Using Decision Sciences Concepts to Think Clearly About Policy Decisions

Ralph L. Keeney
Fuqua School of Business
Duke University

Paper prepared for
"The Irrational Economist" Conference
and Book Writing in honor of Howard Kunreuther

December 4-5, 2008

Chapter to appear in
The Irrational Economist: Future Directions in Behavioral Economics and Risk Management
E. Michel-Kerjan and P. Slovic (eds.) *Public Affairs Press.* (forthcoming 2009)

Using Decision Sciences Concepts to Think Clearly About Policy Decisions

By
Ralph L. Keeney
Fuqua School of Business
Duke University

November 2008

1. Introduction

Policy decisions are important and complex. They are important because the differences in the possible consequences of contending alternatives are often large and can affect numerous people in significant ways. They are complex for several reasons including multiple parties are involved, large uncertainties are relevant, multiple objectives are of concern so the achievement of different objectives must be balanced, and often the scientific, economic, and political substance is sophisticated. There is naturally a desire to make good policy decisions, but there is often a lack in understanding of what is good in any particular decision. Frequently, making a good decision is interpreted to mean making decisions in a rational and objective manner. But, even being rational and objective in making good decisions poses many challenges, because there is no consensus about what is rational and there are disagreements on the meaning of and/or desirability of being objective.

Thoughtful and clear analysis can provide useful insight to help policy makers make better policy decisions. However, whether or not an analysis is available for a specific policy decision, many of the concepts that are essential to critical thinking about the policy decision are either not known or not understood by those making the decisions. Many of these concepts have been well developed in the decision sciences, which consistent with the broad definition in Kleindorfer et al. (1993), I mean here to include aspects of economics, behavioral decision theory, decision analysis, negotiation analysis, and so forth.

When there is a policy analysis, there frequently are multiple analyses done or funded by groups interested in the choices in the decisions. Needless to say, the objectives in these analyses do not necessarily coincide directly with what is best for the public or for the country. This raises two key questions: (1) how can we decision scientists enhance the skills of policy makers to appraise and use any policy analysis, and, more fundamental, (2) how can we best facilitate making better policy decisions?

I believe that if the knowledge that is generally understood in the decision sciences were more broadly understood and used more frequently in making policy decisions, the quality of policy decisions would improve. Indeed, this may lead to greater improvement than providing more or more sophisticated policy analysis. My logic for this claim is the following. The only purposeful way to influence the future is by decisions and these decisions are inherently complex. Many individuals involved in policy decision processes are experts in their policy area, but do not have a conceptual framework to logically exam the overall complexity in a sound or clear manner. For this, an understanding of decision sciences concepts is essential.

The assertion above raises three key research questions:

1. What are the key decision science concepts relevant to policy decisions?
2. Why don't policy makers understand these concepts?
3. How can we educate policy makers to understand these concepts?

This paper is outlined as follows. Section 2 discusses some practical aspects of rationality and objectivity and their relevance to policy decisions. Sections 3-6 illustrate concepts of decision science that seem not to be well understood using several, mainly informal, experiments that I have conducted over the past several years. The list discussed is illustrative as there are numerous others that likely also are not well understood. Section 3 concerns structuring policy decisions, Section 4 concerns assessing, understanding, and communicating about uncertainty, Section 5 concerns articulating and understanding value judgments, and Section 6 discusses the amplification of risk. Section 7 concerns possible steps to educate policy makers about these concepts, with the hope that it would improve their understanding of the implications of policy choices and ultimately lead to making more informed and better choices.

2. Rationality and Objectivity

It is not hard to imagine that most individuals in positions to influence policy think that it is important that policy decisions be made rationally and objectively. However, there is no unique definition or prescription for rationality or objectivity. Numerous aspects of rationality have been studied in depth in economics (see for instance Hogarth and Reder, 1987; Simon, 1990). For policy analysis, I consider there to be a continuous scale of the degree of rationality applied to make a certain decision. One could think of this as practical rationality. Aspects of this include making choices consistent with the objectives articulated for the decisions and understanding how well the various alternatives measure up in terms of those objectives. It should address the inherent complexities of the problem such as uncertainties and value tradeoffs and balance the effort spent to understand the decision problem being faced, creating alternatives for that problem, and choosing the best alternative. Practical rationality would also account for the fact that the resources to make the decision are not infinite, so resources not used on one decision could be used for other useful purposes. In this sense, one would not strive for optimality, whatever that may mean, for one decision at the expense of being forced to make other decisions more poorly because of the complete lack of time and resources. Essentially, the guidelines above for practical rationality are the objectives for the decision of what process to use in making a policy decision.

