NOTICE: This is the author's version of a work that was accepted for publication in the California Management Review. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms, may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in California Management Review, 2010; 52(4): 125-141, DOI: 10.1525/cmr.2010.52.4.125.

Prediction Markets: A New Tool for Strategic Decision-Making

Adam Borison and Gregory Hamm; April 19, 2010

Introduction

Uncertainty is at the center of many of the most important strategic decisions made by private businesses and public agencies.

In the private sector, uncertainty dominates decision-making in key areas such as R&D, product introduction, market entry, project development and technology choice. Toyota's well-known introduction of the Prius is a good example. During development and even early in the introduction process, there were a variety of commercial and technical problems. Success was by no means inevitable, but Toyota's CEO Watanabe supported the car's introduction despite this uncertainty. As he said in a 2006 Forbes interview "I did not envisage such a major success….Some thought it would grow rapidly, and others thought it would grow gradually. I was in the second camp." The leader of this effort, Takehisa Yaegashai, has been quoted as saying the likelihood of success at the time was estimated at 5%. (Slywotzky, p. 24) Of course, as it has turned out, the Prius is now regarded as a considerable market success.

In the public sector, uncertainty plays a pivotal role in many policies and regulations having far-reaching effects - ranging from climate change to health care to national security. Perhaps the most memorable and controversial recent example is the invasion of Iraq, which was justified largely based on assessments of the likelihood that Iraq possessed weapons of mass destruction. There appeared to be substantial evidence, of varying qualities, to support this view. As White House Press Secretary Ari Fleischer said in 2003 "…we have high confidence that they have weapons of mass destruction…." (Van Natta, Jr. and Johnston, 2003) It is (perhaps surprisingly) difficult to find any quantitative likelihood estimates from before the invasion. However, an interesting retrospective analysis by two military authors assigns a probability at the time of 75% that Iraq had WMD based on the evidence then available. (Wong and Roederer, 2006) Of course, subsequent events have proven otherwise.

Not all strategic decisions are as significant (or as well publicized) as these. But strategic decisions are regularly made that have substantial impacts on the decision-maker, on the decision-maker's organization, and on the organization's stakeholders. In the private sector, these stakeholders include not only employees, but customers and suppliers, stock and bond holders, and regulators. In the public sector, these stakeholders can include many, most or even all citizens. In making these decisions, responsible decision-makers seek out the best possible information to support their choices. And affected stakeholders are increasingly demanding this as well.

Some strategic decisions are largely data-driven; that is, there is a wealth of historical data on which to base assessments of the uncertainties that underlie them. Decision-makers then face the time-consuming but relatively straightforward task of collecting and interpreting this historical data. Good examples are weather and commodity prices. There is a wealth of detailed historical data in both these cases. If an important decision depends on the weather conditions at a specific location and season…such as siting a hydroelectric facility, a great deal of historical data can be brought to bear on that decision. Although not without complications, this is the easy case.

Other strategic decisions are largely judgment-driven; that is, there is little or no directly-relevant historical data on which to base assessments of the uncertainties that underlie them. Decision-makers then face the challenging task of eliciting and processing judgment – their own or the judgment of others. Good examples are technology development and social change. There may be past analogues of varying degrees, but there is no database of equivalent events. If an important decision depends on the performance of a new technology such as carbon sequestration, judgment rather than data is the focus.

Of course, this distinction between data-driven decisions and judgment-driven decisions is a bit extreme. Judgment plays a role even where there is a large amount of data. For example, the weather data noted above must be interpreted in light of judgment about climate change. And data plays a role even where judgment dominates. For example, data from other past technological innovations may be applied to the performance of carbon sequestration technology noted above. Recognizing the importance of both data and judgment, research has been conducted not only on obtaining the best data and the best judgment, but combining judgment with data formally. See for example. See Blattberg and Hoch, 1990 for an example of combining judgment and data from model results, and Lawrence et. al., 1986 and Armstrong and Collopy, 1998 for discussion of methods for combining statistical data and judgment in the context of time series forecasting.

Whether alone or in combination with data, judgment plays a critical role in most important decisions in both the public and private sector. And that role appears to be growing if one examines the issues that dominate the current business and policy environment: the cause, rate and response to climate change, the impact of globalization, the prospects for recovery from our current economic/financial crisis, the relative roles of the public and private sectors, the risks of regional conflicts, the extent of environmental legislation, or the advancement of technology.

This paper is designed to help improve the way that these important decisions are made. First, we briefly review current practice for assessing judgmental uncertainty in strategic decision-making. Then, we suggest how an emerging tool - prediction markets - can improve this practice and thereby improve the quality of decisions that affect us all.

Current Practice

How is judgmental uncertainty currently treated in strategic decision-making? Some decision-makers deal with this uncertainty informally. They rely on intuition, experience or perhaps luck rather than analyzing uncertainty in a rigorous fashion. Some decision-makers take a somewhat more rigorous, yet still qualitative approach. They rely on tools such as scenario planning, contingency planning, stress testing or sensitivity analysis that describe alternative futures but not assign relative likelihoods or probabilities.

These informal and qualitative approaches certainly have value, and many decision-makers who rely on these approaches are successful. However, it is difficult to see how they can be fully relied upon and defended when decisions depend not just on the possibility of an event, but on its probability. Of course, in the extreme, this is true of all decisions affected by uncertainty. For example, the wisdom of entering a new market generally depends on the likelihood that the entry will be successful. The wisdom of devoting substantial resources to the development of a new drug depends on the likelihood that the drug will be safe and effective. The wisdom of committing to a new technology depends on the likelihood that the technology will work economically.

