City of Austin
WATERSHED PROTECTION

A Proposal to Train and Develop a Group of "Superforecasters" in the Environmental Resource Management Division

SR-17-03, January 2017

Abel Porras, P.E.

City of Austin
Watershed Protection Department
Environmental Resource Management Division

Abstract

Growth in the City of Austin has become a cause for concern among both city leaders and its citizens. Traffic, affordability, and quality of life are the major concerns driving stakeholders to reduce the impacts of growth. Additionally, with growth comes an increasing interaction between society and the natural environment. This interaction is complex and difficult to predict mostly due to the often indiscriminate nature of human beings, but also because of the complexity of the ecosystem itself. Nevertheless, as Austin grows, it's becoming more necessary to forecast the response of the interaction between the human agents and the natural environment. To that end, this report is proposing that a team of "superforecasters" be developed and trained in the Environmental Resource Management Division to answer some of the more complex questions surrounding the socio-ecological system.

The term "superforecaster" comes from a project sponsored by the University of Pennsylvania where researchers were attempting to determine whether certain traits in humans made them more likely to forecast events pertaining to geo-politics. Their research found that a few select people were able to assess the geo-political events better than our intelligence communities. Furthermore, the researchers found that the characteristics found in these select people are not innate, but could be developed in anyone willing to learn. This research strengthens our understanding of recent advances in cognitive psychology made by Daniel Kahneman and Amos Tversky. They found that in making decisions from ambiguous data, human beings tended to rely on intuition, which can sometimes be biased. Nevertheless, this bias can be diminished over time with some training and careful thinking. This report is proposing to use these training techniques to identify biases in our understanding of social and environmental behavior.

Introduction

The Environmental Resource Management (ERM) Division of the Watershed Protection Department is often expected to answer questions in two general realms: the first pertains to physical phenomena, while the other consists of a more social nature. For instance, ERM employs various scientists to address citizen concerns and questions ranging from groundwater and surface water quality to riparian area functionality to salamander community health and husbandry. Similarly, ERM utilizes engineers to design and analyze controls and practices to best mitigate the impact of man-made pollutants on receiving waters. These areas encompass the physical aspect of ERM's services. Additionally, ERM is also called upon to address questions in support of the physical environment that may consist of a social aspect. For example:

1. What percent of homeowners in a watershed will likely install Green Stormwater Infrastructure if given an incentive?
2. What is the likelihood that a certain educational campaign will reach citizens in a certain manner?
3. What is the likely future demographics/land use of a certain watershed given current population growth?
4. What is the likelihood that a certain political strategy will work on persuading other surrounding municipalities?
5. How do we navigate the inherent conflict within the department in managing vegetation to enhance water quality while not increasing flood risk?
6. How do we communicate with, educate, and positively engage in citizens to tolerate and even advocate for less vegetation maintenance?
7. How do we best allocate the resources needed to construct projects to support water quality?
8. How do we best allocate resources to help an endangered species?

These types of questions, which are typically asked, are difficult to answer. Part of the reason for this difficulty may be due to the wide range of topics that involve social issues, each of which may require its own specialist. Other reasons may be due to the complexity of human nature and its linkage to ecology or the ambiguity of any data that may inform or support an answer. Regardless, addressing these and other similar questions in a satisfactory and consistent manner can hold the key to ERM operating in an effective and efficient manner in the future.

In the past two decades, there has been an influx of publications (both of an academic and of a popular nature) discussing advances in the area of cognitive psychology. The lessons of these advances have impacted other fields such as finance[1], marketing[2], professional sports[3], public policy[4], and many more. The crux of this material is that human beings have overly relied on intuition to make decisions in a world becoming increasingly complex. In the past, intuition (or pattern recognition accumulated over years of experience) has played a significant role in guiding most human endeavors. For example, the decision to buy a car, invest or trade in a certain company, or hire a ball player was often made without an appropriate frame of reference

[1] The Black Swan: The Impact of the Highly Improbable (Taleb, 2010)
[2] Phishing for Phools: The Economics of Manipulation and Deception (Akerlof and Shiller, 2015)
[3] Moneyball: The Art of Winning and Unfair Game (Lewis, 2004)
[4] Nudge: Improving Decisions about Health, Wealth, and Happiness (Thaler and Sunstein, 2009)

or model. As a result, individuals making the decision would rely on superficial or shallow assessments. Without any data or detailed knowledge of the situation, the individual making this decision would fall back on some default position unaware of any implicit biases. This made those individuals prone to mistakes (such as investors in various speculative markets) or manipulative techniques (such as those practices often stereotypically associated with used-car salesmen) that would translate to a cost to that individual.