There are different definitions of objectivity. One definition of objectivity is based on logic and clear thinking. A second definition is the opposite of subjectivity, meaning devoid of values and judgments. For policy analysis, or any decisions worthy of serious thought, I would like my choices to be based on logic and clear thinking. I would never want them to be void of values and judgments. Indeed, it is logic and clear thinking that demand that decisions incorporate values and judgments (Kunreuther and Slovic, 1996).

The entire foundation for a policy decision depends on values and judgments. Somehow, people must decide that there is a policy decision worthy of thought. Then they must frame that decision by articulating problem characteristics like its objectives, uncertainties, and alternatives. All of these require values and judgments. Then choices must be made about who to involve in helping make the decision, what information to gather, and how the choice should be made. Again, these aspects are based on values and judgments. At some stage if an analysis is to be done, it may be a quantitative analysis that some might refer to as objective, as it is essentially mathematics. But it should be realized that the underlying foundation for that analysis is built on values and judgments. In essence, the only reason to ever consider a policy decision is that we care about it. And that care is based on our values.

3. Structuring Policy Decisions

This and the following three sections describe some simple informal experiments that have interesting insights that indicate a lack of understanding of critical decision sciences concepts and the subsequent implications for policy decision making.

In 1987, I gave a two-hour presentation on probabilities relevant to the occurrence of earthquakes as part of a two-day workshop for earth sciences professionals at the University of California. Just before a coffee break preceding my presentation, I asked all 81 participants to fill out a questionnaire that asked the following. Is it true or false that there is a reasonable chance of a moderate to large earthquake in the San Francisco Bay area in the near future. Then I ask for separate definitions of a reasonable chance, a moderate to large earthquake, the San Francisco Bay area, and the near future. I perused the results during the coffee break and was able to begin my presentation with the following.

"You are in complete agreement on the main question: everyone indicated that the statement was true. However, there were significant discrepancies among the definitions. A reasonable chance ranged from 5 to 92 percent. Does the term moderate to large earthquake include very large earthquakes? About half of you thought so and the other half didn't. Some defined this as a Richter magnitude in the range 5-6 and others had a magnitude from 7.5 and above. Because the Richter scale is a log scale, the difference between the magnitudes of these earthquakes is 100 fold. The San Francisco Bay area was defined as within 5 miles of downtown San Francisco at one extreme and over 140 miles in all directions on any of fourteen different fault zones at the other extreme. And as everyone knows, the near future ranges from 3 weeks to 100,000 years."

The point is that if one were characterizing a policy decision about earthquakes, terminology such as presented above is not very effective for defining a policy decision. Yet the use of such general terms is often how policy issues are discussed. The decision sciences fields have long known there is a tremendous range in the meaning of qualitative terms for uncertainty such as a reasonable chance (Kent, 1964; Beyth-Marom, 1982). Hence, we might expect this to be the case for other qualitative descriptions such as moderate to large earthquake or the near future. We know that quantification clarifies the meanings of qualitative terms and reduces ambiguity in many situations and this knowledge would be useful if understood in the policy arena.

A related situation occurred in discussing the potential risks associated with an earthquake fault, called the Ghost Dance Fault, that crosses through the designated nuclear repository site at Yucca Mountain in Nevada. In the early 1990s, I was working with select group of seven seismologists who were the world's experts on that fault. In addition, each was familiar with probability assessment. A key issue was whether the fault was active. I said to the individuals, "Without communication with each other, write down your probability that the fault is active". Then, before any discussion, I asked each to write down the definition of active, which is one of the most basic words of earthquake science. I thought we might get two or maybe three definitions that varied somewhat. Instead, we got six different definitions. Some referred to what happened on the fault in the past, some concerned what might happen in the future, and others concerned what would happen in the future. I explained that such differences occur because qualitative terms like active are not precise and are ambiguous, but everyone was surprised at the range in the difference of the meanings.