Decision-makers rely increasingly on formal quantification of judgmental uncertainty to support their strategic decisions. In part, this is because the technology for collecting and processing information has gotten much better and has become much more widespread over the years, making quantification more feasible. In part, this is because evidence is accumulating that formal analysis, and quantification in particular, improves decision-making, making quantification more desirable. Russo and Schoemaker (1989) in particular describe many of the "traps" that intuitive decision-makers fall into that can be avoided with more formal analysis.

When it comes to developing quantitative estimates of judgmental uncertainty, the current state of the art is based on advances in decision theory starting in the middle of the last century . Over the past several decades, decision theory has produced two well-established techniques for this purpose. The first is to use a structured probability encoding process to assess the opinion of individual experts. The second is to use a structured expert aggregation process to combine these individual opinions. These two processes have been used in a wide range of applications both in the public and private sectors, particularly when faced with "bet the organization" decisions. In the case of probability encoding, there is even a formal manual for the process. (See Stael von Holdstein and Matheson, 1978).

Probability Encoding

Assessing judgmental uncertainty is not just a matter of asking an expert "what do you think the odds are of X?" Individual judgment may be difficult to capture formally, and typically suffers from a variety of cognitive (unintentional) and motivational (intentional) biases. For example, on the cognitive side, most people unintentionally overweight memorable recent experience in their assessments; this is called availability bias. On the motivational side, many salespeople "low ball" their assessments intentionally because they are rewarded for exceeding them. See Tversky and Kahneman (1974) for a more in-depth discussion of biases.

Probability encoding is a formal process for eliciting judgmental probabilities from individuals that was developed and codified beginning in the 1960's, largely in recognition of these biases. The formal encoding process helps ensure that the individual understands the need for the assessment and the importance of a high quality assessment. It also helps ensure that the probabilistic event being assessed is well-defined and, if necessary, that the event is structured as better-understood component pieces. Finally, it helps ensure that biases are uncovered and minimized, and that the individual will stand behind the assessment. For more information on probability encoding, see for example, Brooks and O'Leary, 1983, Spetzler and Stael von Holstein, 1975, and Wallsten and Budescu, 1983.

As described by Spetzler and Stael von Holstein, probability encoding is a five-step process:

Motivating. Rapport with the subject is established and possible motivational biases are explored.

Structuring. The structure of the uncertain quantity is defined.

Conditioning. The subject is conditioned to think fundamentally about his judgment and to avoid cognitive biases.

Encoding. The subject's judgment is quantified in probabilistic terms.

Verifying. The responses obtained in the encoding are checked for consistency.

Research continues on refinements and enhancements to the probability encoding process as outlined above. Some work focuses on the types of questions that should be asked. For example, a recent article by Abbas et. al. (2008) compares probability assessment based on fixing a value first and asking for a probability with fixing a probability first and asking for a value. Other work focuses on scoring rules based on actual event outcomes that provide incentives for honest estimates. See for example, Winkler (1996). Despite these more recent developments, the fundamentals of the probability encoding process have not changed substantially for more than twenty years.

Figure 1 shows a real example of the output of this process for a continuous variable - the future cost of a new environmental control technology. One can easily imagine how this uncertainty could play a significant role in both business strategy and public policy. In this figure, each asterisk (*) represents a response to a particular question, such as "Which is more likely – that I roll a die and get a six or that the new technology costs less than $20M? Once all the responses are collected, a cumulative probability distribution is fitted to the data and verified with the individual. Once the encoding process is complete, the resulting judgmental probability distribution serves as input into the strategic decision-making process through decision analysis, real options, simulation and the like.

