Handling Uncertainty in Engineered Systems

Azad M. Madni and A. Terry Bahill

Contents

A. M. Madni
Systems Architecting and Engineering, University of Southern California, Los Angeles, CA, USA
e-mail: azad.madni@usc.edu

A. T. Bahill (✉)
Systems and Industrial Engineering, University of Arizona, Tucson, AZ, USA
e-mail: terry@sie.arizona.edu

G. S. Metcalf et al. (eds.), *Handbook of Systems Sciences*,
https://doi.org/10.1007/978-981-13-0370-8_22-1

Abstract

This chapter identifies essential activities in the tradeoff study process that are responsible for addressing uncertainty and describes how to handle these uncertainties. These activities all involve human decision-making, which is the source of common mistakes arising from confirmation bias, severity amplifiers, and framing. This chapter explains how these activities are affected by mental mistakes such as biases, simplifying heuristics, cognitive illusions, emotions, fallacies, and psychological traps. This chapter presents examples for handling uncertainty by managing mistakes caused by uncertainty in the problem statement, the evaluation criteria, and weights of importance. It also shows how certainty factors and sensitivity analyzes can help handle uncertainty.

Keywords

Trade-off study · Trade study · Tradeoff study · Uncertainty · Systems engineering process · Decision analysis and resolution · Sensitivity analyses

Introduction

Uncertainty is being in doubt about the value of a variable, the state of a system, the status of a component, or the characteristics of a design. Some uncertainties are measurable: however, certain classes of uncertainties, such as the likelihood and time of occurrence of a future event, are not. Uncertainties are value-neutral and are not necessarily bad. Their causes can be numerous and are often not readily identifiable. Uncertainties produce risks, which are handled by risk mitigation measures that influence or shape outcomes.

Different types of uncertainty affect the design and operation of complex systems. Uncertainties can have many causes such as unobservability, volatility in the environment, rapidly changing threats, unclear needs, fluctuating demand for resources and products, changing market conditions and budget shortfall. There are many methods to cope with uncertainty. They range from rolling-up component reliabilities, to calculating system reliability, to employing physical and functional redundancies, to compensating for disruptions. Even though the need for uncertainty handling is well-recognized, the tools for handling uncertainty tend to be immature, and methods for achieving robust design continue to be in their infancy. This recognition provides the impetus for developing a consistent framework for handling uncertainty. In this chapter we will limit our discussion to handing uncertainty in the tradeoff study process.

Tradeoff Study Process

As part of overall responsibilities, a systems engineer or tradeoff analyst may have to elicit and capture the values and preferences of a customer (a decision maker) (Madni 2014). This is an important step for all stakeholders, to develop confidence in subsequent decisions. However, biases, cognitive illusions, emotions, fallacies, and the use of simplifying heuristics make the capture of decision maker preferences quite challenging (Madni et al. 2005). The timely and judicious use of tradeoff studies can help people in making rational decisions. But this approach can only work as long as tradeoff studies are not beleaguered with mental mistakes. Fortunately, lessons have been learned in how to minimize mental mistakes while creating tradeoff studies (Smith et al. 2007).

Tradeoff studies, also referred to as trade studies, are an effective method for choosing rationally among alternatives. Tradeoff studies involve computation of multiple evaluation criteria in parallel for several alternatives simultaneously (Parnell 2016). In the absence of a tradeoff study, limited attentional capacity invariably leads people to consider criteria sequentially. Tversky and Kahneman showed that humans tend to anchor on certain values before evaluating probabilities (Tversky and Kahneman 1981) or numbers (Tversky and Kahneman 1974). This tendency leads to mistakes and flawed decisions. These findings appear to be robust and are confirmed by recent research that has shown that different areas of the brain are involved in different types of decisions, and that framing and bias show neurobiological correlates (Martino De et al. 2006). Framing of hypotheses is a general problem for human decision-making (Petroski 2003). Fixation on short-term rewards is also a well-documented problem in the absence of a tradeoff study (Ainslie 2001).

Tradeoff studies are probably as old as numbers. The basic procedure of a tradeoff study is to account numerically for empirical values and preferences. When this accounting is formalized, it provides a stable base for choosing among alternatives. Aristotle (384–322 BC) developed logic supported by empirical observation that illuminated the fact that humans are capable of deliberating when making a choice among alternatives. An early description of a tradeoff study that "summed weighted scores" is given in a 1772 letter from Benjamin Franklin to Joseph Priestley. He writes that when people draw out the consideration of multiple criteria over a long period of time, their thoughts about the different criteria become separated. "To get over this" problem, he suggested looking at the pros and cons simultaneously (Bahill and Madni 2017, p. 461).

A simple, hopefully self-explanatory, example of a tradeoff study is given in Table 1. Here is the defining scenario. Imagine that while reading this chapter, you experience an irresistible urge for chocolate chip cookies and a glass of milk. You can simply state the problem as, "I want chocolate chip cookies." You begin to explore how to get hold of chocolate chip cookies to satisfy this urge. You quickly discover that there are no chocolate chip cookies in your house, but you do have yogurt. Unfortunately, yogurt won't do today. You begin to explore your options. You could head over to a bakery and buy chocolate chip cookies. But, wait! That would cut into valuable study time. You simply cannot afford to do that! How about

Table 1 Tradeoff study matrix for the Chocolate Chip Cookie Acquisition system (from Bahill and Madni 2017)

Evaluation criteria	Normalized evaluation criteria weights	Normalized sub criteria weights	Alternative 1: Do nothing, forego the cookies			Alternative 2: Ask mom to bake cookies			Alternative 3: Use Pillsbury's Chocolate Chip Cookie Dough			Alternative 4: Buy cookies at bakery		
			sc	wt × sc	wt × sc	sc	wt × sc	wt × sc	sc	wt × sc	wt × sc	sc	wt × sc	wt × sc
Audible signal for cookies are ready	0.43		0.00		0.00	0.60		0.26	0.50		0.21	0.40		0.17
Lost study time	0.33		1.00		0.33	0.80		0.27	0.50		0.17	0.30		0.10
Nutrition	0.24		1.00		0.24	0.69		0.17	0.48		0.11	0.34		0.08
Calories		0.47	1.00	0.47		0.70	0.33		0.50	0.24		0.40	0.19	
Fat, grams		0.21	1.00	0.21		0.80	0.17		0.40	0.08		0.30	0.06	
Carbohydrates, grams		0.32	1.00	0.32		0.60	0.19		0.50	0.16		0.30	0.09	
Alternative rating					**0.57**			**0.69**			**0.50**			**0.35**
Column sum	1.00	1.00												

Abbreviations: *sc* stands for score, *wt* stands for weight, × is the multiplication sign settling for having a pizza delivered? No! You don't want pizza. You want *chocolate chip cookies*. Frantically, you start rummaging the kitchen. Lo and behold, you find a tube of chocolate chip cookie dough in your refrigerator! You are going to make chocolate chip cookies! However, is that the best alternative? Perhaps you should do a tradeoff study to determine the best alternative.

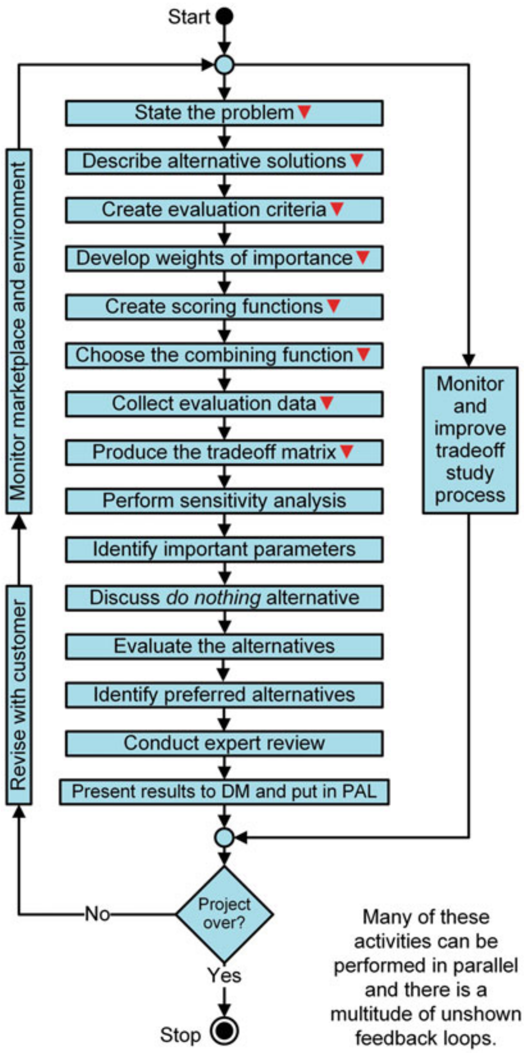
The seminal academic research findings on tradeoff studies appeared in Keeney and Raiffa (1976) and Edwards (1977). Edwards focused on the numerical determination of subjective values, an activity that the cognitive sciences have shown to be difficult. Keeney and Raiffa are best known for their axiomatic derivation of value and utility functions from conditions of preferential or utility independence. In practice, it is expensive and difficult to design and implement experiments that demonstrate these independence conditions, or measure input values among the evaluation criteria of the alternatives. Moreover, a search for alternatives usually produces complex alternatives that are so different from one another that it is difficult to invoke these elegant mathematical theorems. Haimes (2004, p. 208) notes that some discussions of optimality conditions for tradeoff approaches are "strictly theoretical, however, since it is difficult or even impossible to implement the utility function approach in practice." This is so because "it is extremely difficult or impossible to actually determine the decision maker's utility function – that is, to assign numerical utilities to the various combinations of the objectives" Haimes (2004, p. 209).

