FLEXIBILITY IN DESIGN

(Working title, subject to change)

by

Richard de Neufville
Engineering Systems Division
Massachusetts Institute of Technology

and

Stefan Scholtes
Judge Business School and School of Engineering
University of Cambridge

PART 2

*"…in this world nothing can be said to be certain, except death and taxes***."**

Benjamin Franklin (1789)

Part 1 of this book gives an overview of the value of flexibility in design. It first presents the reason why is it necessary to recognize uncertainty explicitly: to do otherwise leads to erroneous calculations and inferior results. It then shows the benefits of suitable flexible designs: they lead to great improvements in expected value, compared to designs based on deterministic forecasts. This overview is useful to all readers. It may be enough for busy leaders at the top of the organizations involved in large-scale technology projects; it may convince them of the need to have their teams recognize uncertainty explicitly and look for good flexible designs.

Part 2 provides detailed descriptions of the major techniques for finding good flexible designs. It assumes the reader is conversant with basic analytic approaches to complex issues, such as modeling, statistical analysis, Monte Carlo simulation, and discounted cash flow analysis. Appendices on these and related topics complement the chapters in Part 2. This part of the book is therefore for current or future implementers of flexible design. You may be the senior engineers and managers, or junior staff or students aspiring to leadership in system design.

The goal is to enable you to ask the right questions, to direct your staff and consultants in the right direction, and to interpret and communicate the results of their technical analyses to a non-technical audience up or outside your organization. The goal is also to help you do some "back-of-the-spreadsheet" prototype modeling yourself, to help you understand on a small, simplified scale what goes on in the large scale analyses that you will ask your staff to undertake. Conversely, we do not aspire to train you to become a statistician or expert mathematical modeler. We assume that you will delegate routine technical aspects of the work to specialists in your team or consultants with appropriate expertise. The goal of Part 2 is to enable you to have a meaningful discussion, based on solid analysis, with modeling experts on the one hand, and with senior leaders and stakeholders on the other.

Analytically, Part 2 presents a range of new approaches suitable to analysis and evaluation of systems under uncertainty. The objective in this respect is to enhance traditional modeling techniques, not to replace them. In each chapter we build on the established methodology, for example a discounted cash flow model, and enhance it in such way that any new features can be "turned off" to recover the traditional model. Our intent is to make it easy for practitioners to adopt the concepts of flexible design by integrating the new concepts easily into proven established practices.

Our approach is pragmatic, directed toward practitioners who will be designing, managing and implementing real projects. We aim to improve practice as much and as easily as possible. More academically inclined colleagues may think we should instead be presenting theoretically correct methods that “get it right”. However, we believe that our modest aspiration to “get it better” is more likely to improve practice. Indeed, the concept of “getting it right” is difficult to defend, once one accepts that modeling the performance of socio-technological systems is an art as well as a science.

In detail, Part 2 covers the main elements needed for the development, selection and implementation of flexible designs. We present a systematic process to follow for each topic. The idea is to provide a coherent framework for integrating the several relevant analytic approaches involved. The organization is as follows:

- Chapter 4 focuses on estimating the distribution of future possibilities. This section deals with the issues associated with recognizing the uncertainties around complex projects. It extends the usual methods of forecasting, which seek to develop the most likely outcome, to procedures for developing the range of possibilities that might occur. Getting a good vision of the range of what may actually happen is essential to developing designs that will deal effectively with what the future brings.
- Chapter 5 develops ways to identify candidate flexibilities. It provides a process for identifying the kinds of flexibilities that can add the most value to a project. This is necessary because it is generally not obvious, in a complex system, what kinds of flexibilities may be most useful. This section shows how we can effectively use simplified models of our systems in combination with Monte Carlo simulation to identify productive forms of flexibility.
- Chapter 6 presents effective ways to evaluate and select preferred designs. It builds upon standard discounted cash flow analyses to provide graphical and tabular presentations of the multi-dimensional trade-offs designers and managers inevitably have to make between increased value, higher risk, and other criteria of good design.
- Chapter 7 focuses on ways to increase the possibility of effective implementation of flexibilities in design. It provides a notional checklist of steps that designers and managers can take that will increase the likelihood that they will eventually be able to use the flexible capabilities designed into a system when they want to do so.
- Chapter 8 is an epilog. It closes Part 2 with an overall assessment of what we can already achieve in flexibility in design, and of the further practical work we need to do to improve our capabilities and procedures. While basic principles and numerous case examples indicate that flexible design can greatly improve expected value, the approach is new and we can certainly improve it.

CHAPTER 4
ESTIMATING DISTRIBUTION OF FUTURE POSSIBILITIES

"I mistrust isolated trends (…). In a period of rapid change, strategic planning based on straight-line trend extrapolation is inherently treacherous (…). What is needed for planning is not a set of isolated trends, but multidimensional models that interrelate forces – technological, social, political, even cultural, along with the economics."

Alvin Toffler (1985)

"The only constant is change, continuing change, inevitable change; that is the dominant factor in society today. No sensible decision can be made any longer without taking into account not only the world as it is, but the world as it will be."

Isaac Asimov, building on Heraclitus (5th century BCE)

This chapter focuses on the challenge of estimating what may happen. This is indeed a challenge. We have to recognize that unpredicted events shape both our lives and that of our projects. Things happen. New technologies change our world, as penicillin radically increased life expectancies and population growth. Political events upset expectations and demands, as the creation of the European Union has reshaped markets and standards over the continent. Economic booms and busts create and reduce demand for services and products drastically compared to previous trends. Nonetheless, we do need estimates we can work with in order to design projects and systems. We need somehow to move ahead, while being modest about how accurate we can be. We do not have the omniscience to predict precisely what may happen. We need to be satisfied with good estimates of the distribution of possibilities.

We need two kinds of estimates of future distributions, as Figure 4.1 suggests. One concerns the outside circumstances that drive the performance of the system. These involve such factors as the price of oil, the economic environment and demand for products, the prevailing regulations, and so on. These elements generally do not depend on the operation of the system we are designing. The other kind of estimate relates to the internal workings of the system itself, to its performance metrics such as the level and cost of production. Estimates about the performance of the system require an understanding of the design and operation of the system; they require an adequate model of its operation and dynamics. The difference between these situations requires system designers and managers to adopt different approaches for each case.

[Figure 4.1 about here]

Experts other than system designers usually generate the estimates of outside circumstances. For example, economists will generate forecasts of economic activity, the overall demand for products and so on. Similarly, geologists normally provide estimates of possible quantities of ore in a mine or of petroleum in a field (see Figure 2.1). In this regard, system designers need to understand the assumptions behind these forecasts, and need to ask the right questions so that they can obtain estimates that are most appropriate for the effective design of a system. Such estimates should recognize the range of possible uncertain future paths. In practice, we can place these probabilistic estimates in spreadsheet worksheets that provide inputs to the Monte Carlo Analysis that explores the possible effect of various scenarios (see Appendix D).

Designers and managers have a more difficult task when it comes to estimating the distribution of the performance of their own systems. Before they can carry out the forecasting exercise, they need to develop a suitable model of their system. To obtain good estimates of performance, they must link together the interaction of the outside circumstances with their design to generate measures of key performance for the system. Good models of systems combine both technical and social considerations (such as economics). Thus, their development can be a special challenge for the professional experts in the system. Doctors expert in the functioning of a hospital, for instance, often are neither trained nor otherwise prepared to understand how their work affects overall performance of the hospital, as measured by costs, average delays in service, or readmission rates of persons who were not fully cured. System designers and managers need to give special attention to the modeling of their operations.

Note that the emphasis on estimating the distribution of outcomes expands considerably on current practice. Designers traditionally prefer to work with fixed specifications. This makes their work easier. With fixed specifications, they do not have to worry about many different combinations of events. Indeed, accepted practice for system design starts with the definition of fixed requirements.[1] This approach calls for specific, precise forecasts of future contexts and needs, known as “point forecasts”. Unfortunately, such point forecasts are generally “wrong” in that the actual results are different from best estimates (see Box 4.1).[2] The realistic approach is different; it is that we cannot know the future precisely; we can only estimate it over a range.

[Box 4.1 about here]

Moreover, this traditional approach focusing on most likely futures leads to faulty estimates of performance and value, due to the flaw of averages. It also ignores the reality that the system will operate in a wide range of circumstances, and makes no provision for adapting the system to these conditions. The traditional approach does not incorporate flexibility and fails to develop the value it provides. To obtain the increase in expected value we can obtain through flexibility, we must focus on estimating distributions of outcomes.

The process for estimating the distribution of future possibilities consists of four steps:

- Identify the key performance drivers. This first step focuses on the choice of environmental variables that we should worry about, the key drivers that are most crucial for the future performance of the system we are about to design. As the number of different drivers that can affect the performance of any large-size system can be huge, it is crucial to prioritize effectively, to reduce the number of variables to a handful of key performance drivers that are mission-critical for future performance. This prioritization requires expertise and insight into the wider operation of the system from engineering, economic and management perspectives.

- <u>Analyze historical trends</u>. The idea is to establish the historical trends of the key performance drivers and the historical variability around this trend. What do we know about the situation today? How did we get to where we are? In the early phases of planning a large system, one will hear many speculative arguments that attempt to explain how the current situation emerged. These often stem from anecdotal evidence colored by personal experience and responsibility. It is important to use as much hard data as possible to challenge unfounded assumptions that might otherwise distort the analysis.
- <u>Identify trend breakers</u>. The next step is to ask: How long will historical trends continue? What are the potential trend-breakers? How likely are they? Responses to these questions are more speculative, based on judgment rather than evidence, and therefore prone to biases. We can minimize these biases through careful analysis of historical data, to impel people to start their arguments from facts; and the establishment and management of a constructive scenario planning process that can challenge assumptions. It is important to involve major stakeholders in scenario planning, both those accountable for decisions and relevant domain experts, for example doctors in the case of a hospital. The idea is to develop a realistic set of future scenarios so that the design can anticipate these eventualities.
- <u>Establish forecast (in)accuracy</u>. We also need to be realistic about our collective ability to predict accurately in our particular field. A priori, our past performance provides a good indication of how well we can hope to do for our next project. If an automobile company's previous forecasts for sales of new models have differed from reality by X percent on average, this represents a reasonable estimate of our likely error in the forecast sales for the next model. We should thus document our record of accomplishment by comparing previous forecasts with what actually happened. As Chapter 2 indicates, the fact is that our forecasts are "always wrong" in that there normally are significant discrepancies between what we or our predecessors predicted and what actually occurred.
- <u>Build a dynamic model</u>. Finally, we put all three steps together and build a model, or several models, to allow us to generate many different evolutions of the system environment. These dynamic models of the environment replace the traditional single-number projections of input variables. Engineers and system architects should optimize their designs against the fluid and unpredictable futures that these models produce.

This chapter develops these steps sequentially. It illustrates them using a range of examples. To demonstrate the entire process, it ends with a case study of an important practical example involving the development of major hospital facilities.

Step 1: Identify the key performance drivers

Usefully identifying the key performance drivers of a system is more difficult than it seems. The obvious answers are often insufficient. For example, developers often present the need for a project in the following terms: demand for some product is growing, soon it will exceed the available capacity, and therefore we need additional capacity. This is the classical argument.[3] Promoters routinely use this reasoning to justify all kinds of public and private projects such as airports, factories, highway projects, power plants, steel mills, etc. The obvious driver of performance of the system thus seems to be demand – for air travel, for factory products, etc. This understanding is not wrong, but it is not enough.

The obvious identification of overall demand for a service as a driver of performance is not sufficient because it does not recognize that demand inevitably comes in many forms. These are not equal, and have different implications for the performance of the system. For example, the demand for power in the off-peak and the peak hours has very different implications for both the design and the performance of an electric power plant. A nuclear reactor is very efficient at meeting continuous demands that continue day and night (the so-called base loads), but cannot ramp up its production quickly to meet peak loads. Conversely, turbines driven by diesel engines can quickly deliver power for peak loads, but are not economical for providing base load power. In short, demand is complex. We need to define such drivers in detail.

In determining the useful drivers of performance, we need to be concerned with three elements:

- <u>Type</u>, what are the elements that will affect system performance?
- <u>Level of aggregation</u>, what is the best way to specify each type? and
- <u>Variability over time</u>, how does the driver fluctuate? And how does this uncertainty enhance the value of flexibility?

We also need to decide:

- <u>Criteria of usefulness</u>, which of these variables do we really need to include in the analysis?

We now discuss how to deal with each issue in turn.

Type of driver: We need to identify the factors that influence system design and its performance. These may be economic, technical, regulatory and others. In general, they may be much broader than system designers might imagine initially, before they think hard about their project. There is no convenient checklist. We need to think deeply about our system and identify the factors that will influence its performance. This process requires the collaboration of a range of experts on the topic. The objective of this phase is to develop a primary list of important drivers for consideration.

Consider the development of a power plant for example. As suggested above, the various forms of demand for power (peak and off-peak) strongly influence the desirable design and value of a plant. Technical uncertainties will come into play too, concerning the resistance of the soil to earthquakes for instance. Possible future regulations about the need for carbon capture will also influence design and system performance. The identification of the important drivers of course depends upon the circumstances. As of 2010 in Europe, regulations about carbon emission – and the uncertainty about their level – appear to be a major driver of power plant design. That was neither the case in China at that time, nor in Europe a decade earlier.

Level of aggregation: It is often useful to think in more detail about the drivers, as suggested by the discussion on the types of demand. In the case of designing a power system, it is crucial to recognize that the demand for power at different times has significant implications for what society builds. Moreover, the demand for power comes in different forms; higher voltages are necessary for transmission and some industries, lower voltages for household consumption.

Complementarily, it can be useful to think also of higher levels of aggregation. When considering the development of an oil field, for example, it is clear that both the price of oil and construction costs affect overall profitability of a project. We can recognize however that the overall economic situation may affect them both: boom times increase the demand and thus the price of oil; this in turn increases demand for drilling rigs and pipelines and thus the cost of building platforms. Thus in some circumstances it might be useful to focus on some overall drivers, such as the state of the overall economy in this case.

Box 4.2 illustrates the identification of performances drivers at various levels of aggregation.

[Box 4.2 about here]

Variability over time: It is also important to disaggregate drivers over time, to consider when they occur. This is particularly important when the driver may vary significantly over time, seasonally for example. Thus, the demands for power at the peak time of day in the peak season may drive the determination of the capacity of a power system.

As variability creates opportunities for flexibility, thinking carefully about variability over time is particularly important when several different drivers require similar facilities. If the peak demands for different services occur at different times, the facilities for the peak requirements of one service may be available for the other service at its peak. This flexibility can be a great source of improved, more efficient design, as Box 4.3 illustrates.

[Box 4.3 about here]

Criteria of usefulness: When developing a final choice of variables, it is important to be pragmatic. We need to keep in mind that our objective is to perform an effective analysis that will help us understand the issues and develop improved designs. We should thus focus on a set of variables that we expect to have impact, that are sufficiently few, and that are simple.

Most importantly, the variables selected as drivers of the system should matter. The uncertainty in their level should have design consequences. If the uncertainty in a variable would not affect the desirable design materially, then it is not a priority for consideration at this stage. Referring to the airport example in Box 4.2, there is uncertainty in the proportion of the travelers who do not speak the local language. This information may be useful to consider when someone has to think about interior signs, but almost certainly would not influence the overall design of the system noticeably. Thus, designer should not consider this variable as a driver for the design of an airport terminal.

We also need to focus on a few variables. Otherwise, the analysis for the overall design will become too complicated. The fact is that the number of scenarios to be considered increases exponentially with the number of driving variables. That is, if we want to consider 3 levels of one variable and 3 levels of another, then we have to think about 3x3 = 9 combinations. This means that we can only properly consider a few variables as drivers of design, even with our fast modern computers. We elaborate in Chapter 5 on this important feature. There are tools, such as Tornado diagrams (explained in Appendix D on Monte Carlo simulation) that help prioritize a long list of uncertain drivers. As with all tools, we need to use them to complement, not to substitute for expert judgment.

Finally, it is crucial to keep the choice of variables simple and intelligible. To be able to make good judgments about the design, we need to be able to develop an understanding of how it works. In that context, complex technical factors may be confusing. We will not be able to form a clear idea of what causes what, and will not be able to plan the design coherently. Furthermore, designers ultimately have to be able to explain and justify their project to decision-makers who will not be familiar with the details. They will need a convincing argument for others to understand and trust. If the model of the system is too complex, neither the designers nor the managers will find it fully acceptable.

Step 2: Analyze historical trends

The purpose of analyzing historical trends is to develop our understanding of what has been happening, with a view to estimating future possibilities. A proper analysis thus should go through several phases, each building on the other. We need to:

1. <u>Understand the data</u>. How good are they? What do they represent?
2. <u>Develop an appreciation for the overall pattern</u>. What is the central line?
3. <u>Assess the uncertainty in the trend</u>. How steady has it been? How has it been fluctuating away from the central line?

Once we have completed these phases, we will be in position to estimate future distributions, through an examination of how trends may break, and then more generally of the several factors that shape the evolution over time.

Understand the data: The first part of the analysis of historical trends is to assess the nature and quality of the data. The fact is that institutions record data for a variety of purposes, under definitions, procedures and circumstances relevant to their immediate situation. These conditions may not be those appropriate for planning and designing a facility.

We should first check the definitions. They often do not mean exactly what one might think. For example, the US Federal Aviation Administration (FAA) keeps a record of aircraft delays. However, this measure only represents a small fraction of what you and I might think of a delay. By convention, the FAA only starts counting delays once they are more than 15 minutes over the scheduled time. The airlines moreover, knowing that “on-time performance” is one of their metrics, pad their schedules to account for all the time their aircraft have to wait for take-off, landing, and so on. So the scheduled flight time from Boston to Washington is now about 30 minutes longer than it was 30 years ago. This increase is due to all the time aircraft have to wait or slow down due to crowding and lack of capacity, which constitutes real delay compared to the earlier situation. Yet despite the reality of delay in the system, the flights can be "on time" and thus have no official delay. Additionally, of course, definitions change and are often not the same from place to place, from one organization to another.[4] Bottom line: it is useful to discuss the meaning of data with relevant domain experts.

We also need to check the data quality. Data may contain errors or be systematically biased. Thus in the United States doctors appear to have systematically under recorded deaths from AIDS, as deaths from alcoholism were in the early 20th century, because of the social shame associated with these conditions. On the other hand, persons whose wages depend on the amount of their work have been known to inflate their workload records.

[Box 4.4 about here]

Develop appreciation for overall pattern: The standard process for estimating the trend of data over time is a regression analysis. This is a process for finding the line or equation that best fits the data. It is usually available as a function in a spreadsheet.

Regression analysis usually calculates the trend as a straight-line function of time t of the form:

Value of factor (t) = a (t) + b

The trend may also be exponential, to represent a constant rate of growth:

Value of factor (t) = b (e)$^{a\,(t)}$

If time t refers to years, its count is usually nominal (1, 2, 3…) rather than the actual calendar year. It is generally sensible to use expert judgment to choose a reasonable trend, which can include more complex patterns such as cycles. Regression is used to determine the parameters, such as "a" and "b" above, by fitting the curve most closely to available data. Appendix E gives examples of regression analysis.

The issue with developing a trend is that the result is sensitive to the period analyzed. If the period considered started in a recession and ended with a boom period, the trend might be increasing rapidly, going from relatively low to relatively high. If one takes a much longer period to cancel out the effect of cycles of activity, one may reach back into times when trends were completely different. In short, the results of a regression analysis depend on arbitrary choices of the length of time and data we consider. Box 2.6 and Figure 2.4 in Chapter 2 illustrate this fact. Thus the mathematically precise regression analysis inevitably leads to imprecise, assumption dependent, estimates of trends.

Assess the uncertainty in the trend: Regression analysis provides an immediate way to estimate the uncertainty in a trend. The statistical analysis generates the so-called "R-squared" metric, which characterizes the goodness of fit between the trend and the observations. By construction, R-squared can only vary between 1.0 (all the observations match the trend exactly) and 0 (the data do not fit any trend at all). Its complement, (1 - R-squared) is a measure of the variance around the trend.[5] The regression analysis thus provides an automatic assessment of the historical variability between observed reality and the trend line.

We can generate another and often more useful measure of variability of the trend line by carrying out regression analyses for data over different periods. The result is a set of different trend lines that provide an immediate sense for the range of possibilities. Figure 4.2 shows the kind of results possible. Such efforts provide a clear indication of the uncertainty in trends.

[Figure 4.2 about here]

Step 3: Identify trend breakers

The analyses in the previous steps are mechanical. Step 1 validates the data and Step 2 summarizes it in trends using standard mathematical analysis. Analysts are mostly organizing data; they do not have to think about it. However, step 3 requires careful thinking beyond the mathematical analysis.

It is important to identify “trend-breakers”. As Chapter 2 indicates, these events disrupt the smooth continuation of what has been happening in the recent past. They come in all kinds of forms: economic crises, political reconfigurations, new technologies, new discoveries, and new market conditions. Trend-breakers happen routinely. Any particular one may come as a surprise; but the record shows that we must expect such surprises. They regularly disrupt long-term forecasts and create new conditions that the designers and managers of systems need to face. The fact that they are routine and significant means that designers should pay attention to them.

Scenario analysis is the basic way to anticipate trend-breakers.[6] It attempts to identify possible “scenarios”, that is, sets of coherent major developments that might affect the industry and the system to be installed. The focus is on a general qualitative rather than a detailed quantitative description of possible futures. A scenario, in this context, is a single concrete and plausible future path that the project and its environment might take. It is largely narrative and it is thus accessible to a wide audience of decision-makers and stakeholders. The set of scenarios normally features one that is largely an extension of “business as usual”, complemented by others that represent different patterns of evolution, generally resulting from some disruptive events that break current trends. Box 4.5 illustrates some alternative scenarios for a particular case.

[Box 4.5 about here]

The development of scenarios implicitly identifies possible trend-breakers. A convincing narrative about a possible future has to include some rationale for the dynamics or chains of events that lead to a scenario that differs from a simple extension of the current situation. Once we focus on these trend-breakers, we can think about how plausible they might be. In Box 4.5 for example, the trend-breaker leading to the “unreliable energy market” scenario would be a war in the Middle East. Based on the record, this is unfortunately quite plausible.

Weighing the relative probability of scenarios is problematic, however. Indeed, it is unlikely that the future actually evolves according to one of the few scenarios. Scenario planning does not predict the future. It helps planning teams identify dynamics that can lead to trend-breakers which, in turn, could lead to different futures without having to go into the technicalities of quantitative assessment. The scenarios themselves are useful markers that indicate the range of possible future developments in a way that is accessible to a large audience.

A successful scenario planning exercise unblocks the thinking that recent trends will continue. It makes participants recognize that the world could be a very different place. It should expand the vision of the range of possible futures, far beyond the range associated with uncertainties in trend lines which are simply possible extensions of "business as usual". Scenario planning is a learning exercise for the planning team. It teaches them to recognize that their system may have to perform in an environment which could be quite different from the one that now seems most likely. Scenario planning thus leads designers to consider ways to enable their projects to transition effectively to possible futures; that is, to be flexible.

Step 4: Establish forecast (in)accuracy

We need to be realistic about our collective ability to predict accurately. Indeed, it is a professional responsibility. To fail to do so, to pretend that a point forecast is precise, is deceptive. Being careful to place reasonable ranges on forecasts is similar to scientists placing error bars on measurements. It should be routine good practice.

Unfortunately, forecasting professionals do not discuss the accuracy of their forecasts. This is because forecasts are, in practice, "always wrong" in that there is a significant gap between the anticipated and actual levels. As Chapter 2 illustrates, these discrepancies can easily be 30% or more over a decade. This is an uncomfortable truth both for clients, who want precision when they pay great sums for the forecasts, and for the forecasters who sell their services expensively on the basis of their great skills and advanced degrees. Forecasting consultants prefer to talk about statistical metrics, which are normally excellent.[7] However, excellent statistics on historical data is clearly not the same as make accurate assessments of future conditions.

The way to assess realistically our ability to predict is to compare forecasts made with the actual outcomes. The way to judge the skill of darts throwers is to see how close they come to the bull's eye. Conceptually, this process is straightforward. One compares the before and after numbers, that is, the prediction and eventual outcome, and notes the differences. In practice, this comparison may require considerable effort to find earlier predictions, often made a decade or more before. For example, the simple comparisons charted in Figure 4.3 took a professional several months to compile from the archival records of the relevant organization.

[Figure 4.3 about here]

The comparisons of past forecasts and reality provide a good indication of how well we can hope to do for our next project. If past forecasts were off by X percent on average over 10 years, then this number provides a reasonable estimate of the average inaccuracy of our current forecasts for the next decade.

Note that the process of comparing previous forecasts with subsequent reality can start at any time. Because of the possible difficulties in finding the results of previous forecasts, it may be expedient to start this process early, to allow as much time as possible to get it done.

Step 5: Build a dynamic socio-technical model

The value provided by any large infrastructure system depends on the demand for its services and the effectiveness and efficiency of their delivery. In financial terms, changing customer needs and competing services drive the top line; economic and operational variables such as availability, cost and productivity of labor and cost of goods drive the bottom line. In general, a system design transforms a flow of input variables into a flow of performance variables.

To estimate the value of a system, we need models that combine both the technical realities of a system and the social, economic and regulatory factors that will influence system performance. In short, we need socio-technical models.

The concept: The purpose of socio-technical modeling is to represent meaningfully the dynamic transformation of performance drivers, both technical and social, into performance indicators and relate it to design parameters. It is thereby to help designers explore alternative configurations of the system and thus position them to select the preferable solutions. The development of a good model can thus be an essential part of the design process. Poor models can lead the design process astray and destroy value, as the modeling for Boston’s sewage system demonstrates (see Box 4.6).