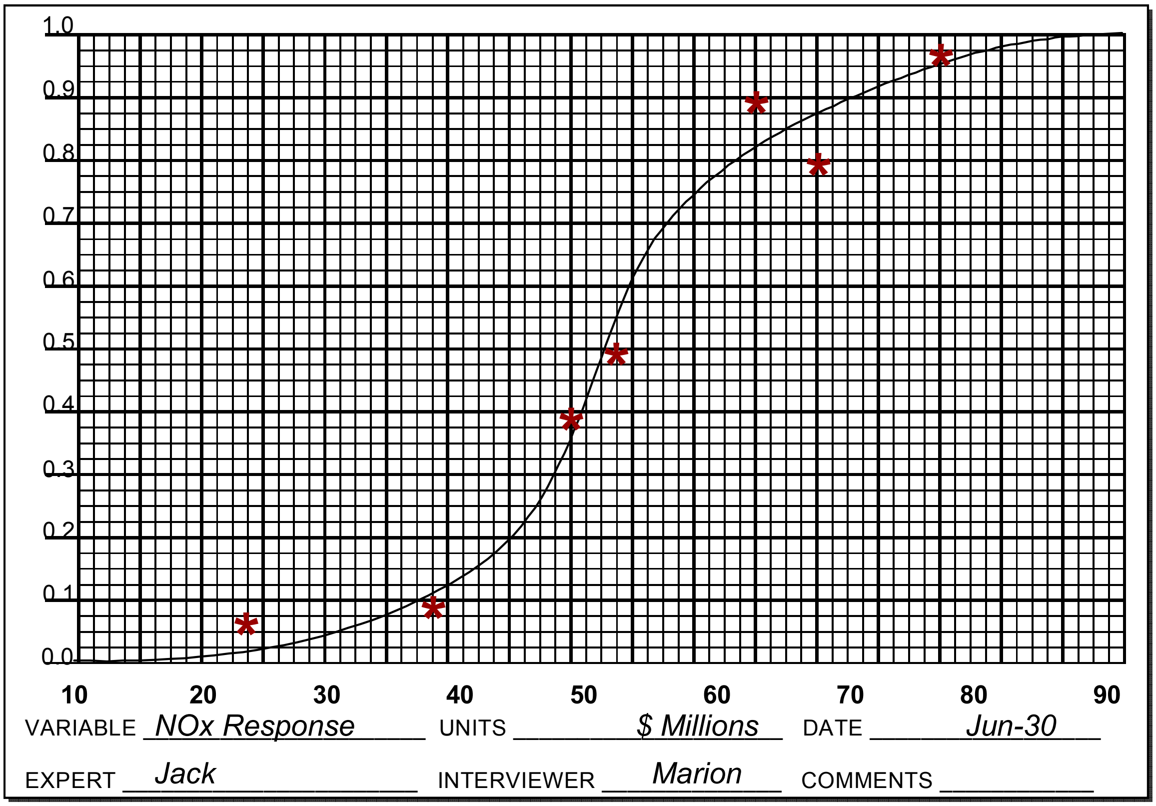

Figure 1: Encoded probability distribution on the cost of a new environmental control technology

A similar process is also used for events with discrete rather than continuous outcomes. In this context, a good example might be passage of a particular piece of environmental legislation.

Expert Aggregation

The probability encoding process outlined above is appropriate and sufficient when a single individual has been identified as the source of information on the judgmental uncertainty in question. Often however, decision-makers and other stakeholders will not want to rely on a single individual. Collecting information from multiple individuals will better reflect the true state of knowledge, and will be easier to justify and defend. Under these circumstances, the probability assessments of multiple individuals are combined using expert aggregation.

In many ways, expert aggregation is a more complex issue than probability encoding. The theory behind combining multiple assessments into a single group assessment is not straightforward analytically, and managing interactions among multiple individuals can be challenging organizationally. Consequently, several approaches have been developed for expert aggregation in different contexts. This work started largely in the 1960's and continues to this day. For more on expert aggregation, see for example, Clemen and Winkler, 1999; Osherson and Vardi, 2006; and Wallsten et. al., 1997.

As explained by Clemen and Winkler (1999), expert aggregation approaches generally fall into one of two categories – mathematical and behavioral. In the former, individual probability assessments are collected, and then combined mathematically according to a set of rules and parameters. In the latter, a group of individuals interact according to a set process, share information, and revise their probability assessments accordingly. In some cases, a consensus is achieved and a group probability assessment emerges. In other cases, no consensus is achieved and additional mathematical aggregation is required.

Again according to Clemen and Winkler (1999), mathematical approaches have a slight edge over behavioral ones. Among the mathematical approaches, a simple average of the individual probability assessments appears to perform the best or nearly so. Among behavioral approaches, the results are less clear. The well-known Delphi technique appears to perform as well but no better than other behavioral approaches.

Figure 2 shows a real example taken from Clemen and Winkler (2006) of the inputs and outputs in a mathematical expert aggregation process using a simple average. Note that the variable in question – peak ground acceleration during an earthquake – is different from Figure 1. The inputs are the probability assessments of the individuals. In this example, the assessments of three experts are shown. The output is the group probability assessment that combines the individual views. As with individual probability distributions, this aggregate distribution can be used directly to support strategic decision-making.

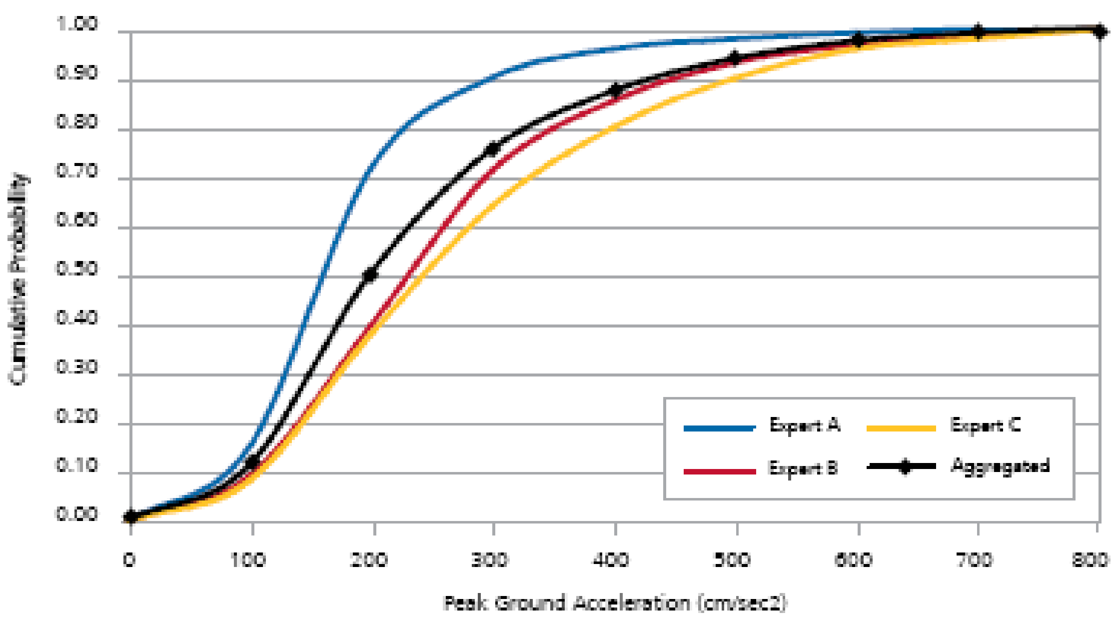

Figure 2: Individual and aggregated probability assessments

Limitations

The decision theory approach described above is effectively the state of the art for quantitative assessment of judgmental uncertainty for strategic decision-making. Despite this, the approach suffers from some important limitations.