As a result, one of the seismologists suggested that it would be interesting if they each wrote down their definition of the Ghost Dance Fault. This was done without conversation and to everybody's surprise, we had two Ghost Dance Faults. The problem stemmed from the fact that a north-south fault segment ran from north of the repository southward under the repository to where it was covered up by a large landslide several thousand years old. There was also a smaller fault segment that ran under that landslide from the south-west. It turned out that some of the seismologists thought both segments were part of the same fault, which they refer to as the Ghost Dance Fault. Others felt that they were separate faults and that the Ghost Dance Fault was the larger north-south segment. Clarifying that there were different characterizations of the Ghost Dance Fault greatly facilitated substantial communication about the possible likelihood and magnitude of any earthquake on the fault.

On consulting assignments over the last few decades on numerous policy decisions, I have noticed that the people in charge of recommending choices or actually making the decisions have a difficult time articulating all of their relevant objectives. More recently in a series of research projects designed to investigate whether people can identify objectives for problems of importance to themselves, we found people typically can identify only about half of their objectives and the ones that they miss are as important to them as those that they identify (Bond et al., 2008).

In a recent study to examine how to help individuals better articulate their objectives, we asked doctoral students the following question: "Suppose you have identified five potential dissertation topics and now have to choose one of them, list all objectives that are relevant for your selection". After their responses, we showed them a complete list from which they could check all the objectives that mattered. Comparing their first list of objectives and those later checked, many significant objectives were missed by a majority of the 23 participating doctoral students. The missing objectives often included the following: helps me build a coherent future research program, helps me balance my career and personal life, and addresses problems that are important (Bond et al., 2009). Essentially all of these students selected those three objectives from the master list presented at the end of the experiment and subsequently indicated they were relatively very important. Good advice to decision makers follows from a related experiment. After first-year MBA students listed their objectives for a summer internship, we informed thm that they surely had missed at least as many objectives as they listed, and then challenged them to think harder to identify more objectives. On average, the number of objectives generated per participant essentially doubled.

It is well known that people use anchoring in estimating probabilities or proportions (Tversky and Kahneman, 1974). Furthermore, they often anchor on information not directly relevant to what is being asked. Anchoring also can occur in characterizing the decision to be addressed. To illustrate this, imagine the important personal decision concerning how much may one should save to live comfortably in retirement. This will depend strongly on how much money would be needed annually to live comfortably in retirement.

Recently I asked 300 MBA students in a university program the following: "Is $x annually, in current dollars after tax, enough to support you during your retirement years? Please answer yes or no". Then I ask, "what is your best estimate of the annual amount, in current dollars after tax, that you will need to support yourself during retirement?"

The students were randomly placed into three groups, which were given either $x = $60k, $100k, or $150k. The average annual amounts needed for these three groups were $158k, $224k, and $378k. Clearly, the personal policy frame to end up with $150k per year is much different than one to end up with $300k per year. There are likely many analogous ways that various individuals with some responsibility for framing a policy problem could be influenced by anchors. Good advice to policy makers would include spend a lot of time to carefully frame the problem and have that definition contributed to by multiple individuals.

4. Assessing, Understanding, and Communicating About Uncertainty

There have been numerous experiments in the last several decades indicating that educated individuals, including those with a course in probability or statistics, have great difficulties in assessing, understanding, and communicating about uncertainties. These difficulties are exacerbated for low-probability events (Kunreuther, 2002).Thus there is every reason to suspect that people who can influence policy have these same shortcomings. The following three informal experiments give support for this assertion.

In 1988, I was a member of a working group sponsored by the US Geological Service to specify probabilities of major earthquakes occurring on segments of various California faults. The other eleven members of the committee had substantial experience and knowledge about California earthquakes. My role was to help them come up with internally consistent sets of probabilistic estimates. A major earthquake was defined as a Richter magnitude 6.5 or greater. There were five segments on the two major faults in the San Francisco area, namely the San Andreas Fault and Hayward Fault. A 45 kilometer segment of the San Andreas Fault was called the Loma Prieta segment. As the result of several meetings, the probabilities of a major earthquake over the next 30 years on each of the five segments was assessed to be 0.2 except for the Loma Prieta segment, which was assessed at 0.3 for the next 30 years (Working Group on California Earthquake Probabilities. 1988). On Oct 17, 1989, a magnitude 7.1 earthquake on the Loma Prieta segment occurred leading to significant damage and 62 deaths in the San Francisco area.