The aim of this report is to propose a solution by addressing questions in the social aspects of protecting the environment that mitigate various sources of bias in our intuition using techniques found in cognitive psychology. This solution involves training a group of employees within the ERM to address these questions in an unbiased manner. The basis of this training will include research from cognitive psychology to remove or ameliorate any biases inherent in our thinking. This group can also be retained to address questions of a more physical or experimental nature.

Background

Like the beginnings of most modern thought, questions in psychology started with the ancient Greeks. Socrates, Plato, and Aristotle all argued to what extent rationality and impulsiveness commanded human action (Gottlieb, 2000). However, it wasn't until the twentieth century that psychology developed into a proper science. Initially, many psychologists adhered to behaviorism, which postulated that all behaviors are reflexes produced by a response to certain stimuli or that individual's history. The behaviorism approach was later criticized and modified with a more sophisticated understanding of behavior to include mental processes (Mandler, 2002).

Beginning in the 1940's, the war effort in the United States brought various academics together to understand and improve the performance of pilots and soldiers. Behaviorism offered little insight into these topics. Thus, the Office of Naval Research organized research into the 1950's, which brought together various psychologists who had begun to abandon the old ideas of behaviorism (Mandler, 2002). As was the case with the physical sciences, it was through improved quantification and measurement of observations that led to the current advances in psychology. In this case, signal detection theory (Green and Swets, 1966), which measured the control of stimuli over behavior, started a major contribution to psychology.

From that advance, Daniel Kahneman and Amos Tversky were able to conduct numerous experiments into the way humans make decisions and judge risk. This work earned Tversky a McArthur "Genius" Grant and Kahneman a Nobel Prize in Economics for his work in questioning the predominant assumption in economics of rational agents in a marketplace.

Kahneman (2012) writes at length about his research with Tversky and its implications. He suggests that the human brain operates under two systems, what he calls System 1 and System 2. System 1 is quicker, automatic, and relies on past experience. System 2 is slower, but more methodical and controlled in its assessments. System 2 requires more time and energy to maintain, and as a consequence, is occasionally used. Nevertheless, both systems serve a purpose. System 1 is useful when the conclusions are likely to be correct, when there is little to no risk in jumping to conclusions, or when that jump to the conclusion saves abundant time and resources.

Looking at mental processes as a composite of these two systems can provide insight into the way humans make decisions. Because of its automatic and quick nature, an overreliance on System 1 will frequently lead to bias and inaccurate conclusions. But when checked and processed with the more methodical System 2, biases can be detected and conclusions can be adapted to control for these biases. This insight into the human brain can put into focus how ERM can apply these understandings to think about and answer local questions pertaining to specific concerns about the links between society and the state of the environment in Austin, Texas. Using the tools from cognitive psychology, a team specifically assigned to address the more ambiguous questions from the department can effectively guard against internal biases to arrive at a solution to the question.

Methods

This section discusses the insights gained from psychological research as well as methods found to mitigate bias. This section will focus on three of the ways in which the human mind can get misled per Kahneman. This will be followed by research applied later by Tetlock and Gardner (2015) to determine whether certain traits can predict whether a certain individual will be misinformed.

Associative Minds

First, humans utilize a process called *associative activation*. That is, individuals automatically (and without any effort) will associate two things together once they have been activated together, which in turn activates other ideas almost simultaneously. This process (System 1 thinking) can be efficient in breaking complicated problems down to simpler problems. This is in a sense pattern recognition and is quite useful when thinking about the different parts of a system. However, under conditions such as *cognitive strain,* where an excess of higher level thinking is involved*,* and *ego deflation*, where an individual has completed an exhaustive mental task, this associative activation will supersede attempts to check the logic of the association. When these conditions are present, this can result in an individual making selfish choices, using sexist language, or (more importantly to this report) making superficial judgements. The researchers note that these type of biases are not due to lower intelligence, but rather stem from some cognitive threshold that is being reached. Once that threshold is reached, the slower, more methodical type of thinking (System 2) is replaced by the quicker, more intuitive type of thinking. The latter is often more prone to mistakes because of the short cuts it takes in addressing the question at hand. This can be observed by presenting the following problem (Kahneman, 2012) to colleagues:

A bat and ball combined cost $1.10. The bat costs $1 more than the ball. How much does the ball cost?

