate, the more informative it is; but when it gets too narrow, it becomes more likely to miss the true value. Hence, a forecaster must make a tradeoff between being as informative as she can be (by narrowing the range included in the interval estimate) and being accurate by increasing the likelihood that the confidence interval will include the true value (by making it wider). This tradeoff is often difficult, if not impossible, to resolve. A confidence interval that is too wide is not very helpful for the decision maker to decide which course of action to take. But an estimate that misses the true answer is also obviously problematic.

Research consistently finds that forecasters set intervals too narrowly.[iv] If people accurately determine their ranges, then 90% confidence intervals should include the correct answer 9 out of 10 times, 80% confidence intervals should hit the mark 8 out of 10 times and so on. Unfortunately, this is not the case, and the difference between expectation and reality is sometimes shocking: 90% confidence intervals tend to include the correct answer only about half of the time, while the average hit-rate of 80% confidence intervals is typically around 35%!

Why are we so bad at producing accurate and well-calibrated confidence intervals? Research has identified two reasons. One is the tendency to focus excessively on the most salient options, or the ones we think are the most probable, and ignore all other options.[v] While one possible outcome may be more likely than any other, it may not always be more likely than all other outcomes combined. Consider, for example, the question of who will win the next World Series. Many baseball fans, especially in New York, start the season expecting the Yankees to be the next champions. After all, the Yankees are the richest team in the league, have more all-stars than any other team and have the most World Series appearances (and wins) in the last 20 years. So if you have to bet on one team, it would not be unwise to put your money on the Yankees. But are the Yankees more likely to win a championship than all other teams *combined*? Probably not: while the World Series has been won by the Yankees 5 times in the last 20 full seasons, it was won by a team other than the Yankees 15 times. So while the Yankees are a better bet than any other team, betting that they will *not* be the champions is still smarter than betting that they will. The trouble is people are not good in combining the likelihoods of all other events and adjusting their estimates accordingly.[vi]

Another reason for the poor performance of confidence interval estimates is that even smart people err in creating them. When someone says they are 90% sure of something, what do they actually mean? According to probability theory, 90% confidence that an event will occur means that if there existed 100 parallel universes, the event in question would occur in 90 of them, or that if the current conditions for this event could be replicated 100 times, the event would occur in 90 of these replications. Grasping this notion is not intuitive, and many studies have indeed found that human intuition operates according to principles other than those prescribed by probability theory.[vii] In a series of experiments we conducted, many of our participants reported a certain level of confidence, e.g., 80%, in a prediction, but then preferred to take a bet with a lower chance of winning, e.g., 67%, over betting on their prediction. If someone's stated level of confidence does not match the probability of occurrence they report, then how should we interpret their forecasts, and what does this tell us about how to make them better?

Improving forecast ranges

Research on forecasting finds that people's statistical intuitions can be highly accurate, but their judgments are drastically influenced by the methods they use.[viii] Several researchers have tested different ways to improve confidence intervals. Jack Soll and Joshua Klayman suggested the forecaster ask herself two separate questions – which value is so high that the true value will most likely fall below it and which value is so low that the true value will most likely fall above it? For example, an interval estimate of 80% confidence should stretch between a value that is 90% likely to be below the true outcome and a value that is 90% likely to be above it. This way, they argued, the forecaster's focus shifts from her best guess to a separate assessment of best-case and a worst-case scenario.[ix]

A curious finding has emerged from forecasting research. A number of studies have found that, while people are consistently overconfident when estimating value ranges for a given level of confidence, this overconfidence largely disappears when performing the reverse task – estimating confidence in the accuracy of ranges.[x] For example, you ask someone for a 90% confidence interval for the high temperature a month out, they will give you a range that misses the mark more than half the time. But if you take that same range of temperatures and ask, "How likely is it that the high temperature a month from today will fall inside this range?" the chance that they will estimate for this range to include the actual temperature will probably be substantially lower than 90%.[xi] This difference is called *format dependence*, and it basically means that matching an interval estimate for an uncertain quantity with its predicted likelihood of including this quantity produces different results, depending on the format of the question.[xii] This insight leads to the following promising idea for improving the accuracy of confidence intervals: First, ask the forecaster for the high and low bounds of a plausible range, without specifying a probability. Next, ask them to estimate the chances that the outcome falls inside the range.[xiii]