About a month after the earthquake, seismologists affiliated with the working group published an article in Science (U.S. Geological Survey Staff, 1990) that included the following statements: The earthquake "fulfilled a long-term forecast", "two decades of research…allowed an accurate long-term forecast of both the occurrence and consequences" of the earthquake, and the earthquake occurred "where it was anticipated". Obviously there were significant misunderstandings about the meaning of probabilities and their validation. First, the earthquake was less likely to occur than not in the next 30 years (probability 0.3) on the Loma Prieta segment. Second, if an earthquake occurred, it was more likely that it would occur on one of the other four segments. So how did this earthquake fulfill a long-term forecast? And why was it where it was anticipated? Indeed, the quality of probabilistic forecasts is a complex concept and accuracy cannot be determined from a single event (Winkler 1969, 1994, 1996).

In the early 1980s, I did a much less formal experiment by asking many individuals about the probability that recently elected president Ronald Reagan would be reelected in 1984. Many of these individuals were people I assumed to be somewhat knowledgeable about political processes. As President Reagan was extremely popular then, many people gave me probabilities of 95% and upward. Then I would ask them some questions about the likelihood that president Reagan would run for reelection and about the likelihood he might not be alive at election time, either for natural reasons or due to assassination. Then I would ask them about the likelihood he might be too ill to run. Finally, I would ask them about the likelihood of major political scandals that might mean he may not win the election in 1984. Often, these probabilities accounted for at least a 20% chance, which was obviously a contradiction to the individual's original estimates. The point is that I think many unstated assumptions are made, and not even realized, by the people making probabilistic estimates and by most people interpreting them. This same shortcoming, unstated and unrecognized assumptions, likely plays a role in affecting judgments in many important policy decisions.

Interpreting probabilities even when there is a significant amount of data is often perplexing. It has been shown many times that ones' intuition in such situations can often be far from accurate. A couple of years ago at a military academy, in a questionnaire prior to a seminar, I asked the following: "Suppose a rare medical condition is present in 1 of 1000 adults, but until recently there was no way to check whether you have the condition. However, a recently developed test to check for the condition is 99% accurate. You decide to take the test and the result is positive indicating that the condition is present. What is your judgment of the chance in percent that you have the condition?" Almost half of the over 300 respondents thought the chance they had the disease was 99%, approximately a quarter had an answer between 3%-15%, and the other quarter were spread thinly over the entire range from 0 to 95%. The correct answer is 9%.

An instructor of many of the students told me he had a similar problem a few weeks earlier on an exam that many of the respondents had taken. In essence, it involved a dog sniffing cars for car bombs. There was a low chance that any car had a car bomb and the dog was 98% accurate. As you might expect, a majority of the students were way off in estimating the likelihood that a car identified as one with a car bomb by the dog actually had a bomb. This problem of interpreting and manipulating data is complex and likely the source of important misinterpretations on major policy problems.

5. Articulating and Understanding Value Judgments

As discussed in section 2, values are essential to policy decisions. If we did not have any values relevant to a decision, then we should not care about it and of course we should not be spending any time thinking about it.

One class of policy decisions where values are critical, and yet greatly misunderstood, concerns policies intended to save lives. These saved lives are what we refer to as statistical lives because at the time the policy is made, one cannot know whose lives will be saved by alternative choices. There has been a tremendous amount of research on an appropriate value of a statistical life (Zeckhauser,1975; Linnerooth, 1979; Zeckhauser and Viscusi, 1990; Keeney,1990). The notion is now well understood in the decision sciences fields and by some in government, such as policy makers in the Office of Management and Budget. Yet numerous people are both disturbed and baffled by the notion of trading off economic costs versus statistical lives.

In the early 1980s, I took part of in some programs intended to educate members of the media (e.g. television networks, newspapers) about various aspects of environmental and health and safety risks. I prepared a one-page questionnaire for the participants prior to my talk. One of the earlier questions asked "In evaluating risks to lives of the public, judgments should or should not be made about the relative of importance of the lives of different individuals." The large majority of the responses of over 100 participants replied "should not". A later question asked "Do you agree or not that in allocating government funds to reduce risks to citizens, the saving of a life of any individual should be counted equally". In this case, almost all participants responded that they agree. I pointed out that counting individuals equally was a value judgment and that the responses on these two problems conflicted. This caused some consternation. Then I ask about the relative evaluation of children versus the elderly and people become a little more disturbed.