The quick, intuitive answer is that the ball costs ten cents. The mind will quickly associate the bat with $1, and thus, the ball with ten cents. However, upon checking the answer, it's clear that the ball cannot cost ten cents. To address this problem and others, which are more complicated, the key would be to utilize both Systems 1 and 2. The benefits of System 1 can be realized by breaking the complicated problem down to its associative elements. However, one must also be

willing to check one's answers prior to making a final decision or assessment or refrain from addressing this problem if a cognitive threshold has been reached. The willingness to engage in System 2 is a necessary and important characteristic in mitigating the impacts of bias.

Availability

The second type of bias comes from an *availability heuristic*. This is where an individual bases the answer to a question on the amount of information that has been available to him or her. So, a person might judge that death from shark attacks are more likely than death from diabetes, if that person has been exposed to a preponderance of news on the latest shark attacks.

Stereotypes, which Kahneman calls *representativeness*, often fit into this heuristic. When representativeness is combined with the plausibility of a story, the result can be biased. Consider this request from Kahneman to evaluate the probabilities of the following occurrences:

> *A massive flood somewhere in North America next year, in which more than 1,000 people drown.*
> *An earthquake in California sometime next year, causing a flood in which more than 1,000 people drown.*

He then states that the "*California earthquake scenario is more plausible than the North American scenario, although its probability is certainly smaller. As expected, probability judgments* [from participants in a study] *were higher for the richer and more detailed scenario, contrary to logic. This is a trap for forecasters and their clients: adding detail to scenarios makes them more persuasive, but less likely to come true.*"

As a point of comparison, Kahneman asks:

> *Which alternative is more probable?*
> *Mark has hair.*
> *Mark has blond hair.*

Both questions (the flood and the hair question) have the same logical structure, but the hair question is often answered correctly. Why? Kanheman states "*the more detailed outcome* [pertaining to the hair question] *is only more detailed – it is not more plausible, or more coherent, or a better story.*" For the flood question, humans almost automatically construct a representative story or a correlation between an earthquake in California and its resulting flood, neglecting the fact that there are more floods in North America than floods from earthquakes in California. Again, it is the fast thinking of System 1 that quickly reacts to see the world as more coherent than it really is.

An appropriate manner in which to mitigate *representativeness* is by establishing a *causal base rate*. That is, the estimate of a probability of an outcome based on a plausible baseline statistic. Kahneman also warns against the propensity of humans to find a causal story to fit the data. He highlights this story:

> *You see a person reading* The New York Times *on the New York subway. Which of the following is a better bet about the reading stranger?*

She has a PhD.
She does not have a college degree.
Representativeness would tell you to bet on the PhD, but this is not necessarily wise. You should seriously consider the second alternative, because many more non-graduates than PhDs ride in New York subways.

Thus, in this story, one should start by estimating a *causal base rate* of people with PhDs riding the subway (say 10% of total ridership) and a *causal base rate* of people without a college degree riding the subway (say 50% of total ridership). To address the question of which set of people one is more likely to encounter on the subway, choose the larger set of people.

Once these base rates have been established, if one has additional information that might change the probabilities, then these base rates should be adjusted accordingly. Thus, if one believes that there is a better correlation between *New York Times* readers and PhDs than between *New York Times* readers and non-college graduates, then the probabilities can be determined by modifying the casual base rates (10% probability of the stranger being a PhD might increase to 20%) depending on the veracity/validity of the information on the correlations. Therefore, by finding a causal base rate and questioning the extent to which the data received can be verified will also assist in decision making and assessments.

Statistics

The final type of bias occurs in thinking about statistics and probability. Researchers have found that humans are not particularly adept at thinking statistically. Specifically, people, including the most quantitative of us, have a difficult time thinking about inferences about extreme events taken from a small sample size. Kahneman (2012) states the following:

- *Large samples are more precise than small samples.*
- *Small samples yield extreme results more often than large samples do.*

The first statement has a clear ring of truth, but until the second version makes intuitive sense, you have not truly understood the first.