[Box 4.6 about here]

A good system model will correctly represent, to the extent possible, the way different drivers of the system affect the others and overall performance. It will be a “causal model”, that is, one that shows how its elements affect each other. A good system model will combine both technical and social aspects. It will represent the technical features (such as the way greater loads affect the size of the design), and it will represent social features (such as the way the cost of production raises prices and decreases demand for services). A good system model represents the complex socio-technological system, that is, the combined interaction of technical models with the socially driven environment. For example, a proper model for analyzing the development of an oil field will combine geology, structural knowledge and operating features on the one hand, with an understanding of the markets for oil and gas and the regulatory climate that may favor cleaner petroleum products.

Socio-technical models are extensions of traditional engineering models that change their nature profoundly. In contrast to our aspirations for engineering models, we have to admit at the start that we may never get the right answer from a socio-technical model. We cannot take the socio-technical system into the laboratory and test it in detail to validate the contribution of each component. This is different from models that seek to represent accurately technical performance, such as the flow of air through a jet engine.

Socio-economic modeling is both an art and a science. It is a science in that it uses statistics to enforce mathematical rigor in developing the models. It is a science in the sense that it enforces logical coherence by exposing the same encoding of all design alternatives to the same assumptions. It is an art because it requires creative simplifications of design alternatives and meaningful simplifying assumptions of the complex environments within which the system operates. It is an art in that statistics cannot prove a model correct.

The inability of statistics to prove a model is a fundamental concept. A useful mantra is: "Correlation is not causality". Just because a model offers good correlation with historical observation does not mean that it correctly displays the way variables interact with each other. Box 4.7 illustrates the point. It is routinely possible to develop several, often contradictory models from the same data. For example, in working on models of the use of water in Boston, our review identified over a dozen doctoral dissertations and other sophisticated analyses that proposed quite different models. Users should choose models that make sense.

[Box 4.7 about here]

Developing good models: To develop a good model we need to incorporate our understanding of the causal mechanisms as well as possible. We need to combine theoretical concepts (such as the idea that price increases generally reduce demand) and practical understanding of the mechanisms at work. Some of these will be broadly applicable, for example that the flu season increases the demands on emergency services at hospitals. Some mechanisms result from local conditions concerning the regulations and other practices that induce various behaviors.

Good blending of the technical and social factors requires careful statistical analysis. At this point, it is advisable to consult a professional statistician or econometrician to help analyze causal relationships rigorously. In fact, it may be tempting, and is indeed common practice, to outsource this analysis wholesale to a consulting firm. This can be a big mistake, as Box 4.8 illustrates. This practice gives away important learning opportunities. It is important that some of the planners and designers with wider responsibility remain closely involved in the analysis, in fact lead the analysis, and try to understand and interrogate the work of the statistics professionals. If the senior designers withdraw from the analytical work, the activity easily becomes a statistical exercise devoid of important context-specific insights.

[Box 4.8 about here]

Good models also capture uncertainty. They should certainly include the factors that are the drivers of design, as well as the major potential trend-breakers determined in the scenario analysis.

If our models do not account for uncertainty, they will not enable designers to estimate the value of any design accurately, due to the flaw of averages. Moreover, they will not enable them to determine the value of flexible designs, which only have if there is uncertainty.

Use of models: Designers use planning models to test alternative designs under conditions that are as realistic and as fair as possible. The practice is to expose design alternatives to the same set of assumptions about the system environment, and to calculate how these simplified designs fare in the created simplified worlds. A financial cash flow analysis, for example, will encode projects using associated cost and revenue flows over time, and compare alternative investments on the basis of common assumptions concerning demand and discount rate for instance. This is a fair comparison in some sense. We do not want to compare one design using a 10% discount rate operating under low demand to an alternative design using a 5% discount rate with high demand. Neither do we want to compare a project that includes headquarter overhead with another project that excludes these costs. However, the fact that the models lead to fair comparisons says nothing about their accuracy.

Accuracy of models: No model can expect to predict the performance of a complex socio-technological system accurately. This is important to understand. The accuracy of a model's performance projection depends crucially on the assumptions about the causal relationships and the simplifications made to set it up. Therefore, the results of models can only be interpreted relative to these inputs – and never in absolute terms, as in "the NPV of the project is X".

Once we accept this, it should be clear that the development of the model cannot be delegated to a team of accountants, statisticians or consultants, even though they may well be indispensable participants in the process. The designers and decision-makers themselves, those who are ultimately accountable for a project, have a responsibility to get involved in developing the models and its assumptions. It is surprising, how often analysts neglect this advice in practice.

We can estimate the accuracy of the model by considering the accuracy of its inputs and its correlation with historical data. One of the useful side products of the statistical analysis used to validate the model is the set of statistics, in particular the R-squared metric, that indicate how closely the models fit known situations. Put another way, these statistics provide a measure of the variability around the model, that is, an estimate of its likely accuracy. Additionally, we can estimate the relative accuracy of the variables in the model. For example, we can use comparable statistics to estimate the accuracy of economic predictions or of estimates of the size of petroleum fields. In practice, we use these estimates as inputs into the Monte Carlo simulations that will generate estimates of the distribution of the results of the model.

Hospital capacity planning case study [being reworked]

Take-Away

- Recommended procedure
  - Sufficiently broad ranges
  - Constant tracking of developments
- Take-away: Change process -- Don’t ask for the “right forecast” Go for range and scenarios

Box 4.1

Some examples of point forecasts that served as the basis for major designs:

- In 1988, forecasters predicted that there would be 1 million customers for satellite telephone service in 1998. This was a driving criterion for the design of the Iridium satellite communication system launched a decade later. In 1998, Iridium had only about 50,000 customers (see Box 1.1).
- In 1982, airport planner for the new international airport for Lisbon projected that it would serve 23 million passengers in 2010. By that date, the traffic was only slightly over half that prediction. It is unlikely that the government will have a new airport built as planned. (NAER, 1982)
- Xxxx Some example from Yun's research on A. hospital

Box 4.2

Identification of variables for airport planning in 2010 in Europe

The traditional drivers for airport development are the number of aircraft operations and passengers, separated into international and domestic, where applicable.[8] These drivers obviously connect directly to the design of the airside (the runways and taxiways) and the landside (domestic and international terminal facilities) of an airport.

A little thought indicates that designers should consider these drivers in more detail. On the landside, we should distinguish between passengers flying with traditional and low-cost airlines. This is because these types of airlines require different kinds of terminal buildings. The low-cost airlines typically want simpler facilities that will be cheaper. On the airside, we should distinguish between airlines operating many connecting services ("hub" airlines) and those that do not (point-to-point airlines). Connecting airlines want to concentrate arrivals and departures together to reduce connecting times for passengers and thus provide convenient service; whereas point-to-point airlines spread their flights more evenly over the day. Thus, the type of airline drives the amount of runway capacity needed.

Similarly, airport designers may in some circumstances want to think of drivers at a more aggregated level. In regions where airlines are merging, such as in Europe in the early 21st century, designers need to consider the possibility that their major airline might disappear, which would greatly affect the need for airport facilities.[9]

Box 4.3

Flexible designs for variable airport traffic

Conventional design for airport terminals creates facilities to meet a reasonable peak for the several kinds of drivers.[10] Airport designers may thus calculate the size of the facilities to meet the needs of international and domestic passengers, and sum up these spaces to plan the overall building size.

Better practice recognizes that the peaks of international and domestic passengers can occur at different times. Domestic passenger traffic often peaks during the morning and evening rush hours, while intercontinental traffic peaks in mid-afternoon or late at night due to the constraints of time zones. Designers can thus save significantly on the overall size of the building if they make a significant fraction of the space flexible, so that it can serve domestic passengers when their flow peaks, and then international, intercontinental passengers during their peaks. Creating this kind of “swing” space reduce the overall requirements by a third or more, thus correspondingly increasing design efficiency.

Canadian airports, Vancouver in particular, are notably efficient in using flexible space for international and domestic passengers. They have created partitions and doors that allow them to secure aircraft gates or passenger lounges for either use. Similarly, airport designers can create common passenger lounges serving many flights, and greatly reduce the overall space required compared to that required to serve each flight individually.[11]

Box 4.4

Water use for Boston

Designers based the original plans for the $6 Billion sewage treatment facilities in Boston Harbor on historical data from the Metropolitan Water Resources Agency. These showed steadily increasing consumption then already over 300 gallons (over 1100 liters) of water per person per day. Accordingly, the original designs extrapolated future daily use to about 350 gallons per person.

Did the data mean that Bostonians actually washed, drank and flushed that much? Actually not. It did mean that the Agency sent that much water from the Quabbin Reservoir into the Boston system. However, about 1/3 of this seems to have been lost through leaks in the old distribution system. The sewage plant therefore could be much smaller than initially planned.[12]

Did the trends demonstrate that usage would continue to grow? Not either. Once the authorities built the facility, they raised the price of water to cover the costs of construction. Households went from paying about $20 to around $500 a year. Higher prices encouraged them and industry to repair leaks and install water saving devices. The overall consumption of water thus decreased about 20% compared to the earlier baseline, to around 70% of basis for design.

Box 4.5

Possible scenarios for power system in Europe

In considering major investments in electricity generating plants in Spain, three scenarios seemed relevant. They illustrate the meaning of scenarios. In short form, they were:

Business as usual: No major departures from current trends. Supplies of fuels remain reliable. Driven by steadily higher worldwide demand, fuel prices increase faster than inflation. Greater concern for global warming leads to strong incentives to invest in carbon capture. Regulations continue to be favorable to private investments and adequate returns.

Unreliable energy markets: Unrest in the Middle East disrupts deliveries of LNG (Liquefied Natural Gas), and leads to much higher prices for unreliable supplies. These circumstances enable Russia to extort extraordinary prices for guaranteed steady supply. The European Union steps into this crisis and limits price increases by the suppliers of electric power, effectively taking over their assets at major discounts.

Plentiful nuclear power: Worldwide investments in nuclear power provide electricity reliably at a manageable price. This leaves LNG producers with excess capacity and drives their prices down substantially. Meanwhile, the switch to nuclear reduces CO_2 production and lowers the price of carbon permits. Lower costs of fuel and carbon permits make LNG power plants very profitable.

Each scenario that differs from “business as usual” embodies one or more trend-breakers. For example, either chaos in world energy supplies or a crash of the LNG markets could break the existing trends in power production in Europe, and greatly shift production opportunities for electricity producers.

Box 4.6

Modeling for design of Boston's sewage system

The design of Boston's $6 billion sewage treatment system rested on a simple model of the interactions between the social and technical conditions. It was that:

- Sewage to be treated = function of (Population in service area and per person water use)

This was inadequate. It left out the crucial contribution of economics. The result was a poor design, oversized and configured for the wrong mix of inputs. The system its construction was unnecessarily expensive build, and its operations are inefficient.

An appropriate model would have recognized that the implementation of the new sewage system would greatly increase the costs to users, which depend on their metered use of water. Customers would then reduce consumption by fixing leaks, installing low-flush toilets, and so on. Moreover, the installation of water saving devices would necessarily take years. In addition, the price of water would rise over time since the requirement to pay for the fixed costs of the sewage system would force the operators to raise the price per unit as usage decreased over time. Finally, these adjustments would change the nature of the sewage and thus the operation of the treatment plant: a lower proportion of water in the mix, for example, affects the pumping requirements and changes the chemistry of the biodegradation process.

A suitable model would have had at least the following elements:

- Sewage quantity = f(Population; Price of water changing over time, Usage changing over time)
- Sewage quality = f(Price of water changing over time)
- System costs = f(Sewage quality changing over time)
- Price of water = f(Annual costs of system, Usage changing over time)

Box 4.7

Correlation is not causality

It is obvious that the number of firefighters and fire trucks at a fire is a good predictor of the amount of fire damage. These factors correlate closely. We can confidently assume that something big is going on if we see many firefighters at a fire.

This model is however a terrible model for the design of a system to control fires and damage. We should definitely not conclude that sending in fewer firefighters would reduce the damage from a fire – quite the contrary!

This is an example of spurious correlation. The correlation is real, but the causality does not exist. The actual situation is that “Big Fires” cause “Big Damage” and trigger “Dispatch many firefighters”. The correlation between “Big Damage” and “Many firefighters” is there, but the latter does not cause the former.

Box 4.8

Designing emergency rooms for a hospital

The ER facilities at a New England hospital were consistently crowded. The hospital decided to fix this problem and called upon a consultant to develop a model of ER use, and to propose a solution. These California consultants applied a national model to the situation and recommended a large expansion of about 25 beds.

The consultants’ model was not appropriate for the hospital. They developed it without detailed input from the hospital directors and staff, and thus without an understanding of the local context. If they had done so, they would have realized that in this particular situation much of the crowding of the ER was due to the local administrative habits for discharging patients from the hospital. Indeed, the local practice was to discharge patients late in the day. This meant that ER patients needing to be admitted to the hospital systematically crowded the ER waiting to move on. As the hospital recognized this causal chain, and thus moved to earlier discharges, much of the crowding in the ER disappeared.[13] The consultants’ model thus led to a wasteful over design.

The moral of the story is senior decision-makers and experts in the system should be part of the process for developing system models.

Figure 4.1. Elements in chain of uncertainty in a system

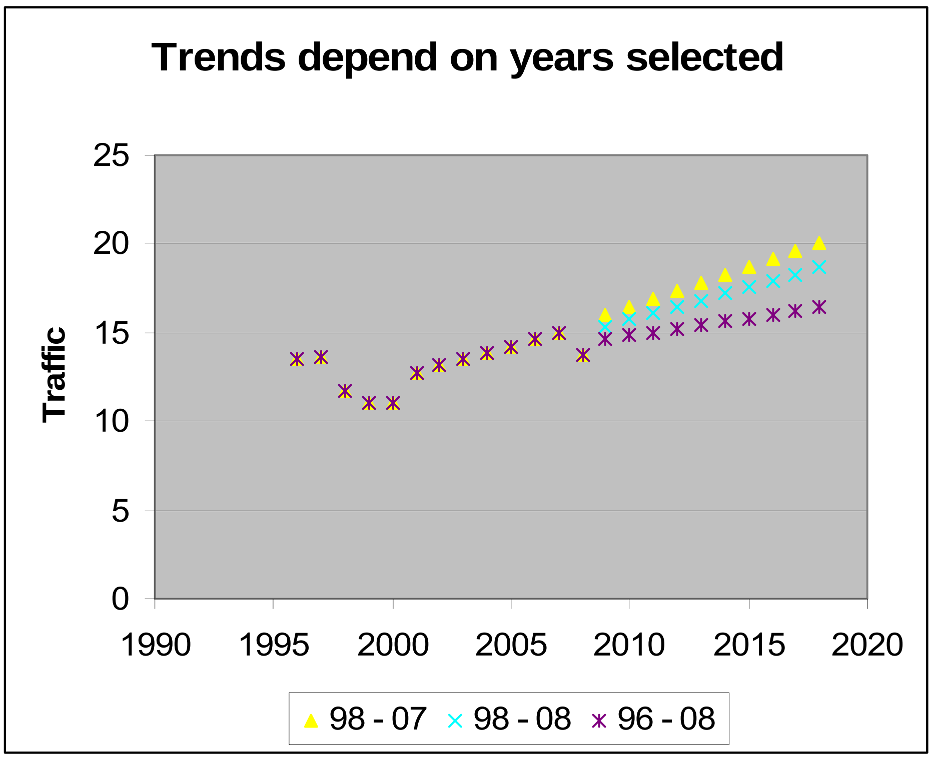


Figure 4.2. Traffic data for a major US airport (reprint of Figure 2.4)
(Source: Based on data from the US Federal Aviation Administration)

Find suitable figure (or a graph) from Yun

Figure 4.3. Comparisons between forecast and actual data type for Addenbrooks Hospital in Cambridge, England. (Source: Lee, 2009

CHAPTER 5
STEP 2: IDENTIFYING CANDIDATE FLEXIBILITIES

"Everything should be made as simple as possible, but no simpler."[14]

Attributed to Albert Einstein

This chapter describes ways to identify the most valuable kinds of flexibility for a system. This is an important topic. Having accepted that the future is uncertain, and thus concluded that it would be good to have a flexible design, the question is: How should we implement flexibility? We need to know: What parts of the system should be flexible? How flexible should they be?[15]

It is not obvious which flexibilities will add the most value to a project. The answer depends on many interacting factors, such as the:

- Nature of the system: What it takes to make an automobile plant responsive to changes in the market demand for different types of vehicles is different from the kind of flexibility that would be useful in developing a copper mine. The automobile factory might benefit from having multi-purpose robots; the copper company might best choose to access the ore deposit with an open pit rather than through tunnels.
- Kinds of the uncertainties: Are these mostly on the supply side, such as the state of our technology? Or are they on the demand side? In detail, for example, are we mostly concerned with the size of the market, as we might be for a new product? Or with the price we might get for the product, as when we develop a commodity such as oil? Are we thus interested in the flexibility to expand easily, or to redeploy capabilities and equipment to parts of an oil field that might only be profitable at high prices?
- intensity of the uncertainties: Is the system well-established and slow to evolve, perhaps like a rail transit network in a big city? Or are we dealing with a situation that is fluid and subject to rapid changes, such as the air transport industry in a deregulated environment? Are we thus interested in long-term expansion options, such as the ability to double-deck a bridge? Or should we focus on short-life design that enables us to reconfigure a project for different types of users or aircraft?
- Cost of implementing the flexibility: How much might it cost to create some form of flexibility in the first place? How much effort is needed to take advantage of this adaptability later on when it might be useful?

We can in general incorporate many different kinds of flexibilities into the design of a system. A number of these will improve value. But which are likely to add the most value to the system? Which should we focus on?

This chapter provides a practical guide to best current practice for determining the better kinds of flexibility to build into a system. The recommended approach is to use screening models to examine the design space rapidly, and thus to identify attractive design concepts that we can explore in depth. The chapter first presents the three ways to develop screening models: bottom-up, simulator, and top-down.

Each of these can be useful. The choice between them depends on the details and the nature the problem. The chapter concludes with a case study of the use of screening models to identify desirable flexibilities in the configuration of automobile factories. For further guidance on how to identify interesting possible flexibilities in specific fields, please consult the case underway, involving a wide variety of case studies, and the field is developing rapidly.

Concept of Screening Model

The recommended approach is to use “Screening Models” for preliminary identification of the most desirable candidate flexibilities. Screening models are simple, understandable representations of the performance of the system or project under development. The chapter first defines the concept of screening models and shows why we need them. It then presents in detail the main approaches to developing screening models: bottom-up descriptions of detail, “simulator” models, and top-down conceptual representations. Finally, it explores how we should use screening models to identify desirable candidate flexibilities.[16]

A screening model is a simple representation of a system that can be used for quick analysis of the performance of a design. For example, the screening model used to describe the economics of alternative designs for oil platforms might take a couple of minutes to estimate the value of a specific design under a single scenario. In contrast, the full oil-and-gas model used for final design might take a day or more to do a complete analysis of the same case. The complete model combines detailed descriptions of the design of the platform, the behavior of the flows in the oil field, and the economics over time. Different expert groups will have developed each sub-model and they will not have fully integrated them with the others. Such complex models are common in the design of technological systems. Obviously, we cannot use models taking a day or more to examine a single design for a single scenario to examine and compare several designs over thousands of scenarios – there are not enough days in the year. By contrast, screening models allow us to look at thousands of alternatives very quickly, typically in a day or less. Speed is and must be a primary characteristic of a screening model. Speed enables a screening model to fulfill its function of considering how many different designs perform over many different scenarios.

Box 5.1 about here

We need fast models so that we can usefully identify the best opportunities for creating value. This is how we can explore the extraordinarily large number of combinations of possible designs and paths of evolution of uncertainties over time (see Box 5.1).[17] We need efficient means for searching for good designs, to complement our intuition or experience with what worked in the past. Intuition and creativity are important assets, but can often lead us astray, especially about complex combinations that we have not experienced. Researchers have repeatedly shown that our intuition on probabilistic effects is particularly weak. Standard psychological tests, of the kind used in introductory classes on behavioral science, demonstrate our collective difficulty in dealing with uncertainly. People rapidly fix their thoughts rather arbitrarily on notions that often turn out to be wrong.[18] Likewise, experience about what has functioned elsewhere, under other circumstances, is not necessarily a solid guide to what may be best in a specific situation. As designers, we need rigorous analytic processes to complement our valuable intuition and experience, and to help us to examine the many alternatives. Because of the large number of possibilities, we need fast models to do this job.

We also need simple models. Because the initial design phase combines both intuition and analysis, we need models whose workings we can understand. Large complex models, put together by many different experts (some of whom may have left the organization years ago) are unlikely to enable us to enhance our intuitive thinking.[19] Conveniently, we can also analyze simple models quickly.

Screening models should also be understandable. A simple structure makes it easier for planners and analysts to see how the parts affect each other, and thus to appreciate how various decisions may effect the overall performance of a system. An understandable model helps build confidence in the analysis process. Decision-makers are more likely to accept a model, and consequently the resulting analysis, when they can see how and why it works. By contrast, a complex model whose interactions are unclear, may appear like a “black box” that cannot be trusted – often rightfully so!

Screening models are thus both qualitatively and quantitatively different from large complex models. They should be easier to understand from an overall perspective so that system planners and designers can develop confidence in them and believe their guidance. They should be simple so that they can rapidly scan the wide range of possible designs and thus identify configurations that are worth examining in detail.

We use screening models to find attractive possibilities for configuring and operating complex systems. We use them to explore quickly the range of design possibilities and scenarios of uncertainty. If we think of all the possibilities as an unexplored continent, screening models enable us to fly over the terrain at high altitude and identify interesting peaks. Once we have found some high-points, we can investigate their possibilities in detail. The name for screening models describes their function: they screen or filter out interesting possibilities.

Screening models complement the complete, complex models needed to design a system in detail. They are not substitutes. Screening models are most useful in the early stages of the design process, when we are trying to decide the overall features of some system or subsystem. They help us investigate what kinds of flexibilities we might best incorporate into the design, how we might organize the system into modules, and how we might phase the development over time. Screening models help us understand and define interesting configurations or ‘architectures’ for the system. Once we have used screening models to identify interesting design possibilities, we then have to examine them in detail with more complete, detailed analyses. Figure 5.1 shows the sequence.

Figure 5.1 about here

Best practice uses screening and detailed models together. In the first stages of a project, we should use the screening models more, to develop a rough understanding of the values of a broad range of designs. Once we have narrowed down the range of design alternatives, we use the complex models to test and validate the general results of the screening models. Later in the design process, the detailed models refine and modify the design configurations suggested by the screening models. We may also use the screening models to explore particularly interesting features that have emerged in the design process. The screening models thus have the more prominent role at the start, and the complete detailed models dominate as the project nears final design.

Development of Screening Models

A good screening model should have two principal features: it must be fast and should rank alternatives reasonably correctly. A screening model must necessarily be fast, so that it can provide a quick way to look at designs under many different scenarios. A good screening model should also rank designs so that the short list represents excellent candidates for detailed examination. To fulfill its function of identifying good candidate designs, it must correctly identify which are most preferable, in comparison with the others.[20]

Note carefully that screening models do not have to be accurate, in the sense that they measure the performance of a design precisely and correctly. That is not their function. Their role is to define a short-list of possibilities. The detailed models will refine the design specifications and will use them to provide exact assessments of the performance of the designs. We need to recognize and accept that screening models will not be precise -- they are stripped down representations of the performance of the system, and are necessarily approximate.

The distinction between the requirements for screening models and detailed design models is important. Professionals who normally work with the detailed design models, who probably have contributed to the refinements and precision of these models, will naturally resist the use of simple models. They will consider the screening models to be poor substitutes for the “real thing”. They will want to “improve” the screening models by making them more like the detailed models. We must resist the tendency to complicate the screening models unnecessarily. We need to stress that screening models are not substitutes for the fully detailed models; they are complementary assistants. The screening models are scouts, as it were, looking for good opportunities for the troops to exploit with a detailed design.

Types of Screening Models

It is useful to think of screening models as belonging to three generic types:

- Bottom-up models – simplified versions of the detailed descriptions of the system, built up from understanding of the parts;
- Simulators – that mimic detailed descriptions of the system; and
- Top-down models – conceptual representations of the overall pattern of behavior of the system.

These categories refer to how analysts create screening models. They indicate different paths experts take to arrive at the goal of creating simple models they can use to understand a system and define good opportunities for flexibilities that will increase value.

In practice, a screening model for a particular problem may be a combination of elements. A top-down process might describe the interactions between major modules of the screening model of a system, and a bottom-up process might represent the working of particular modules. For example, analysts might set up a top-down model to describe the overall interactions between the emergency and in-patient facilities at a hospital, the way patients transfer out of emergency care and get admitted to regular wards, and how admissions patterns, health care policies, and operational constraints determine the availability of beds in the hospital. Within this context, a bottom-up model might detail the operation of emergency facilities, for example, how doctors triage patients and move them on to the next stages of care. Analysts can construct a useful screening model, one that provides the means for rapid exploration of alternatives, in many ways.

Bottom-up Screening Models

Bottom-up screening models have been more common. This is because they are easier to develop. They build directly upon the detailed technical knowledge of the professional teams responsible for the development of a system. The development of a screening model from the bottom up thus starts with considerable knowledge about the system. The challenge is to simplify meaningfully the already existing knowledge of the system complexities.

A major obstacle in creating a screening model from the bottom up is the possible difficulty of meshing the detailed knowledge about the parts of a project into a coherent whole. We may face this problem when the project we are considering is an assembly of unrelated elements, or when it is unprecedented – such as the rehabilitation of a large area for the development of the London Olympics. In such cases, we have to create an overall model for the complete system. The difficulty then is that the different specialists associated with the project will have computer models based on different kinds of data, organized at different levels of detail, and intended for different purposes. Often, these specialists have not worked together closely, so they have not coordinated their efforts. In this case, the development of a suitable screening model for the system may require a significant effort, as is the situation for developing integrated models of forest fires (see Box 5.2).