Another question asked: "if it could be uncontroversially proven that an improvement in the maintenance of air pollution control equipment on coal-fired power plants would save two statistical lives a year, what is the maximum dollar amount that should be expended for this maintenance?" Possible responses were $100k, $1million, $10 million, $100 million, and whatever it costs. The responses were rather evenly spread across these options. Many people, including many involved in setting policy, do not seem to understand that value tradeoffs between economic costs and statistical fatalities are an inherent part of numerous policy decisions.

One very informal experiment I did in the late 1980s with technical professionals taking a summer program at MIT was to ask individuals to evaluate the relative importance of a statistical death of various ages by drawing a line of that relative importance for ages 0-100. Most of the lines were basically flat up to a certain age and then gradually decreased to somewhere between 20% and 50% of the importance in earlier years. The interesting point was that, for almost all of the participants, the relative importance began to decrease at an age that was 10-15 years older than the participant who was expressing his or her values.

A major error in prioritizing objectives is that most individuals are willing to do it without understanding how much of each of the objectives they are prioritizing. For the media questionnaires mentioned earlier, I asked: "In considering the clean-up of old hazardous waste sites, rank the following in order of importance: economic cost of the clean up, potential human life lost or sickness due to the hazard, and potential damage to the natural environment." Everyone who responded to this question had economic costs as least important and every respondent but one had potential human life lost or sickness as most important. But this is essentially meaningless, as very few of these individuals seemed to think that an increase in the clean-up cost of a few billion dollars would be worth avoiding a few people being sick for a week and then recovering completely.

6. The Amplification of Risk

What is now referred to as the social amplification of risk is relevant to numerous public policy decisions. Eight years prior to the introduction to the social amplification of risk framework (Kasperson et al.,1988), I conceptualized an amplification model, although I did not use the word amplification, that separated the direct personal impacts of fatalities due to a specific cause and the induced indirect societal impact of those fatalities. Then, I constructed a value model and did assessments to preliminarily examine the importance of such amplification (Keeney, 1980). Specifically, my model was concerned with any government agency that had responsibility for public safety.

In conceptualizing the model, I stated that "there should be two major governmental concerns regarding impact of fatalities on the public. The first reflects the direct personal impacts of pain, suffering, and economic hardship. These fall most heavily on only a very small percent of the total public, those individuals who are the fatalities and their friends and relatives. The second concern involves indirect societal impacts, which include general political, social, and economic turmoil which may occur as a result of the fatalities." Furthermore, I stated "Formalization of the distinction between personal and societal impacts is appropriate for major problems, such as the public risks from power plant accidents."

The value model that I built to evaluate consequences separately included personal direct impacts and societal indirect impacts and the contributors to each of those were differentiated as to whether they pertained to fatalities due to involuntary or voluntary risks. To examine the relative importance of the direct and the indirect impacts, the model included a parameter for the relative weight given to the indirect societal impact of the first fatality due to involuntary risks relative to the direct personal impact of that fatality. In subsequent assessments for decisions that involved the possible misuse of stolen nuclear material with four knowledgeable individuals, the relative weight given to the indirect societal impact versus the direct personal impact was 10, 27, 45, and 118 respectively.

There were three important aspects of this model and the assessments. The first is that it made a clear distinction between the direct personal impacts of fatalities and the indirect societal impacts they induced in important policy areas. Second, it built an explicit value model that allowed one to incorporate these different impacts and keep them distinct from each other. Third, the assessments demonstrated that the societal impact seemed significant at least to a few people who thought hard about the situation.

7. Educating Policy Makers about Decision Science Concepts

In many government agencies, a high proportion of its staff and managers influence policy in some way. I believe that a major shortcoming that hinders quality policy making is that many individuals involved in policy making do not understand many of the concepts relevant to the complex decisions facing them. It would be useful to investigate whether this belief is true. If it turns out to be true, there are significant interesting research questions about this issue.