Kahneman (2012) is pointing toward the phenomena of jumping to conclusions under little information and associating what might be an extreme event with that small sample. Consider the experiment of flipping a fair coin three times and then flipping it five times. One is more likely to arrive at three heads from three coin flips (12.5% chance) than arriving at five heads from five coin flips (3.125% chance). Even though one might think that getting three heads out of three coin flips is an extreme event, it is still probable. This is in contrast to getting five heads out of five coin flips, which is even rarer, and underscores the potential for misinterpretation of extreme events due to a small sample size. Thus, Kahneman (2012) instructs the forecaster to include higher uncertainty in their estimates for small sample sizes or data that is suspect.

These are just three examples of misjudgments made by individuals of all walks of life. Other types of errors involve forgetting how observations will often regress to the mean, or the overconfidence in one's beliefs. However, by following the heuristics above, one can make a significant impact on mitigating bias.

Superforecasting

Other researchers have applied these findings in an attempt to determine which characteristics leads people to refrain from following these biases. These findings were recounted in the book *Superforecasting: The Art and Science of Prediction* (Tetlock and Gardner, 2015). Here, researchers drew upon a random sample of people throughout the United States to estimate the likelihood of various geo-political events occurring. These events ranged from the likelihood of North Korea launching a nuclear missile in the next year to the likelihood of an increase in the interest rate in the next six months. These researchers found that the best of these forecasters were not explicitly the smartest, or the most quantitative, or the most engaged in current events. Rather, these group of people shared another set of characteristics that could be learned and acquired.

Tetlock and Gardner summarize these set of characteristics and paraphrased here. The "superforecaster" is a person who is willing to learn to:

1. Break complex problems into smaller, simpler problems. This is one of the short cuts used by System 1 in addressing questions, but can be used as an advantage to create a rough outline in breaking down the complicated problem. However, this step must be followed up with a System 2 check.
2. Find (or estimate) causal base rates for each of the simpler problems.
3. Adjust your base rates accordingly based on current or new evidence.
4. Have a nuanced view of the problem. Tetlock and Gardner (2015) advise potential superforecasters to downgrade one's overconfidence in an answer or data set. They advise to question the data, the sample size, and the model constructed. Include any uncertainty into the answer and adjust the probabilities downward accordingly.
5. Think about statistics (or uncertainty) in more than three settings (impossible, maybe, or certain). This will also help the assessors in scoring the superforecasters. One possible scale for quantifying the likelihood of an answer is to use:
    a. 0-5% rare
    b. 5-15% unlikely
    c. 15-25% possible, but not likely
    d. 25-40% possible
    e. 40-60% even odds
    f. 60-75% good chance
    g. 75-85% fairly likely
    h. 85-100% near certain
6. Look for contradicting pieces of data and/or small sample sizes, which can increase uncertainty and shift the settings accordingly.

The ability to superforecast can be improved upon by working as a team with a diverse set of viewpoints. To maximize the potential, Kahneman (2012) proposed the following rule. "Prior to the discussion of an open question, have the members write out a summary of their solution and position. This will allow a diversity of ideas to be discussed while preventing others from lining up for or against an initial position and also encourages uncorrelated errors."

Proposal

We propose to develop and train a team of "superforecasters" to answer any questions of a socio-ecological nature. We will follow the methods summarized out in Tetlock and Gardener (2015). The first step would be to call forth any willing participants from the division. After this, the participants would be given weekly (or bi-weekly) practice socio-ecological questions. The question would be in the form of a likelihood given a time constraint. Samples questions might ask "what is the likelihood that City Council will approve *x* in the next six months?" Or, "what is the likelihood that population growth in Austin will be between 2 and 3% in the next year"?

Participants would submit an answer. Then, they would meet to discuss their respective thinking and analysis with other participants, which would then be recorded. After which, the participants can adjust their answers at any time and draw upon whatever information they find on that topic until a day before the question has expired. The answers would be tracked electronically and once the answer is known (because the event either happened or it didn't happen), the participants' likelihood answers would be assessed by using the Brier scoring system. After about six months of these type of questions, the participants would be ranked and the top ten would then be selected to become "superforecasters".