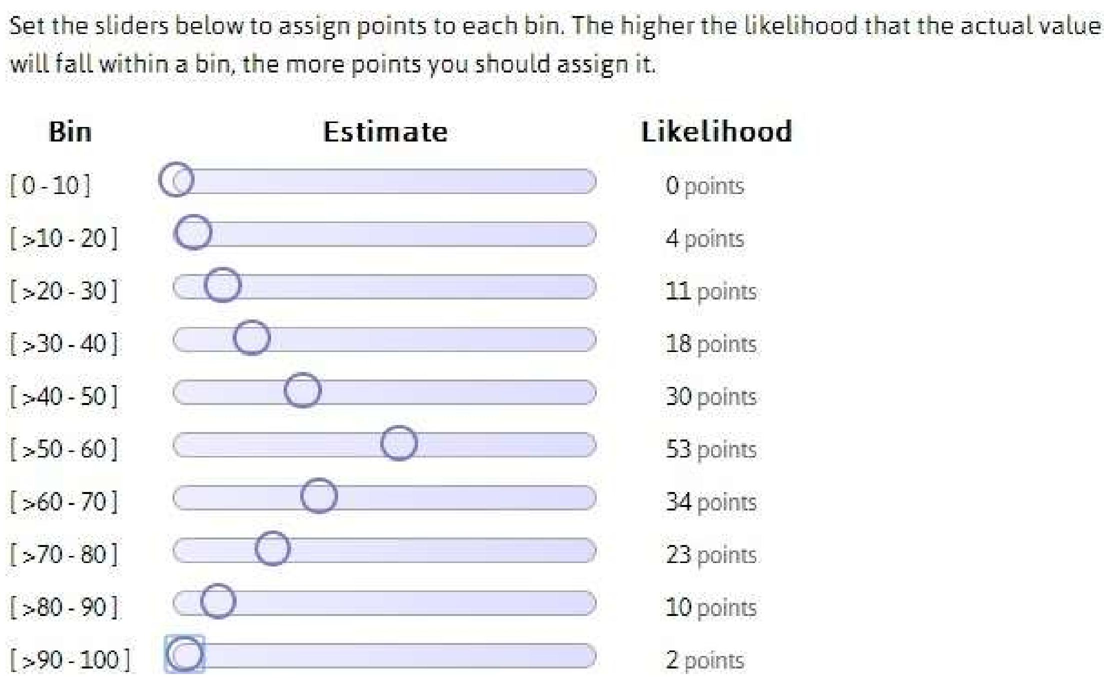
This is a step in the right direction for improving confidence intervals. The problem is that likelihood estimates are not as useful for decision makers as are value estimates. A product manager might be interested in the chance that demand for the product will be between 15,000 and 20,000 units this quarter, but she probably will be more interested to know what the demand is likely to be. But perhaps these likelihood judgments can be transformed somehow to interval estimates. This is the question which we attempted to answer in our research. And we think we may have found the answer.

Introducing the SPIES Method – Subjective Probability Interval Estimates

Based on the prior findings outlined above, we have developed a novel approach to eliciting forecasts. This method, called SPIES – short for Subjective Probability Interval EStimates, provides a simple way to produce forecasts and predictions that offer both better accuracy and greater flexibility for the decision maker.

The foundation of the SPIES method consists of a number of principles: first, forecasting should be as simple as possible. Second, the forecaster should take into account all possible outcomes; This prevents some scenarios from simply being overlooked. Also, a graphical presentation of the entire distribution prior to the estimate has been found useful in improving accuracy.[xiv] Third, building on the findings of format dependence research, likelihood judgments should be generated by the forecaster rather than built into the question. Finally, the likelihoods of the possible outcomes should be weighed relative to one another, rather than generated as precise probabilities. This will free the forecasters from calculating their estimates so that they sum up to exactly 100% and minimize the biasing effects of thinking in probabilities.[xv] Based on these principles, the SPIES method includes a graphical interface, which presents the forecaster with the full range of possible outcomes. This range is divided into a number of bins. If one forecasts a candidate's vote share in an upcoming election, then the forecaster sees a range that extends from zero to 100, and this range can be divided into bins of 1-10%, 11-20%, 21-30% and so on. See Figure 1. For each one of these bins, forecasters estimate the likelihood that the true outcome will fall inside it. For simplicity's sake, they do this by allocating points to each bin. The more probable the forecaster thinks a bin is to include the true outcome, the more points she assigns it. An attentive reader would realize immediately that these points are akin to probabilities. But unlike probabilities, points are not constrained to equal 100 or to represent an exact likelihood. The absolute number of points given to a single bin is unimportant, as long as bins get points in proportion to their likelihood. Of course, not every bin must be assigned points; the forecaster can think that the true outcome falling in a certain bin is impossible and give that bin zero points. To see a SPIES forecasting tool in action, visit http://fbm.bgu.ac.il/lab/spies/spies.html.