Box 5.2 about here

The usual situation is that the professionals concerning with a project already have some integrated set of models to describe the system. Typically, the project team has been working on similar projects for some time, and has developed ways to coordinate these models into a suite that functions together (see Box 5.3). These models bring together knowledge about:

- Technical possibilities, which the engineers and planners can implement;
- Environment factors displaying the effects of climate or the dynamics of flows over time; and

- The economics or values that may be generated by a project.

Such suites of models provide a good basis for the development of a suitable screening model.

Box 5.3 about here

The development of a bottom-up screening model is relatively easy if we can start from an existing model of the system. We can then implement them without much special effort. The idea is to reduce the complexity of the inputs, which in turn simplifies the description of the problem and thus reduces the time it takes to do an analysis. For example, analysts can replace complex input dynamics with time-averages or simple linear trends. They can also aggregate periods of analysis, working with years rather than weeks or months. When they apply such simplifications, they cut down the size of the problem by reducing the number of variables (by over 90% if we move from monthly to annual data), and correspondingly the number of associated constraints. Since the time to carry out an analysis is roughly exponentially proportional to the square of the number of constraints, these simplifications speed up the analysis considerably. Moreover, the smaller problem is much easier to specify and input into the machine, which often is a critical consideration for the analysts.

When simplifying a complex model, we have to be careful not to toss out detail that may be crucial for the value of the project. We must not throw the baby out with the bathwater, so to speak. We need to think about what we are doing; the loss of critical information typically seems to result from lack of attention, either individual or institutional. Standard procedure in one oil company illustrates the point. The geologists in this corporation have detailed models, backed by Monte Carlo distributions, of the uncertainty associated with the production of a field. Yet when they pass their information over to either the finance or the design departments, they simplify their data to a single number, such as the most probable value. As uncertainty about the amount of oil in a field is a crucial driver of value, its systematic elimination is clearly an excessive simplification.

Simplified bottom-up models are common. Overall, the ease of creating the simplified models, combined with the effectiveness of this approach, accounts for their widespread use.

Box 5.4 about here

Simulator Models[21]

“Simulator” models provide an excellent way to develop screening models. The idea is to represent the output of a suite of models representing a system through a formula or simpler set of models. In contrast to the simplified models, that change the inputs to the integrated model, the simulator approach focuses on the outputs.

The development of a simulator model is largely a statistical exercise. The idea is to fit a simple model with a few parameters, to the output of a complex model with many variables. The process has two steps:

- Identifying the most important variables (hopefully few); and then
- Fitting a simple functional form to a series of model responses.

In detail, there are two ways to develop simulator models: direct and indirect. The direct approach is a process for replicating the output of the complicated, detailed model. It ignores the interior mechanics of the system and focuses on mimicking the overall results. The indirect approach is more subtle. It develops new, approximate models for the components of the complex model, and accepts them if they produce overall responses that are sufficiently similar to the results of the complex models.

.

Direct Approach: The direct approach process for developing a simulator model goes through the following steps:

1. Select a set of system requirements and parameters the system must respect, define the range of values that these might have in different situations, and specify a limited number of values for each of these parameters over these ranges. For example, the price of oil is an important parameter for determining the value of an oil field. The analyst might choose to assume that it will vary over the range between 20 and 140 $ per barrel, and fix on considering specific values $20 apart, that is prices of $20, 40, 60, up to 140 per barrel.
2. Run the integrated system model to find the output of the system for all the possible combinations of the chosen values of the parameters. For the oil field example, this would be for all the combinations of the price of oil, the estimated size of the reservoir, the capacity of the pumping facilities and so on. (This is a “full factorial” exploration of the response of the system.) The result, for instance, would be the net present value for each combination of parameters.
3. Develop an equation that most closely correlates the input parameters to the results. This step involves a standard statistical analysis.
4. The resulting equation is the simulator screening model.

Box 5.5 presents the results for an application in the configuration of oil platforms.

Box 5.5 about here

The direct approach to developing a simulator model is particularly attractive because of its ease of development and use. The development process does not require detailed knowledge about the inner elements of the system. Since the statistical analysis is elementary, all we need to develop a simulator model is the ability to run the integrated model and the time to do so for enough combinations to provide a satisfactory representation of the overall response of the model. (What is enough depends on whether the values generated change rapidly or smoothly, and on the cost of running the models through more combinations of the parameters.) The simulator could not be simpler to use: It is an equation with a limited number of parameters – something that a computer can calculate in “no time”.

The drawback of the direct approach is that is offers no possible insight into the inner workings of the system. The simulator may tell you, for example, how the cost of a system varies with its size – but the equation gives no clue as to why costs vary as described. The simulator does not allow you to look into the system – the final equation is all there is.

A simulator model has the valuable advantage of allowing the analyst to explore the overall behavior of the system and verify that this makes sense. This is a particularly important capability. Indeed, it can easily happen that the integrated model of the system gives wrong, even impossible answers – but that nobody notices. The structure of an institution may create this kind of myopia. This can occur when the area specialists developing the detailed model have focused on the components within their particular competence, and did not get to appreciate the overall behavior of the integrated model. The experts in a limited area are responsible for only a small part of the model and may never get to see the big picture. Similarly, the analysts working with the entire complex model may be too busy using it in a project that they do. They may not have time to explore its overall features. In any case, it is desirable to verify that the overall complex model of the system performs as it should. A simulator model provides an easy means to accomplish this validation.

Errors in the overall model may occur when its components embed assumptions appropriate to a limited context but not to the overall system. The possibility that the integrated model may have fundamental flaws is counter-intuitive. Analysts often assume that an integrated model must be accurate since it combines the detailed knowledge of specialists. It does, of course, but the integrated model may lack an appreciation of the overall behavior of the system. (See Box 5.6.) A simulator model may then provide a valuable check on the modeling process, and may point out the need to revise the complex detailed model of the system.

Box 5.6 about here

Indirect Approach: The indirect approach to developing a simulator model requires much more effort than the direct approach. It also requires substantial understanding of technical, scientific and social elements that are involved in the system or project. Correspondingly, its results may provide a much deeper understanding of performance.

The essence of the indirect approach is to develop and use simple models that we can substitute for the complex models used for the detailed design. These new simple models are not simplified versions of the complex models, developed by dropping out complexities or linearizing non-linear relationships. The analysts develop these models from higher-level considerations (such as the requirement for mass balance), and adjust their parameters so that the simple models satisfactorily replicate the behavior of the detailed models used in design. Box 5.7 explains this by example.

Box 5.7 about here

Top-down Screening Models

Top-down models provide an overall view of the system. They show how its major parts influence and interact with each other over time. A top-down model can usefully reveal complex interactions. In a hospital for example, the construction of new wards might reduce financial losses – save money – by reducing the way emergency surges (for example, during the flu season) force the cancellation of elective procedures and thereby increase the utilization of operating theaters and expensive specialist personnel and equipment.

Top-down models are particularly useful to the extent that they provide insight into the dynamics of a system. As the hospital example suggests, a change in one area of the system frequently has complicated knock-on impacts on quite different areas, often with some delay. Indeed, changes to a complex system rarely only lead to immediate effects, such as caused by flipping a switch and turning on a light. Normally, and especially when humans are an integral part of the system, developments in one area change the conditions for participants in other areas, which in turn ultimately affects their behavior. It is important to understand these ripple effects. We cannot plan intelligently without thinking about the sequence of potential consequences of our actions.

It takes care and effort to develop insightful top-down models. The work requires two types of investigation. First, we need to document the chains of physical interactions to answer the question: how do changes in one area affect the operations in others? Complementarily, it is necessary to develop an understanding of how the humans in the system – such as the doctors, nurses and patients associated with the hospital – react to the physical changes. For example, cancellations of elective procedures can give a hospital a bad reputation for elective care and depress demand. More patients might choose to use the hospital once they learn that services there have improved. For instance, they might go immediately to emergency when they break a finger, instead of waiting to go to their doctor's office during the day. In general, good top-down models require an integrated perspective on both the mechanical and behavioral aspects of the system.

Systems dynamics models have proven to be a good way to present top-down models. Systems dynamics is a modeling procedure based on 50 years of development and supported by a range of standard software.[22] Its design makes it easy to describe the interactions between different elements of a system, both mathematically for the computer and visually so that planners and designers can appreciate the way that changes in one part of the system affect other parts. (See Box 5.8.)

Box 5.8 about here

Use of Screening Models to Identify Candidate Flexibilities

We can use screening models to identify candidate flexibilities in three general ways: conceptual, optimization and patterned search.[23] The conceptual approach relies on the simplicity of the model. It exploits this feature to enable analysts and decision-makers to think through the issues and agree on possible designs for in-depth analysis. The optimization approach uses mathematical tools to home in on likely best designs. It is feasible in the special cases when the structure of the models makes it possible to use optimization techniques. Patterned search uses simulation to explore many alternatives systematically. It is the more general technique, which can be widely applied.

Which approach is best depends on the problem and situation. We can also combine the approaches, for example as when the conceptual approach provides guidance to a patterned search. The following sections illustrate the use of the approaches.

Conceptual Approach: The conceptual approach is particularly useful in getting planners and designers to "think outside their box". An overall perspective on the functioning of a system inherently shows the interactions between the different functions of the systems. It thus brings these interactions into focus for the specialists and experts for the particular functions, and allows them to appreciate how their performance depends on that of other functions. It thereby opens up opportunities for exploration of new design opportunities and of candidate flexibilities.

Some examples illustrate how the conceptual approach works. (See Box 5.9). As these show, the experts in the several functions of the system often know about the specific linkages between the functions in advance. These linkages may be common knowledge to the participants. Yet in practice, analysts often ignore these linkages because they are outside the participants' immediate areas of expertise or responsibility. In such situations, the role of the screening models can be to spotlight the importance of the interdependencies between functions.

Box 5.9 about here

Optimization Approach: The optimization approach works best when applied to systems for which a single coherent model exists to which we can apply some optimization method. Such models exist in many fields, in particular those that we can represent by networks of flows. A supply-chain network connecting factories, warehouses and end users is one example. Others involve the distribution of electric power or the development of a river basin.

The situations for which the optimization approach works well contrast with other fields in which the system model is an assembly of quite different models. The latter are not only common but may be prevalent. The oil-and-gas model for the evaluation of oil field developments is one example: it combines distinct models for the flow of oil and gas through porous media, for the operation of the oil platforms and related mechanical processes, and for the economic evaluation of the value of the entire system. Similarly, the models for communication satellites combine distinct technological elements for the use of the information bandwidth and the deployment of the satellites, and for the estimation of future demands for service and the overall value of the network.

When we can describe the system with a coherent model suitable for mathematical optimization, the procedure to identify interesting candidate flexibilities is:

1. Optimize the design for one set of contextual conditions (such as estimates of future demand and prices), normally chosen in the mid-range of the range of values expected.

2. Repeat the optimization process for several levels of the major contextual conditions, to observe how the optimal solution changes. This step is similar to running a "data table" command in a spreadsheet.
3. Observe which elements of the design change when the contextual conditions change – these are the ones in which there should be the flexibility to adjust the design. Conversely, the design elements that are insensitive to changes in the contextual conditions do not need to be flexible.

The planning for the development of a river basin in China illustrates this approach. (See Box 5.10.)

Box 5.10 about here

Patterned Search: A patterned search is similar to the optimization approach, insofar as it systematically tries out different types of design. The difference is that the patterned search is not directed by a set of procedures that systematically lead it to an optimum solution; it does not benefit from a mathematical process to determine which designs might be optimal. The designs explored by a patterned search depend upon guidance developed from conceptual models, the design team experienced in the area, or from similar cases for projects that seem comparable.

Absent any specific guidance, the default is to consider standard alternatives for providing flexibility. This is a straightforward process that tries out flexible designs that have proven to be effective in other situations. These include such possibilities as:

- Phased design: developing capacity for production in smaller units enables the system operators to limit capacity if it turns out not to be needed, and to time its development for later according to the growth of demand. This approach has the advantage of deferring costs, if not eliminating them entirely, and thus of increasing present values. Building modules also reduces costs as the developers learn how to produce units more effectively.[24] As a counterweight, building smaller may forgo the economies of scale of larger construction.
- Modular Design: designing the system for "plug-and-play", that is, with the capability to accommodate the addition of new features through simple connections. For example, USB ports allow computer users to add all kinds of capabilities to their existing systems.
- Design for Expansion: creating the system with the built-in capacity for expanding its size. Prime examples of this are bridges originally built with the strength to carry an eventual second deck should the demand arise, such as the George Washington Bridge across the Hudson in New York City, and the Ponte 25 de Abril across the Tagus River in Lisbon.
- Platform Design: creating a "platform", that is a basis on which the system managers can create many different designs, already conceived or yet to be developed. This approach is well-known in manufacturing for major consumer goods. For example, automobiles manufacturers routinely put several different styles of chassis on top of a common base and wheel train.[25] Likewise, makers of power tools often have a standard core, consisting of a grip and motor, to which they attach different kinds of business ends.

- Shell Design: in which designers create the capacity for some future use, without immediately dedicating it to any particular use. In a number of cases, hospitals have built extra space for which they have no immediate use, but which they can later fit out for doctors' offices, wards for patients, or even operating rooms. They recognize some sort of future need, but are reluctant to commit to expensive completion until future requirements become evident.

Patterned searches often deliver excellent results in terms of increasing the expected value of the system. This is because each of these flexible designs systematically provides system managers with the ability to adapt efficiently to what they may eventually need, while avoiding the losses associated with investments that turn out to be unneeded. The flexible designs both increase the expectation of gains and reduce the possible losses. Flexible designs thus tend to produce net gains in expected value.

Example Application: Configuration of Automobile Factories

A team from the MIT Materials Systems Laboratory developed a screening model for the production of automobiles in plants of various sizes. In this case, the models representing the cost of production consist of sub-models for each of the different processes in the plant, each worked out in detail through years of collaborative research between the MIT and major car manufacturers.

The simulator model for the cost ??? of the output of xx plants was of the form

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

Develop this case form Yang's dissertation.

Take Away

The recommended approach to identify the most valuable kinds of flexibility for a system is to use "Screening Models". Screening models are simple representations of the performance of a system that require little time to run. They thus contrast with the complex models used for detailed design, which may take days to run, and which thus constrain designers to examine only a few alternatives under limited conditions. Screening models thus enable designers to analyze the performance of alternative developments under all kinds of conditions. Screening models provide the way to determine which flexible designs may offer the greatest value.

Screening models come in different forms. They may be bottom-up versions that are simplified versions of the complex, detailed models used for design. They may be simulator models that statistically mimic the complex models. They may be top-down representations of major relationships between the parts of the system. The most appropriate approach depends on the situation and what is available.

Analysts will identify good candidate flexibilities by exercising the screening models. Sometimes the analysis will be conceptual. Analysts, system managers and other stakeholders in the system will in this case use the screening model of the system as the basis for thinking outside of their professional boxes to see how flexibilities in the overall system could be beneficial. Most often, designers will use the screening models to test the possible flexibilities over the many combinations of possible uncertainties.

The overall result is to define good candidate flexibilities, that is, a short list of possibilities. These candidates developed using screening models, which are deliberately simple and thus incomplete, need validation. We need to use the standard processes for detailed design to examine each of the short-listed possible flexible designs.

Box 5.1

Large Number of Possible Combinations

The number of possible designs for any system rapidly becomes huge. Consider the simple problem of developing warehouses at 10 locations, which can either be “large” or “small”, over 3 periods. The number of possible designs for this simplistic problem is 10 squared to the power of 3 = $(10^2)^3$ = 1 million.

The number of possible development paths for some uncertainty over time is likewise very large. Consider for example the development of an open pit copper mine over 20 years. Its overall value depends on variation of the price for copper over time. If copper sells for a high price early in the development, this will lead to higher present values than if high prices occur later. Furthermore, different prices will lead to different development strategies, known as ‘mine plans’: When prices are higher, it is rational to focus on immediate exploitation of richer deposits, to take of advantage of immediate opportunities. When prices are lower, it may be better to concentrate on removing overburden and to position the mine for rapid access to deposits later on when prices are higher. Therefore, we should consider the effect of possible sequences of prices. If we simplistically think that prices can either rise or fall some amount (say 10%) from year to year, then the total number of possible distinct paths is 2 to the power of 20 = (2^{20}) = over a million!

In general, it takes several hours to develop a development strategy for a mine for a single sequence of prices. This is because the mine plan is the solution to a very large optimization problem. Analysts divide the volume of the into cubes, perhaps 20 meters on a side. There may be 100,000 of them in a major mine. Each of these blocks has an estimated content of ore. The mine plan thus defines the most profitable sequence of removing these blocks, subject to physical constraints (blocks on top have to come out first) and capacity constraints of the mine equipment. It is not practical to run this optimization thousands of times. We thus need some simpler model to enable us to consider the effect of the range of possible prices sequences over time.

Box 5.2

Creating a System Model: Forest Fire Abatement

Forest fires are a major concern in many parts of the world, especially in the large timber growing areas in dry regions such as Australia, Indonesia, Portugal, Russia and the Western part of the United States. "At the beginning of the 21st century, the…Earth … annually experienced a reported 7 to 8 million fires with 70,000 to 80,000 fire deaths.[26] Forest fires also often destroy towns and have great regional impacts. These impacts motivate the search for effective strategies to reduce the impact of forest fires.

The design of a regional fire abatement program needs to integrate several major components such as the:

- Climate and Meteorological patterns of wet and dry seasons;
- Topological descriptions of the terrain, which affects how fires spread;
- Descriptions of the composition of the forest, as different species burn differently;
- Detailed models of the dynamics of fire fronts advancing through a forest; and the
- Effects of firebreaks and other actions to limit the spread of fires once started.

Furthermore, we have to model all these elements over time, as forests mature and ground cover changes.

Many experts have developed these elements of the system, typically in different organizations, often in different countries, certainly in different ways. Integrating them into a coherent whole requires a special effort.[27]

A useful screening model for this system will have to simplify many of these models. For example, much of the detail about how a fire front advances will not be useful in defining how we should develop the most effective regional fire abatement strategy. A simplistic fixed advancement rate may well be preferable at this high level of decision-making. Note carefully however that using an average instead of a distribution of possibilities could lead us to the Flaw of Averages unless we are careful!

Box 5.3

Examples of Integrated System Models

River Basin Development: Planners for the development of river basins use models that integrate:

- A “river flow model” that describes the downstream progress of water both naturally and as released by dams and irrigation channels;
- A “hydrologic model” of the varying patterns of rainfall, snow melt and run-off; and
- An “economic model” to calculate and optimize the cost and net value of flood control, agriculture and electric power that the dams and other projects generate.

Each of these can be more or less detailed, depending on whether the suite of models considers detailed weekly or monthly flows, or works with average flows over a season or a year.

Oil and Gas: Developers of oil fields use some form of “oil and gas” model. Each organization will have its own version, but the overall elements are the same. The system model combines:

- A “facilities model” that defines the sizes and details of the many elements needed to pump the crude from the field; to separate the oil, gas, and water; to compress the gas; store the products for transshipment, etc.
- A “field model” that describes the dynamics of the flow of oil, gas, and water through the ground to the wells in response to both natural pressures and those the operators of the field create through injection wells; and
- An “economic model” that translates the cash flows of capital expenses, operating costs and revenues into net present values and other measures.

Each of these models is likely to be highly complex. They also need to pass information back and forth between each other. A change in the pumping capacity for example, can change the flow in the oil field, which in turn affects the pumping capacity that the operators might need and use. Because of this complexity, the definition of the best design for the development of an oil field for a single scenario may take a day or more. Hence the need for simplified models for the screening process.

Box 5.4

Simplified Bottom-up Screening Model

River Development in China: Working with a Chinese investment bank, the research team analyzed the development of a major river in South-west China, whose potential for hydropower, flood control and irrigation were undeveloped. The plans included a set of dams, diversion tunnels and irrigation projects. The design of the overall project had to specify both the size of each project (such as the height of each dam), and the sequence and timing of the development.

The team developed a screening model for the river basin. It was a simplified version of the detailed models designers were using to specify the details of the development. It worked with average annual flows rather than stochastic seasonal flows, and eliminated a number of possible developments from consideration, since previous analyses indicated they were not desirable. These simplifications drastically reduced the size of the stochastic, mixed-integer programming problem. The simple version had less than 1/10th the number of constraints and variables than the complete problem. The screening model could develop an answer in a few hours on a laptop computer, rather than the several days for the complete suite of detailed models run on large machine.[28]

Box 5.5

Simulator – Direct Approach

Oil field development: Designers for deep-water oil platforms use oil-and-gas models to define the details of a prospective platform. These models combine detailed descriptions of the design of the platform, the behavior of the flows in the oil field, and the economics over time. In practice, it may take a day or more to use an oil-and-gas model to do a complete analysis of a single design for a specified set of parameters.

It is straightforward to develop a simulator that reproduces the overall value of a platform calculated by the detailed suite of models. First, we assume the principal drivers of value. For example, the analysts might agree that these are the size of the field; the capacity of the platform; and the price of oil. We then run the detailed model for combinations of these drivers of value. For each combination, we get a value for the system. Finally, we do a statistical analysis to find the best fit for an equation such as:

$$\text{Value} = a(\text{size of field}) + b(\text{capacity of platform}) + c(\text{price of oil}) + e$$

If the value function appears non-linear, we might prefer an equation such as

$$\text{Value} = e + k(\text{size of field})^a(\text{capacity of platform})^b(\text{price of oil})^c$$

In any case, the procedure is similar.

Box 5.6

Simulator Check on Integrated Models

Simulator models can reveal inconsistencies in the complex, detailed models. Working on a design of a major, multi-billion dollar project, a research team used the simulator model to determine that the integrated model the designers were using had fundamental flaws and led to major expensive design errors.

Our team examined the simulator to see how the detailed integrated model generated overall characteristics of the system. In particular, we looked at what the simulator reflected in terms of how the overall cost of the system depended on its size. The general relation between optimal cost for any is the so-called “cost function”. Economists routinely approximated this relationship with the equation:

Cost of System = A (Capacity of System)^{B}

If B is less than 1, then the costs do not grow as fast as size, and we talk of economies of scale. [If we increase capacity by 1%, cost increases by B %.] B is the “economies of scale factor”. The lower B is, the greater the economies of scale. Both experience and theory indicate that the lower limit on the economies of scale factor B is about 0.6.[29]

Economies of scale are crucial factors in many industries. Significant economies of scale drive designers to create the largest economically reasonable facilities. This is because economies of scale mean that the average cost per unit of production capacity decreases substantially as we build larger plants. Economies of scale are common for production systems using tubes or containers, physical processes in which the cost increases with the surface area but the capacity increases with the volume. Examples include transportation (ships and aircraft); the production of electric power; and chemical processes as petrochemical facilities.

When the team looked at the response model for the system, it found that the detailed model implied that the economy of scale factor for the system was about 0.3! This looked miraculous – and was indeed too good to be true. It turned out that there was an error in the complex model. The actual economies of scale were much less. The consequence was that the design team, led astray by their integrated model with its implied tremendous economies of scale, was systematically overbuilding the size of their facilities.

The cause of the problem was that one of the components of the model assumed that a particularly important piece of kit only came in one size. This size of kit was normal for a particular, very large overall system, but inappropriate for smaller versions of the system. The assumption that this element only came in one size led to excessively high total costs for smaller systems and thus the apparent enormous economies of scale.

Box 5.7

Simulator – Indirect Approach

Oil Field Development: The research team developed a simulator model for the analysis of the development of an oil field using the indirect approach. The idea was to create simple models of the component parts of the oil-and-gas model. These distinct sub-models corresponded to each of the major elements in the system: the reservoir, the facility for exploiting the field, and the project economics. This box illustrates the procedure using the example of the reservoir sub-model.

The basis for the simulator sub-model of the reservoir behavior over time was a simplified representation based on material balance equation assuming homogeneity of the reservoir. To achieve a suitable fit with reality, in this case represented by a mature oil field for which 15 years of data were available, the analysis examined various correction coefficients to adjust the model for the flow of water and gas through the reservoir. The preferred model was the one that minimized cumulative error in terms of the use of water injection and the resulting production of oil and gas. As Figure 5.2 indicates, the resulting model closely tracked the complex, non-monotonic behavior that exists in practice. This is ultimately the test of validity for any simulator model.[30]

Figure 5.2 here

Box 5.8

Top-Down Systems Dynamics Model

Electric Power in Kenya: The development of electric power in Kenya has been problematic. The national power company has found it difficult to finance enough production capacity and the distribution grid to provide reliable power. This encourages industry to build duplicate back-up facilities to ensure production. Customers meanwhile invest in solar energy and diesel generators, thus reducing the potential markets for the national company and discouraging it from connecting villages to the national system. The fundamental question is: how should Kenya best develop its electric power system?

Steel developed a top-down model of the Kenya power system.[31] She did this based on 6 months of field study, assembling technical information about the system and conducting interviews to understand the behavior of power producers, industrial users, and individual customers. Figure 5.3 shows portion of her model. It maps out the factors that describe the way industry in Kenya could decide on their strategy for obtaining power, either from the national grid or their own sources.

Figure 5.3 about here

Steel used the complete model to show the effect of different development strategies on the future market for electric power in Kenya. Some scenarios might lead toward an integrated national grid; others would favor a decentralized pattern of local production and use. This kind of information helps planners and developers to chose the kinds of projects that will best add value, and thus that they should develop. Figure 5.4 shows one of her forecasts.

Figure 5.4 about here

Box 5.9

Conceptual Approaches to Using Screening Models

Hospital Design: The emergency facilities at a major hospital were consistently congested. The responsible physicians clamored for additional facilities to solve this problem. They and the hospital administration generally assumed this need was evident and the only question was how large an expansion should be.

A simple model review of the hospital as a whole brought out the linkage between the emergency facilities and regular wards. This made evident the fact that minor operational changes in the management of the regular wards, such as discharging patients in the morning rather than in the afternoon as was normal, would greatly reduce the congestion in the emergency room that built up as the day progressed. The screening model showed that under reasonable assumptions, there was in fact no need to expand the emergency facilities!