Research should identify the key decision science concepts useful for policy analysis. These concepts range from sunk costs to all the decision heuristics to quantifying likelihoods to indirect impacts to value tradeoffs. I think almost all of our fundamental concepts are relevant, but no doubt some of them are more useful than others.

It would be important to understand why policy makers do not seem to understand these concepts, or even possible reasons why they would not be able to understand them. There are many possible explanations for this that would be interesting to investigate. These include all of the following: the concept was never explained, poorly explained, not learned when explained, there is a lack of motivation, the concept was forgotten, there was no time to learn the concept, learning was not a priority, the individual thinks qualitatively and the concept is quantitative (or vice versa), an individual thinks the concepts are irrelevant, the individual is a deterministic thinker and the concepts are probabilistic, or human evolution is simply too slow.

There are also a series of research issues that could be addressed on how best to communicate decision science concepts to policy makers in a way they can understand and want to use in their decision making. This would involve detailed experiments about how one could best get policy makers to utilize the concepts and gain the insights that could be developed by their use. If possible, it would then be desirable if one could do some research to understand if better decisions did result from a better understanding and use of the decision science concepts.

References

Beyth-Marom, R. 1982. How Probable Is Probable? Numerical Translations of Verbal Probability Expressions. *Journal of Forecasting*, **1**, 257-69.

Bond, S.D., Carlson, K.A., and Keeney, R.L. 2008. Generating Objectives: Can Decision Makers Articulate What They Want? *Management Science*, **54**, 56-70.

Bond, S.D., Carlson, K.A., and Keeney, R.L. 2009. Generating Better Sets of Objectives for Decisions, in review.

Hogarth, R.M. and Reder, M.W.,editors. 1987. Rational Choice: The Contrast between Economics and Psychology. University of Chicago Press, Chicago.

Kasperson, R.E., Renn, O., Slovic, P., Brown, H. S., Emel, J., Goble, R., Kasperson, J.X., & Ratick, S. (1988). The Social Amplification of Risk: A Conceptual Framework. *Risk Analysis*, **8**, 177-187.

Keeney, R.L. 1980. Evaluating Alternatives Involving Potential Fatalities. *Operations Research*, **28**,188-205.

Keeney, R.L. 1990. Mortality Risks Induced by Economic Expenditures, *Risk Analysis*, **10**, 147-159.

Kent, S. 1964. Words of Estimated Probability. *Studies in Intelligence*, **8**, 49-65.

Kleindorfer, P.R., Kunreuther, H.C., and Schoemaker, P.J.H. 1993. Decision Sciences, An Integrated Perspective. Cambridge University Press, Cambridge.

Kunreuther, H.C. 2002. Risk Analysis and Risk Management in an Uncertain World. *Risk Analysis*, **22**, 655-64.

Kunreuther, H.C, and Slovic, P. 1996. Science, Values, and Risk. *Annals of American Academy of Political and Social Science*, **545**, 116-125.

Linnerooth, J. 1979. The Value of Human Life: A Review of the Model. *Economic Enquiry*, **17**, 52-74.

Simon, H.A. 1990. In Rationality in Action, P.K. Moser, Ed, Rationality in Action, Cambridge University Press, Cambridge, 189-204.

Tversky, A., and Kahneman, D. 1974. Judgment under Uncertainty: Heuristics and Biases. *Science*,**185**,1124-1131.

U.S. Geological Survey Staff, 1990. The Loma Prieta, California Earthquake: An Anticipated Event. *Science*, **247**, 286-293.

Winkler, R.L. 1969. Scoring Rules and the Evaluation of Probability Assessors. *Journal of the American Statistical Association*, **64**, 1073-1078.

Winkler, R.L. 1994. Evaluating Probabilities: Asymmetric Scoring Rules, *Management Science*, **40**, 1395-1405.

Winkler, R.L. 1996. Scoring Rules and the Evaluation of Probabilities, *Test,* **5**, 1-60.

Working Group on California Earthquake Probabilities. 1988. U.S. Geological Survey Open-File Report 88-398.

Zechhauser, R. J. 1975. Procedures for Evaluating Lives. *Public Policy*, **23**, 419-464.

Zeckhauser, R.J. and Viscusi, W.K.1990. Risk within Reason. *Science*, **248**, 559-564.