First, any method that relies on individual judgment must address the well-documented biases that are associated with that judgment. As noted earlier, Tversky (1974) reported years ago on groundbreaking in identifying and characterizing these biases. More recently, other authors have described the foibles of individual judgment in a variety of contexts. Loftus (1981) focuses on the problems that individuals have in remembering and reporting even the most basic facts. Camerer and Johnson (1991) and Tetlock (2003) show that even so-called “experts” don’t appear to fare much better in medical and political contexts respectively. Formal probability encoding is designed to overcome many of the biases in individual judgment. However, there is really no significant theory underlying the level of quality of the resulting assessments – particularly their degree of calibration or consistency with actual event outcomes. There are no principles regarding the degree to which biases are reduced through specific encoding procedures or the extent of procedures required to achieve a specified level of quality. This lack of a strong foundation is viewed as an important shortcoming and area of improvement, even by leading academics and practitioners in the field. “…[S]ubjective probability elicitation and combination needs a firmer theoretical structure if…[it is]…to be used with confidence.” (Ferrell, p. 448)

Second, the empirical evidence - as distinct from the theoretical foundation - regarding the quality of the assessments is relatively sparse and mixed. There is certainly empirical evidence, although not as much as you might expect, that formal methods are an improvement over informal ones. For example, Browne et. al. (1997) report that formal probability encoding tools produce demonstrably better results. However, there is little evidence about the resulting quality of these assessments (as opposed to the improvement). Furthermore, much of the evidence involves very specific applications and tools, making it difficult to draw general conclusions. Roger Cooke (1991) conducted one of the few broader reviews on the quantitative assessment of judgmental uncertainty, largely in the context of science-based probabilistic risk analysis. His observations are instructive. He reports on one review of 84 empirical studies where “47 reported poor performance, while 37 reported good performance.” He notes that literature citations show a much greater prevalence of poor performance (poor citations outweigh good citations “by a factor of 6.”) but cautions against relying too heavily on these results because bad news typically dominates good news. His overall conclusion is that “[e]xpert opinions in probabilistic risk analysis have exhibited extreme spreads, have shown clustering, and have led to results with low reproducibility and poor calibration.” Despite this conclusion, he remains a “glass half full” supporter of the approach.

Third, perhaps due to the two limitations noted above, the internal credibility and external defensibility of the approach is a matter of concern. Internally, decision-makers have difficulty justifying the use of judgmental assessments. This is particularly true in contexts where decision-making is typically based on (more credible) historical frequency data. Externally, these judgmental assessments have difficulty getting widespread support among stakeholders. Speaking of delays in the use of probabilistic risk analysis (PRA) in nuclear power regulation, for example, Bier (1999) notes. “…the subjectivity of PRA results is largely responsible for the delay.” On the legal side, there has been a debate for at least a hundred years on the role of expert testimony, specifically probabilities, in both civil and criminal cases. In the late 1800's, for example, the Supreme Court of Mississippi made its feelings about subjective probability assessments quite clear (Rogers, 1891).

The courts…have gone quite far enough in subjecting the life, liberty and property of the citizen to the mere speculative *opinions* of men claiming to be *experts* in matters of science, whose confidence, in many cases, bears a direct similitude and ratio to their ignorance. We are not disposed to extend this doctrine into the field of hypothetical conjecture and probability, and to give certainty as evidence to that which, in its very nature, must be wholly uncertain and unsatisfactory.

More recently, there has been considerable study and debate on the role of probabilities and probabilistic analysis in court, including high-visibility criminal cases. See, for example, Kaye (2003). The greater the scrutiny faced by public and private decision-makers, the more important the ability to justify and defend the approach taken to quantifying judgmental uncertainty.

These limitations underlie an argument in favor of current practice that is remarkably similar to the argument in favor of democracy as a political system; that is, it is imperfect but the best we can do. This is clearly the tone taken in Morgan and Henrion (1992, p. 137). "…one can only proceed with care, simultaneously remembering that elicited expert judgments may be seriously flawed, but are often the only game in town." This situation leaves many decision-makers and stakeholders in a quandary. They would like to approach strategic decisions involving judgmental uncertainty with more rigor, but they are reluctant to rely on an approach that suffers from potentially serious limitations. This reluctance is amplified where decisions are made, reviewed or evaluated in a public context.

Prediction Markets

Until recently, the decision theory approach – probability encoding and expert aggregation – was generally the best that strategic decision-makers and stakeholders could do with respect to judgmental uncertainty - despite the limitations. However, a new approach has emerged in the past few years with the potential to put decisions based on judgmental uncertainty on a better foundation. The term associated with this approach is most commonly "prediction markets," although predictive, information and decision markets are used as well. For an excellent review of this approach, see Wolfers and Zitzewitz, 2004.

The fundamental principle behind prediction markets is to create a market where a number of individuals can place bets on a specific outcome of interest, such as a political or social event. The market is closed when the event occurs, the outcome is revealed and the participants win or lose based on that outcome. The bets establish a price for a contract tied to the outcome, and a group probability assessment of the outcome, a prediction, can be obtained directly or indirectly from this price. For example, a prediction market could include a bet on whether the President elected in 2012 will be a Democrat. To bet on this outcome, the participant buys a contract that pays one dollar if a Democrat wins the 2012 election. Until the outcome of the 2012 election is revealed, the price of this contract will be something less than (or equal to) a dollar because the most it can return is a dollar. Under suitable conditions, the current equilibrium market price of the contract, say \$0.60, represents the group assessment of the probability that the President elected in 2012 will be a Democrat, 60%. (This figure is hypothetical.)

Based on archeological discoveries, betting appears to be a human activity dating back thousands of years. Furthermore, there are examples dating back hundreds of years of organized efforts to use betting as a kind of probabilistic forecasting tool akin to a prediction market. The modern history of prediction markets, however, goes back a little over twenty years to 1988 when the University of Iowa created the Iowa Electronic Market (IEM) for betting on the outcome of the upcoming Presidential election. The IEM has continued to function and expand modestly over the years.