As it is not possible to predict the precise effect of management changes, the better design appeared to be a small expansion that only kitted out a limited number of extra emergency beds, that had the flexibility to increase the number of emergency wards in the future.

The staff of the emergency ward was of course fully aware that their facilities were crowded with patients waiting for beds in the regular wards. However, their responsibilities and specific functions did not condition them to appreciate how changes elsewhere could improve their performance, and did not lead them to suggest of such operational solutions instead of expensive expansion.[32]

Oil Field Development: We observed a similar pattern in planning for the development of an oil field in deep water. In this case, the team developing the platforms focused on the design of the platform. However, a larger view of the situation indicated that the performance of the field, and thus of the platform, depended significantly on the way the undersea wellheads connected with each other. These tiebacks between wells allow the managers of the oil field to distribute the flows from the wells according to their quantity and viscosity and thereby improve the performance of the oil field, and thus the value of the platforms. In this case, incorporating the consideration of tiebacks in the design helped increase the expected value by 78%![33]

As with the hospital case, the designers of the platform knew that management of the tiebacks had a strong effect on the performance of their platforms. However, until the MIT team had developed the screening models, and made the influence specific, the design process for the platform kept to its “box”, and did not consider the effect of these underwater elements on the performance of the platform. The screening model enabled the design team to “think outside the box”.

Box 5.10

Optimization Approach to Using Screening Models

River Development in China: In this case there was great uncertainty in the long-term price for hydroelectric power. China was shifting to market pricing from a regime in which government agencies set prices with little consideration for the cost of production or the willingness to pay. Correspondingly, it was expected that the schedule of prices would shift to some new norms. Future demand for power was then doubly uncertain, because uncertainty about prices increased the difficulty in estimating regional growth over the long-term.

The analysis team used the screening model to identify what part of the design should be flexible. They specifically explored the effect of uncertainty in the price of electricity. Variability in this important factor not only influences the value of the project but also, as is generally the case, significantly changes the optimal design. For this river basin, it affects the optimal size of the development at Site 3, as the Table 5.1 shows. Conversely, the designs for Sites 1 and 2 dams were not sensitive to the price of electricity. The analysis thus successfully screened the plans for the design element that should be flexible, that is, Site 3.

Table 5.1 about here

Table 5.2 completes the story. The screening process focused attention on keeping the design for Site 3 flexible until some future time when the uncertainties about the prevailing price of electricity resolved. In addition, the analysis demonstrated that the value or the design was sensitive to time, so that the developers should be flexible as to when they would start the elements of the project.

Table 5.2 about here

Table 5.1:

Price of Electric Power Renminbi / KWH	Value of Project Renminbi. 10^6	Optimal Characteristics of Dams					
		Site 1		Site 2		Site 3	
		Power MW	Volume 10^9 M^3	Power MW	Volume 10^9 M^3	Power MW	Volume 10^9 M^3
0.10	0	0	0	0	0	0	0
0.13	367	3,,600	9600	1,700	25	0	0
0.16	796	Same	Same	Same	Same	0	0
0.19	853	Same	Same	Same	Same	1,564	6,593
0.22	1,607	Same	Same	Same	Same	1,723	9,593
0.25	2,196	Same	Same	Same	Same	1,946	12,242
0.28	2,796	Same	Same	Same	Same	1,966	12,500
0.31	3,396	Same	Same	Same	Same	1,966	12,500

Source: Adapted from Wang (2005) p. 188.

Table 5.2

Site	Source of Option Value	
	Timing	Design
1	Yes	NO
2	Yes	NO
3	Yes	Yes

Source: Adapted from Wang (2005) p. 189.

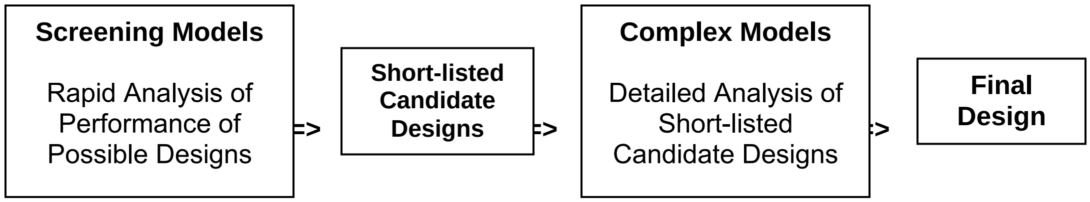


Figure 5.1: Screening Models Precede and Complement Detailed Design Models

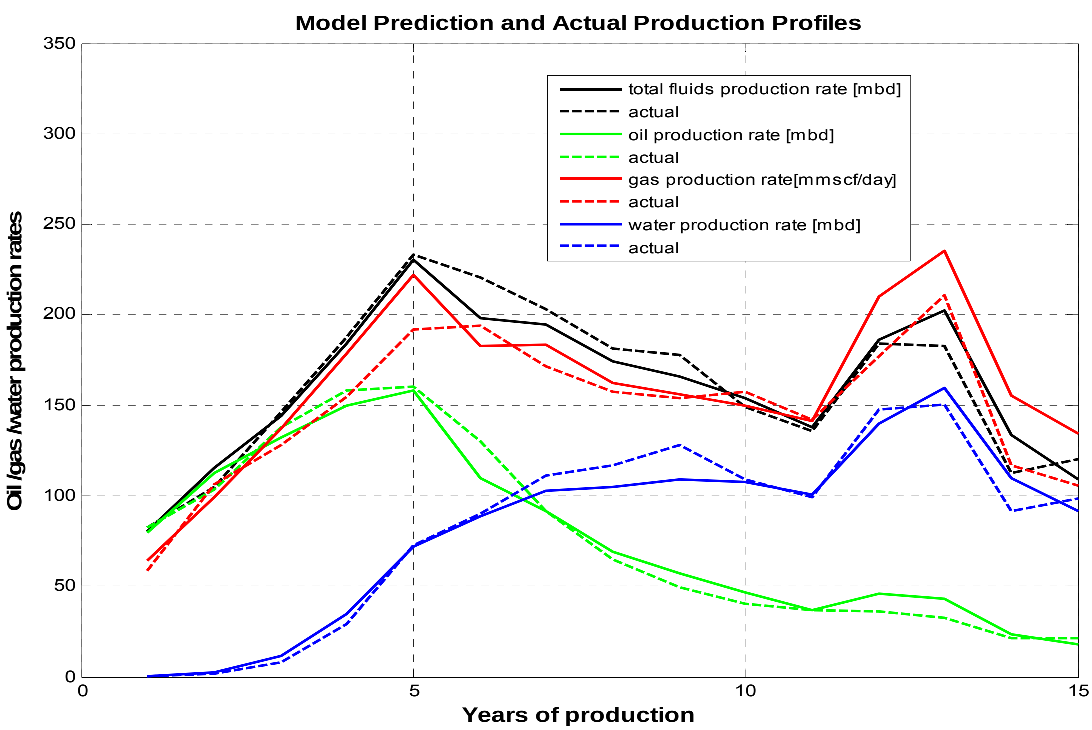
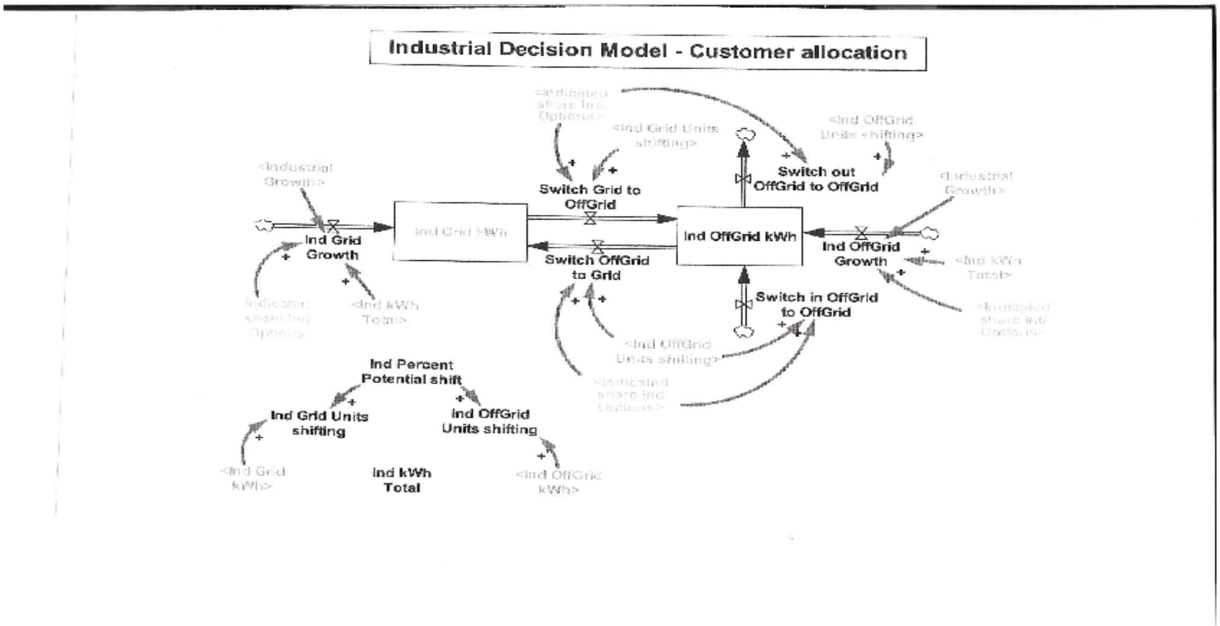


Figure 5.2: Comparison of simulator and detailed models. Source: Lin (2009)



FF

igure 5.3: Systems dynamics model of industrial decisions concerning electric power in Kenya. Source: Steel (2008), p. 199.

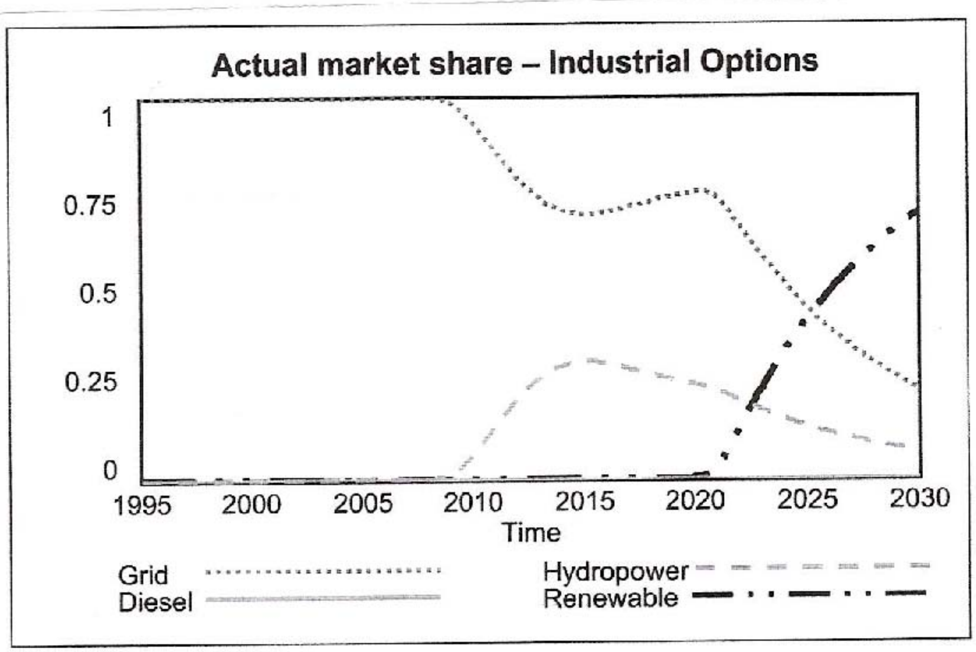


Figure 5.4: Possible future market shares of electric power sources in Kenya. Source: Steel (2008), p. 169.

CHAPTER 6
STEP 3. EVALUATING AND CHOOSING FLEXIBLE DESIGNS

"Asking the right question is half the answer"

Old saying

This chapter provides a process for choosing the preferable solution in the context of uncertainty. It expands upon the standard economic procedure for evaluating projects. Discounted cash flows and net present values are at its core. These metrics, however, are neither accurate nor sufficient once we recognize that future contexts and outcomes are uncertain. Thus, the proper evaluation of projects with uncertain outcomes extends the evaluation procedure beyond the simple formulas that assume that the future is known (see Appendix B).

The process for evaluating and choosing designs in the context of uncertainty extends standard evaluation procedures in three ways. It recognizes that:

- The evaluation must consider the range of scenarios, not only the most likely futures. It thus avoids the "flaw of averages" (see Chapters 1 and 2 and Appendix A) and obtains a more accurate assessment of value.
- The value of project is a distribution of possibilities. We must therefore think in terms of <u>expected</u> net present values (ENPV) over a distribution, rather than a fixed single net present value (NPV) result.
- The complete evaluation needs to consider several factors. This is because two very different distributions of value may have the same expected value. A fair comparison of projects must then characterize the underlying distributions in some way, for example by describing upper and lower extremes for each project. A good evaluation should be multi-dimensional.

The process for evaluating and choosing the preferable project or design incorporates these considerations.

Note the stress on choosing "preferable" solutions, rather than "best". It is important to keep in mind that it is not realistically useful to think of best designs, once we understand that the goodness of a project has several distinct attributes. This is because different decision-makers define "better" according to their own preferences. For example, one person may prefer a project with higher rewards even though it is more risky than the alternative, while another may find it necessary to avoid risky endeavors. The notion of which project is "best" is thus not absolute, but depends on the preferences of the persons involved. Note furthermore that when a project concerns several stakeholders or decision-makers, the decision they agree to (and thus "prefer" to others in operational terms) depends on their relative power and on the procedures by which they arrive at their agreements or compromises. In general, whenever we choose between projects with multi-dimensional values, we should think in terms of preferred choices, and recognize that each stakeholder has their own.

The process for evaluating and choosing designs consists of three steps:

- Evaluation of individual designs. In the context of uncertainty, this analysis has two special features. First, it involves the consideration of the various distributions or scenarios of possible futures. Second, it requires the analyst to define the circumstances in which system operators will use the flexibility in a design.
- Multi-dimensional comparison of designs. This analysis calls for clear presentation of the different values of each project. The idea is to help decision-makers focus on the issues that are most important to them, and to help them determine their choices.
- Validation by sensitivity analysis. As estimates of future uncertainties are themselves imprecise, responsible analysis will explore the consequences of different assumptions about the future. We need to develop confidence, to the extent possible, that choices rest reliably on the analysis.

This chapter develops these steps sequentially. The chapter ends with a case study of an important practical example involving the design and development of a major oil field.

Evaluation of Individual Projects

Distribution of Outcome Values: We need to analyze each candidate design over the range of important uncertainties. This is essential. If we do not do this, if we consider alternatives only from the perspective of the expected or most likely set of future circumstances, our calculations of system value will almost certainly be wrong, often seriously so. Failure to examine the performance of designs under the range of circumstances is to fall into the trap of the “Flaw of Averages”. If we fixate on the most likely circumstances, our analysis will miss the effect of the uncertainties and we are thus likely to settle on inferior choices.

The analysis considering distributions of events will give a distribution of possible values of a project. Combinations of most likely events should give more likely outcomes. Combinations of favorable circumstances should lead to particularly good results. Unfavorable combinations of events should likewise lead to bad results. The results of combinations of favorable and unfavorable circumstances will lead to results somewhere in between in ways that are difficult to anticipate intuitively. A careful analysis must be done to get an accurate assessment of the distribution of value for any project.

Monte Carlo simulation is the recommended process for analyzing the range of possible outcomes for design alternatives. This is a standard process. It consists of evaluating the performance of a proposed design many times, for the range of possible uncertainties according to their estimated probability of occurrence. The important result is that the analysis develops a distribution of the possible performance for the trial design.

Box 6.1 about here

Mathematically speaking, Monte Carlo simulation calculates the performance of each alternative considering the joint distribution of the uncertainties. It does this in two phases. First, it samples from the distributions of possible circumstances (such as future demand, prices, and so on). This gives one possible result of the project. Second, the Monte Carlo process repeats the sampling process a great many times (for example, 1000), giving each possible future circumstance its appropriate chance of being sampled. It thus creates the distribution of the performance of the design that is consistent with the joint distribution of possible circumstances.

The simulation keeps a record of everything that it does, so that the sensitivity analyses can examine the results in detail, as discussed further on. Appendix D describes Monte Carlo simulation in detail. Software for carrying out Monte Carlo simulations is widely available, inexpensive, and easy to use.[34]

Use of Screening Models: Monte Carlo simulations routinely involve 1000 analyses of a design, often many more.[35] We need many values to get a good representation of the distribution of the values associated with a design. The consequence of this is that, in order to carry out the calculations in reasonable time, we want to work with models of the system that take little time for a single simulation. Monte Carlo simulations thus predominantly use screening models, the simpler, mid-fidelity representations of the system described in Chapter 5.

Monte Carlo simulations based on screening models are appropriate in the initial phase of the design, when we are developing a short list of the best possible forms of flexibility to include in the architecture of the system. In this phase, when we are identifying overall concepts for the design of the system, we do not have to worry especially about detailed accuracy, since we will later carry out high-fidelity analyses of the short list of possibilities.

When we come to the detailed design, we are unlikely to be able to use the standard Monte Carlo analysis. This is because the models used in detailed design take too long to calculate the value of a design to allow for 1000s of simulations. In this case, we will have to be satisfied with far fewer tests of the design, and will have a cruder description of the performance distribution of the alternative design choices. However, this fact does not alter the fundamental reality that we need to compare distributions of values for each design.

Rules for exercising flexibility: The evaluation of a design with flexibility differs from that of a fixed design. The analysis of a fixed design is simple. This is because the fixed design is passive; it simply reacts to future circumstances. By definition, by construction, a fixed design does not change over its lifetime. However, system operators can change a flexible design. Management can and should actively take advantage of the flexibility to adapt the flexible design to the circumstances. That is the whole point of a flexible design. The evaluation of a flexible design is thus more complex than that of a fixed design; it has to account for the possibility of design changes and their effects.

The issue in evaluating a flexible design is that we cannot know in advance when the management of a system may change the design. If the demand for project services increases rapidly, management may rapidly exercise the flexibility to expand the system. On the other hand, if the demand increases slowly, management may decide to expand slowly, late, or never. In short, we can expect management to take advantage of flexibility in a design when circumstances are "right". The Monte Carlo evaluation of a flexible design has to incorporate some procedure to identify when the possible future circumstances are "right", and then change the design so that the evaluation can mimic intelligent management of the system.

The "rules for exercising flexibility" provide the mechanism for identifying when and how intelligent management would take advantage of flexibility and change the design. These "rules" monitor each possible future scenario or forecast developed by the Monte Carlo simulation, look for situations that call for design changes, and instruct the automated analysis to adapt the design and system performance in the basic evaluation model. The rules for exercising flexibility should mimic what the management of the system would do if ever they had to deal with a situation. Good rules for exercising flexibility are thus essential for the assessment of the value of a flexible design.

A "rule for exercising flexibility" is a small logical element that, as its name implies, determines when (and how) the system should change in response to changing context and opportunities. Using terms from computer programming, the rules for exercising flexibility are "IF, THEN" statements. For example, if management only wants to expand capacity once demand for additional space appears robust, its rule might be to expand only after observed demand exceeds capacity for at least two consecutive years. A suitable rule concerning the possibility of adding more floors to a parking garage would then be:

IF…. demand exceeds capacity for two consecutive years,

THEN add floors to meet demonstrated demand.

This rule would keep track of the relationship between demand and capacity for two years before triggering an expansion, the associated costs of construction and operation, and the benefits of increased performance. Such rules are easy to program into a spreadsheet as indicated in Appendix D on Monte Carlo simulation.

The simulation process calls upon the rules for exercising flexibility each time it generates new values for the uncertain variables, that is, in each period of the simulation. At that point, it consults the rule for exercising flexibility and asks: "are the circumstances right for exercising the flexibility?" If the answer is "No", then the simulation does not change the description of the system (the spreadsheet for example). If however the answer is "Yes", then the rule triggers the change in the system and carries on the subsequent analysis with that change in place. In Chapter 3 we use an example of a parking garage that could be expanded if demand were high. An appropriate rule for exercising flexibility would, for example, check if the simulated demand in the previous two years exceeded capacity, and add extra floors if that were the case.

In line with what could happen in reality, any rule for exercising flexibility leads to a wide range of development paths. For the parking garage for example, in some scenarios the traffic does not grow, and the garage does not expand. In others, demand grows rapidly, and the simulation accordingly adds floors early in the life of the project. In general, the rules lead to a distribution of developments, each with their own net present value.

It should be noted that different rules for exercising flexibility can apply at different times over the life of a project. For example, management might be ready to expand a project during the early part of the life of a project, but not want to spend money on changing the system as it nears the end of its useful life. Analysts can easily program different rules into their simulations. The essential thing is to anticipate management choices as reasonably as possible.

Target Curves: A “target curve” is a convenient way to present the distribution of possible values associated with any design. It shows the percent probability of results being lower than any specified level or target of performance. It derives directly from the results of a simulation. We achieve this by sorting the generated performance values in ascending order. Thus, Figure 6.3 is the target curve associated with the distribution in Figure 6.2.

Figure 6.3 about here

It is standard to indicate the average value of the distribution as a vertical reference line on the graph. Figure 6.3 thus shows that the Expected Net Present Value (ENPV) of the project is $0 for this case. [As a semantic caution, note carefully that this value is “expected” in the mathematical sense that it is the average value. It is not necessarily the most likely value. The expected value may even never happen. For example the average value when rolling a six-sided die is 3.5 { = (1 + 2 + 3 + 4 + 5 + 6) / 6}, a number which of course never shows up!] Note that, as in Figure 6.3, the expected value does not necessarily occur in the middle of the distribution. Because the distribution of outcomes is generally asymmetric, we can anticipate that the expected value to be at some other point than the median, the target with a 50% probability.

A target curve presents a lot of information in a compact form. Referring to Figure 6.3, it shows:

- The probability of breaking even (EPNV = 0) or doing better on the project, this is the complement of the probability of doing worse (in this case the probability of at least breaking even is 60%);
- Lower Extremes, such as P_{10}, one measure of risk (a loss of 15 million of present value);
- Upper Extremes, such as P_{90}, (a gain of slightly more than 9 million);
- The Range of results, reflecting the dispersion in outcomes (a span of 44 million);
- The Risk of the downside of any specified level (sometimes referred to as the Value at Risk). For example, there is a 10% chance of losing more than 15 million, and a 90% chance that the results will be less than a gain of 9 million. These 10% and 90% results are also called the P_{10} and P_{90} values. These values are often preferred as measures of the range because they are statistically more stable than the absolute minimum and maximum values generated by a simulation; and
- The Difference between the Median Value at a cumulative probability of 50% (about a gain of 2 million) and the Average Value (0) caused by the asymmetry in the distribution, in this case skewed toward great downside losses.[36]

Depending on the context, some of these characteristics of the distribution of results may be among the set of dimensions that the decision-makers will want to consider when choosing between designs.

Table 6.1 represents the same data as a target curve, as well as other factors of interest. A table may also be preferable for audiences or managers who are not comfortable with graphs. For the graphically inclined, however, the target curve is a useful and effective way of comparing many of the values of alternative designs.

Table 6.1 about here

Using target curve as a guide to design: Flexibility in the design generally enhances performance in two complementary ways. Flexibility can:

1. Reduce the downside consequences; and
2. Increase the upside opportunities.

Target curves provide a visible guide to what flexible designs can do to improve the performance of initial designs, as Figure 6.4 indicates.

Figure 6.4 and Box 6.2 about here

Actions to reduce the downside consequences increase the expected value of a project, by minimizing the downside tail of poor results. They may also increase the minimum result. They act like insurance or "put" options in financial terms.[37] A typical way to reduce the downside risk is to invest in a relatively small project at the beginning, and to defer expenses until the need for them has been validated by future developments. For example, with reference to the garage case, it may be better to build only a four storey garage, rather than the 6-storey garage that might be best if we had to choose a fixed design. Obviously, when we limit our investments, we can limit our losses in a project.

Likewise, actions to increase the upside also increase the expected value of a project. They may also increase the extreme upper results, such as the P_{90}. Financially, such efforts act like "call" options. A standard way to increase the upside opportunities is to build features into a project that enable it to morph into a state that better meets future demands or requirements. These can be actions that allow for future expansion, such as the early investment in extra strength for the Ponte 25 de Abril across the Tagus, in Portugal, which allowed for later double decking of the bridge when and as needed. Alternatively, flexible designs may enable a shift from one mode of operation to another, such as when power plants install the capability to burn either natural gas or oil, and can thus take advantage of the advantageous changes in the relative prices of these fuels.[38]

Comparing Target Curves: A target curve for a design A may be completely to the right of the target curve for a design B. In this case, A represents an improvement over B insofar as A reduces the chance of missing any desired performance target. Figure 6.5 shows an example of this for alternative designs for an oil platform complex. Development strategies 11 and 12 are each unambiguously to the right of the base case Strategy 1. The designs whose target curves are completely to the right of another have a lower probability of failing to meet any target and thus offer a higher expected value.

Figure 6.5 about here

Note that a design with a target curve completely to the right of another does not necessarily represent a better choice. This is because a target curve does not represent all the important attributes of a design. For example, the design whose target curve is to the right, and that thus offers better performance, may also cost more in that it requires a greater initial capital investment (Capex). Decision-makers might prefer the design that costs less even though it also performs less; they might appreciate a higher ratio of performance to investment. In general, target curves represent the risk profiles of the performance of designs, but not other factors – such as environmental performance, that might be crucial to the final selection of a project.

Note also that a design whose target curve is completely to right of another may not perform better under all circumstances. The design represented by the target curve to the right may perform exceptionally well in some limited cases, perhaps precisely in those circumstances in which the alternative does not perform well. Target curves represent overall risk profiles over all cases, not a head-to-head comparison of performance in specific circumstances.