The INCOSE Systems Engineering Handbook (2004) provides a consensus of the fundamentals of tradeoff studies. Tradeoff studies are prescribed in industry for choosing and ranking alternative concepts, designs, processes, hardware and techniques. Today, tradeoff studies are broadly recognized and mandated as *the* method for choosing alternatives by simultaneously considering multiple criteria. They are the primary method for choosing among alternatives given in the Capability Maturity Model Integration (CMMI 2015) Decision Analysis and Resolution process.

We note that even if the mathematics and utility curves are done correctly, care needs to be exercised in doing a tradeoff study, because it is difficult to overcome mental mistakes. We will discuss mental mistakes in tradeoff studies and offer suggestions for ameliorating their occurrence. Systems engineers can exploit this knowledge to sharpen their decision-making, and institutions can incorporate this knowledge in their documented decision-making processes.

Figure 1 presents the Tradeoff Study Process. The very first step, State the Problem, has several sources of uncertainty. It is inevitably the case that the initial problem statement is imprecise and the trade-off space initially defined is incomplete. These sources of uncertainty need to be addressed before proceeding with tradeoff studies. Probing the statement of the problem, reformulating queries, and identifying new variables that need to be included in the trade-off space are the means for reducing uncertainty in the problem statement. Thereafter, steps are relatively straightforward. Almost all activities contribute to uncertainty, however, in Fig. 1 the *major* contributors of uncertainty have marked with a red triangle (▼).

Fig. 1 The Tradeoff Study Process. Activities marked with a red triangle (▼) are the *major* contributors of uncertainty. DM represents the Decision Maker and PAL is the Process Assets Library



SIMILAR Process

The tradeoff study process of Fig. 1 can be generalized in terms of the SIMILAR process (Bahill and Gissing 1998), which comprises seven key activities: **S**tate the problem, **I**nvestigate alternatives, **M**odel the system, **I**ntegrate, **L**aunch the system, **A**ssess performance, and **R**e-evaluate. These seven activities are conveniently summarized using the acronym SIMILAR (see Fig. 2). We use this process to provide the overall context for problem-solving during system design. At the outset, we want to

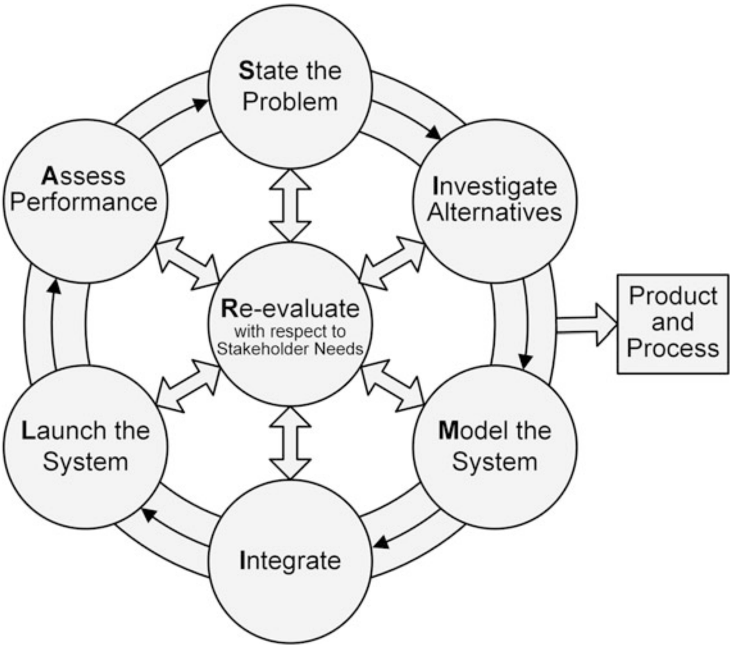


Fig. 2 The SIMILAR process

clarify that the activities in the SIMILAR process are performed iteratively and in parallel. Each activity in the SIMILAR process is described next.

State the Problem

"The beginning is the most important part of the work" Plato, *The Republic*, fourth Century B. C.).

The problem statement contains many tasks that are performed iteratively, many of which can be performed in parallel. The following tasks fit into the problem statement activity.

- Understanding customer needs is the first and foremost task.
- Identify stakeholders such as end users, operators, maintainers, suppliers, acquirers, owners, customers, bill payers, regulatory agencies, affected individuals or organizations, sponsors, manufacturers, etc.
- Where do the inputs come from? Requirements come mainly from the customer and the systems engineer. Evaluation criteria and proposed alternatives for tradeoff studies come from the design engineer. Risk events are identified and described by the risk analyst.
- Describe how the system works using stories and use case models. The use case models provide requirements and test cases.

- State the problem in terms of *what* needs to be done, not *how* it must be done. The problem statement may be in prose form or in the form of a model.
- Develop the incipient architecture.
- Initiate risk analysis. Yes, the risk analysis of the system should begin at the same time as the requirements discovery and tradeoff study processes.

Investigate Alternatives

The following tasks fit into the Investigate Alternatives activity.

- One should investigate alternative requirements, designs and risk events using evaluation criteria such as performance, cost, schedule, and risk.
- For quantitative analyses, identify attributes of requirements, evaluation criteria for tradeoff studies, and the likelihood of occurrence and severity of consequences for risk events. Assign them weights of importance to show priorities.
- Scoring (utility) functions are mandatory for tradeoff studies, but are optional for requirements and risks.
- Select evaluation methods, that is, methods for combining the data. State the combining function that will be used. Usually, this will be the Boolean AND function for requirements, the sum of weighted products for tradeoff studies and the product function for risks.
- Finally, one must collect evaluation data and use it to assign values to attributes for requirements, weights and scores for tradeoff studies, and attributes for risk analyses.

Model the System

The following tasks fit into the Model the System activity.

- Abstract models are typically created for most requirements, alternative designs, and risk events. These models are consistently elaborated (Wymore 1993) (that is, expanded) throughout the system life cycle. A variety of models can be used.
- Requirements can be modelled with use case models, textual shall statements, tables, spreadsheets and specialized databases. The requirements must be clarified, decomposed, allocated, and derived.
- Tradeoff studies are usually modelled with tradeoff matrices implemented with spreadsheets. The alternative designs within them are modelled with UML diagrams, SysML diagrams, analytic equations, computer simulations, and mental models.
- Risks are modelled with tables containing values for the likelihood of occurrence and severity of consequences and figures displaying these data.
- Everything must be prioritized. You should prioritize the requirements set to find the most important requirements. For tradeoff studies, you identify preferred

alternatives with a tradeoff matrix. You adjust the ranges for likelihood and severity for risk events in order to find the greatest risks.
- The results of a sensitivity analysis can be used to validate a model, flag unrealistic model behavior, point out important assumptions, help formulate model structure, simplify a model, suggest new experiments, guide future data collection efforts, suggest accuracy for calculating parameters, adjust numerical values of parameters, choose an operating point, allocate resources, detect critical evaluation criteria, suggest tolerance for manufacturing parts and identify cost drivers.

Integrate

The following tasks fit into the Integrate activity.

- Integration means bringing elements together so that they work as a whole to accomplish their intended purpose and deliver value. Specifically, systems, enterprises and people need to be integrated to achieve desired outcomes. To this end, interfaces need to be designed between subsystems. Subsystems are typically defined along natural boundaries in a manner that minimizes the amount of information exchanged between the subsystems. Feedback loops between individual subsystems are easier to manage than feedback loops involving densely interconnected subsystems.
- Evaluation criteria should trace to requirements. Risks should trace to requirements, particular documents, or brain storming sessions. Requirements should trace to higher-level requirements and should link to risks. Requirements and risks might link to technical performance measures (TPMs).

Launch the System

The following tasks fit into the Launch the System activity.

- System launch means either deploying and running the actual system in the operational environment, or exercising the model in a simulated environment to produce necessary outputs for evaluation. In a manufacturing environment, this might mean buying commercial-off-the-shelf hardware and software, writing code, and/or bending metal. The purpose of system launch is to provide an environment that allows the system or its model to do what it is being designed to do.
- The outputs of these processes are a requirements specification, preferred alternatives and the risk register. One should continually monitor the requirements (in the requirements database), alternative designs (in the process assets library, PAL) and risks (in the risk register) looking for possible changes and bring these to the attention of the decision makers. One should continually monitor the market place

looking for new requirements, products, designs and risks and bring these to the attention of the decision makers.

Assess Performance

The following tasks fit into the Assess Performance activity.

- Test, validation and verification are important tasks for all processes.
- There should be both regularly-scheduled and problem-initiated expert reviews. The results of these reviews are presented to the decision maker (DM) and are put in the process assets library (PAL).
- Evaluation criteria, measures, metrics and TPMs are all used to quantify system performance. Evaluation criteria are used in requirements discovery, tradeoff studies and risk analyses. Measures and metrics are used to help manage a company's processes. TPMs are used to mitigate risk during design and manufacturing.