Since the IEM was founded, the academic and commercial interest in prediction markets has grown and the number and scope of such markets has increased dramatically. There are now many variations on the basic theme. Some markets involve real money, others involve virtual or play money. Some involve simple contracts such as the one mentioned above, others involve more complicated conditional contracts. (See Plott 2000 for a discussion of design variations and their impact on market information-gathering performance.) One very important distinction is that some markets are public and others are private.

There are currently a handful of public prediction markets. These are open to all comers, thereby truly working to take advantage of the “wisdom of crowds” or the view that such markets work best with large numbers of diverse participants (see Surowiecki, 2004). As public markets, they generally focus on political, social and economic events of broad general interest. Public markets that use real money are subject to considerable regulatory oversight…particularly in the United States, This has restricted their development and created some turmoil in this new industry. Many public market developers are motivated by the information content which they then use for professional or business reasons. Others are motivated directly by the financial rewards of the market itself. Although it is difficult to tell, participants in public markets appear to be motivated by a mixture of curiosity, pride and money.

In addition to the IEM, three good examples of public markets are Intrade, the Hollywood Stock Exchange and the CFO Prediction Market. Intrade is an Ireland-based company that is currently the leader in public prediction markets. In fact, Intrade’s tag line is “The Prediction Market.” Members wager real money, and Intrade offers a very broad range of potential bets in categories such as art, business, current events, entertainment, politics, science, weather and more. For example, it includes bets on the passage of CO2 cap and trade legislation and the confirmation of cold fusion. Like most exchanges or brokers, Intrade profits by earning a small fee on each transaction. Following up on the hypothetical example at the beginning of this section, Figure 3 shows the actual market data for the Intrade bet on the 2012 U.S. election. Based on this market, the current group probability (as of October 25, 2009) is roughly 65% that a Democrat will win the 2012 election. The Hollywood Stock Exchange has a much narrower focus. Participants use play rather than real money to bet on the success or failure of movies, and they can redeem their winnings for prizes. The owners of the Hollywood Stock Exchange sell the information to the entertainment industry. The CFO Prediction Market is run on CFO.com by CFO magazine. Naturally, it focuses on variables of interest to a financial executive audience, such as economic growth. This value to the organizer is primarily publicity. Winning participants are awarded modest prizes, although the primary motivation appears to be personal pride at being smarter than the experts.

Figure 3: Intrade "Price" for Democrat Winning 2012 Election

Private prediction markets are considerably more common than public ones. These markets are open only to members of a specific group, usually employees of a particular company or a subset of those employees with particular qualifications. As such, they focus on events of interest to that specific group, such as the completion date of an important project or the introduction of a new product by a competitor. Market developers are motivated entirely by the information provided, and this information is communicated directly to the appropriate decision-makers. Again, it is difficult to tell, but participants are typically motivated by loyalty to their group, as well as pride and curiosity. These markets involve play money and prizes; it appears that prizes are rarely a strong motivation in these private markets.

Several organizations have been quite public about their use of private prediction markets. These markets seem to be particularly popular among technology firms in Silicon Valley. According to Bo Cowgill, a product manager who oversees their prediction markets, Google has run hundreds of such markets internally (Dye, 2007). The major topics are company performance, new product performance and industry trends. Google also runs “fun” markets on sporting and other events primarily for entertainment and familiarization. BestBuy, General Electric, HP, Yahoo and many other prominent firms have also publicized their use of prediction markets. In government, the most famous (or infamous) prediction market was the one set up by the Department of Defense on political events that was aborted because of the public and Congressional reaction to the idea of betting on terrorist attacks, assassinations and the like.

Applications to Strategy

There is a great deal of popular enthusiasm over the potential for prediction markets for strategic decision-making. Recent articles in *California Management Review* (Ho and Chen, 2007), *The Economist* (2009), the *McKinsey Quarterly* (Dye, 2008) and *Risk* (Patel, 2006) are clear evidence of this. As is the Harvard Business School case study involving prediction markets at Google. There is also a small but growing body of formal research, at least at a conceptual or experimental level, on the design and implementation of prediction markets to support strategic decisions. See for example, Green et. al. (2007) for a comparison of prediction markets and the Delphi method.

As noted above, academic and commercial interest in prediction markets has grown substantially since the 1980's. Considerable research and experimentation has been conducted on these markets, and what conditions make them more or less successful in different contexts. Based on this work, three criteria appear to be most important specifically in the context of strategic decision-making.

The first criterion deals with the design of the market itself and contracts in that market. There are many potential "flavors" of prediction markets. There are different methods for determining the type and magnitude of potential bets, for matching buyers and sellers and for determining the payoffs to winners. In the context of strategic decision-making, care must be taken in this design so that contract prices can be interpreted without ambiguity as a group probability assessment, and the probability assessment is as accurate and credible as possible. Some of the simple designs that are used in existing markets, and may work adequately in those contexts, may not be suitable.

The second criterion deals with the selection of variables for the market. These variables must be defined in a way that meets two, potentially-conflicting goals. They must be clear and relevant to the participants – otherwise, the quality of the results will be poor; and they must directly fit the decision-making context – otherwise, the usefulness of the results will be limited. In existing public prediction markets, only the former consideration is paramount. In existing private prediction markets, the latter criterion has dominated and markets have sometimes suffered from lack of interest and activity as a result. A good example of this challenge involves long-term events. From a strategic decision-making point of view, an organization may be very interested in the long-term potential of a new technology or product. On the other hand, participants may be largely uninterested in betting on an event whose outcome remains unresolved for years. In this particular case, selection of a portfolio of short-term and long-term variables, along with careful market design, can help overcome this problem.