In general, the target curves for alternative designs do not dominate each other. Figure 6.6 illustrates the phenomenon. It shows the target curves for designs with different numbers of levels of the parking garage. Comparing the curve for the 7-level garage (yellow) with that for the 4-level structure (blue), we can see that they cross, at the level of 50%. Until that point, the curve for the 4-level garage is below and to the right of that for the 7-level, meaning that the smaller structure has less chance of delivering low values of performance. This is of course what one would expect: having spent less to build the smaller structure, the owners have less to lose. Conversely, above the crossover point, the curve for the 7-level garage is to the right of the curve for the 4-level structure, which becomes vertical. This means that the limited capacity of the smaller garage limits its maximum value, whereas the larger structure can deliver much more value if the conditions are right. This kind of crossover effect is common.

Figure 6.6 about here

Choosing between Target Curves: Sometimes the choice between designs seems obvious. Consider the comparison between the target curves for the 4 and 5-level designs in Figure 6.6. The curve for the 5-level design (mustard) is almost entirely to the right of that for the 4-level structure (blue). The 5-level design offers almost a uniform improvement on the 4-level design in terms of its ability to meet targets. Its chance of 10M or more is 2%, but about 6% for the 4-level design. The 5-level design has a 65% chance of breaking even or doing better, whereas the 4-level design has no chance. Overall, the Expected Net Present Value of the larger fixed garage is about 0 compared to a loss of 2M for the smaller project. It thus seems easy to prefer the 5-level design over the 4-level.

However, the choice between designs is often not obvious. While the 5-level design dominates the 4-level design, the choice between 5 or 6 levels (brown) is not as clear cut. The corresponding curves cross each other in the middle. The 6-level design has considerably higher upside, with a 20% chance of achieving a value of 8M, which is not possible for the 5-level design. However, this increased upside comes with an increased downside; the 6-level design has a 15% chance of losing 10M or more, against about a 6% chance for the 5-level design. The choice between these two designs depends on how the systems owners or designers feel about these risks. In general, we need to face up to the fact that choices between designs require us to consider conflicting objectives.

Multi-dimensional comparison of projects

Decision makers generally want a project to perform well over several criteria. For example, they might want the system to provide good value (as represented by ENPV), to minimize the risks (perhaps as defined as the P_{10} value), and to perform well environmentally. The comparison and eventual selection of projects involves many dimensions.

Trade-offs: In choosing between projects with different levels of attainment, stakeholders balance the advantages and disadvantages of one alternative compared to another. For example, they might compare the extra return or ENPV of an alternative and the extra risk it might entail. That is, they "trade-off" the different dimensions of projects.

The trade-off decision-makers are prepared to make between two dimensions of choice implicitly defines their relative value for these dimensions. For example, we can interpret the extra risk they might be willing to incur for an increase in expected value as the price they are willing to pay for risk. It is possible to imagine that if there might be some way to define the relative value decision-makers ascribe to each dimension of choice, it would thus be possible to define an overall measure of value for all the dimensions of choice. This approach is not practical, however.[39]

Good practice considers the several criteria for choice on their own. It is not practical to give consistent and accurate monetary (or other) values to the many dimensions of choice or otherwise to blend them into a single measure of goodness. It is therefore not possible to define an objective function suitable for overall optimization. Consequently, it is not possible to determine the preferred design through a purely mathematical procedure that ranks projects unambiguously. In comparing alternatives, decision-makers need to confront several criteria and make the trade-offs appropriate to their situation and preferences.

Concept of multi-dimensional choice: In comparing alternatives, it is helpful to divide them into two categories. Dominated alternatives are those that are demonstrably inferior because they do not perform as well as some other alternative on all criteria of choice. We can discard these from consideration.

Obviously, if we know of a design which is better on all counts (for example, provides more value, is less risky, and requires less investment) we will prefer it without further analysis.

Our attention should focus on the dominant alternatives. These designs cannot be improved in any dimension without giving up performance in some other dimension. Conversely, a dominated alternative is one for which there is another alternative that is better in all criteria. With reference to Figure 6.7, the dominant alternatives are those on the outer edge of the feasible region of alternatives, that is, on the line A-B. This boundary is the “Pareto frontier” and defines the best available designs. Note that, by convention, the axes all plot benefits, so that better performance is to the right and upward from the origin. Thus alternatives to the left and towards the origin of line A-B, such as point C, are dominated designs. By construction, there are no known designs to the right and above the Pareto frontier. The concept of the Pareto frontier is general. While Figure 6.7 shows it in two dimensions for simplicity, it can have as many dimensions as necessary. As can be seen on the graph, any increase in one benefit for a dominant design (that is, one on the Pareto frontier) comes at the expense of a decrease in some complementary benefit. The Pareto frontier defines the tradeoffs that available for any dominant design.

Figure 6.7 about here

The choice between dominant designs depends on the currently prevailing preferences of the decision-makers and the stakeholders. As noted earlier, people’s preferences depend on their circumstances. For example, if they are secure they may be ready to accept more risk, but they may be risk-averse if they are vulnerable. Whether their choice is A or B (or some other design in between), depends on the relative importance the system managers attach to the criteria of choice in their circumstances, to the trade-offs they are then willing to make. For example, how much of the monetary benefits of A might they be willing to trade off to obtain greater complementary benefits, such as those represented by B?

As a practical matter, it is generally not possible to determine reliably and in advance how anyone will want to choose between trade-offs. Managers may not be willing and able to define their preferences until confronted with the choices. Moreover, important benefits may be difficult to define precisely. For example, a national Energy Minister responsible for the implementation of an electric power system will probably need to balance the economics of national investments against the need to distribute the benefits across the country. A system that provided reliable power only to rich areas and disregarded poor rural areas might not be sustainable, politically or otherwise. The relative value of such benefits is difficult to define under the best circumstances. It is highly unlikely that anyone can express these values accurately concerning speculative cases.

The choice of the design that is preferred overall requires some sort of process that allows the decision-makers to select the solution they jointly prefer from the set of best designs. Box 6.3 gives an example.

Box 6.3 about here

Capex Issue: Developers of major systems need to pay special attention to the amount of money they must provide to build the first phase of a project. This amount is the initial capital expenditure (Capex). Although it is included along with all other costs in the calculation of net present values, it differs substantially from the other costs that occur over the life of the system.

The initial Capex for a new system:

- Is a large amount of money -- the cost of an automotive factory; the launch of a new aircraft or satellite system; the discovery, validation and marketing of a new pharmaceutical drug; development of a major oil field; or the construction of a new highway, bridge or railroad line; may easily exceed a billion dollars;
- Has no off-setting revenues – in contrast to operating or expansion costs that are committed during the life of the project and are thus partially offset and paid for by ongoing revenues;
- That cannot be adjusted to the ongoing success of the project – again in contrast to subsequent operating costs that largely vary with the demand, and expansion costs which only occur if sufficient thresholds are met; and
- Is especially risky – because we do not yet know the desirability of the future product or service.

In short, the initial Capex represents an important initial barrier to the development – and thus to the choice of a design alternative.

A good analysis should feature the level of initial Capex. This requires special attention. The Monte Carlo evaluation of a system does not automatically calculate Capex. It is not part of the net present value or of the distribution of results as presented by the target curves. It is latent in the analysis and we need to draw it out and make it explicit.

The obvious way to bring Capex information into the evaluation is to present it, along with the other characteristics of the evaluation, in a comparative display such as Table 6.2. This will allow the decision-makers to consider Capex as they think about the tradeoffs involved in making their choice. All else being equal, we can expect that a design with lower initial Capex will be preferred.

Table 6.2 about here

Case of Uncertainty: In the context of uncertainty, the preferred designs are those that simultaneously increase expected value, particularly compared to fixed designs, and best manage the risks and opportunities in the project. The choice of the preferred design embodies several tradeoffs. At a minimum, these are between:

- Initial capital expenditures (Capex) and more speculative later benefits;
- Minimizing downside risks and maximizing upside opportunities; and
- Prospective benefits and possible retrospective regrets about having made the wrong choice.[40]

Different organizations, different leaders and managers, will feel differently about these tradeoffs, and their views are likely to depend on their circumstances, their competition, and their other major investments.

Illustrative Example: This example illustrates the evaluation over many criteria in the context of uncertainty. The case concerns the development of a major oil field. It focuses on comparing a base case against two flexible alternatives:

- A "Rigid" design, sized in the standard way that optimizes the system based on the assumption that the future quantities and price of oil will be as projected;
- A "Lower Flexible, Lower Cost" design that provides some flexibility and does not deliver the highest expected net present value – but whose initial Capex cost is about 30% less; and
- A "More Flexible, Higher Cost" alternative that delivers the maximum expected value by exploiting full flexibility to adapt to the actual quantities and price of oil that actually occur.

Table 6.2 presents the performance of these choices, according to important criteria. As it suggests, the choice between the three designs is not obvious: each design is best in at least one category, as the bold face type indicates.

Table 6.3 about here

Both flexible designs have advantages and together seem to dominate the rigid design. But the case is not definitive. The rigid design, "optimized" to a particular set of circumstances, performs best when these occur. In this case, the rigid design provides the highest possible performance (1286 in Table 6.2) compared to the flexible alternatives that incur the cost of providing flexibility. In the many other circumstances that may occur, the rigid, "optimized" design is inferior. Being unable to adapt to circumstances, it can neither avoid the losses on the downside nor take advantage of upside opportunities. Thus, the Expected Net Present Value of the rigid design (822) is lower than those of the flexible designs (900 and 929). Moreover, since the rigid design commits all of its capital costs at the beginning, it is by far the most expensive in terms of the expected present value of costs (1006 compared to 688 and 969). The rigid design neither delays any of the costs (savings in present value) nor avoids them if certain design capacities are not needed (savings in avoided costs). Thus although the rigid design may be "optimized", it is in fact often an inferior design considering the possible scenarios.

The More Flexible design might appear most attractive, since it provides the highest Expected Net Present Value (929). However, it is over 40% more expensive than the Lower Flexible design (969 compared to 688, based on the Expected Net Present Value of total costs that occur over time, for example when management exercises the flexibility.

The Lower Flexible design is thus arguably preferable. The More Flexible design requires an extra 281 million in expected present value of costs, for a gain of only 29 million in terms of overall Expected Net Present Value. The small gain from the upgrade from the Lower to the More Flexible design does not seem to justify the extra costs that are about 10 times larger. That is, while the More Flexible design is perhaps "best", it may be a "best" that is not worth paying for.

However, the More Flexible design in this case is more reliable in terms of delivering results. The standard deviation of its Net Present Value is by far the lowest (165, compared to 285 and 334 for the alternatives). In this sense, its performance is most “robust”.[41] Put another way, its minimum result (788) is by far the best, over twice the value of the alternatives (358 and 308). This result is not surprising since the More Flexible design is more able to adjust the design to the circumstances that may prevail; it is best able to avoid downside results, and capable of taking advantage of upside opportunities.

The choice between the More Flexible and Less Flexible alternatives is thus not obvious. Depending on the needs, preferences, and risk perceptions of the decision-makers, they might choose one or the other. The role of the analyst is to bring out the issues, and to help the project leaders make the decisions that are best from an overall perspective.

Difference Curves: A difference curve provides a useful way to think about the choice between two alternatives. It plots the difference in the target measure of performance between two designs. It helps us think about questions such as: Is the higher Expected Net Present Value achieved by a design worth the extra cost and risks associated with the upgrade from another design? In contrast to target curves that present overall performance, difference curves highlight the change between two designs.

A difference curve illustrates the relative value of a flexible version of a fixed design. Consider the example of the parking garage in Chapter 3. There are two designs for 4-level parking garages. The fixed one has no possibility of expansion. The flexible one can extend vertically up to 8 levels, because its design features stronger columns and footings that will take the added weight of additional levels if they should be added. The flexible design costs more than the fixed design of the same number of levels – this increases its possible downside. The flexible design can however expand its capacity and take advantage of higher demands when they appear. The target curve for the flexible 4-level design is thus far to the right of the fixed design in terms of maximum gains. The comparison of the target curves for these two designs in Figure 6.9 brings out this advantage of the flexible design. The difference curve in Figure 6.10, constructed by plotting the difference between these target curves, makes the effect salient.

Figures 6.9 and 6.10 about here

Difference curves likewise compare two different designs, for example the flexible 4-level design and the most attractive fixed design. Assuming this is for 5-levels, we obtain the difference curve for the possible gains from the flexible 4-level design shown in Figure 6.11. Reading the figure, we can see that the flexible design in this case:

- Delivers a gain in Expected Net Present Value of 1.9;
- With an 80% chance of performing better than the fixed design, as indicated by the break-even point in Figure 6.11; and
- Has a 40% chance of outperforming the fixed design by 2.5M or more, which would seem to outbalance the 20% chance of underperforming the fixed design.

Figure 6.11 about here

Validation by Sensitivity Analysis

Upside-Downside Curves: Upside-Downside curves provide a convenient way to bring out the different levels of risk associated with alternative designs. These compare performance of different designs both on average (that is, their ENPV) and at selected equal levels of probability such as P_{10} and P_{90}. They provide a way to compare the average and spreads of performance.

Figure 6.12 shows the Upside-Downside curves for the different designs of the parking garage. The Upside line shows the probability (in this case 5%) of achieving a profit as high as or higher than the profit on the line. Conversely, the Downside curve shows the probability (in this case also 5%) of having a loss as large as or larger than depicted on the line. The middle line shows the average or Expected Net Present Value of the design.

In this particular situation, the comparison of spread of performance and average values indicates that the 5-level design delivers about the same level of average performance, of ENPV, as the 6-level design, but with far less uncertainty – a smaller spread and standard deviation – about the results. Depending on the preferences of the decision-makers, this might be a sufficient reason to prefer the 5-level fixed design to the 6-level fixed design.

Figure 6.12 about here

Regret Plot: Project leaders may also want to think about possible regrets they might have about their choice between two alternatives, between the fixed and flexible design for example. They might ask: "If we choose the flexible alternative, what are the chances that we would ever think that we made the wrong or bad choice?"

We can answer this question by creating a regret plot. This device compares the performance of the two alternatives in each of the scenarios generated by the Monte Carlo simulation. The vertical axis corresponds to the value of the prospective preferred design (the flexible 4-level design in this case), while the horizontal axis is the value by which this preferred design exceeds that of the alternative. All the situations for which the preferred design performs better than the alternatives are to the right of the black vertical line in the plot. All the times the preferred design delivers a positive net present value are above the horizontal black line. Figure 6.13 is the regret plot comparing the flexible 4-level design and the fixed 6-level design.

Figure 6.13 about here

We can use the regret plot in Figure 6.13 to make the following useful observations about the flexible design. It generally:

- Outperforms the fixed design – most of the points in the plot are to the right of the vertical split (the Monte Carlo simulation can give you the exact number if desired); and
- Delivers positive value even when the fixed alternative would have provided more – these are the points above the horizontal split, and to the left of the vertical line. In these cases, the flexible design is not a bad choice, although the fixed alternative would have been better.

Later Capital Expenditures: System managers may want to anticipate when they might find it desirable to commit to second and subsequent phases. It normally requires further capital expenditure to take advantage of flexibility, and it is useful to plan for this eventuality. Knowing when they might want to take advantage of flexibility in design helps their financial planning.

Knowing when it might be desirable to implement flexibility may also help the system managers choose their initial design. If the analysis indicates that it is very probable that they should implement some flexibility soon, they may want to incorporate this feature in the initial development, to avoid whatever disruption might be associated with implementing the flexibility. For example, adding more levels to a parking garage is likely to require the owners to close it for some time and, at the very least, work in the midst of operations associated with this structure, such as a shopping mall or office blocks.

The Monte Carlo simulation provides information about when system managers might use the flexibility. Each simulation has a history of when its rules called for using the flexibility. Keeping track of this history over all the simulations provides distributions of when and how to use flexibility. For example, Figure 6.14 indicates that the probability of wanting to expand the 4-level flexible parking garage in the first two years is over 40%. This information might push the decision-makers to build a larger facility (such as a 5-level flexible garage) from the start, and thus avoid disruption so soon after opening.

Figures 6.14 and 6.15 about here

Monte Carlo simulation likewise can provide information about how much flexibility to use, and provides a means of validating the use of flexibility in the design. Thus, Figure 6.15 shows the distribution of the final height of the garage. In this case, it shows that the probability of ending up with an 8-level structure is over 70%. However, the analysis also shows that, from the prospective of making a choice at the beginning, an 7-level or higher garage is a losing proposition in the sense that its expected net present value is negative at the start, as Figure 6.13 shows. Putting these observations together means that:

- It would be wrong to build the 8-level or bigger facility at the beginning,
- Yet it is highly probable that it would be advantageous to end up with such a facility; and
- Thus validates the concept of design flexibility, that is, the ability to move from a good starting point to a different end.

Sensitivity Analysis: Assessments of uncertainty are also uncertain. Our evaluations can thus benefit from an analysis of the sensitivity of the results to the description of the uncertainty. This is part of good practice. All analyses should incorporate a sensitivity analysis, that is, an examination of the way results depend on their assumptions.

From the perspective of design, we must be most concerned with the robustness of our design choices. The question is: would reasonable changes in our assumptions lead us to change our preferred designs? We already know that, in an uncertain world, we do not know what the actual value of our designs will be. The real question is whether we are making the right choices. If design A is preferred to design C for a wide range of assumptions, then we can confidently choose it. If however the choice between these designs depends on the assumptions made, then we have to be more careful. In any case, the object of the sensitivity analysis is to check out that our preferred design is in fact better across reasonable assumptions.

Difference curves provide a convenient way to present the results of a sensitivity analysis. Since they display the effect of changes, they immediately provide the kind of sensitivity information we can use. For example, the analysis of the parking garage assumed that the uncertainty involved a range of 50% variation up or down around the base case demand parameters. We can examine the sensitivity of the results, specifically the conclusion that the flexible 4-level design was overall best, by redoing the analysis with different assumptions. Figure 6.16 shows the results for examining the difference in results if we assume that the range of variation is 25% and 75% instead of 50%.

Figure 6.16 about here

We interpret these results by seeing to what extent changes in the assumptions about the uncertainty would change the decision. In the case of Figure 6.16, we see that the flexible design leads to a net increase in ENPV over the entire range of possible uncertain examined (that is, from +/- 25 to +/- 75%). Whereas the evaluation leads to different values for different assumptions, the ENPV is strongly positive in all these cases (from about 1.8 to 2.2M). We thus see that the selection of flexible design can be sustained, based on ENPV, even though we might not know precisely, given uncertainties about the range of variation in demand, what the ENPV might be.

Figure 6.16 also illustrates the fundamental fact that the average value of flexibility increases with increased uncertainty. This reflects a basic truth: the greater the uncertainty, the more flexibility has value. Put another way, uncertainty is the great driver of the value of flexibility. This is why flexibility is so important in the design of large-scale, long-term projects: these are the most uncertain projects, and thus the ones where flexibility is most desirable.

In the case of the parking garage, the value of flexibility increases from 1.8M for 25% variation to 2.2M for 75% variation. In other words, if the world were more uncertain than anticipated in the original analysis, the flexible design will be worth more, on average. The intuitive reason for this is that the flexible 4-level structure exploits demand upside but avoids exposure to demand downside. Therefore, when the variation around the base case demand is larger, the upside is more likely and we are more likely to use the flexibility.

NEED TO MAKE A BIG DEAL ABOUT IDEAS THAT FLEXIBILITY IS DESIRABLE WHEN UNCERTAIN. THE UNK UNKS INCREASE VALUE OF FLEX. IF THERE ARE MORE UNCERTAINTIES OUT THERE, THEY DELIVER MORE VALUE (DO NOT DESTROY VALUE)

Example Application: Development of a Deep Water Oil Field

This example concerns a real-world case developed by a collaborative effort involving Lin (2009), MIT faculty members, and a major oil company. We developed this material to demonstrate the value of flexible designs for major real-world projects. The application was very successful: it demonstrated a 78% increase in expected value for the project, while reducing initial capital expenditures (Capex) by about 20%.

The physical situation: The oil field lies in deep water, off the coast of Angola. The field appears to be a series of reservoirs located across a wide area. It is relatively isolated in that it is not close to other operating or former fields, and does not have access to existing underwater pipelines or storage facilities. The development of the field will thus be self-contained. The company will extract oil and gas using one or more platforms. These will support facilities for primary treatment of the product (such as extracting water and compressing gas), and will store it until it is delivered to ships that will carry it to refineries.

It is impossible to know the effective size of the field until long after the original platform is built and production starts. This is because the effective size depends on factors that cannot be fully determined until production is well under way. Indeed the quantity of oil that it is economical to access depends on the permeability of the deposits, on the detailed nature of fractures in the field (that alter the pressure fields that drive the flow of oil and gas), on the viscosity of the crude, and the amount of gas and water in the deposit. Moreover, what is economical to extract obviously depends on the price of the product. If the oil price is high, it will be worthwhile to spend more effort pumping fluid into the field to increase pressure and the amount extracted. As Figure 6.17 indicates, estimates of the quantity of recoverable oil in a field can be very uncertain.

Figure 6.17 about here

Base case design process: A base case design was available at the beginning of the effort to develop flexible design alternatives. The design teams within the oil company had developed it along conventional lines, standard in the major oil companies we have worked with.

The design teams used deterministic forecasts to develop the base case design. Although the geological side of the company estimated the quantity of recoverable product[42] within a broad range (the P_{10} and P_{90} values were about +/- 50% around the P_{50} value, as in Figure 16.17), the design team focused on a single value. Similarly, the design team stuck with the instructions developed by their upper management to assume that the price of oil would be fixed over the life of the project, at the level management had accepted (at that time) to be the long-term average price of oil.[43]

These deterministic criteria for design naturally burdened the design process with a substantial handicap. Most obviously, they led to erroneous assessments of expected value, as discussed in Chapters 1 and 3 and in Appendix A on the Flaw of Averages. Specifically, it channeled the design toward a fixed capacity, which will cut off benefits from flows that might be higher than anticipated, and not protect from the losses if the flows are less than anticipated. Evaluations based on most likely estimates will necessarily give incorrect answers, since they ignore the reality that constraints shape actual returns. Furthermore, this tunnel vision fixation on average values leads the design team to ignore interesting potential. If you assume that the price of oil will always be $50 a barrel, then of course you ignore opportunities that might be profitable at $60, $70 or $80 a barrel, prices that have occurred and are likely to do so again.

The fact that the base case design used deterministic estimates of recoverable oil and of its price, which in fact are so uncertain, provides great assurance that some flexible designs could provide far greater expected value. Indeed, because the long-term prospects were so very uncertain, and because uncertainty drives the value of flexibility, the case study team had every reason to believe that flexible designs would be enormously valuable – as they turned out to be.

Development of screening model: The design team for the oil company worked with a highly detailed representation of the system. This "oil-and-gas" model consists of three major parts:

- <u>A geologic model</u> describing and optimizing the flow of oil and gas through the oil fields under different designs and operating strategies. This is highly complex, non-linear and extensive. It takes significant time to run over the lifetime of the oil field.
- <u>A model of the man-made developments</u> including platforms, the wells, and the sub-sea pipelines and storage facilities. This is also highly complex, representing hundreds of significant sub-systems that designers can combine in many ways. It includes optimization routines to automate the selection of appropriate sub-systems into a coherent design. This model is also highly complex, non-linear and extensive -- and takes significant time to run.
- <u>An economic model</u> representing the cash flows of expenses and revenues and ultimately calculating the net present value of a design operating over the lifetime of the project. Comparatively, this is simple and direct.

To analyze a particular design, analysts had to hook all three models together. A single pass through this apparatus could take days.

The highly detailed model was inappropriate for a Monte Carlo simulation. Thousands of simulations of a single configuration would take years. In this situation, as often the case for significant projects, and as described in Chapter 5, it was necessary to develop a simplified model that the simulation could run quickly, within minutes.

The case study team thus developed a screening model. It was "mid-fidelity", in contrast to the "high-fidelity" detailed "oil-and-gas" model the design team for the oil company normally used. As indicated in Chapter 5, the development of the screening model started with a response surface for the high-fidelity model, that is, a representation of its overall behavior. With suitable fine-tuning, we ended up with an acceptable mid-fidelity screening model that enabled us to run Monte Carlo simulations in minutes.

Flexible designs: The base case design featured a large oil platform, optimized for the assumed size of the oil field and price of product. It performed well under these conditions. If however quantities, oil, and prices were low, then the design was a poor investment, being much too big. If quantities of oil and prices were high, then the base case design was too small to take full advantage of these opportunities. In short, the optimized base case design performed poorly over a wide range of possible scenarios. This is a prevalent phenomenon. Solutions optimized for one particular future scenario will typically perform poorly when other futures occur. Conversely, solutions that appear sub-optimal will often perform much more robustly across a range of futures.

The initial flexible designs targeted the weaknesses of the base case design. They used smaller modules (which limited the downside consequences) that the company could expand (to take advantage of good opportunities, should they occur). Some of the alternative architectures did not increase value – the modules were too small to take advantage of economies of scale in construction. However, experiments with different designs led to a configuration that increased expected net present value by almost 50% (strategy 5 in Figure 6.18).

Figure 6.18 about here

Additional interesting opportunities for flexible design emerged from detailed technical conversations with the oil company. As frequently happens, understanding and working with the engineering details enables designers to uncover particularly clever designs. In this case, a flexible design of the undersea pipelines between the several fields significantly improved the overall performance. The reason is that much of the profitability of an oil field depends on the flows of crude that the operators can achieve. For example, if the product is too viscous, it may effectively cap production. As can be imagined, the quality of the oil that operators may recover is uncertain. In this situation, a network of interconnecting undersea pipelines (technically known as tiebacks) provides flexibility to control the quality of oil in the pipelines, by suitably mixing flows from several wells. This operational flexibility combined with the flexibility in constructed capacity led to a 78% increase in expected net present value over the base case (strategy 8 in Figure 6.18).