Re-Evaluate

The distinction between an engineer and a mathematician is the use of feedback in design. For a century, engineers have used feedback to control systems and improve performance. It is one of the most fundamental engineering concepts. Re-evaluation is a continual feedback process with multiple parallel loops. Re-evaluation means observing outputs and using this information to modify the inputs, the system, the product and/or the process.

The SIMILAR Process (Fig. 2) shows the distributed nature of the Re-evaluate function in the feedback loops. However, it is important to realize that not all loops will always come into play. The loops that are used depend on the problem to be solved, and the problem context. Re-evaluation includes formal inspections, expert reviews and reviews with the customer. A very important and often neglected task in any process is monitoring and improving the process itself. This self-improvement process is shown explicitly in Fig. 1 with the Monitor and improve the tradeoff study process.

Contributors to Uncertainty

The study of human decision-making reveals that the presence of cognitive biases can never be ruled out (Smith et al. 2007; Mohanani et al. 2018). This is also the contention of the economic school of heuristics and biases, which produced Prospect Theory (Kahneman and Tversky 1979), a theory that describes how people respond to choices under risk and uncertainty. Innate human biases, and external circumstances, such as the framing or the context of a question, can compromise decisions.

It is important to note that subjects maintain a strong sense that they are acting rationally even when they are exhibiting these biases (Kahneman 2011).

In this chapter we have marked with a ▼ the actions in Fig. 1 that are the biggest contributors to uncertainty. They all deal with human decision-making rather than uncertainty in the weather, climate, solar variability, geology, political actions or interpretation of experimental data. In subsequent paragraphs we will explain how the activities of Fig. 1 are affected by uncertainty. Most reasons involve confirmation bias, severity amplifiers, and framing. Therefore, we will first discuss these three decision modifiers.

Confirmation Bias

Arguably, the most important cause of fallibility in human decision-making is confirmation bias. Humans hear what they want to hear and reject what they do not want to hear. Humans filter out information that contradicts their preconceived notions and remember things that reinforce their beliefs. Confirmation bias causes decision makers to actively seek out and assign more weight to evidence that confirms their hypotheses and ignore or underweight evidence that could disconfirm their hypotheses. For example, mothers emphasize good deeds of their children and de-emphasize their bad deeds. This is why we often hear the mother of a terrorist crying out, "My boy is innocent. He could never have killed all those people." People who think that they have perfect memory and perfect recall tend to ignore instances when they forgot something and tend to be secure in long-term memory instances when they correctly recalled events and facts. Senior citizens often believe that they are good drivers despite tests that show that they have poor vision and slow reflexes. Thirty years ago, most cigarette smokers were in denial about the hazards of smoking. Some people say, "There must be a storm coming, because my arthritic joints are hurting."

Social media is making this worse. Not only do you filter what you see and hear, but also Facebook filters what you are exposed to. They present to you things from the friends you care about. These friends are probably ideologically like you, which accentuates the filtering process.

Nickerson (1998) reported many common instances of confirmation bias. In one, the subjects were given a triplet such as (2, 4, 6) and were asked to guess the rule that was used to generate the triplet and then try to prove or disprove that rule. After each guess, they were told if they were right or wrong. For example, if the subject's mental model for the rule was "successive even numbers," they might guess (10, 12, 14) or (20, 22, 24), triplets that would confirm their mental model, but they would seldom guess (1, 3, 5) or (2, 4, 8), triplets that might disprove their mental model. He also presented another example of confirmation bias: witches.

The execution of 40,000 suspected witches in seventeenth century England is a particularly horrific case of confirmation bias functioning in an extreme way at the societal level. From the perspective of the people of the time, belief in witchcraft was perfectly natural and sorcery was widely viewed as the reason for all ills and troubles that could not otherwise be explained. In one test of a woman being a witch, the mob tied the suspect to a chair with a rope and threw her into a river. If she floated, it was proof that she was a witch and she was executed. If she sank, well, too bad.

Until the nineteenth century, physicians often did more harm than good because of confirmation bias. Virtually anything that could be dreamt up for the treatment of a disease was tried and, once tried, lasted decades or even centuries before being given up. It was, in retrospect, the most frivolous and irresponsible kind of human experimentation. They used blood-letting, purging, infusions of plant extracts and solutions of metals, and every conceivable diet including total fasting. Most of these were based on no scientific evidence. How could such ineffective measures continue for decades or centuries without their ineffectiveness being discovered? Probably, because sometimes patients got better when they were treated, sometimes they did not and sometimes they got better when they were not treated at all. Peoples' beliefs about the efficacy of specific treatments seem to have been influenced more strongly by those instances in which treatment was followed by recovery than by those instances in which there was no recovery. A tendency to focus on positive cases could explain why the discovery that diseases have a natural history and people often recover from them with or without treatment was not made until much later.

Most people react to news articles with confirmation bias. If a liberal reads a news story about a scientific study that showed how effective it was to give money to poor people, he might think, "That's interesting. I'll remember that." However, if one of those liberals reads about a new study showing that giving people money when they are unemployed just makes their lives worse, then he might start looking for flaws in the study. If a person has a long-felt belief that the income gap between the rich and the poor in the World is too large and is growing too fast, then a new study that challenges this belief might be met with hostility and resistance. However, if that person readily accepts a study that reinforces his belief, then that is confirmation bias.

People do not think like scientists: they think like lawyers. They form an opinion and then emphasize only evidence that backs up that opinion.

Severity Amplifiers

Interpersonal variability in evaluating the seriousness of a situation depends on the framing. That is, the circumstances surrounding the event will affect how a person responds to it. An evaluation may depend on factors such as how the criterion affects that person, whether that person voluntarily exposed himself to the risk, how well that person understands the alternative technologies and the severity of the results. The following are severity amplifiers: lack of control, lack of choice, lack of trust, lack of warning, lack of understanding, being manmade, newness, dreadfulness, fear, personalization, ego, recallability, availability, representativeness, vividness, uncertainty and immediacy.

Examples of these severity amplifiers include: *Lack of control*: a man may be less afraid driving his car up a steep mountain road at 55 mph than having an autonomous vehicle drive him to school at 35 mph. *Lack of choice*: we are more afraid of risks that are imposed on us than those we take by choice. *Lack of trust*: we are less afraid listening to the head of the Centers for Disease Control explain anthrax than listening to a politician explain it. *Lack of warning*: people dread earthquakes more than hurricanes, because hurricanes give hours or days of warning. People in California follow strict earthquake regulations in new construction. People in New Orleans seem to ignore the possibility of hurricanes. *Lack of understanding*: we are more afraid of ionizing radiation from a nuclear reactor than of infrared radiation from the sun. In the 1980s, engineers invented nuclear magnetic resonance imaging (NMRI). When the medical community adopted it, they renamed it magnetic resonance imaging (MRI). They dropped the adjective *nuclear* to make it sound friendlier.

Manmade: we are more afraid of nuclear power accidents than solar radiation. *Newness*: we are more afraid when a new disease (e.g. swine flu, SARs, Ebola and Zika) first shows up in our area than after it has been around a few years. *Dreadfulness*: we are more afraid of dying in an airplane crash than of dying from heart disease. *Fear:* if a friend tells you that, a six-foot rattlesnake struck at him, how long do you think the snake was? We suspect three feet. But, of course, the length of the snake is irrelevant to the harm it could cause. It is only related to the fear it might induce. *Personalization*: a risk threatening us, is worse than that same risk threatening you. *Ego:* a risk threatening our reputations is more serious than one threatening the environment. *Recallability:* if something can be readily *recalled*, it must be more important than alternatives that are not as readily recalled. We are more afraid of cancer if a friend has recently died of cancer. We are more afraid of traffic accidents if we have just observed one. Recallability is often called *availability.* Something that is readily *available* to the mind must be more important than alternatives that are not as readily available.

Representativeness: the degree to which an event is similar in essential characteristics to its parent population increases its seriousness. In a series of coin tosses THTHTH would not be representative of randomly generated coin tosses as it is too well ordered. So, it would not merit much concern. *Vividness of description*: an Edgar Allen Poe story read by Vincent Price will be scarier, than one that either of us read to you. *Ambiguity or uncertainty:* most people would rather hear their ophthalmologist say, "You have a detached retina. We will operate tonight." than, "You might have a detaching vitreous, or it could be a detaching retina, or maybe its cancer. We will do some tests and let you know the results in a week.*" Immediacy*: a famous astrophysicist was explaining a model for the life cycle of the universe. He said, "In a billion years our sun will run out of fuel and the earth will become a frozen rock." A man who was slightly dozing awoke suddenly, jumped up and excitedly exclaimed, "What did you just say?" The astrophysicist repeated, "In a billion years our sun will run out of fuel and the earth will become a frozen rock." With a sigh of relief, the disturbed man said, "Oh thank God. I though you said in a *million* years."

Framing

Utility is a subjective measure of happiness, satisfaction or reward a person gains (or loses) from receiving a good or service. Utility is considered not in an absolute sense (from zero), but subjectively from a reference point, established by the Decision Maker's (DM) perspective and wealth before the decision, which is his frame of reference (Kahneman 2011). Framing (the context of a question) could affect his decision. The section on Severity Amplifiers stated that interpersonal variability in evaluating the seriousness of a situation depends on framing. That is, the circumstances surrounding the event will affect how a DM responds to it. An evaluation may depend on factors such as how the criterion affects that DM, whether that DM voluntarily exposed himself to the risk, how well that DM understands the alternative technologies and the severity of the results. In the previous section we gave over a dozen severity amplifiers that would affect the framing of a problem.