The third criterion deals with the selection of participants in the market. Many existing markets, both public and private, rely on large numbers of participants…hundreds or thousands…to provide both liquidity and diversity. For strategic decision-making, this may not be feasible or desirable. The challenge then is to find a relatively-small group of participants…as few as twenty…that is both sufficiently knowledgeable and sufficiently diverse. In practice, the implication is that markets aimed at strategic decision-making should not be open to all comers, as in existing public markets, nor should it be limited only to a particular organization, as they are in existing private markets. Participants instead should include a mix of qualified individuals with different backgrounds, experiences, perspectives and affiliations.

Our firm, National Economic Research Associates (NERA), recently established a form of prediction market we call a *Risk Information Market.* This market is based on software from

Crowdcast, a leading vendor, and includes a mix of NERA and non-NERA participants. This market is intended to generate information that clients can use for risk management and strategic planning. It is currently focused on issues of interest to the energy industry, such as nuclear and solar technology, carbon-related policy and resource prices. The design and management of this market is based on the criteria outlined above, and we are working to address the resulting challenges. Figure 4 shows the current status (as of October 25, 2009) of one market variable – the amount of nuclear capacity under construction worldwide at the end of 2011. This market is constructed specifically to generate information in the form of a probability distribution. This particular distribution shows that the consensus forecast among our participants is a little over 54000MW, and there is roughly a 1% chance that the amount will be as low as 14,000MW or as high as 95,000MW.

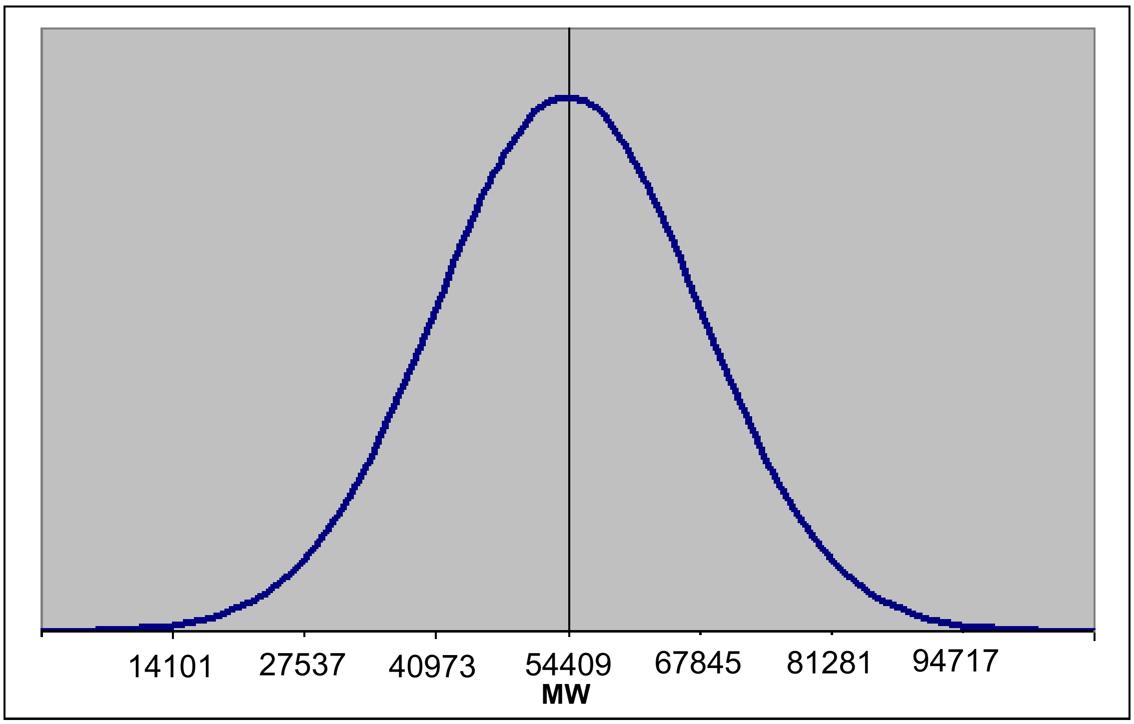

Figure 4: Nuclear capacity under construction worldwide at end of 2011 (NERA Prediction Market)

Advantages

With careful design and execution along the lines mentioned above, prediction markets have three advantages that appear to address the limitations discussed earlier regarding current practice built on decision theory. Specifically, prediction markets have advantages in their theoretical foundation, empirical evidence and credibility/defensibility.

On a theoretical level, prediction markets have three important features that distinguish them from current practice. First, they are built on real rather than hypothetical choices. In current practice, participants indicate hypothetically how they would choose or would allocate something of value if they had the choice. In prediction markets, participants make real choices and make real allocations of something of value to those choices. They “walk the walk.” In principle, this will provide a better measure of the true beliefs of each participant. Second, the relative confidence or strength of opinion of each participant is captured directly and measured explicitly in the amount of money they are willing to bet. In current practice, confidence is either not measured at all or is self-reported. In principle, this will provide a better measure of the relative knowledge of each participant. Third, prediction markets can draw on the power of “rational expectations” to argue that long-run results from such markets must not be systematically biased. Otherwise, of course, there would be opportunities for risk-free return. No such argument can be made for current practice. In principle, this will provide for higher-quality assessments over the long-run.