Figure 1. A SPIES elicitation showing 10 bins and points assigned them reflecting their rated likelihood.



By rating all the bins, a forecaster reveals his or her subjective probability distribution. This probability distribution is rich in information. It can tell us which value range the forecaster thinks is most probable, how uncertain the forecaster believes the estimate is, which outcomes we should expect and which we can invest fewer resources in preparing for. But SPIES can also produce confidence intervals. By combining the estimated probabilities of the bins, and with a few additional assumptions, we can use a SPIES report to construct any confidence interval the decision maker may find useful. This provides great flexibility to the decision maker, which other forecasting formats have trouble with. Suppose a product manager requests an interval estimate of the next quarter's demand for her product. She is initially interested in a safe estimate, with a confidence level of 90%, but this estimate turns out to be too broad and not informative enough. The manager can then compute a more focused 75% or a 50% confidence interval using the same SPIES report. There is no need to ask the forecaster to redo the forecast.

Having a sense of the full probability distribution is essential for planning purposes. For one thing, it allows firms to navigate the twin risks of producing too much vs. producing too little. In order to decide how to balance these risks, the costs of excess inventory must be compared with the costs of unfulfilled demand (including lost business and angry customers). If your inventory is costly to produce, the product is highly perishable, and profit margins are thin, then you might decide to err on the side of producing less and risking stockouts. On the other hand, if you have plenty of warehouse space, overstock lasts indefinitely, and marginal profits are high, then it might be worth producing enough that stockouts are unlikely. Understanding the full probability distribution is essential for planning how best to handle these two opposite risks.

In addition to the flexibility it offers, SPIES is simply a better method for producing accurate forecasts. A growing body of research documents tests of the quality of the confidence intervals computed from SPIES, comparing them with intervals produced in the traditional method. The SPIES method outperformed traditional confidence intervals every time. For example, in one study we asked participants to estimate the temperature in Washington DC in a month. While people who forecasted with 90% confidence intervals were correct only 29% of the time, SPIES produced 90% confidence intervals that hit the true answer at a nearly 74% rate. In another study, participants estimated years of historical events using either confidence intervals or SPIES. Once again, SPIES improved accuracy over the traditional method, from 54% to 77%.[xvi]

More recent research has also found that forecasting with SPIES can lead to better decisions. In a simulated supply chain and order decision experiment, participants who used SPIES to forecast future product demand achieved 3.1% higher profits than those who made their forecasts using 90% confidence intervals.[xvii]

The SPIES method is easy to implement and does not require an advanced degree in decision analysis. It has been put to use successfully in some real forecasting tasks, including in our work for the Good Judgment Project forecasting geopolitical events for the Intelligence Advanced Research Projects Activity as a part of their ACE forecasting tournament. (Visit www.goodjudgmentproject.com for more information.) However, the method is not without its challenges.

Challenges with using SPIES

The leading cause of forecasting failures is our limited ability to fully integrate and accurately weight the different pieces of information we use to make a forecast. The SPIES method does nothing to fix people's flawed cognitive processes, but rather works around them to produce estimates that are more informative and more likely to be accurate. But similar to traditional forecasting methods, forecasts produced using SPIES are still made by humans and are dependent on the forecaster's assumptions and beliefs. For example, suppose a forecaster insists that the S&P 500 will not fall by more than twenty points in one day and sets the lowest bin's minimum value to twenty (implying a zero probability of any value below it). A SPIES forecast will also reflect this error. But it reduces the chances of more common, routine errors. Current research now looks at how the bins themselves should be set. The most common approach divides the range into bins of equal width. However, the forecaster herself can set the width of each bin dynamically within the range. Initial tests of this option have not yielded any improvements in forecast accuracy.

Making forecasts using SPIES is unfamiliar to most people. Probabilities, confidence intervals, and probability distributions are not, for most of us, the most intuitive ways to think about uncertainty. Completing a full SPIES probability distribution takes longer than simply making a point prediction. However, research suggests that this extra time is helpful in getting people to think about the issue and make more accurate forecasts as a result. When asked for a simple point prediction, our research shows that people can respond quickly with a forecast, and will gamely claim that they are 80%, 90%, even 100% sure that their estimate is about right. The evidence suggests that these speedy intuitive assessments, while they may be easier to come up with, and may feel more "right," do not tend to be accurate, and are associated with reckless degrees of overconfidence. SPIES is useful as a tool that helps us sidestep some of the biases to which human intuition is vulnerable, but that does usually entail more work.