In contrast to defining framing in passing, as we have done so far, we will now explain framing directly, based on Beach and Connolly (2005). The DM has a vision, a mission, values, morals, ethics, beliefs, evaluation criteria and standards for how things should be and how people ought to behave. Collectively these are called *principles*. They are what the DM, the group or the organization stands for. They limit the goals that are worthy of pursuing and acceptable ways of pursuing these goals. These principles are difficult to articulate, but they powerfully influence the DM's behavior. They are the foundation of the DM's decisions and goals; actions that contradict them will be unacceptable. The utility of the outcomes of decisions derives from the degree to which these decisions conform to and enhance the DM's preconceived principles.

Goals are what the DM wants to accomplish. The goals are dictated by the principles, the problem, the problem statement, opportunities, desires, competitive issues or gaps encountered in the environment. Goals might seed more principles. Goals should be SMART: specific, measurable, achievable, realistic and time-bound.

The DM has *plans* for implementing the goals. Each goal has an accompanying plan. Each plan has two aspects: (1) tactics are the concrete behavioral aspects that deal with local environmental conditions and (2) forecasts are the anticipation of the future that provide a scenario for forecasting what might result if the tactics are successful. The plans for the various goals must be coordinated so that they do not interfere with each other and so that the DM can maintain an orderly pursuit of the goals. The plans are also fed back to the principles; therefore, they might foment more principles.

Framing means embedding observed events into a context that gives them meaning. Events do not occur in isolation; the DM usually has an idea about what led up to them. This knowledge supplies the context, the on-going story that gives coherence to experiences, without which things would appear random and unrelated. A frame consists of the principles, goals and plans that are deemed relevant to the decision at hand and that fixes the set of principles that influence that decision.

The DM uses contextual information to probe his or her memory. If the probe locates a contextual memory that has similar features to the current context, then the current context is said to be recognized. *Recognition* defines which principles, goals and plans are relevant to the current context and provides information about the goals and plans that were previously pursued in this context. If a similar goal is being pursued this time, then the plan that was used before may be used again.

In summary, framing means describing all aspects of the problem, the problem statement and the DM's mind that will affect decisions.

Factors that Affect Actions in the Tradeoff Study Process

We now examine the activities in Fig. 1. Specifically, we identify human psychological factors that can adversely influence human decision-making when dealing with uncertainty. We only give short phrases listing these factors. They are explained in detail in Smith et al. (2007) and Bohlman and Bahill (2014).

- State the problem. This activity tends to be affected by severity amplifiers and framing. Additionally, it is affected by poor problem stating, incorrect phrasing, attribute substitution, political correctness, and feeling invincible.
- Identify stakeholders. This activity is affected by framing.
- Understand customer needs. This activity is affected by confirmation bias, severity amplifiers, and framing.
- Identify and analyze risk events. This activity is affected by confirmation bias and severity amplifiers.
- Describe alternative solutions. This activity is affected by confirmation bias, severity amplifiers, and framing.
- Create evaluation criteria. This activity is affected by severity amplifiers. Additionally, it is affected by the evaluation criteria hierarchy, relying on personal experience, the Forer Effect (Forer 1949), and attribute substitution.
- Develop weights of importance. This activity tends to be affected by severity amplifiers. Additionally, it can be affected by whether the weights are the result of choice or calculation.
- Create scoring functions. Mistakes here include mixing gains and losses, not using scoring functions and anchoring. The biggest mistake is stating output scores with false precision.
- Choose the combining function. Lack of knowledge is the key problem in this activity. There are several appropriate combining functions. One of the oldest and most studied means for combining data under uncertainty is the certainty factor calculus employed by the Mycin expert system at Stanford University in the 1980s (Buchanan and Shortliffe 1984). Other valid approaches are the Bayesian Belief Networks (Cooper 1990), Dempster-Shafer Theory (Zadeh 1986), and Fuzzy Logic (Zadeh 1965).
- Collect evaluation data. Assign values to (1) attributes, (2) weights and scores and (3) likelihood and severity. All three of these activities can be adversely affected by confirmation bias, severity amplifiers, relying on personal experience, magnitude and reliability, and judging probabilities poorly.

- Produce the tradeoff matrix. Prioritize alternatives. This activity is affected by confirmation bias, severity amplifiers and framing. This activity can be degraded by serial consideration of alternatives, isolated or juxtaposed alternatives, conflicting evaluation criteria, adding alternatives, maintaining the status quo, and uneven level of detail. The order in which the alternatives are listed has a big effect on the values that humans give for the evaluation data. Therefore, a tradeoff study matrix should be filled out row by row with the status quo being the alternative in the first column. This makes the evaluation data for the status quo the anchors needed for estimating the evaluation data for the other alternatives. This is good choice because the anchoring alternative is known, is consistent, and you have control over it.
- Perform sensitivity analyses. Done right, this should not create uncertainty.
- Monitor the marketplace and the environment. This activity is typically affected by severity amplifiers. Additionally, tunnel vision can throw off the analysis. Therefore, to avoid tunnel vision, it is imperative that the environment be a part of the framing.
- Conduct expert reviews. Formal inspections and expert reviews are done entirely by humans. Therefore, every human limitation such as cognitive biases and misconceptions/preconceptions must be addressed.
- Review with customer and other stakeholders and revise. The most common mistake in design projects is failing to engage stakeholders and consult with experts in universities and local industries (Bohlman and Bahill 2014). It is imperative to engage all stakeholders especially in upfront engineering to avoid the likelihood of extraneous design iterations and rework.

Handling Uncertainty in the Problem Statement

Now that we have identified sources of uncertainty in the tradeoff study process, we will present examples of some techniques for handling uncertainty in the tradeoff study process. The first and most important step in performing a tradeoff study is stating the problem (see Figs. 1 and 2). Uncertainty can cause mistakes in the problem statement. This section is based on Diogenes (Bahill 2012). These are some of the related tasks that were described in the State the Problem paragraph of the SIMILAR Process section of this chapter.

- Explain what the system is supposed to do.
- Understand customer needs.
- Identify stakeholders.
- Discover the inputs and their sources.

In the beginning of any system design, what the system is supposed to do is uncertain. Functions, requirements, and desirements may have been stated, but incomplete understanding, mistakes, unknown technology, and improvement opportunities usually change the preconceived functioning of any system. To understand

and explain what the system is supposed to do and how it works, we use a multitude of stories and use case models.

It turns out that all of these activities involve human decision-making. Therefore, most of the mistakes caused by uncertainty will be found in the system documentation.

Understanding customer needs, identifying stakeholders, and discovering the system inputs are all affected by uncertainty, confirmation bias, severity amplifiers, framing, and many other mental mistakes (Smith et al. 2007).

Overcoming Problem Statement Mistakes

The primary reason that these mental mistakes are so important is that people do not realize that they exist. And the people that know of their existence believe that these mistakes do not affect *their* decision-making. However, when results of these mistakes are pointed out, most people are willing to rewrite to eliminate their undesirable effects. So, the best way to get rid of such mistakes is to bring them out in the open.

We will now present our process for ameliorating such mental mistakes. To handle uncertainty in the problem statement, all of the work products must be available for public review, must be subjected to formal reviews, must be approved in expert reviews, and all of these activities must be in a feedback control loop with frequent small iterations.

The first step in our process is to prepare a document that explains confirmation bias, severity amplifiers, and framing, as well as the mental mistakes of poor problem stating, incorrect phrasing, attribute substitution, political correctness, and feeling invincible (Smith et al. 2007). This chapter could serve this purpose. All people involved in the process must read this document in advance.

Common Problem Statement Mistakes

In this section, we examine four specific mental mistakes that the inspection team is likely to encounter when inspecting problem statement documents (Smith et al. 2007).

Not Stating the Problem in Terms of Customer Needs

Uncertainty about the customer needs, might lead the design engineer to commit to a class of solutions rather than customer needs. Identifying the true customer needs can be difficult because stakeholders often refer to both problem and solution domains – whichever comes most naturally (Mannion and Hermann 2006, p.26). The initial problem statement must be written before looking for solutions (Wymore 1993).

Recommendation: Communicate with and question the customer in order to determine his or her principles, values, and needs. State the problem in terms of customer requirements (Bahill and Dean 2009; Daniels and Bahill 2004; Hooks and Farry 2001; Hull et al. 2005). Later, after gaining a better understanding of evaluation criteria and weights of importance, create solutions that match the requirements.

Incorrect Question Phrasing

The way a question is phrased may determine the answer you get. Alluding to the problem of formulating public policy, Kahneman and Ritov (1994) showed their subjects (1) brief statements that looked like headlines and (2) proposed methods of intervention. Some subjects were asked to indicate their willingness to *pay for* the interventions by voluntary monetary contributions, while other subjects were asked which intervention they would rather support.

Issue Mammals:

Problem: Several Australian mammal species are nearly wiped out by hunters.
Intervention: Contribute to a fund to provide a safe breeding area for these species.

Issue Workers:

Problem: Skin cancer from sun exposure is common among farm workers.
Intervention: Support free medical check-ups for threatened groups.