On an empirical level, there is a limited but growing body of evidence that prediction markets can be superior to other forecasting techniques. It is widely reported that the IEM has been more accurate that most opinion polls. (See Berg et. al., 2008) This is encouraging but leaves open the question of accuracy versus political experts using formal probability encoding procedures. Evidence that private prediction markets outperform experts at firms such as Google and HP has also been reported (Chen and Plott, 2002). This is encouraging, although it is unclear how the assessments from experts were obtained. Statistical evidence is more limited than anecdotal evidence, but is also encouraging. For example, Pennock et. al. (2001) report on the results of research on public prediction markets in a letter to *Science* indicating the assessments from these markets are very well-calibrated; that is, they are highly consistent with actual outcomes.

Third, prediction markets have important characteristics that may make them more credible and defensible thank current practice. Current practice relies heavily on one-on-one discussions with an individual or a group of individuals. Courts, regulators and the public have concerns with probabilities that emerge from such one-on-one discussions, and arguments that these assessments are “the best we can do” carry only modest weight. On the other hand, despite the current financial crisis, there appears to be considerable belief in market mechanisms and market prices for generating forecasts. Consequently, although it is too early to tell, probability assessments that emerge from prediction markets may have greater support both within and outside of organizations than assessments based on current practice.

Limitations

Of course, prediction markets suffer from their own set of limitations. The widespread application of these markets is relatively new, and there are a variety of conceptual and practical issues that have not been fully addressed at this stage.

One limitation is simply the “other side” of the market coin. While there may be many good empirical and theoretical reasons to believe in markets as good predictors, there are also good reasons to believe that markets can fail in small and large ways on this front. With respect to traditional markets, there is a growing body of work in behavioral economics that is devoted precisely to this observation. Thaler (1993) touches on many of the most important topics in this field. There is also considerable work on predictive errors in other relevant contexts, such as pari-mutuel betting at race tracks. See Thaler and Ziemba (1988). Overall, the appropriate response appears to be caution – markets typically generate useful information but they are not perfect and can stray significantly from the "truth."

A second limitation is not that markets may be subject to unintentional bias, but that they can be intentionally manipulated. Again, there is both empirical and theoretical research to support this view. In traditional markets, the incentive for manipulation is direct and substantial. The manipulators stand to make a lot of money from the market. In virtually all prediction markets, this incentive is typically absent or minimal…since the money involved is small (or even nonexistent). However, there can be other reasons for manipulation. If information from a prediction market, public or private, is going to be used for business or policy decisions, it may be to the advantage of affected organizations to manipulate that information. In the case of a private prediction market within a firm, for example, individuals or groups may have a strong incentive to bias the assessment of the release date or projected sales of a particular product. In the case of a public prediction market, one firm may have an incentive to provide false information if they believe a competitive firm may use the market information in its decision-making. While the incentive for manipulation is generally smaller than in traditional markets, the difficulty is likely to be smaller as well. Again, the appropriate response appears to be caution.

Finally, a third limitation is that these markets can be most problematic for precisely the long-term forecasting that is critical to much strategic decision-making. Prediction markets are best suited to events that are both near-term (say, resolved within a year) and discrete (say, involving only "yes" and "no" outcomes). Near-term means that participants are attentive and motivated, and that bets can be thought of as being fundamentally on the event outcome…not the direction of the market prior to learning that outcome. Discrete means that the probabilistic information can be extracted from a single bet, where the price of this contract is the consensus probability. It is more difficult to design markets that address long-term events, say the price of nuclear technology in 2020 rather than the price of solar panels in 2010. It is also more difficult to design markets for continuous variables, say the price of CO_2 allowances rather than the passage of CO_2 legislation. This represents a challenge for prediction markets in the strategic decision-making context; one that further research and experience may address.

Conclusion

It is always exciting when a "new mouse trap" comes along. And it is particularly exciting when this new mouse trap has the potential for significant positive impact. This appears to be the case with the prediction markets, and their application to strategic decision-making. While it is important not to be carried away by the latest fad and to recognize the limitations of this new tool, our own judgmental probability assessment is that prediction markets will play a major role in improving the quality of strategic decisions in the future.

References

J. Scott Armstrong and Fred Collopy, *Integration of Statistical Methods and Judgment for Time Series Forecasting: Principles from Empirical Research*, International Journal of Forecasting, 1998, pp. 69-293.

Ali E. Abbas, David V. Budescu, Hsiu-Ting Yu and Ryan Haggerty, *A Comparison of Two Probability Encoding Methods: Fixed Probability vs. Fixed Variable Values*, Decision Analysis, Volume 5, Number 4, 2008, pp. 190-202.

Joyce E. Berg, Forrest D. Nelson and Thomas A. Rietz, *Prediction market accuracy in the long run*, International Journal of Forecasting, Volume 24, 2008, pp. 285-300.

Vicki M. Bier, *Challenges to the Acceptance of Probabilistic Risk Analysis*, Risk Analysis, Volume 19, Number 4, 1999, pp. 703-710.

Robert C. Blattberg and David J. Hoch, *Database Models and Managerial Intution: 50% Model + 50% Manager*, Management Science, Volume 36, Number 8, pp. 887-899, 1990.

Daniel G. Brooks and Timothy J. O'Leary, *A Comparison of Encoding Techniques*, OMEGA: The International Journal of Management Science, Volume 11, Number 1, 1983, pp. 49-55.

Glenn J. Browne, Shawn P. Curley and P. George Benson, *Evoking Information in Probability Assessment: Knowledge Maps and Reasoning-Based Direct Questions*, Management Science, Volume 43, Number 1, 1997, pp. 1-14.

Colin F. Camerer and Eric J. Johnson, *The Process-performance Paradox in Expert Judgment: How Can Experts Know So Much and Predict So Badly?* in Towards a General Theory of Expertise: Prospects and Limits (K. A. Ericsson & J. Smith, eds), Cambridge Press, New York, 1991, pp. 195-217.