Multi-dimensional evaluation: Although strategy 8 looks like a winner from the perspective of maximizing value, the ENPV does not tell the whole story. The team therefore conducted several multi-dimensional comparisons, such as the one in Table 6.3. This shows that while the expected value of the flexible design is indeed 78% better than the base case design (257 compared to 146), it also costs more to initiate (a minimum of 115 compared to 100). The flexible design is also likely to end up costing more over its lifetime (167 to 100), but a large part of these capital investments would come once the upside potential had been validated through experience, and so would be much less risky than the initial investment. Considering all these features, the most flexible strategy may indeed be the preferred choice.

Sensitivity analysis: Analysts should of course ask themselves if they are assuming their conclusions, that is, if their conclusions are simply a product of some key assumption. In this spirit, the design team examined the sensitivity of the choice of strategy 8 to various assumptions. Table 6.4 shows the results of the sensitivity analysis with respect to the cost of acquiring the flexibility, in particular by investing extra capital in the sub-sea tiebacks. The original analysis assumed that the extra cost of the tiebacks would be about 4% of the Capex of the base case, as indicated by the bold face in Table 6.4. As expected, strategy 8 looks better and better as the cost of flexibility decreases. However, if the cost of flexibility becomes too expensive, strategy 8 would under-perform the inflexible base case. If the flexibility cost 20% of the base case, strategy 8 would deliver lower ENPV (91 compared to 100) at a higher cost!

Such sensitivity analyses are useful to decision-makers because they indicate the range over which a choice, such as for strategy 8 over the base case, might be desirable. In this case, for example, one might conclude that strategy 8 would be a preferable choice so long as flexibility cost no more than 12% of base case Capex. So long as one believed that to be true, then one could confidently choose strategy 8 even if one were unsure of how much the flexibility would cost exactly.

Tables 6.3 and 6.4 about here

Take Away

This chapter presents, and demonstrates by example, a three-step process for evaluating and choosing designs in a context that recognizes uncertainty and examines flexible design alternatives.

The first step consists of using Monte Carlo simulation to determine the distribution of possible outcomes associated with any possible design. The application of the simulation proceeds along usual lines. But analysts have to take special steps in order to evaluate flexible designs properly. Specifically, they have to insert “rules for exercising flexibility” into their model of the system. These indicate when the management should implement flexibility associated with a flexible design. This step develops distributions of performance of the alternatives, conveniently presented as target curves and tables.

The second step involves a multi-dimensional analysis of the several benefits and costs associated with projects. This process distinguishes the evaluation of designs in an uncertain context from conventional evaluation that supposes a deterministic context. The process recognizes that project leaders will have to choose among projects using criteria that they cannot usefully collapse into a single metric. At the least, they will have to balance risks and rewards in several ways. The comparison of alternatives involves comparisons between target curves, usefully highlighted by difference curves. Comparative tables are also most helpful, especially to incorporate metrics, such as capital costs (Capex) that are not part of the target curves.

The third step validates the evaluation and choice by detailed analysis of the possible timing of investments in implementing flexibility. It also examines the sensitivity of the choice of a design to the upside and downside of a project, possible regret, and to assumptions about the nature of the uncertainty.

The chapter illustrates these steps by specific applications. It presents detailed analysis of the simple case of the parking case. This example has the merit of being easy to understand intuitively. To demonstrate how the process can be successfully applied to major projects, the chapter ends with a thumbnail sketch of the results and benefits of the approach to the design of the development of a major deep-water oil field.

Box 6.1

A Simple Simulation

Consider the financial evaluation of the parking garage mentioned in Chapter 3. Its future annual demand is uncertain. The standard model is a spreadsheet of revenues, costs, net revenues and present values. The basic evaluation process inserts the demand for each year into the spreadsheet, which generates the revenues, operating costs, net cash flows and a net present value.

The Monte Carlo simulation uses the basic process as, for many times as desired, perhaps a thousand. Each time it generates and applies a set of possible demands for parking, for each year over the life of the project, as generated by the distribution of possibilities. For example, the analysis might assume that year on year changes in traffic are sampled from a bell-shaped distribution (also known as the Normal distribution). The simulation would then generate a sequence of possible future scenarios, such as those shown in Figure 6.1.

Figure 6.1 about here

From each set of yearly traffic flows, the simulation generates an NPV. From the distribution of possible traffic flows, it thus generates a distribution of the results, as Figure 6.2 shows. Note that this distribution is not bell-shaped like the input. As usually happens, the physical features of the project distort the results. In the case of a garage with limited capacity, the higher levels of traffic do not result in correspondingly higher NPV beyond that associated with a full garage because capacity is limited.

Figure 6.2 about here

Box 6.2

KEY PARADIGM

IMPROVING A SYSTEM IN AN UNCERTAIN ENVIRONMENT MEANS MOVING ITS TARGET CURVE TO THE RIGHT,
MAKING THE ACHIEVEMENT OF TARGETS MORE LIKELY.

DESIGNS OFTEN HAVE A TARGET CURVE THAT WE CANNOT MOVE ENTIRELY TO THE RIGHT.
DECISION-MAKERS MUST MAKE A TRADE-OFF BETWEEN
MOVING PARTS OF THE CURVE TO THE RIGHT, MAKING THESE TARGETS MORE ACHIEVABLE,
AND ACCEPTING A LEFTWARD SHIFT OF OTHER PARTS OF THE CURVE,
MAKING THE CORRESPONDING TARGETS LESS ACHIEVABLE.

Box 6.3

Development of Rio Colorado in Argentina

The Rio Colorado flows through Argentina from the Andes to the Atlantic. Its development involves structures to promote irrigation in the mountain foothills, to control floodwaters, and to provide water for industry and coastal populations. Designs that allocated water primarily to the richer industrial regions (such as A) contributed the most to the national economy. Those that distributed the water most evenly (such as B) did not create substantial financial benefits.

Figure 6.8 about here

If it had been practical to tax the coastal regions to transfer their benefits to the interior regions, the A designs might have been acceptable. However, this was not possible. Somehow, the designers had to mediate the claims of the various regions. Their solution, which provided the preferable balance between the objectives of financial benefit and fairness of distribution, was around the “knee” of the Pareto Frontier, as Figure 6.8 indicates. This was agreed to be the best design because, compared to A, it greatly increased the fairness of the distribution of benefits, without sacrificing economic benefits excessively.[44]

Table 6.1: Table describing measures of distribution of outcomes of a design for a project

Metric	Present Value
ENPV	0 million
Probability of Breaking even	60 %
10% Value at risk or P_{10}	15 million loss
90% Value at risk or P_{90}	9 million gain
Minimum result	34 million loss
Maximum result	10 million gain
Range of results	44 million span
Difference between Median and EPNV	2 million

Table 6.2: Results, in millions, for three possible designs of an oilfield. Best performance in each category highlighted in boldface. (Source: adapted from Hassan and de Neufville, 2006)

Evaluation Metric	Three Possible Designs		
	Rigid	Less Flexible Lower Cost	More Flexible Higher Cost
Expected Net Present Value (EPNV)	822	900	**929**
Standard Deviation NPV	285	334	**165**
Low-end Result (P_{10})	358	308	**788**
High-end Result (P_{90})	**1286**	1134	1148
Expected Present Value of Costs	1006	**688**	969
Initial Capital Expenditures (Capex)	625	**419**	589

Table 6.3: Multi-dimensional comparison of selected design strategies for development of an oil field. (Adapted from Lin, 2009)

Strategy	Values, as % of Base Case ENPV or Capex						Expected	
	NPV			Capex				
	ENPV	Min.	Max.	Expected	Min.	Max.	Reserves	Tiebacks
Base Case	146	-99	400	**100**	**100**	**100**	100	0
5	204	-83	463	131	93	161	146	3.8
8	**257**	**-71**	**578**	167	115	193	**178**	6.3

Table 6.4: Example sensitivity analysis for development of oil field: effect of different costs of flexibility on the value of the Strategy 8 compared to base case. The original analysis assumed that the cost of flexibility was about 4% of Base Case Capex. (Adapted from Lin, 2009)

Metric (as % of Base Case ENPV or Capex)	**Strategy 8** **Flexibility Cost (% of Base Case Capex)**							**Base Case**
	0	2	**4**	8	12	16	20	
ENPV	198	188	**177**	155	134	112	91	100
Min ENPV	43	33	**22**	2	-19	-40	-61	-66
Max ENPV	328	327	**335**	315	295	274	253	251
Expected Capex	169	173	**177**	185	192	200	207	100

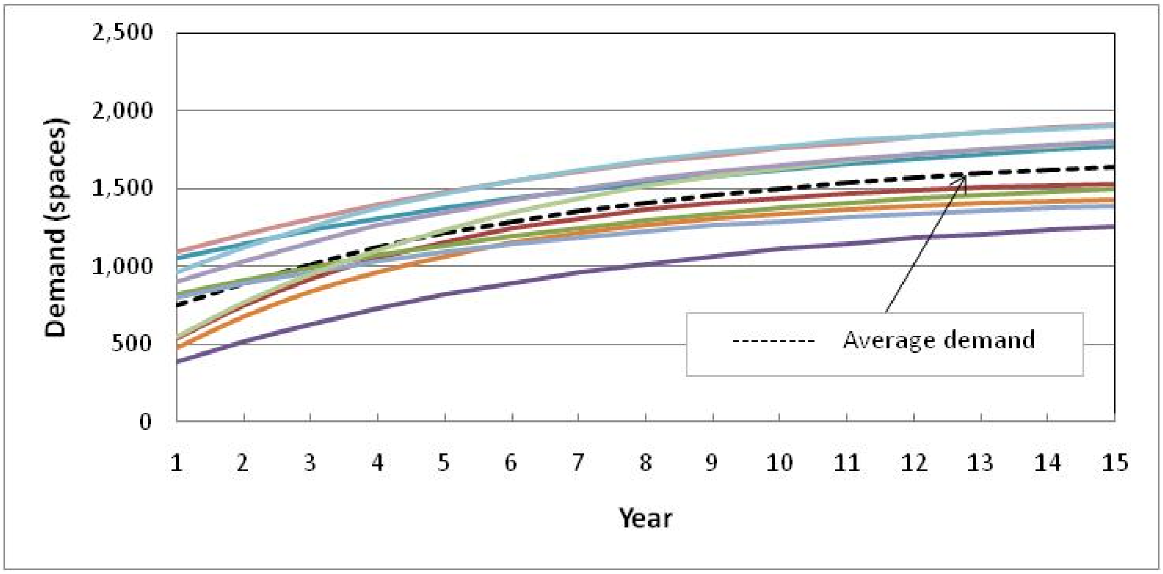


Figure 6.1: Possible future demands for parking spaces over the 15-year life of the garage, generated by the Monte Carlo analysis.

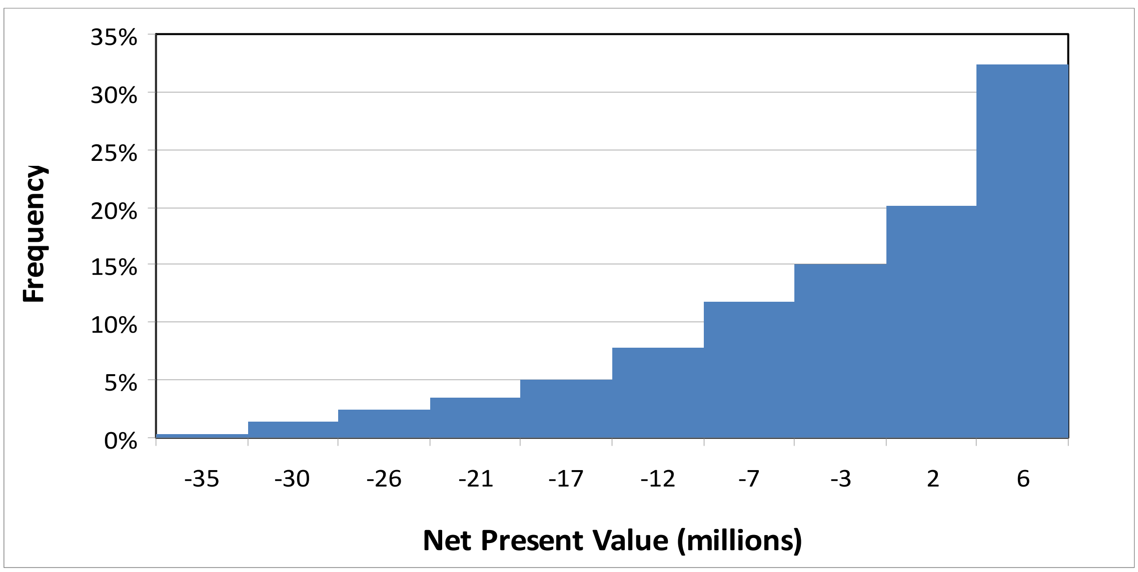


Figure 6.2: Histogram showing the distribution of possible values of the fixed, 5-story parking garage, considering the possible uncertainty in demand. (Same as Figure 3.5)

Note – the numbers on the horizontal axis should delimit the ends of the ranges.

We need to fix this.

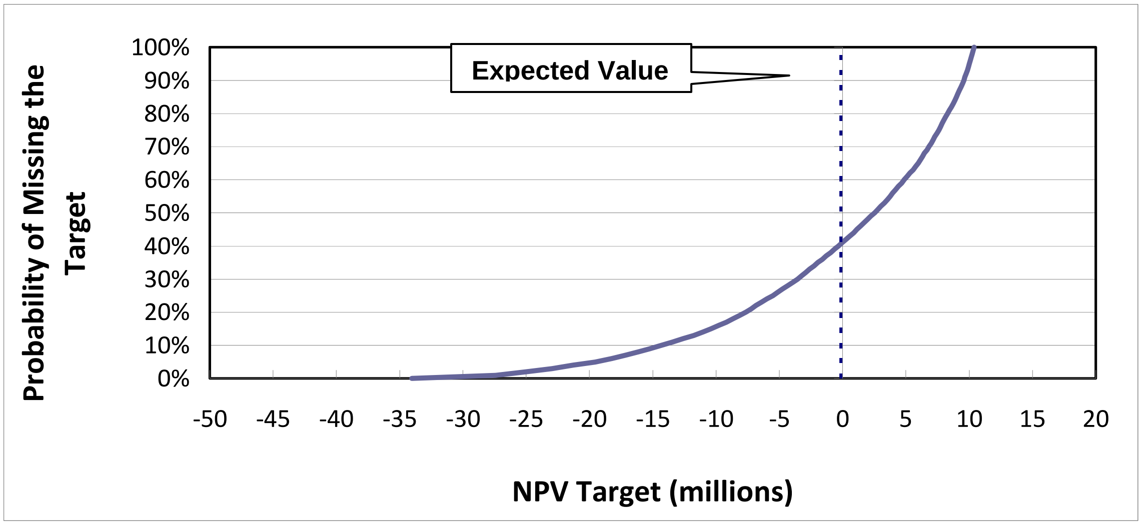


Figure 6.3: Target or cumulative percentile curve for fixed, 5-story parking garage.

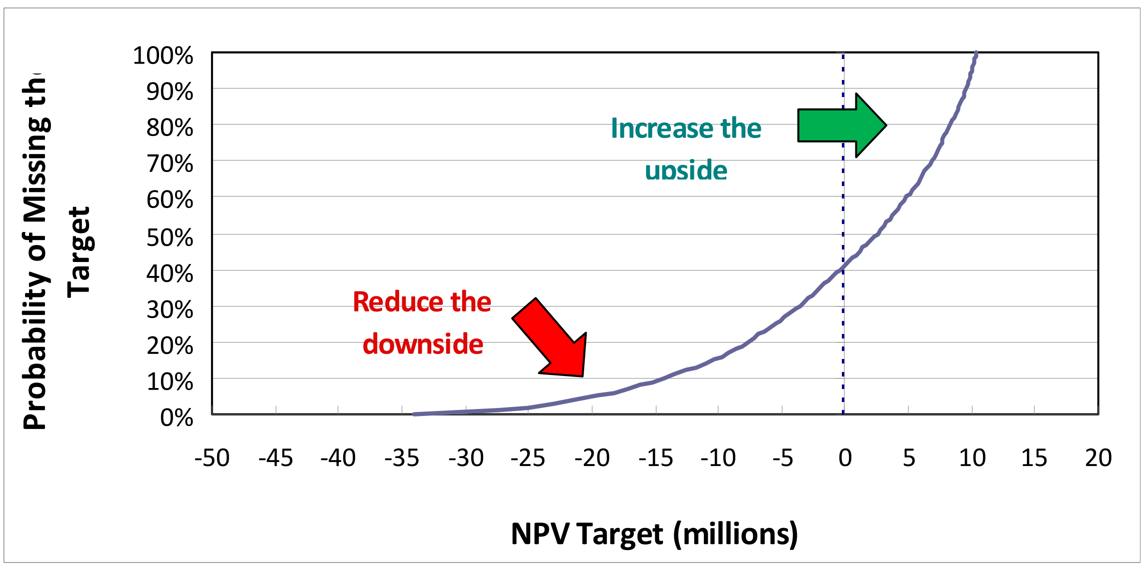


Figure 6.4: Graphical view of desirable ways to improve design by using flexible designs to shift the target curve to the right

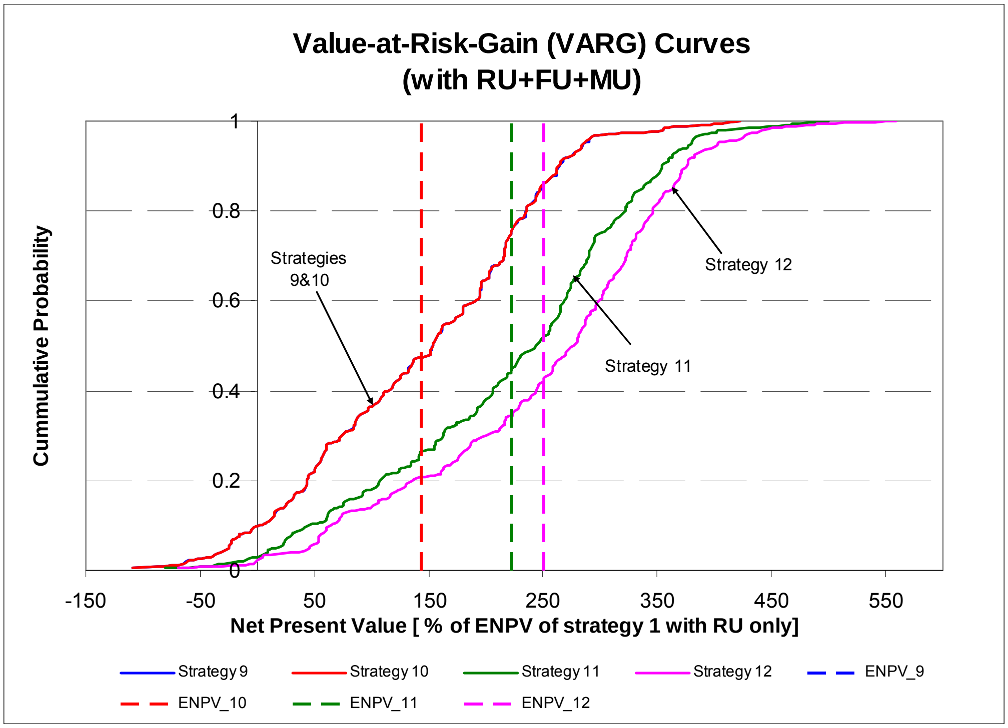


Figure 6.5: Dominant target curves. Strategies 11 and 12 for design of oil platform complex dominate Strategy 1. (Horizontal normalized to preserve commercial confidentiality) (Source: Lin, 2009)

We need to eliminate title at top of figure; to rename the vertical axis and generally make all labels consistent across sets of like figures; and to Label strategies 9 and 10 as Strategy 1.

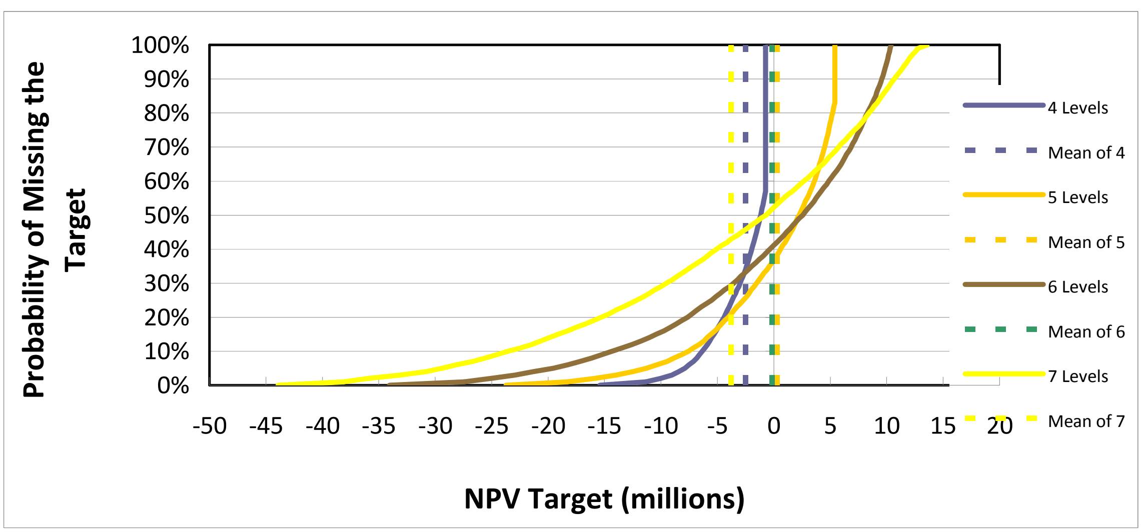


Figure 6.6: Overlapping target curves associated with fixed, that is inflexible, designs of parking garage with different number of levels. Michel – the color for the 6-level should not change color

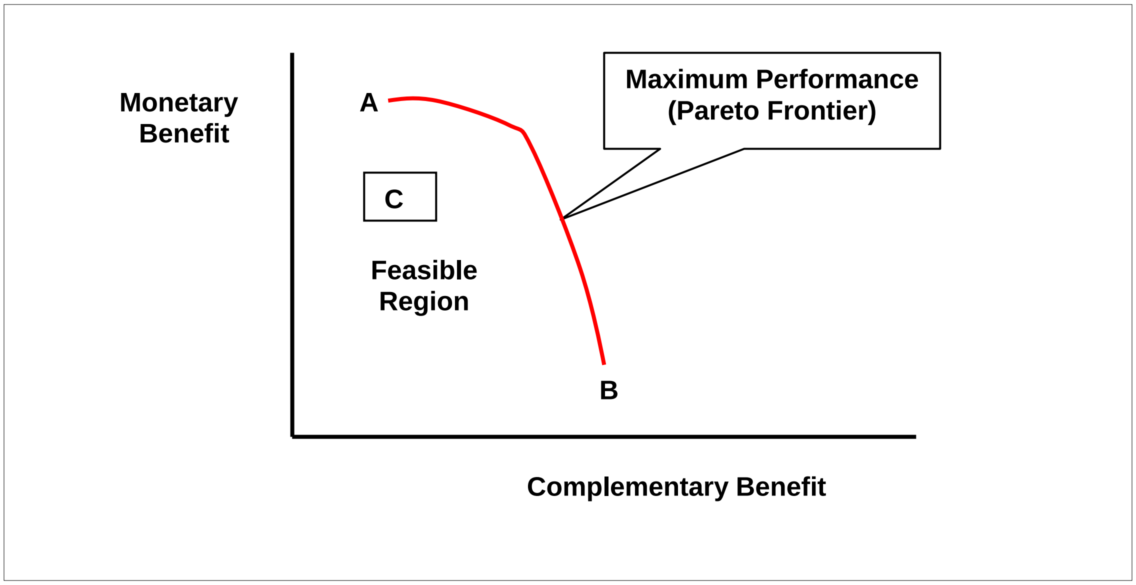


Figure 6.7: General diagram of the performance of designs when all axes represent benefits. In this case, the set of best designs, known as the Pareto frontier, is furthest away from the origin.

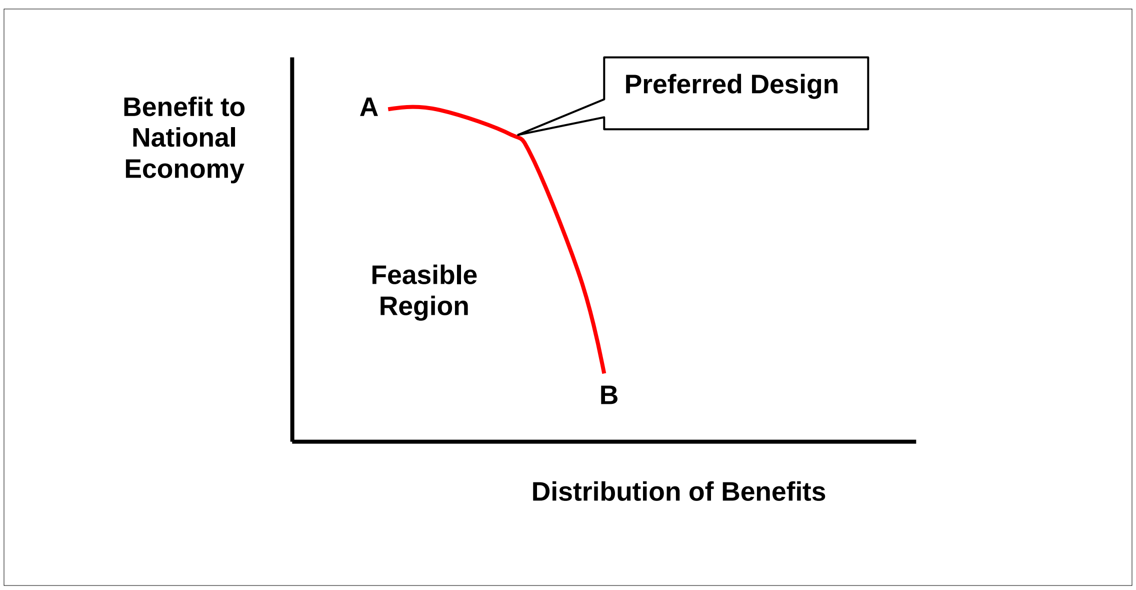


Figure 6.8: The preferred design selected from the set of best designs for the development of the Rio Colorado in Argentina. (Adapted from Marks et al.)