On being asked how much money they would be willing to contribute, most subjects indicated that they would *contribute more money* to provide a safe breeding area for the Australian mammal species than they would to *support* free medical check-ups for the threatened farm workers. However, when the subjects were asked which intervention they would support, they indicated that they would rather support free medical check-ups for threatened workers.

If a problem statement is vague (such as "work for the public good"), proposed solutions could vary greatly, and derive support for very different reasons and in different ways. If a problem statement is poorly written or ambiguous, dissimilar alternative solutions could remain in the solution pool, obfuscating their rational consideration, especially if the rationale for the different psychologically attractive values of the alternative solutions are not well understood (Keeney 1992).

The above example of phrasing the question is subtler than the following one. When asked which package of ground beef they would prefer to buy, many more people chose the package labelled "80% lean," than the one labelled "20% fat."

Recommendation: Questions designed to get a value for a criterion should be tightly coupled to the criterion.

Substituting a Related Attribute

Attribute substitution occurs when a subject is assessing an attribute and substitutes a related attribute that comes more readily to mind. In effect, "people who are confronted with a difficult question sometimes answer an easier one instead" (Kahneman 2003, p. 707). For example, when confronted with a choice among alternatives that should properly be decided by a full tradeoff study, there is a strong tendency to substitute a seemingly equivalent yet much simpler decision question in place of the tradeoff study process.

Recommendation: Sponsors of tradeoff studies should realize that a premature reduction of a tradeoff study process to a simpler decision question is a common error that prevents consideration of the original multi-objective decision.

Political Correctness

Political correctness often makes top-level decision makers afraid to state the problem clearly and concisely. Furthermore, research funding is a zero-sum game. If I give more money for your research, then I have to take money away from someone else's research: and political correctness might give their research a higher priority.

Recommendation: Be aware of political correctness. Never include something strictly because it is politically correct. Point out sections that obfuscate because of political correctness and rewrite them clearly but tactfully.

Handling Uncertainty in Evaluation Criteria and Importance Weights

Creating evaluation criteria and Developing weights of importance are two activities of the Tradeoff Study process of Fig. 1 that are subsumed in the Investigate Alternatives activity of the SIMILAR Process of Fig. 2.

Designing Evaluation Criteria

When evaluating systems and investigating trade-offs, the analyst and the customer must jointly select measures that encompass the customers' preferences and values regarding system designs with respect to the problem at hand. This involves a lot of uncertainty. Simple, objective, and quantitative techniques are needed to analyze alternative system designs and compare attributes such as performance, cost, schedule, and risk. Selecting a technique for conducting these analyses is typically handled with approaches such as multi-attribute utility theory. The general tasks involved in multi-attribute utility measurement are structuring objectives, choosing evaluation criteria, eliciting single-attribute scoring functions, eliciting weights, and selecting combining functions. This section deals with creating evaluation criteria.

When designing systems, the analyst and the customer must identify the customers' principles and values. With these metrics, one can infer the overall desired quality or performance of the system as judged by the customer. We call such measures evaluation criteria. Evaluation criteria are specific items that need to be quantified to determine how well the system under study satisfies the design requirements from the stakeholders' points of view. Consider the simple example of evaluating personal computer systems. When comparing computer systems, one may use evaluation criteria such as cost, processor speed, amount of memory and hard disk size, to determine how well each computer system meets the requirements associated with these metrics.

In summary, evaluation criteria need to be established and maintained. Each criterion should link to a tradeoff requirement, that is, a requirement whose acceptable value can be more or less depending on quantitative values of other requirements. It is recommended that companies have a repository of generic, hierarchically arranged, evaluation criteria. The top-level evaluation criteria might be performance, cost, schedule and risk. Management should prioritize these four evaluation criteria at the beginning of the project and then make sure that everyone is made aware of these priorities. Finally, each criterion must have a weight of importance.

Weights of Importance

Uncertainty in the evaluation criteria is handled with weights of importance, through which priorities among evaluation criteria and levels in the evaluation criterion hierarchy are established. That is, evaluation criteria with higher importance are given more weight in the overall evaluation of the system. Weights of importance are assigned in accord with the five principles of Von Neumann and Morgenstern (1953) namely that fundamental decision analysis is based on probability, order, equivalence, substitution and choice. As a first approximation for the desired weights, the customer assigns a number between 1 and 10. The weights are then normalized in each category. While this method is ad-hoc and without axiomatic basis, it has been proven to be useful as an indicator of stakeholder preferences despite inconsistencies that theoretically could occur with this method. The key point here is to get the decision maker to think about these issues. The exact numbers are less important. After all, decision makers tend to be inconsistent, and they do change their minds.

Creating Importance Weights

There are several uncertainty issues that should be brought to the attention of a decision maker to help him or her in assigning weights. These include: organization commitment, critically to mission success, architecture, business value, priority of scenarios (use cases), frequency of use, benefit, cost, benefit to cost ratio, time required, risk, safety, complexity, implementation difficulty, stability, dependencies, and reuse potential (Botta and Bahill 2007).

Deriving Values for Weights of Importance

A dozen methods exist for deriving numerical values for the weights of importance for the evaluation criteria (Botta and Bahill 2007). These methods can be used by individuals or teams. The following is the most common method for deriving weights of importance. The engineer derives strawman weights of importance for all the evaluation criteria. These weights of importance are numbers (usually integers) in the range of 0 to 10, where 10 is the most important. He or she presents these weights of importance to the decision maker and helps him or her adjust them. Then he or she meets with the customer (however many people that might be). For each criterion, the systems engineer leads a discussion of the evaluation criteria and tries to reach a consensus on the priority of each criterion. On the first pass, each stakeholder might be asked to evaluate each criterion and then the engineer computes the average value. Ultimately, the customer's opinion prevails. If the customer only considers one criterion and declares the criterion to be a 10, then it's a 10. While we would not expect a domain expert to assign a weight of 0, a weight of 0 can be given to evaluation criteria that have no effect on the output, but whose consideration should be made prominent. When the output values are used for numerical comparisons in complex high-risk situations, then additional quantitative methods might be useful. When the weights are to be assigned using both the decision makers' relative importance and also the expected range of input values then the method of swing weights can be used.

The Method of Swing Weights

The swing weight method will now be explained using a new car selection example. For evaluation criteria, we will use Five-year Life Cycle Cost, Horsepower and Safety. The Five-year Life Cycle Cost (in US dollars) includes purchase price, taxes, licenses, loan interest, insurance, gasoline and maintenance. Horsepower is the peak SAE net horsepower. (However, the Horsepower to Weight Ratio might have been a better criterion.) The Safety rating is 0 to 5 stars based on the National Highway Traffic Safety Administration. Here are values for some typical cars.

	Values		
Evaluation criteria	Car A	Car B	Car C
Five-year life cycle cost (US $)	$52,000	$34,000	$22,000
Horsepower (hp)	290	240	170
Safety (stars)	4	5	3

Next, we determine the range of each criterion. As noted earlier, there are several choices for the range. For this example, we will use the real data given above. For the three cars that we are examining the maximum and minimum Five-year Life Cycle Costs are $52,000 and $22,000.

	Value for the worst alternative	Value for the best alternative
Five-year life cycle cost	$52,000	$22,000

Next, we consider Horsepower. The three cars that we are examining have a minimum Horsepower of 170 and a maximum of 290.

	Value for the worst alternative	Value for the best alternative
Horsepower	170	290

Our third criterion is Safety. The three cars have minimum and maximum values of 3 and 5 stars.

	Value for the worst alternative	Value for the best alternative
Safety	3	5

We now have definitions and ranges, measured from worst to best, for each of the three evaluation criteria that matter the most to the customer in this new-car selection example. Other characteristics, such as color or type of transmission, may also be important considerations in the choice of a car. However, in our example we are assuming that on all these other evaluation criteria the differences between the cars are unimportant. This does not mean that these other characteristics do not matter, but only that, in the context of this choice, they are unlikely to vary sufficiently to warrant making explicit trade-offs among them. Or on the other hand, they might be so important that cars without, for example, an automatic transmission, will not even be considered.

Now imagine a *hypothetical* car that is the worst it can be on all three evaluation criteria. In other words, its Five-year Life Cycle Cost is $52,000, its Horsepower is 170 and its Safety rating is 3 stars. Suppose that you can change the value of one (and only one) of these evaluation criteria on this hypothetical car from the worst to the best. This means that you can change only one of the following:

- Five-year Life Cycle Cost, from $52,000 (worst) to $22,000 (best).
- Horsepower, from 170 hp. (worst) to 290 hp. (best).
- Safety, from 3 (worst) to 5 (best).

Which one would you want to change? Suppose you say Five-year Life Cycle Cost. That means that you value a $30,000 drop in price (a change from $52,000 to $22,000) more than you do either an increase of 120 Horsepower or an increase of 2 stars of Safety. This criterion, the one that you most want to change from worst to best, is the one you weight most highly in the context of this problem. Assign it 100 points.

Now, which criterion do you value second most? Let us say it is Horsepower. Ask yourself, "How much less do I value the 120 Horsepower change compared to the $30,000 drop in price?" Suppose that you value it half as much. Then you would assign it 50 points, or half the weight you gave to the most important criterion.