Kay-yut Chen and Charles Plott, *Information Aggregation Mechanisms: Concept, Design and Field Implementation for a Sales Forecasting Problem*, California Institute of Technology Social Science Working Paper 1131, March 2002.

Robert T, Clemen and Robert L. Winkler, *Aggregation of Expert Probabilities* in Advances: Aggregate Probabilities (Draft), 2006, chapter 9.

Robert T, Clemen and Robert L. Winkler, *Combining Probability Distributions from Experts in Risk Analysis*, Risk Analysis, Volume 19, Number 2, 1999, pp. 187-203.

Peter A. Coles, Karim R. Lakhani and Andrew McAfee, *Prediction Markets at Google*, Harvard Business School Case, May 30, 2007.

Roger Cooke, Experts in Uncertainty: Opinion and Subjective Probability in Science, Oxford University Press, 1991.

Renee Dye, *The Promise of Prediction Markets: A Roundtable*, McKinsey Quarterly, 2008, Number 2, pp. 84-93.

William R. Ferrell, *Discrete Subjective Probabilities and Decision Analysis: Elicitation, Calibration and Combination* in Subjective Probability, G. Wright and P. Ayton eds., John Wiley and Sons, 1994, pp. 411-451.

Kesten C. Green, J. Scott Armstrong and Andreas Graefe, *Methods to Forecasts from Groups: Delphi and Prediction Markets Compared*, Foresight: The International Journal of Applied Forecasting, 2007.

Teck-hua Ho and Kay-yut Chen, *New Product Blockbusters: The Magic and Science of Prediction Markets*, California Management Review, Volume 50, Number 1, 2007, pp. 144-158.

D.H. Kaye, *The Laws of Probability* in Evidence, Proof and Facts: A Book of Sources, Peter Murphy, ed., Oxford University Press, 2003, pp. 529-546.

M. J. Lawrence, R. H. Edmundson and M. J. O'Connor, *The Accuracy of Combining Judgmental and Statistical Forecasts*, Management Science, Vol. 32, No. 12, December 1986, pp. 1521-1532

Elizabeth F. Loftus, *Eyewitness Testimony: Psychological Research and Legal Thought*, Crime and Justice, Volume 3, 1981, pp. 105-151.

Granger Morgan and Max Henrion, Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis, Cambridge University Press, 1992.

Daniel Osherson and Moshe Y. Vardi, *Aggregating Disparate Estimates of Chance*, Games and Economic Behavior, Volume 56, 2006, pp. 148-173.

Navroz Patel, *Deciding on a Future*, Risk, February 2006, pp. 33-34.

David M. Pennock, Steve Lawrence, C. Lee Giles and Finn Arup Nielsen, *The Real Power of Artificial Markets*, Science, Volume 291, February 9, 2001, pp. 987-988 (letters).

Charles Plott, *Markets as Information Gathering Tools*, Southern Economic Journal, Volume 67, Number 1, 2000, pp. 1-15.

Henry Wade Rogers, The Law of Expert Testimony, Central Law Journal Company, St. Louis, MO, 1891.

J. Edward Russo and Paul J.H. Schoemaker, Decision Traps: The Ten Barriers to Brilliant Decisions and How to Overcome Them, Simon & Schuster, New York, 1989.

Adrian J. Slywotzky, The Upside: The 7 Strategies for Turning Big Threats into Growth Breakthrough, Crown Publishing Group, New York, 2007.

Carl S. Spetzler and Carl-Axel S. Stael von Holstein, *Probability Encoding in Decision Analysis*, Management Science, Volume 22, Number 3, November 1975, pp. 340-358.

Carl-Axel S. Stael von Holstein and James Matheson, A Manual for Encoding Probability Distributions, SRI International, Menlo Park, CA, 1978.

James Surowiecki, The Wisdom of Crowds: Why the Many are Smarter than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations, Doubleday, New York City, 2004.

P. E. Tetlock, Expert Political Judgment: How Good Is It? How Can We Know?, Princeton University Press, Princeton, 2003.

R.H. Thaler (ed.), Advances in Behavioral Finance, Russell Sage Foundation, New York, 1993.

R. H. Thaler and W.T. Ziemba, *Anomalies: Parimutuel Betting Markets: Racetracks and Lotteries*, Journal of Economic Perspectives, Volume 2, 1988, pp. 161-174.

A. Tversky and D. Kahneman, *Judgment under Uncertainty: Heuristics and Biases*, Science, Volume 185, Number 4157, September 27, 1974, pp. 1124-1131.

Don Van Natta, Jr. and David Johnston, *A Nation At War: Banned Weapons*, New York Times, April 14, 2003.

Thomas S. Wallsten and David V. Budescu, *Encoding Subjective Probabilities: A Psychological and Psychometric Review*, Management Science, Volume 29, Number 2, February 1983, pp. 151-173.

Thomas S. Wallsten, David V. Budescu, Ido Erev and Adele Diederich, *Evaluating and Combining Subjective Probability Estimates*, Journal of Behavioral Decision Making, Volume 10, 1997, pp. 243-268.

Robert L. Winkler, *Scoring Rules and the Evaluation of Probabilities*, TEST, Volume 5, Number 1, June 1996, pp. 1-60.

Justin Wolfers and Eric Zitzewitz, *Prediction Markets*, Journal of Economic Perspectives, Volume 18, Number 2, Spring 2004, pp. 107-126.

Maj. Ernest Y. Wong and Lt. Col. Rod Roederer, *Should the US Have Attacked Iraq? Can Decision Theory Shed Light on Polarizing Debate?*, OR/MS Today, December 2006.

-----------, *An Uncertain Future*, The Economist, February 28, 2009, p. 68.