At this point, real socio-ecological questions can be submitted by the management of ERM for discussion and documentation among the "superforecasters". Sample question might include "what is the likelihood that *x* education strategy would work within the next six months?" "What is the likelihood that an increase in capital expenditures for salamander habitat in the next fiscal year would be acceptable to City Council or the general public?" Or "what is the likelihood that a flood risk will occur in the next 10 years in a certain watershed due to improved riparian functionality?"

One of the key features of this program would be the documentation of the analyses by the superforecasters. Upon learning the answer, the superforecaster would be able to look back at their thinking and see where their bias was and maybe adjust accordingly for the next question. It is anticipated that the "Superforecasters" will be able to use the methods set forth by Kahnemann (2012) and Gardener and Tetlock (2015) to estimate the likelihoods of the questions. Given these estimates by the superforecaster, management can weigh these forecasts in their assessments of policy going forward.

Conclusion

Human beings have evolved to think fast using only their pattern recognition refined by years of experience. While this may have alerted humans to predators in the savannahs and forests, this type of thinking is of limited use in our modern existence of complex social interactions. This complexity is intensified when dealing with ecological systems, as dealt with by the ERM. To

address some of these socio-ecological questions, the biases in our intuitions to answer these questions should be moderated.

Current research in cognitive psychology points to three areas where our biases can adversely impact our decision making processes. The biases include association, availability, and an incapacity to statistical thinking. To combat these weaknesses in our thinking about forecasts, Kahneman (2012) suggests using a preliminary System 1 examination of the problem refined by the System 2 check, looking at causal base rates, looking for new information, and questioning the veracity of the data. Greater uncertainty about the data can lead to adjustments in the causal base rates.

The proposal to train and develop a team of superforecasters for the ERM would be useful in addressing questions of a socio-ecological nature that are inchoate, novel, ambiguous, or encumbered in uncertainty. This same team (or method) could be extended to verify research work coming from ERM or to guide proposed research work. Superforecasters could look at information from the project, such as the sample size, effect size, etc., and make some kind of likelihood determination.

Furthermore, the above proposal is just one application to the latest advances in cognitive psychology. This application has been proposed for use in ERM as a way of addressing questions at the boundary between social policy and ecology. Another application might include using that research to influence citizens for a certain social policy. Priming or nudging the citizen toward a certain stance on an issue has been discussed for some time now in the literature (Thaler and Sunstein, 2009). However, this concept presumes that ERM knows what is in the best interest of society at large and can avoid any unintended consequences.

Finally, ERM can make some of the information collected on the environment widely available to the polity at large in the hope that they would become aware of their own biases and seek to correct it themselves. One way of achieving that would be to develop and train a group of the public to become "superadvocates" (rather than "superforecasters") for water quality. The process would be similar to becoming a "superforecaster" in that they will be given weekly questions on the environment in Austin and asked to assess the likelihood of its truth. The best among the participants could then be called on to reach out to others in the City of Austin.

Despite Kahneman advising us that the research he presents is statistically valid, care must still be taken with any findings. Nevertheless, as with most science, any incremental change is useful, and organizing a team founded on these principles from this research would be a step forward in addressing a difficult problem.

References

Akerlof, G.A, and R.J. Shiller. *Phishing for Phools: The Economics of Manipulation and Deception*. Princeton University Press, 2015.

Green, D. M., and J. A. Swets. *Signal detection theory and psychophysics.* Wiley, 1966.

Gottlieb, Anthony. *The Dream of Reason: A History of Western Philosophy from the Greeks to the Renaissance*. W.W. Norton and Company, 2000.

Mandler, George. "*Origins of the Cognitive (r)evolution*." Journal of the History of the Behavioral Sciences 38 (2002).

Kahnman, Daniel. *Thinking, Fast and Slow*. Farrar, Straus and Giroux, 2013.

Lewis, Michael. *Moneyball: The Art of Winning an Unfair Game*. W.W. Norton and Company, 2004.

Taleb, Nassim Nicholas. *The Black Swan: The Impact of the Highly Improbable*. Random house, 2007.

Tetlock, P.E. and D. Gardner. *Superforecasting: The Art and Science of Prediction*. Broadway Books, 2015.

Thaler, R.H. and C.R. Sunstein. *Nudge: Improving Decisions about Health, Wealth, and Happiness.* Penguin Books, 2009.