Now let us consider the last criterion, Safety. Because this criterion is ranked below Horsepower, it should get fewer points. For example, if you value it two-thirds as much as Horsepower, then give it 33 points. (Note that this also means that Safety, with its 33 points, is only one-third as important for this decision as the Five-year Life Cycle Cost.) This process produces weights of 100, 50, and 33. All that remains is to normalize the weights so that they sum to 1, as shown in here.

Evaluation criteria	Weight of importance	Normalized weight of importance	Car A	Car B	Car C
Five-year life cycle cost (US $)	100	0.55	$52,000	$34,000	$22,000
Horsepower (hp)	50	0.27	290	240	170

(continued)

Evaluation criteria	Weight of importance	Normalized weight of importance	Car A	Car B	Car C
Safety (stars)	33	0.18	4	5	3
Sum		1.0			

We have just derived normalized weights of importance for our three evaluation criteria. This process can be expanded for any other evaluation criteria that still exist. It should be repeated in frequent iterations.

Other Methods for Deriving Weights

In another method for deriving weights of importance, the decision maker rank orders the evaluation criteria. Rank ordering gives ordinal numbers, not cardinal numbers. However, often this technique works because rank ordering is easier for humans than assigning weights of importance. The Analytic Hierarchy Process (AHP) can also be used to derive weights. The engineer helps the customer make pair-wise comparisons of all the evaluation criteria to derive the weights of importance. This approach is feasible when using a commercial software tool. Prioritization is another obvious method that can be used to derive weights of importance (Botta and Bahill 2007).

Many other methods exist for deriving weights of importance, including: the ratio method (Edwards 1977), tradeoff method (Keeney and Raiffa 1976), swing weights method (Kirkwood 1998), rank-order centroid techniques (Buede 2009), and paired comparison techniques discussed in Buede (2009) such as the Analytic Hierarchy Process (Saaty 1980), trade-offs (Watson and Buede 1987), balance beam (Watson and Buede 1987), and judgments and lottery questions (Keeney and Raiffa 1976). These methods are more formal and some have an axiomatic basis.

Cardinal Versus Ordinal

Ideally, weights of importance should be cardinal numbers not ordinal numbers. Cardinal numbers indicate size or quantity. Ordinal numbers merely indicate rank ordering. (This mnemonic may be useful, ordinal is ordering, as in rank ordering.) Cardinal numbers do not just tell us that one evaluation criterion is more important than another, they also tell us how much more important. If one evaluation criterion has a weight of 6 and another has a weight of 3, then the first is twice as important as the second is. However, in practice, ordinal numbers are usually used in tradeoff studies.

Importance Weights Mistakes

For the baseball umpire who needs to call balls and strikes, the Umpire's Assistant (http://sysengr.engr.arizona.edu/UmpireAssistant/index.html) is an intelligent decision aiding system that helps him or her to call balls and strikes accurately, consistently and in real-time. Unlike unassisted human umpires, the Umpire's Assistant uses the same strike-zone standards for all leagues, parks, umpires, batters, and

Table 2 Preliminary list of evaluation criteria for the Umpire's Assistant

Utilization of resources figures of merit requirements	Value	Normalized weights
1. Available money	2	0.02326
2. Available time	2	0.02326
2.1 system design & prototyping by 12/31/15	2	0.02326
2.2 system verification testing by 2/16	2	0.02326
3. Technological restrictions	10	0.11628
3.1 to not significantly alter the dynamics of baseball	9	0.10465
3.2 to comply with local, regional, state, federal laws	10	0.11628
3.3 to comply with FCC rules	10	0.11628
4. Adaptability	8	0.09302
4.1 to comply with Standards & Specifications of MLB	8	0.09302
4.2 to comply with Standards & Specifications of NCAA	8	0.09302

pitchers. Table 2 gives the original list of evaluation criteria for this system (Bohlman and Bahill 2014). It contains the following mistakes. The normalized weights add up to 0.826. They should add up to 1.0 in each category and subcategory. Listing five digits after the decimal point is an example of the mental mistake of implying false precision.

When a group of people is asked to assign a weight of importance for an evaluation criterion, each person might produce a different value. Different weights arise not only from different preferences but also from irrational severity amplifiers (Bahill and Karnavas 2000). These include the factors of lack of control, lack of choice, lack of trust, lack of warning, lack of understanding, manmade, newness, dreadfulness, personalization, recallability, and immediacy. Excessive disparities occur when a person assesses a weight of importance after framing the problem differently. An evaluation may depend on how the criterion affects that person, how well that person understands the alternative technologies, the dreadfulness of the results, etc. As a result, each person might assign a different weight of importance to any criterion. The decision analyst should assign weights to the evaluation criteria so that the more important ones will have more effect on the outcome. Weights are often given as numbers between 0 and 10, but are usually normalized so that in each category they sum to 1.0. These methods can be used by individuals or teams. If pair-wise comparisons of preferences between the evaluation criteria can be elicited from experts, then the weights of importance can be determined through the Analytic Hierarchy Process (AHP). However, performing pair-wise comparisons can lead to intransitive preferences. Therefore, the AHP computes an inconsistency index to warn if the domain expert is giving intransitive responses.

Recommendation: Interpersonal variability can be reduced with education, peer review of the assigned weights, and group discussions. But be aware that people are like lemmings: if you reveal how other people are voting, then they are likely to respond with the most popular answers. It is also important to keep a broad view of the whole organization, so that evaluation criteria in one area are considered in light of all other areas. A sensitivity analysis can show how important each weight is. For

Table 3 Revised list of evaluation criteria for the Umpire's Assistant

Utilization of resources evaluation criteria	Weight of importance	Evaluation criteria normalized weight	Subcriteria normalized weight
1. Available money	2	0.09	
2. Available time	2	0.09	
2.1 system design & prototyping by 12/31/15	2		0.5
2.2 system verification testing by 2/14/16	2		0.5
3. Technological restrictions	10	0.45	
3.1 to not significantly alter baseball dynamics	9		0.31
3.2 to comply with local, state & federal laws	10		0.35
3.3 to comply with FCC rules	10		0.35
4. Adaptability	8	0.36	
4.1 to comply with MLB rules	8		0.5
4.2 to comply with NCAA rules	8		0.5

unimportant weights, move on. For important weights, spend more time and money trying to get consensus: this might include showing the recommended alternatives for several different sets of weights. Table 3 is a suggested revision of the evaluation criteria of Table 2 (Bohlman and Bahill 2014).

Of course, there would be a paragraph explaining each of these short evaluation criteria tags. The abbreviations would be explained in these paragraphs.

Handling Uncertainty with Certainty Factors

Creating scoring functions and choosing the combining function are two activities of the Tradeoff Study process of Fig. 1 that are subsumed in the Investigate Alternatives activity of the SIMILAR Process of Fig. 2.

Different methods exist for combining scores, or values, in a tradeoff study to calculate a numerical measure that can be used to compare alternatives. The combining methods described here are used to combine data at all levels of the evaluation criterion hierarchy.

At the lowest level, when we are dealing with individual evaluation criteria, the scores are given as outputs of scoring functions (Wymore 1993; Bahill and Madni 2017) associated with each of the evaluation criteria, and weights are based on expert opinion or customer preference. When we move to higher levels in the hierarchy, the scores are derived by combining scores at lower levels. Therefore, scoring functions are not needed, because these scores are all already normalized. Again, weights at higher levels in the hierarchy are based on expert opinion or customer preference, perhaps from a different category of stakeholder.

Common Combining Functions

The following functions combine data from one or many evaluation criteria. In these equations, n is the number of evaluation criteria to be combined, x_i is the output score of the scoring function for the i^{th} evaluation criterion, and w_i is the normalized weight of importance for the i^{th} evaluation criterion. Weights of importance are expected to vary from zero to one. If the weights vary from 0 to 100, then the equations would have to be adjusted.

Sum Combining Function

The sum combining function is the simplest and most common method for combining data. It is ideal when the evaluation criteria show perfect compensation, that is, when both criteria contribute to the result and when more of y and less of z is just as good as less of y and more of z. To describe this data combining process, first suppose there are n reasonably independent evaluation criteria to be combined (perhaps they are in the same layer in the evaluation criterion hierarchy). We assign a qualitative weight to each of the n evaluation criteria and then normalize the weights so they add up to 1. Data are collected for the evaluation criterion; each evaluation criterion is then evaluated with its scoring function and the resulting scores (valued from 0 to 1) are then multiplied by their corresponding weights. The final result is the summation of the weight-times-score for each evaluation criterion. This process is commonly used, for example, when computing a student's grade point average.

The *sum combining function* is

$$f(x) = \sum_{i=1}^{n} w_i \cdot x_i, \text{if } n = 2 \text{ then } f(y, z) = w_{\mathrm{y}} y + w_{\mathrm{z}} z.$$

The sum combining function is appropriate when the decision makers' preferences satisfy additive independence (Keeney and Raiffa 1976) which is the case for most industry applications that we have seen. The sum combining function was used in Table 1.

Sum Minus Product Combining Function

The sum minus product combining function for two evaluation criteria, y and z, is given as:

$$f(y, z) = w_{\mathrm{y}} y + w_{\mathrm{z}} z - w_{\mathrm{yz}} yz$$

where w_{yz} must be uniquely evaluated.