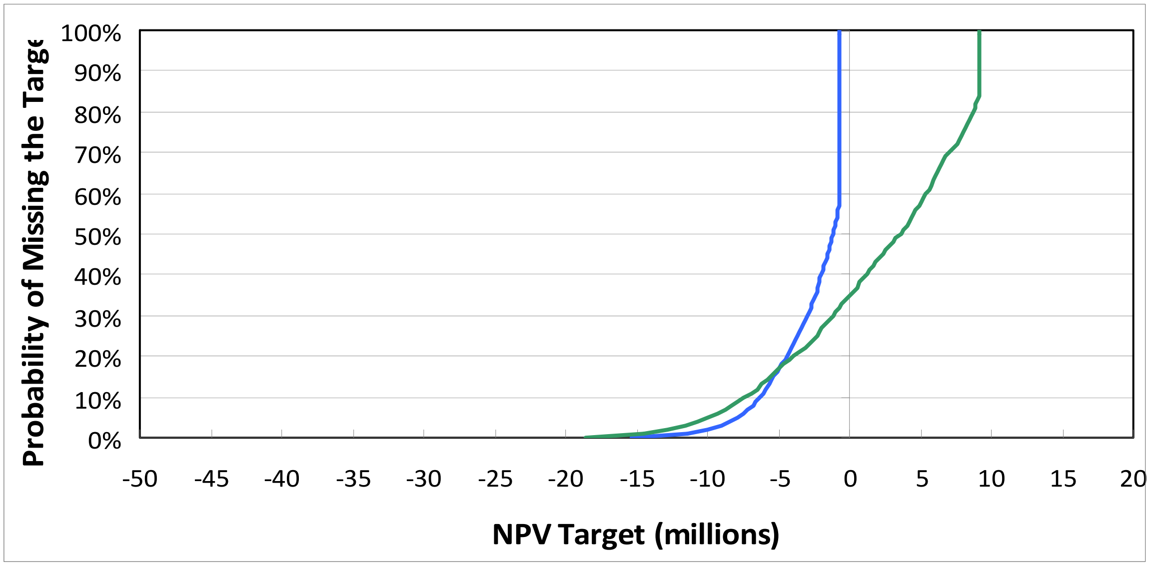


Figure 6.9: Comparison of target curves for fixed and flexible designs of 4-level parking garage.

.

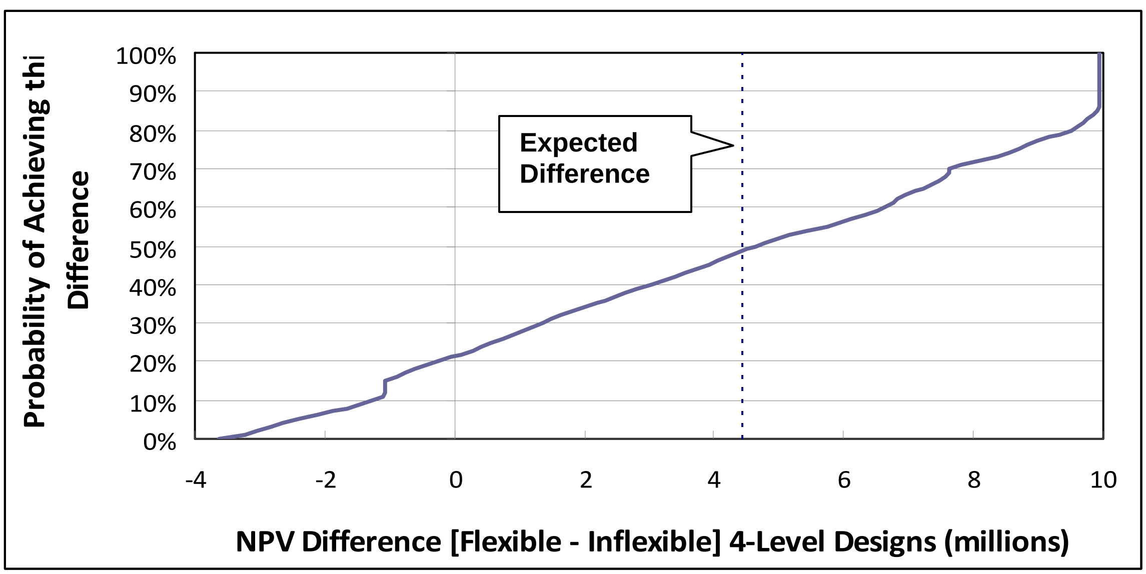


Figure 6.10: Difference curve comparing fixed and flexible designs for the 4-level parking garage.

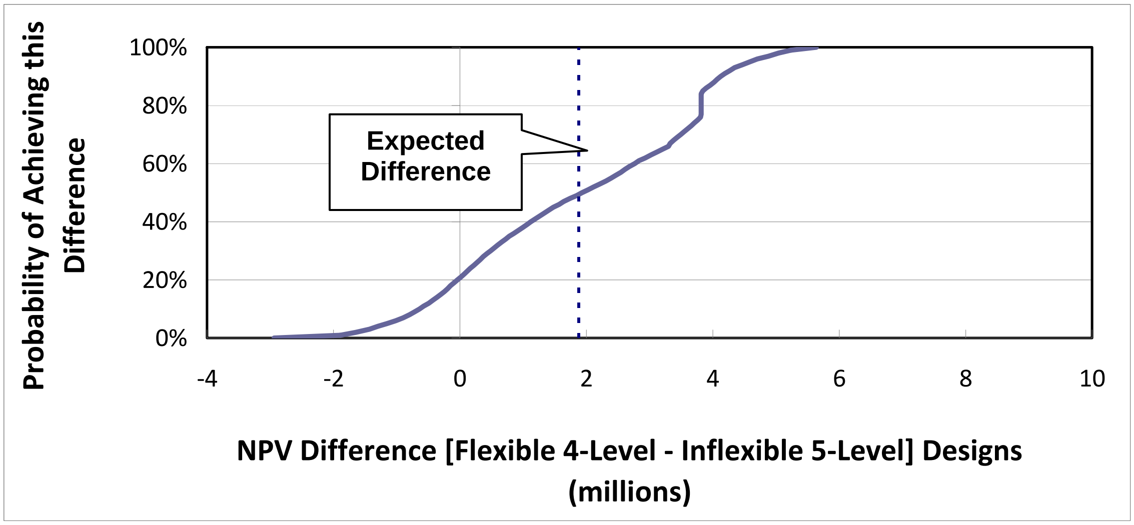


Figure 6.11: Difference curve showing the advantage of the 4-level flexible design over the 5-level fixed design.

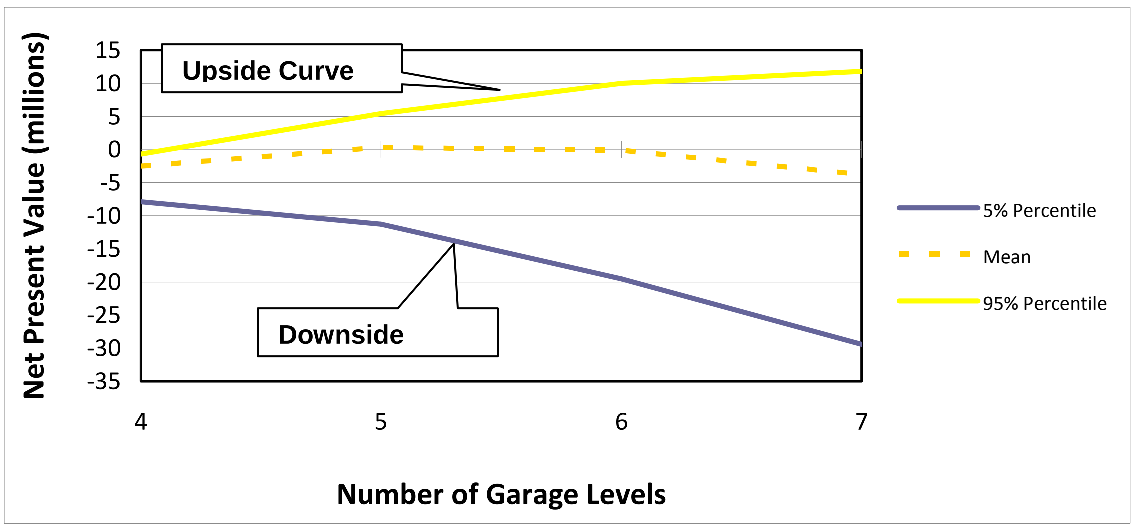


Figure 6.12: Upside-downside curves associated with designs for different fixed number of levels. The upside curve is for a 5% chance of a gain equal or above (the same as a 95% of being below) and the downside curve is for a 5% chance of results less than indicated.

Michel, can be redo? We need to be consistent about using extreme values – 10 and 90 % as in text (and oil industry, for example).

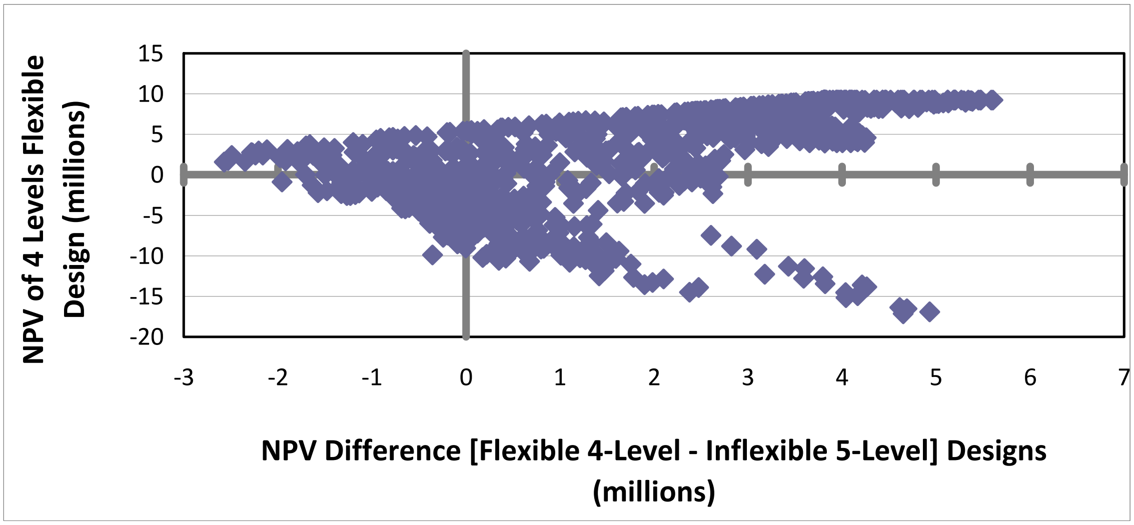


Figure 6.13: Regret Plot: case by case comparison of flexible 4-level flexible design and fixed 5-level design for 1000 demand scenarios.

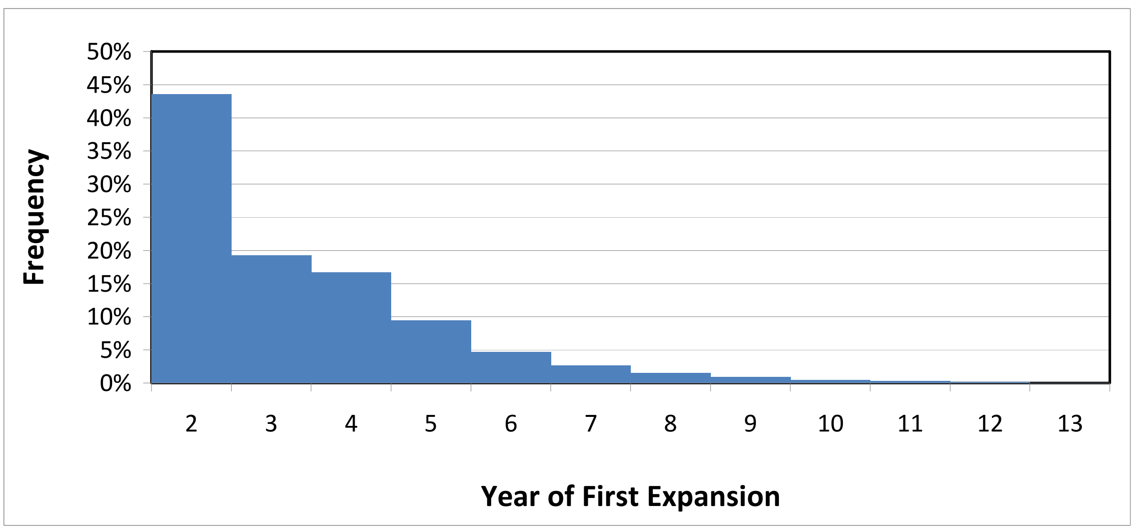


Figure 6.14: Year of first use of flexibility to add levels to parking garage

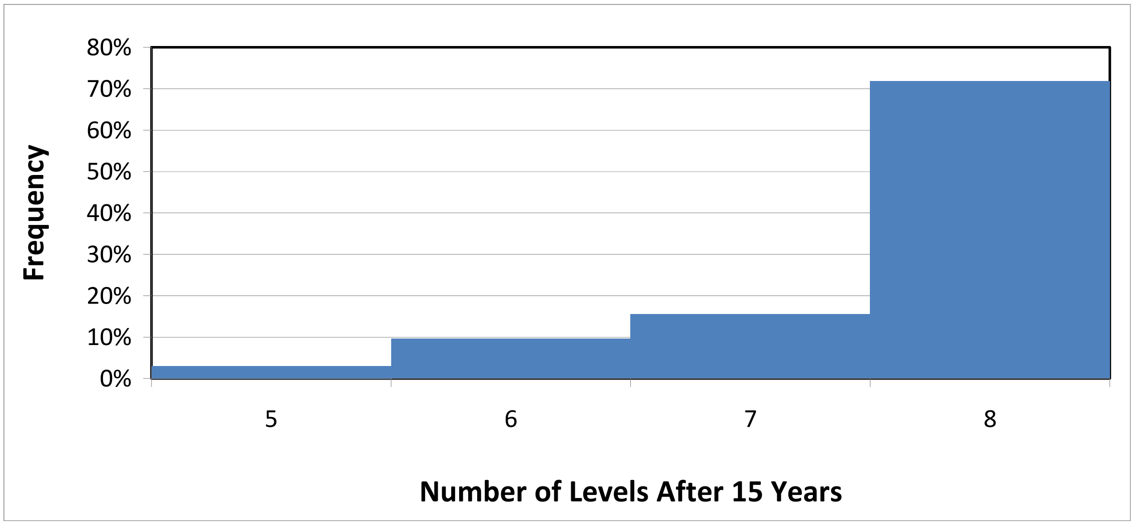


Figure 6.15: Amount of flexibility used over the life of project for parking garage

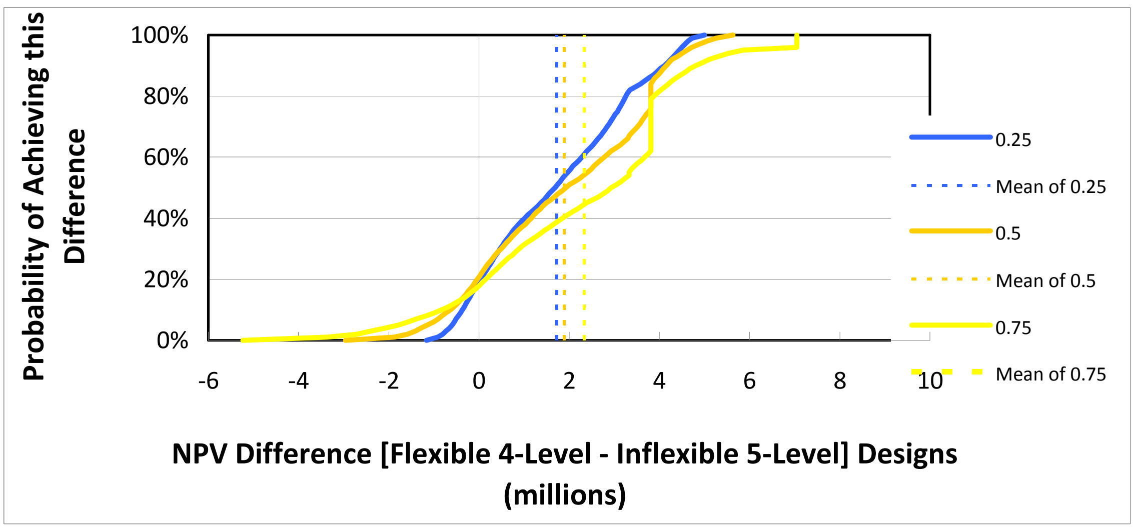


Figure 6.16: Sensitivity of value of flexible design to level of uncertainty (base case compared to higher and lower levels)

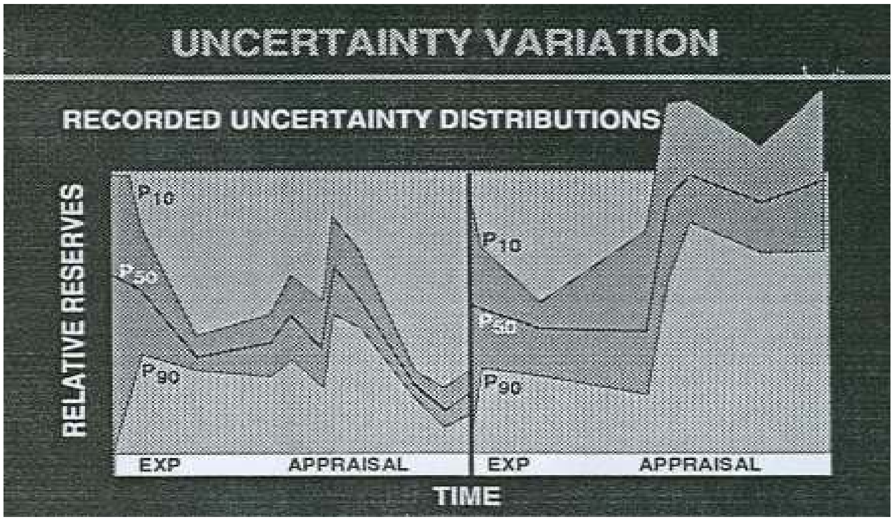


Figure 6.17: Example variation in best estimates of oil that can be commercially extracted from two fields (Source: Lin (2009) from BP sources). (Repeat of Figure 2.1.

This figure needs to reorganized

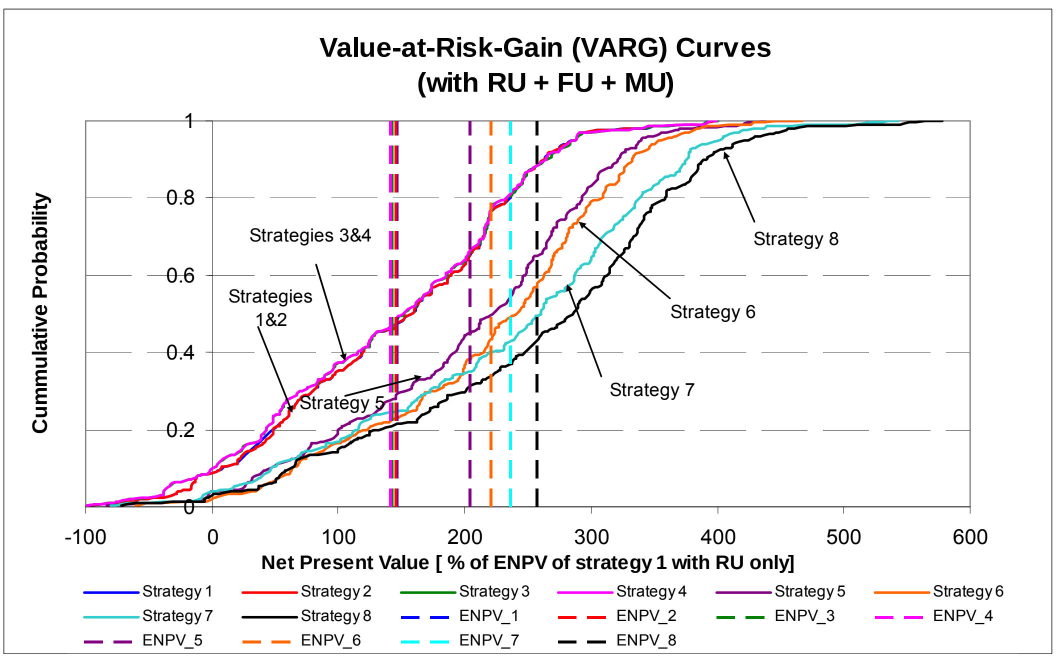


Figure 1: VARG curves for strategies 1~8 (with initial 150 MBD capacity) under RU + FU + MU

Figure 6.18: Target curves for alternative designs for the development of the deep water oil field (Source: adapted from Lin, 2009)

This needs to be redone – new titles, new axes, etc.

CHAPTER 7
IMPLEMENTING FLEXIBILITY

"It's not enough to have an idea; you have to know how to move the furniture around"

Remark by Senator Timothy Wirtz

It is not enough to design flexibility into the system; designers need to make sure that future managers will be able to use this asset when it would be desirable to do so. If the designers do not create the conditions that will make it possible to implement the flexibility as needed, they will destroy its value and waste their effort to create the capability. Designers thus have the responsibility to make sure that the flexibility they develop will continue to be available. They need to do their best to keep the flexibility "alive".

Maintaining the capability to implement flexibility is more than a technological issue, it is a social process. The physical capability to make use of the flexibility designed into a system is unlikely to disappear. The extra strength built into a building to allow for the construction of more floors will continue to be available indefinitely. However, a wide range of institutional, regulatory, or political developments may easily prevent the owners of the building from using this strength. For instance, new zoning codes might change the allowable height for a building on its property. New safety regulations might require unanticipated space for emergency exits for all new construction. Bankers might not be willing to finance the current owners. To create useful flexibility, the design process needs to consider and deal with such matters. Designers need to extend their thinking beyond the purely technical aspects.

Maintaining flexibility is also a process that extends the design period well into the future. Just as the logic of dealing with uncertainty impels us to design for physical developments that we can implement as needs or opportunities arise; it also impels us to think through the implementation of these developments over time. In this context, the design process does not stop when it delivers the original plans to the developers; it continues so long as design flexibilities are there for possible implementation.

This chapter provides guidance on how to make sure that managers can implement flexibility designed into a system as and when desirable. By way of motivation, it first catalogs and illustrates common phenomena that prevent future managers from making good use of flexibility in design. The chapter then proposes both initial preventative and ongoing operational actions to preserve the capability to implement flexibility. It concludes with an illustration of an organization that has put into practice the principles recommended here.

Common obstacles to implementation

There are many obstacles to making good use of design flexibility. While we can organize these in several ways, it seems useful to consider five general categories for this discussion. These are:

- <u>Ignorance</u> – future managers forget about or otherwise ignore that the flexibility exists.

- Inattention – nobody monitors or pays attention to the circumstances that would trigger the appropriate use of the flexibility, and therefore managers miss the good opportunities to use their flexibility.
- Failure to plan – the design process simply did not think through what needed to be done to implement the flexibility, and thus created insurmountable obstacles.
- Stakeholder block – groups using or affected by the system think that the implementation would somehow hurt them, and manage to block the use of the flexibility.
- External developments – regulatory, political or other developments eliminate or otherwise constrain the right to implement the flexibility.

Let's consider each of these.

Ignorance: Future owners or managers of a system may simply not understand the flexibility designed into a system. If the owners or future system operators were not part of the original design process, they may simply not understand the full capabilities associated with their system. This is the professional counterpart to our personal experience: how many of us truly understand the full capabilities of our laptops or cell phones? Obviously, if owners do not understand what they can do, they will not think about using it.

In a similar vein, the process for managing the system may easily lose sight of the flexibility. The persons responsible for this knowledge may retire or move on to different responsibilities. Most commonly, the system will have become the responsibility of new decision-makers through a sale of the property, a merger of companies, or a change in government. Box 7.1 provides a typical example.

[Box 7.1 about here]

Inattention: To implement flexibility well, systems owners have to do it at a suitable time. This means that they have to pay attention to developments. When the flexibility is associated with routine processes, this may be easy to do. For example, equipment manufacturers with flexible production lines easily adjust these facilities according to the orders that come in. Similarly, airport terminals with "swing" gates allocate them to international or domestic passengers over the day as the mix of traffic changes.[45] In such cases, the signal that it is time to implement the flexibility is almost automatic and comes directly to the persons in charge of the facility. The airport or plant managers sense the demand and can make changes.

Difficulties arise when the flexibility addresses uncertainties that are not routine. They compound when the persons or departments that sense the opportunity to exercise flexibility are not those capable of doing so. Consider a fleet of communication satellites designed with the flexibility to reposition themselves to serve different needs or markets. The sales department might know that it would be nice to have some additional capability in an emerging market, but neither have the ability to implement the flexibility to obtain the new capacity, nor perhaps even know that the flexibility to do so exists. Similarly with the development of oil platforms: the production business units may want new capabilities, but neither understand that the design is ready for it, nor have the ability to get project designers to pay attention to their desires.[46] Because of a lack of attention in the right places to the circumstances that call for the implementation of flexibility, managers may not be able to implement it.

Failure to plan: Implementing any significant form of design flexibility takes effort. It is not just a matter of flipping a switch (or, as in the case of financial options, of notifying a stockbroker). Managers have to obtain the budget for the project, coordinate with existing users and stakeholders of the system, obtain local and possible international permissions, secure contractors or others to execute the operation, etc. Implementing flexibility requires careful planning. Managers need to know how "to get from <u>here</u> to <u>there</u>."