Many firms have invested in inventory-management systems that are not built to work with SPIES inputs. These systems routinely take point-prediction inputs and combine them with historical data, benchmarking, or data from relevant reference classes in order to interpolate a probability distribution and compute production and purchase quantities. Sometimes these systems can do an adequate job if they have enough high-quality historical data. However, their usefulness diminishes drastically when the quality and quantity of historical information decreases. When you are considering a new product or a new market, prior data are absent and technological systems that rely on statistical analysis of prior data fall silent. In these circumstances, informed human judgment is the best option, and SPIES is a valuable tool for helping people think through the profound uncertainties associated with novel products and new markets.

Conclusion

Making accurate forecasts of a future outcome is uniquely challenging. A primary reason for this is the difficulty to foresee and consider all of the factors that might influence the future. The frailties of human judgment are exacerbated by the tendency to underestimate the extent of their missing knowledge. Even when there are crucial gaps in our understanding of how the future will unfold, we routinely fail to appreciate the imperfection in our foresight. And when we try to provide organizations with the answers they ask of their forecasters, such as exact future values, gross errors are inevitable.

Research on forecasting has looked for ways to alleviate these problems by investigating the cognitive underpinnings of these biases, developing programs for training forecasters to estimate uncertainty better, and developing new approaches that are less prone to biases. We have presented one such method. SPIES offer a significant improvement in accuracy of range forecasts, compared to the direct range estimation method. In addition to improved performance, SPIES provide managers the flexibility of changing their forecast criteria, such as the width of the confidence interval or the confidence level associated with it, midstream, without having to collect new forecasts. This way, the manager can get a more informative forecast, or get answers to different questions, all from the same information set. Forecasts made with SPIES are richer in information than are traditional forecasts. Thus, SPIES enable the forecaster to incorporate more information into the prediction, but also to convey how much she thinks she knows, and how reliable she thinks her forecast is. Understanding the forecasters' level of confidence may be the most important feature of any forecast.

---

[i] Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In Kahneman, D., Slovic, P., & Tversky, A. (Eds.) *Judgment under Uncertainty: Heuristics and Biases.* Cambridge: Cambridge University Press.

[ii] Makridakis, S., Hogarth, R. M., & Gaba, A. (2010). Why forecasts fail: What to do instead. *Sloan Management Review, 51*(2), 83-90.

[iii] Budescu, D. V., & Du, N. (2007). Does past volatility affect investors' price forecasts and confidence judgements? *International Journal of Forecasting*, *23*(3), 497-511.

[iv] Klayman, J., Soll, J., González-Vallejo, C., & Barlas, S. (1999). Overconfidence: It depends on how, what, and whom you ask. *Organizational Behavior and Human Decision Processes*, *79*(3), 216-247.

[v] Wallsten, T. S. (1981). Physician and medical student bias in evaluating diagnostic information. *Medical decision making*, *1*(2), 145-164.

[vi] Tversky, A., & Fox, C. R. (1995). Weighing risk and uncertainty. *Psychological Review*, *102*(2), 269-283.

[vii] Einhorn, H. J., & Hogarth, R. M. (1985). Ambiguity and uncertainty in probabilistic inference. *Psychological Review*, *92*(4), 433-461.

[viii] Goldstein, D. G., & Rothschild, D. (2014). Lay understanding of probability distributions. *Judgment and Decision Making, 9*(1), 1-14.

[ix] Soll, J. B., & Klayman, J. (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 299-314.

[x] Winman, A., Hansson, P., & Juslin, P. (2004). Subjective probability intervals: How to reduce overconfidence by interval evaluation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(6), 1167–1175.

[xi] Klayman, J., Soll, J. B., Juslin, P., & Winman, A. (2006). Subjective confidence and the sampling of knowledge. In Fiedler, K., & Juslin, P. (Eds.) *Information sampling and adaptive cognition*, New York: Cambridge University Press.

[xii] Juslin, P., Wennerholm, P., & Olsson, H. (1999). Format dependence in subjective probability calibration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(4), 1038-1052.

[xiii] Speirs-Bridge, A., Fidler, F., McBride, M., Flander, L., Cumming, G., & Burgman, M. (2010). Reducing overconfidence in the interval judgments of experts. *Risk Analysis*, *30*(3), 512-523.

[xiv] Goldstein, D. G., & Rothschild, D. (2014). Lay understanding of probability distributions. *Judgment and Decision Making, 9*(1), 1-14.

[xv] Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review 102*(4), 684-704.

[xvi] Haran, U., Moore, D. A., & Morewedge, C. K. (2010). A simple remedy for overprecision in judgment. *Judgment and Decision Making, 5*(7), 467-476;

[xvii] Ren, Y. & Croson, R. (2013). Overconfidence in newsvendor orders: An experimental study. *Management Science, 59*(11), 2502-2517.