The sum minus product combining function has its origins in probability theory: it is appropriate for computing the probability for the union of independent events. It is also used in fuzzy logic systems. It is the function used in Mycin-style decision support systems for computing certainty factors when two or more rules with the same conclusion succeed. Certainty Factors (CFs) have been used in expert systems for 40 years (Buchanan and Shortliffe 1984). The underlying method surrounding CFs, which is based on probability theory, has stood up to mathematical scrutiny. Thus, a vast knowledge base exists for CFs, and a great deal is known about their properties and uses.

Suppose, the certainty factor for the first premise of a rule is $CF_1 = w_1x_1$ where w_1 is the weight (between 0 and 1) and x_1 is the score (output of the scoring function, also between 0 and 1) and the certainty factor for the second premise is $CF_2 = w_2x_2$. The equation for determining the certainty factor after both rules are evaluated as true is

$$CF_{\text{both}} = CF_1 + (1 - CF_1)CF_2$$

In general, the certainty factor combining equation is

$$CF_{\text{total}} = CF_{\text{old}} + (1 - CF_{\text{old}})CF_{\text{new}} \text{ or}$$

$$CF_{\text{total}} = w_{\text{y}}y + w_{\text{z}}z - w_{\text{yz}}yz$$

which of course is the sum minus product combining function. The nature of the rule restricts CFs to the range $\{0, 1\}$.

The CFs for the remaining evaluation criteria are combined using equation $CF_{\text{total}} = CF_{\text{old}} + (1 - CF_{\text{old}})CF_{\text{new}}$ to create an aggregate score for their respective subcategories. When we move to the next level up, the x_i's and w_i's become the weights and scores for the subcategories just calculated, and so on. At the highest level of the evaluation criterion structure, CF_{total}becomes the overall evaluation from which alternatives can be compared.

An advantage of using CFs as a combining method is that the weights (w_i) do not have to be normalized within any rule or section. This means each time the objectives hierarchy is modified, be it the addition or subtraction of a new rule or evaluation criterion, or a new layer in the hierarchy, it is not necessary to re-normalize the weights as with sum combining function. This feature simplifies computations. A disadvantage of the certain factor calculus is that it is cumbersome to apply when more than two rules succeed simultaneously. Furthermore, the sum minus the product combining function is only valid if the evaluation criteria are normalized with a range of zero to one. Whereas, the most convenient expert system certainty factors are those between 0 and 100, so we must normalize those to the range $\{0, 1\}$.

Numbers Should Have Two Parts

Numbers in scientific publications should have at least two parts. The first is the magnitude or value. The second part should indicate the uncertainty, reliability, confidence, likelihood, range of validity, tolerance, direction (for vectors), variance, standard deviation, sample size, margin of error, skewness, or some combination of these qualifiers. Here are some examples. Complex numbers have real and imaginary components. Anytime a statistical analysis gives a mean it should certainly give a variance. In a normal Major League Baseball (MLB) game, the coefficient of restitution (CoR) might *range* from 0.4 to 0.5. MLB rules allow the radius of the baseball to be 1.45 ± 0.02 inches (*tolerance*). The typical launch velocity of a batted ball has a magnitude of 92 mph and a launch angle of 30 degrees (*direction*). Summary statistics often give the mean and *variance* (*standard deviation*) of bat swing speeds. Risk is quantified as likelihood of occurrence times the severity of consequences. A benefit to cost ratio has two parts. In alternating-current, electrical systems being analyzed with Laplace transforms all variables have real and imaginary components that are used to model respectively the steady-state and transient behaviors of the system. In electrical circuits being analyzed with phasors, there are three parts: amplitude, frequency, and phase. In this chapter, we have emphasized giving the magnitude and uncertainty of numbers as shown in Table 4.

Handling Uncertainty with Sensitivity Analyses

Preforming a sensitivity analysis is an activity of the Tradeoff Study process of Fig. 1 that is subsumed in the Model the System activity of the SIMILAR Process of Fig. 2. Validating and verifying the system are important tasks of the Tradeoff Study

Table 4 A method of including certainty factors along with weights of importance for evaluation criteria

Evaluation criteria	Evaluation criteria weight of importance (with certain factor)	Normalized evaluation criteria weight of importance	Evaluation subcriteria weight of importance (with certain factor)	Normalized evaluation subcriteria weight of importance
Audible signals for cookies are ready	9(80)	0.43		
Lost study time	7(70)	0.33		
Nutrition	5(50)	0.24		
Calories			9(90)	0.47
Fat, grams			4(30)	0.21
Carbohydrates, grams			6(30)	0.32
Column sum		1.00		1.00

process of Fig. 1 that are subsumed in the Assess Performance activity of the SIMILAR Process of Fig. 2.

In many system designs, the specification can be inspected to reveal that the major indices of the system (such as performance, cost, schedule or risk) are driven by relatively few input variables and system parameters. Values for these variables and parameter are usually uncertain. A common interpretation of Pareto's rule would state that 80% of the influence could be traced to 20% of the inputs and parameters. That is, variations of the few prominent inputs or parameters can have a substantial impact on the characteristics of the system. If these important parameters also have large uncertainties, then red flags are raised and resources are committed to resolving those uncertainties first. The process of uncovering inputs and parameters that drive the system's properties is called *sensitivity analysis*. This Section is based on Smith et al. (2008).

We recommend that all tradeoff studies incorporate sensitivity analyses. We have used the results of sensitivity analyses to

1. Point out important values whose uncertainties must be investigated.
2. Guide formulation of the model architecture.
3. Choose an operating point,
4. Flag strange or unrealistic model behavior.
5. Highlight important assumptions of the model,
6. Validate the model.
7. Detect critical evaluation criteria.
8. Adjust numerical values of the parameters.
9. Suggest the required accuracy for calculating parameters.
10. Suggest tolerance for manufacturing parts.
11. Suggest new experiments and guide data collection activities.
12. Allocate resources.
13. Pin-point true cost drivers.
14. Reduce risk (Karnavas et al. 1993).

If you show your customer the requirements that are driving system cost, then he or she may relax a requirement and save money.

In analyzing the sensitivities of a tradeoff study, we are typically interested in (1) those inputs, parameters, and evaluation criteria that are most important and deserve further attention, and (2) those inputs and parameters that, when varied, could change the recommended alternative. The first issue can be investigated using a relative-sensitivity measure for each parameter for each alternative. The second can be ascertained by employing a search algorithm using the sensitivity of each parameter (Karnavas et al. 1993). Typical tradeoff study parameters include weights at all levels of the tradeoff hierarchy, and the scoring function inputs and parameters.

A sensitivity analysis reveals which inputs and parameters are the most important and most likely to affect system behavior and/or model predictions. Following a sensitivity analysis, values of critical parameters can be refined while parameters that have little effect can be either simplified, or ignored. In the manufacturing environment, they can be used to allocate resources to critical parts allowing casual treatment of less sensitive parts. If the sensitivity coefficients are calculated as functions of time, it can be seen *when* each parameter has the greatest effect on the output function of interest. This can be used to adjust numerical values for the parameters. The values of the parameters should be chosen to match the physical data at the times when they have the most effect on the output.

Performing Sensitivity Analyses

There are several common ways to perform sensitivity analyses. A partial derivative can be a sensitivity function for a system described by equations. Otherwise, spreadsheets are convenient for doing sensitivity analyses of systems that are not described by equations.

Some systems are not composed of equations or quantitative models. Therefore, the analytic sensitivity functions and numerical estimation techniques are not appropriate. For such systems, we can use the *what-if* sensitivity analysis technique. We simply ask what would happen if a certain event occurred or a certain parameter were changed.

Validation

A sensitivity analysis is powerful technique for validating systems. Validation means ensuring that the system suits the customer's actual needs. If a system (and its model) is very sensitive to parameters over which the customer has no control, then it may be the wrong system for that customer. If the sensitivity analysis reveals the most important input or parameter and that result is a surprise, then it may be the wrong system. If a system is more sensitive to its parameters than to its inputs, then it may be the wrong system or the wrong operating point. If the sensitivities of the model are different from the sensitivities of the physical system, then it may be the wrong model.

In a set of requirements, if you delete a requirement, then your completeness measure (perhaps a traceability matrix) should reflect that change. When you make a decision, you should do a sensitivity analysis and then see if changing the most important decision parameters would change your decision. In a tradeoff study, domain experts should agree that the most important evaluation criteria identified by the sensitivity analysis are indeed the most important evaluation criteria. In a risk analysis, experts should agree that the risks identified by the sensitivity analysis as being the greatest are indeed the greatest risks. After you prioritize a set of items, you should do a sensitivity analysis and discover the most important evaluation criteria. Then the values for those evaluation criteria should be changed to see if they change the prioritization. These are all validation concerns.

Verification

Sensitivity analyses can also be used to help *verify* systems. Verification means ensuring that the system complies with its requirements and conforms to its design.

In a manmade system or a simulation, unexpected excessive sensitivity to any parameter is a verification mistake. Sensitivity to interactions should definitely be flagged and studied: such interactions may be unexpected and undesirable.

In general, we do not say that a model is most sensitive to a certain input or system parameter. Rather we must say that a particular facet of a model is most sensitive to a particular input or parameter at a particular frequency, point in time, and operating point. The chosen facet most likely will be related to the question that the model was formulated to answer. Choosing the facet is the key to the sensitivity analysis.

Sensitivity analyses are especially helpful when modeling systems with uncertainty. The sensitivity analysis shows which inputs and parameters are important and which are not. This allows the selective allocation of resources to experiments that will reveal more accurate values for the most important parameters. Sensitivity analyses help us to understand how uncertainties in the inputs and model parameters affect the model and its predictions. While not addressed in this chapter, sensitivity analyses can also be used to study uncertainty in model architecture, assumptions, and specifications. Sensitivity analyses are used to increase the confidence in the model and its predictions, by providing an understanding of how the model responds to changes in its inputs, the data used to calibrate it, its architecture, or its independent variables.

Computing Sensitivities

To help validate a model, compute sensitivities for all parameters and inputs. Most systems should be more sensitive to their inputs than to any of their parameters. To help verify a system, compute sensitivities. Excessive sensitivity to any parameter is a verification mistake. After you build a model, write a set of requirements or design a system, you should study it to see if it makes sense. One of the best ways to study a model is through sensitivity analysis.

This section has described several ways to do sensitivity analyses. Different kinds of sensitivity analyses typically have different purposes and produce different results. Therefore, when presenting the results of sensitivity analysis, it is very important to state what type of sensitivity analysis was performed. For many sensitivity analyses, this would simply be a statement that the absolute, relative, or semirelative-sensitivity functions were computed either analytically or empirically; and the nominal operating point and the step size of the parameter perturbations would be given.

Sensitivity analyses point out the most important variables and parameters in a system. When these most important properties are also the ones with the most uncertainty, red flags should be raised and time and money should be committed to their resolution.

Conclusions

Uncertainty arises from factors that are both external and internal to the system. Examples of factors that contribute to external uncertainty are changes in market conditions or in the operational environment, new competitors or threats, emerging requirements, unobservability, changes in priorities, and delays in maturation times of promising new technologies. Internal uncertainties stem from unanticipated challenges that surface during program/project execution, system design and implementation, and creating performance requirements. By far the greatest uncertainty is coping with unknown futures. This problem requires designing for alternative futures – the hallmark of resilient design (Madni and Jackson 2009; Neches and Madni 2013).

There are several approaches to dealing with uncertainty depending on context. Uncertainty may stem from incomplete/fuzzy requirements or technology maturation rate. Clearly, if requirements and technologies are both stable in a project, it is relatively straightforward to plan ahead and execute the plan, because there is very little uncertainty to further ameliorate. On the other hand, when a project intends to capitalize on new or emerging technologies, uncertainty is best handled by placing "smart bits" or incorporating real options in both system architecture/design and program schedule (Madni and Allen 2011). Finally, when the technology aspect is relatively stable, but the requirements continue to evolve, then an incremental commitment approach can be pursued (Boehm et al. 2014).

References

Ainslie G (2001) Breakdown of will. Cambridge University Press, New York

Bahill AT (2012) Diogenes, a process for finding unintended consequences. Syst Eng 15(3):287–306. https://doi.org/10.1002/sys.20208

Bahill AT, Dean F (2009) Discovering system requirements, chap. 4. In: Sage AP, Rouse WB (eds) Handbook of systems engineering and management, 2nd edn. Wiley, Hoboken, pp 205–266

Bahill AT, Gissing B (1998) Re-evaluating systems engineering concepts using systems thinking. IEEE Trans Syst Man Cybern C 28(4):516–527

Bahill AT, Madni AM (2017) Tradeoff decisions in system design. Springer International Publishing, Cham

Bahill AT, Karnavas WJ (2000) Risk Analysis of a pinewood derby: A case study. Syst Eng 3(3):143–155

Beach LR, Connolly T (2005) The psychology of decision making: people in organizations, second edn. Sage Publications, Thousand Oaks, CA

Boehm B, Lane J, Koolmanojwong S, Turner R (2014) The incremental commitment spiral model. Addison-Wesley, Pearson Education, Inc.

Bohlman J, Bahill AT (2014) Examples of mental mistakes made by systems engineers while creating tradeoff studies. Stud Eng Technol 1(1):22–43. https://doi.org/10.11114/set.v1i1.239

Botta R, Bahill AT (2007) A prioritization process. Eng Manag J 19(4):20–27

Buchanan BG, Shortliffe EH (1984) Rule-based expert systems. Addison-Wesley, Reading

Buede DM (2009) The engineering design of systems: models and methods, 2nd edn. John Wiley and Sons, Inc, New York

CMMI for Development (2015) ver 1.3 [cited 2015 December]. Available from: http://cmmiinstitute.com/cmmi-models
Cooper GF (1990) The computational complexity of probabilistic inference using bayesian belief networks. Artif Intell 42:393–405
Daniels J, Bahill AT (2004) The hybrid process that combines traditional requirements and use cases. Syst Eng 7(4):303–319
Edwards W (1977) How to use multiattribute utility measurement for social decision making. IEEE Trans Syst Man Cybern 7(5):326–340
Forer BB (1949) The fallacy of personal validation: a classroom demonstration of gullibility. J Abnorm Psychol 44:118–121
Haimes YY (2004) Risk modeling, assessment, and management. Sage AP, editor. Wiley, Hoboken
Hooks IF, Farry KA (2001) Customer-centered products: creating successful products through smart requirements management. AMACOM, New York
Hull E, Jackson K, Dick J (2005) Requirements engineering. Springer, London
INCOSE Systems Engineering Handbook v2a 2004 [March 2005]. Available from: http://www.incose.org/ProductsPubs/incosestore.aspx.
Kahneman D (2003) A perspective on judgment and choice: Mapping bounded rationality. Am Psychol 58(9):697–720
Kahneman D (2011) Thinking, fast and slow. Farrar, Straus and Giroux, New York, ISBN: 978-0-374-27563-1
Kahneman D, Ritov I (1994) Determinants of stated willingness to pay for public goods: A study in the headline method. J Risk Uncertainty 9(1):5–37
Kahneman D, Tversky A (1979) Prospect theory: an analysis of decision under risk. Econometrica 46(2):171–185
Karnavas WJ, Sanchez PJ, Bahill AT (1993) Sensitivity analyses of continuous and discrete systems in the time and frequency domains. IEEE Trans Syst Man Cybern 23(2):488–501
Keeney RL (1992) Value-focused thinking: A path to creative decision making. Harvard University Press, Cambridge, MA
Keeney RL, Raiffa H (1976) Decisions with multiple objectives: preferences and value tradeoffs. Wiley, New York
Kirkwood C (1998) Strategic decision making multiobjective decision analysis with spreadsheets. J Oper Res Soc 49(1):93–98
Madni AM (2014) Generating Novel Options during Systems Architecting: Psychological Principles, Systems Thinking, and Computer-Based Aiding. Syst Eng 17(1):1–9
Madni AM (2015) Expanding stakeholder participation in upfront system engineering through storytelling in virtual worlds. Syst Eng 18(1):16–27
Madni AM, Allen K (2011) Systems thinking-enabled real options reasoning for complex socio-technical systems programs. Conference on Systems Engineering Research
Madni AM, Jackson S (2009) Towards a conceptual framework for resilience engineering. Syst J IEEE 3(2):181–191
Madni AM, Sage AP, Madni CC (2005) Infusion of cognitive engineering into systems engineering processes and practices. 2005 IEEE International Conference on Systems, Man and Cybernetics: Institute of Electrical & Electronics Engineers (IEEE)
Mannion M, Hermann K (2006) Using parameters and discriminants for product line requirements. Glasgow, UK
Martino De B, Kumaran D, Seymour B, Dolan RJ (2006) Frames, biases, and rational decision-making in the human brain. Science 313(5787):684–687
Mohanani RP, Salman I, Turhan B, Rodríguez P, Ralph P (2018) Cognitive biases in software engineering: a systematic mapping study. https://doi.org/10.1109/TSE.2018.2877759.
Neches R, Madni AM (2013) Towards affordably adaptable and effective systems. Syst Eng 16 (2):224–234
Nickerson RS (1998) Confirmation bias: a ubiquitous phenomenon in many guises. Rev Gen Psychol 2(2):175–220

Parnell GS (2016) Trade-off analytics: creating and exploring the system tradespace. Wiley
Petroski H (2003) Framing hypotheses: a cautionary tale. Am Sci (Jan–Feb) 91(1):18–22
Saaty TL (1980) The analytic hierarchy process: planning, priority setting, resource allocation. McGraw-Hill, New York
Smith ED, Son YJ, Piattelli-Palmarini M, Bahill AT (2007) Ameliorating mental mistakes in tradeoff studies. Syst Eng 10(3):222–240
Smith ED, Szidarovszky F, Karnavas WJ, Bahill AT (2008) Sensitivity analysis, a powerful system validation technique. Open Cybern Syst J 2:39–56
Tversky A, Kahneman D (1974) Judgment under uncertainty: heuristics and biases. Science 185:1124–1131
Tversky A, Kahneman D (1981) The framing of decisions and the psychology of choice. Science 211:453–458
Von Neumann J, Morgenstern O (1953) Theory of games and economic behavior, 3rd edn. Wiley, New York
Watson SR, Buede DM (1987) Decision synthesis: the principles and practice of decision analysis. Cambridge University Press, New York
Wymore AW (1993) Model-based systems engineering. CRC Press Inc., Boca Raton
Zadeh LA (1965) Fuzzy sets. Inf Control 8(3):338–353. https://doi.org/10.1016/s0019-9958(65)90241-x
Zadeh LA (1986) A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination. AI Mag 7(2):85–90