The problem is that the designers of the technical system often fail to think through what it takes to implement their flexibility. They need to develop an implementation plan, and to take the corresponding steps to put the appropriate capabilities in place. Sometimes the failure to plan is glaringly obvious, as with building complexes in which the heating and cooling plant were placed as close as possible – and right in the way of the possible expansion, thus making it impossible.[47] More generally, the difficulties can be subtle, as Box 7.2 illustrates.

[Box 7.2 about here]

Stakeholder block: Many stakeholders normally participate in the execution of any important project. Their acceptance and cooperation may be essential to the implementation of flexibility. As it often happens that they can block projects, managers must deal with these interest groups.

Stakeholders internal to the organization may be significant. Groups within an organization may resist some action that, while good for the organization, adversely affects their standing or benefits. Consider the situation of the organization running the fleet of communication satellites: the group in charge of the operation of the satellites may resist requests for repositioning the satellites to expand markets. Why? Perhaps because the company measures the performance of the operating department on their costs, so that this group does not want to spend whatever is necessary to implement the flexibility, which almost certainly is not in their normal budget. Likewise, the operators may be concerned about possible technical difficulties and would prefer not to mess with a system that is operating well. In any case, their cooperation will be essential to the implementation of the flexibility of repositioning the satellites.

External stakeholders are also often significant. They may also be more difficult to work with than internal stakeholders. Within a company or organization there may be a decision-maker who can resolve internal disputes authoritatively, such as might exist between the sales and operating departments of the organization operating communication satellites. External stakeholders by definition are not subject to such power. Somehow, system managers may have to negotiate with them or use the political process to implement desired flexibility. Box 7.3 illustrates the issue.

[Box 7.3 about here]

External developments: Finally, all kinds of external developments can make it impossible or impractical to implement flexibility. For example, new building codes may prohibit what was previously acceptable. Such was the case after earthquakes in California led to changes in requirements for all new construction.

System managers may be powerless to prevent changes that can block their use of flexibility. They can, however, monitor developments and plan to implement flexibility (if desirable) before the changes take place. Thus, the staff of the Health Care Services Corporation (HCSC) carefully monitored the evolution of the zoning codes in Chicago, to be sure that HCSC would be able to exploit its flexibility to expand its building vertically, as Chapter 1 describes. For them, as for other developers, good working relations with the authorities were essential to their ability to implement planned flexibility.[48]

Initial preventative actions

To maximize the likelihood of being able to implement design flexibility, the design process can take both initial preventative and ongoing operational actions. The ones lead into the others.

The initial preventative actions cluster into three major types:

- <u>Integrated project delivery</u> – creating the design with the participation of major stakeholders in the process, and thereby uncovering and understanding many of the issues that might eventually be barriers to implementation.
- <u>Development of game plan</u> – carefully thinking through what would be required for future implementation, thereby avoiding the creation of obstacles while laying the groundwork for easy implementation.
- <u>Anticipating developments</u> – in accord with the game plan, taking actions that increase the potential for implementing of the flexibility, should it ever be desirable.

Integrated project delivery: Integrated project delivery is a non-traditional process for designing and implementing projects. Project developers increasingly use this innovative process, but it is still a novelty. The basic concept is that all the stakeholders involved with the delivery of a project should work together collaboratively and simultaneously. The designers work together with the manufacturers and the prospective operators, to create facilities they can easily build and maintain; and these technical professionals also work with the forecasters and financial institutions to plan coherent phasing and flexibility of the project.[49]

Integrated project delivery contrasts with the traditional process, in which distinct professional teams work independently. The stereotypical version of normal project delivery is that forecasters or others develop requirements for a system; the process passes these on to designers that translate these specifications into drawings; these plans then get turned over to manufacturers who try to build the system as best as they can; and so on. As the description indicates, this process limits the scope for flexibility. For example, the manufacturer has little ability to influence the concept of the design, which engineers have fixed by the time it arrives for implementation. Moreover, the traditional process does not give the participants much opportunity to understand each other's perspective and concerns; they thus have little chance of anticipating and dealing with issues that may hinder eventual implementation of flexibility.

An integrated process of project delivery is useful because it enables the participants to anticipate and avoid issues. For example, if the designers of the Royal Victoria Infirmary had been in touch with the persons organizing the financial arrangements, they could have made sure that the loan documents included a description of possible evolution of the project and thus facilitated onward financing (see Box 7.2). Vice-versa, if the financial process had been aware of the clever way the designers were planning the hospital, they could have taken the initiative to make sure this was clear to prospective lending agencies.

Development of game plan: The development of a game plan for eventual implementation is basic to successful implementation. The idea is simple: designers should lay out the steps that managers would need to take to implement each particular form of flexibility and to anticipate how best to carry them out. The difficulty lies in effective implementation.

A good game plan will correctly anticipate what tasks necessary for implementation. Unless the design process carefully consults with the stakeholders in the project, they are likely to miss key details. Thus with the Royal Victoria Infirmary: the architects and engineers had doubtless thought through the purely technical construction process for implementing the flexibility to expand the hospital. However, whatever they planned seems to wasted because appropriate financing arrangements had not been included in the plan. Thus, some form of integrated project process is essential for the development of a good game plan for future implementation.

Anticipating developments: Following through on the game plan, the design process may anticipate that managers must take some actions now to facilitate future implementation of flexibility. These actions may be as necessary to success as the technical actions that the design has inserted into the system.

For example, US national airport authorities (the Federal Aviation Administration) have been anticipating the possibility that Chicago would need a major new airport, and have been working with regional authorities to develop this plan. In this case, they have generally worked out the design and location. However, this is not enough to achieve implementation should that ever be desirable. So they have also been supporting the purchase of the land that would be needed. Their reasoning is that if they waited to buy the land when they actually had decided to build the airport, it might have been transformed from farms into suburban developments and be impossible – politically or financially – to acquire.[50] Similarly, Boston and Los Angeles have maintained Worcester and Palmdale airports operational, as insurance that they could exploit the flexibility to use their capacity if ever that made sense.

Ongoing operational actions

The owners and operators of a system need to help sustain the ability to implement flexibility for as long as this may be useful. In so doing, they act as extensions of the initial design process, which cannot be considered complete until all possible flexibilities have been used or discarded. Three kinds of actions seem most useful in this regard:

- Maintaining the right to implement – as the ability to implement is often contingent on various legal permissions, it is important to keep these up-to-date and in effect.
- Maintaining the knowledge to implement – effective implementation requires people, and more generally institutions that understand the nature of the flexibility and know how to proceed when the opportunity is attractive.
- Monitoring the environment – this is crucial to knowing when it would be desirable to implement flexibility and to obtain its highest value.

Maintaining rights: The ability to implement flexibility may require legal or other permissions. These can easily evaporate if not maintained. For example, many European countries now require owners to renew patent rights every year. Companies wanting to maintain the right to produce a product with patent protection must thus be careful to keep their patents in force. Likewise, it is frequently the case that planning permissions to build a facility expire if the permit holder does not use them within a specified limit. The watchword in many cases is thus: “use it or lose it”.

One of the more effective ways to maintain the right to do something is to keep on doing it from the start. In most regions, it is normal to accept that owners can maintain existing rights to operate almost indefinitely. A person running a pig-farm in an agricultural area that is undergoing suburbanization can expect to carry on the business – those who have been moving in had to consider this activity. However, a newcomer is unlikely to get permission to start a new pig farm. Similarly, an industrial plant can expect to continue its operations as the city grows around it, even as the political process might stop a new factory. In this spirit, it is essential for Boston and Los Angeles to keep airport operations active at Worcester and Palmdale, if they want to maintain the ability to implement the flexibility to expand at those locations. If they ever allow these airports to cease operations, the neighbors will grow accustomed to the closure, and reopening the airports would be highly problematic.

Maintaining the knowledge: In parallel, it is important to maintain institutional knowledge about the nature of a flexibility to insure the ability to implement it in the future. Most basically, it is necessary to know that the flexibility exists (see Box 7.1). However, this by itself is generally not enough. It is important to know the crucial details. For example, the Tufts Dental School in Boston was able to exploit the flexibility to expand its building vertically because it could access the plans and knowledge of the local engineers who had originally designed the building a generation earlier. This was a piece of good luck for them, as professional firms merge, principal designers move and retire, and requisite knowledge generally may no longer be available.

Organizations planning for the possibility of implementing some flexibility should pay attention to the process of maintaining access to the design knowledge that might eventually be essential. Thus, the staff of the HCSC kept up close relations with its original architects (and with its suppliers) so that when the time came for doubling the height of its building, they had their team in place. If they had not proactively maintained this knowledge base, it is doubtful if they could have implemented the flexibility to expand their building vertically as originally intended.[51]

Maintaining staff continuity can be the essential ingredient for ensuring effective implementation of flexibility. In case of the HCSC, for example, the process was effective because the same person headed their facilities department from the planning of the original building in Chicago, built with the flexibility to expand vertically, to the completion of the eventual expansion almost 20 years later. Because that person understood the flexibility, because he maintained the institutional knowledge through his relations with the project architects, the project could go smoothly in a way that is almost unthinkable without his presence. In this case, as can easily occur elsewhere, the feasibility of a major project depends upon a single or just a few individuals. An important key to implementing flexibility successfully may simply be to maintain professional staff in place – the cost of this insurance can be insignificant compared to the size and value of the project!

Monitoring the environment: Flexibility has the most value when implemented at the right time. It is therefore important to know when the right time might be. Organizations considering the possibility of implementing some design flexibility thus need to set up and maintain a process to track the factors that might trigger the effective use of some flexibility. This process needs to have at least two features. It should:

- <u>Highlight triggers</u> – conditions that indicate when it is desirable to implement a design flexibility, and
- <u>Establish useful information flows</u> – to make sure that the information needed flows to the persons that can use it to decide whether and when to exercise the flexibility.

The design process will normally have identified the conditions that call for the use of the design flexibility. This step is an essential part of the process. To be able to value and thus justify any possible flexibility we might build into our system, we have to know when, under what conditions we might use it. In fact, these conditions define the ‘rules for exercising flexibility’ that we need to embed in the Monte Carlo simulations, as indicated in Chapter 6. These are the triggers that the organization, that its managers need to keep in mind over the life of the system.

The issue is that the design team needs to highlight the conditions for implementing flexibility. It needs to pass the knowledge about them on to those who will be responsible for the design in the future. An understanding of triggers that justify the implementation of design flexibility needs to be brought forward from the depths of the analysis to the top of the organization that will operate the system in the future. As part of its planning for implementation, the design team needs to create signposts to guide future managers.[52]

Highlighting triggers is often not easy. Sometimes the conditions justifying the use of flexibility are obvious, and great efforts are not required. In the case of HCSC, which had a need to keep its staff on one campus, the trigger was closely associated with the number of staff and the difficulty of crowding more employees into the existing facilities. As space became tight, the desirability of expansion became clear. However, the conditions favorable for the use of flexibility are generally not obvious. They often combine several factors. The desirability of expanding the capacity of an oil platform, for example, depends on both the price of oil, the estimates of the amount recoverable, and the speed at which this oil can be extracted. Moreover, the conditions that would reasonably trigger the implementation of the flexibility are likely to combine in complex ways – very high oil prices might compensate for lesser estimates of recoverable oil, for example. Creating and maintaining suitable signposts for future managers can thus be difficult.

Having good signposts is not sufficient, in any case. It is also necessary to make sure that relevant information will flow to the future leaders of a system so that they can compare the signposts and the current information. The essential part consists of confronting the available data and the trigger conditions. This is what is required to create the conditions for thinking about and eventually choosing to take advantage of flexibility in design.

It is necessary to create a process whereby appropriate information is available at the right place. Normally, an organization will have the relevant information available somewhere. The question is whether it is in the right form and the right place. Consider the example of expanding the capacity of an oil platform. We may safely assume that different departments of the oil company will be fully aware of the price of oil, of the estimates of recoverable oil in place, and of the pressure fields that control the pace of extraction. The difficult bit is making sure that such information comes together and usefully becomesof available to those who are in a position to decide whether to implement the design flexibility.

The process of monitoring the environment to know when it might be desirable to implement flexibility is obviously crucial. If we do not have a way to know when to act, it is clear that we are unlikely to act. Unfortunately, there is no simple way to ensure that the information will be available to the right people in good time. Different approaches will be better in different organizations.

Example Application: Dartmouth-Hitchcock Medical Center

The Dartmouth-Hitchcock Medical Center (DHMC) provides a good example of how to design and implement flexibility. The DHMC has both built facilities that it can expand and otherwise alter flexibly and set up an organization that ensures timely and effective use of the design flexibility.

The DHMC is located in the State of New Hampshire in the United States. It is a major regional hospital closely associated with a university with a leading medical school. It thus has to serve four important stakeholders. In addition to providing surgical facilities and hospital beds for major interventions, it must also: cater to medical practices that serve walk-in patients from throughout the area; respond to the clinical needs of around 10,000 (???) students, staff and dependents of Dartmouth College; and provide specially facilities for research carried out under the sponsorship of the US National Institutes of Health, which have their own requirements.

In the 1980s the DHMC had the opportunity to move from a conglomeration of facilities nestled on the university campus to a large new site. The DHMC took advantage of this opportunity to create a flexible design that would allow each of their major stakeholders to expand or change their facilities as and when needed. Also, and this is the point of this example, set up an organization and processes to facilitate effective implementation of the design flexibilities.

Flexible physical design: The essence of the design for the DHMC is that facilities for each of the major stakeholders are spaced along a central spine, which provides for circulation and a variety of common services throughout the DHMC (see Figure 7.1). The buildings serving each of the major clients extend out at right angles to the spine. These facilities each differ in size and content, according to the different needs of their users. The design kept certain elements, such as floor heights, uniform to facilitate communication, utilities and the like. Since these facilities are essentially independent, they can expand or change function at their own speed according to their needs and financial capabilities. At one point for example, the DHMC added extra in-patient beds (at the far left of Figure 7.1). Most recently, towards the right side of the plan, the research group added three floors to its building.

Notice also that the design locates the utility plant (proving heating and air conditioning for the DHMC) far away from the medical buildings. This is unusual. Common practice places the utility plant close to the buildings it serves, in order to minimize the cost of ductwork and to reduce the losses (in heat or cold) along the length of these pipes. In this case, however, the design recognized that a remote location was an essential part of the flexible design. If the utility plant were next to the medical buildings, it would have limited their possibilities for future expansion. Moreover, the design made the utility plant much larger than it had to be to serve the initial facilities. The idea was to facilitate eventual expansion as and when needed.

Design for implementation: The DHMC provides an excellent example of how it is possible to design for implementation of flexibility. In their case, they created a separate organization to manage and develop the facilities. This group is a special-purpose consortium of the major stakeholders. Its day-to-day role is to manage the common utilities and services such as parking. As regards the implementation of flexibility, it has a long-term role to maintain relationships with architects and other designers, to oversee all construction and, in general, be the common agency to construction management for each of the four major stakeholders. In turn, representatives of the stakeholders meet regularly with the facilities group to keep abreast of ongoing issues and their aspirations for further developments.

The DHMC facilities group provides a role model for how to achieve each of this chapter's recommendations for implementation:

- Integrated project delivery -- the DHMC group provided the central node to coordinate and mediate the various desires and issues of the stakeholders, the designers, the construction crews and the local and regional planning authorities.
- Development of game plan – the DHMC clearly has thought through how it would proceed to implement the development of the hospital facilities in its ever-changing technological and regulatory environment. They made sure the physical plan enabled separate expansions according to the several needs and opportunities of the stakeholders and – by establishing the separate facilities group – gave themselves the ability to monitor and respond to events coherently and consistently.
- Anticipating developments – the DHMC not only make sure that the original utility plant was located remotely, but has also carefully sited subsequent supporting facilities. For example, it has built structured parking with restraint, giving great emphasis to preserving ground-level parking near the buildings so as to maintain the ability to develop the site easily.
- Managing rights -- the original design might be described as a smaller set of buildings surrounded by a large area of ground-level parking. Creating this large surrounding space required them to clear and grade much more space than they might otherwise have done. But the benefit of clearing this space and using it from the start is that they have effectively established the right to use the cleared area. They will not have to face public pressure to maintain the wooded areas, as they might have had to if they had only cleared a smaller area.
- Maintaining the knowledge – the DHMC facilities group has managed to maintain the knowledge of flexibility by keeping together a consistent team over more than two decades. Their facilities manager and her team have been in place a long time, and they have been working with the same architects over that period.
- Monitoring the environment – Through their regular meetings with the stakeholders, the DHMC facilities team manages to stay current on the desires of the stakeholders for expansion, their possibilities for financing, and the timing of possible implementations of the flexibility designed into the original system.

Take-Away

Flexibility that that we cannot implement when needed is worthless. It is thus essential for designers to pay close attention to how future system managers could implement any flexibility designed into a project.

Thinking about implementation stretches conventional concepts of design. This is because many of the obstacles that may have to be overcome are social, involve legal requirements, economics and financing, institutional rivalries and of course politics. Moreover, the implementation phase for any design flexibility may last years, often a decade or more. This is not the common view of design practice. Yet if effective design involves getting things done, as we believe it should, then design for flexibility extends practice both over time and into social issues.

To insure the possibility of effective implementation, designers should pay attention both to the immediate design process and to its extension over time. In the immediate, they should encourage collaborative consultation among the stakeholders, in some form of integrated design process, and pay attention to developing a game plan for implementation. Over the life of the project, or as long as design flexibilities may be available, it is important to maintain the right to exercise the flexibility, establish signposts that indicate when flexibility should be implemented, and establish processes that will ensure that timely relevant information flows to the system leaders who have the authority to implement the flexibility. Good designers can do this, although it sounds difficult. Good designs teams are successful at implementation, as the example shows.

Box 7.1

Ignorance of flexibility

Parking Garage: The multi-level garage in a major shopping mall in Britain illustrates how ignorance can prevent the implementation of flexibility designed into the system. When it would have made sense to expand the structure, some five years after its opening when the mall was doing well, the then current did nothing because the then had no idea of what they could do.

Indeed the developers of the garage had sold the property to a new owner, as is usual. Developers are in the business of taking greater risks with the hope of greater rewards, and they routinely sell off properties once they have become profitable. In this case, the new owner was an investment trust with a large portfolio of all kinds of properties. They were financial experts, not designers and did not think of design capabilities.

The capability to expand the garage was not self-evident. Although the engineers had carefully built the facility with the capacity to add on several levels, this was not visible. The foundations and columns much stronger than needed for the original structure, but their strength was hidden. The construction process had obviously buried the foundations, and carefully encased the steel in the columns in protective concrete. So the new owner could not see the capability for expansion when conducting the due diligence inspection of the property. It was a classic case of "out of sight, out of mind."

Box 7.2

Failure to plan

Newcastle hospital: The structural design for the Royal Victoria Infirmary in Newcastle-upon-Tyne provided the strength to add extra floors in the future. This gave the owners the physical ability to adapt the hospital to future needs as required.

However, the hospital was not able to implement this flexibility once the high demand for its services seemed to justify the relatively small expense of the expansion. This was because the design process had not created a workable plan for eventually obtaining the effective cooperation of all the other stakeholders. In this case, the particular obstacle laid in the original documents describing the project to its financing institutions. These agreements did not discuss the possibility of changing the project. This meant that the original terms did not allow the, and thus that the hospital would have to refinance the entire project with new loans in order to do the expansion. The great expense of this operation made the expansion economically impossible.[53]

Box 7.3

Stakeholder block

Houston Metro: The designers of the Houston metro carefully arranged for the possibility of constructing an office building and a parking garage near one of the metro stations close to the Texas Medical Center. They correctly estimated that the location abutting the hospital complex was ideally convenient and that the development could be highly profitable.

The hospitals also recognized such value, and arranged to build medical offices and convenient parking garages in and around the Texas Medical Center. These facilities were indeed profitable, and became a major source of funding for the hospitals. The hospitals thus became important stakeholders in any future development around the metro station: this would become competition and thus undercut their finances.

When developers tried to implement the flexibility designed around the metro station, they found it impossible to do so. The hospitals appear to have pleaded with the political authorities, asking them not to create competition that would threaten their role in the community. In the end, the hospitals corralled enough political support to prevent the metro system, itself governed by a political process, from going ahead with the project.[54]

The case illustrates the fact that stakeholders affected by the implementation of some flexibility can often organize to prevent such a development.

Figure 7.1: Schematic plan view of Dartmouth-Hitchcock Medical Center, showing how the design enables independent expansion, at right angles to the spine, of the facilities serving the several major stakeholders. (Source: DHMC and Shepley Bulfinch architects)

[1] See for example Shishko et al. (1995). Their Systems Engineering Handbook sets out the accepted practice, which begins with defining the requirements. Similarly, architects typically work to architecture programs that define the amount and types of spaces they have to provide for a facility.

[2] See Chapter 2, especially the section on “The standard forecast is “always wrong”.

[3] See the seminal text by Manne (1967). As presented in the Appendix C on the Economics of Phasing, his approach leads to interesting insights. However, his implicit concept of overall demand as the appropriate system driver is overly simplistic.

[4] For example, countries define socially important measures politically. Thus to count officially as “unemployed” in the United States, a person has to have been employed previously and to be actively looking for a job. The French tend to focus on whether you have a job, even if you have given up looking for one. In Japan, the Government pays companies to keep people salaried even if there is no work for them; this is functionally equivalent to an unemployment dole, but Japan does not classify the recipients as unemployed. See (Innes, 1989)

[5] Technically, R-squared is the ratio between the reduction in the variance in the data with respect to the trend line, and the amount of variance in the data absent the trend line. The complement to R-squared is thus a measure of the differences between the data and the trend line.

[6] Scenario analysis forms the basis of strategic military planning. Its use in business goes back to the oil crises in the 1970s. Subsequently petroleum companies, most notably Shell, developed and popularized this planning tool for civilian use. See Lingren and Bandhold (2009). Schoemaker (1991) provides a useful description of how one can organize scenario planning for a group.

[7] Statistical analyses of historical data series are normally excellent, with R-squared values close to 1, for two reasons. The first is that any factor that changes exponentially (that is, at a given growth rate) correlates well with any other exponentially changing factor. Consequently, it is easy to find good candidates for correlation. Secondly, it is a mathematical fact that you can always improve the overall correlation by adding more variables to the equation. Thus, forecasters do not generally find it difficult to get good statistical metrics if they want to do so.

[8] The traditional view is available in the publications of the International Air Transport Association (IATA, xxxx) and the US Federal Aviation Administration (FAA, (zzzz)

[9] There are many examples of airlines disappearing and changing the need for airport facilities: The collapse of Swissair and TWA left Zűrich and St. Louis airports with large empty terminals. Similarly, the reorganizations of US Airways and Alitalia caused these airlines to abandon their hub operations at Pittsburgh and Milan.

[10] It is generally not economical to design for the absolute peaks in demand. The facilities might be unused for the rest of the year. Usual practice is to design for some fraction of the absolute peak demands, and thus to put up with crowding, delays or otherwise poor service during the absolute peaks. Designers thus make trade-offs. What is “reasonable” depends upon the circumstances.

[11] de Neufville and Odoni (2003), and de Neufville and Belin (2002) provide detailed explanations.

[12] See Nababan (1993) for a case study Boston's experience.

[13] For details, see Maseda (2008).

[14] The original quote is "It can scarcely be denied that the supreme goal of all theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience." (Einstein, 1933)

[15] The question of identifying where designers should insert flexibility in the system is crucial for engineering design. By contrast, the topic does not normally arise in the real options literature based in finance. Financial analysis normally assumes that the existence of an option is obvious, and that the essential problem is to value this option.

[16] Screening models have a long history in the design of projects. Jacoby and Loucks (1972) and in de Neufville and Marks (1974), published one of the first major applications. They used simple linear models of the dynamics of river basins to identify plausible development strategies for the Delaware River. They then investigated those possibilities with detailed stochastic models of the hydrologic flows.

[17] The number of development paths increases exponentially with the number of periods because system operators react to prices and change the system as prices change. This reality contrasts with the use of binomial expansions in the financial analysis of options, according to which the total number of end states equals the number of periods plus one. The use of the binomial expansion assumes "path independence", that is, that where you end up is independent of how you get there. Path independence can make sense for prices. For example, the price of a ton of coal now is what it is, regardless of previous prices. Path independence is not an appropriate assumption to describe a system: the configuration of a project generally depends very much on the previous history. For example, if prices and demand have been high, plant managers are likely to have expanded or upgraded a factory, as they might not have if prices and demand have been low.

[18] Gigenrenzer (2002) provides an excellent description of these issues.

[19] The lack of acceptance of financial analysis of options among designers is an example of how people often distrust complex models. Black-Scholes and similar financial processes for evaluating the value of options are indeed difficult to understand. When presented with results of these models, decision-makers find themselves in the position of having junior staff tell them what to do based on impenetrable assumptions. Moreover, the assumptions used by the financial analysts – essential to the operation of their analyses – are often inappropriate. For example, the financial analysis of options normally assumes that future uncertainties derive from "stationary", "martingale" processes whose mean and variance does not change. In reality, this assumption is not valid over long periods, over the life of interesting systems in particular.

[20] Note that we do not talk about the best solution. It is impossible in all but the simplest cases to identify the true optimal design. This is because the model is a simplification, and our assumptions about the distribution of the uncertainties can hardly ever be fully justified through external evidence. Moreover, as

Chapter 6 explains, the notion of what is "best" is clear only in the simplest cases. In general, any design has many qualities that we cannot legitimately reduce to a single scale. Furthermore, the different stakeholders for a system will prize different attributes and prefer different solutions.

[21] In mathematical jargon refers to simulator models as "surface response models". The name derives from the idea that value of a system is a function of the design parameters, and that this value function is a surface in n-dimensional hyperspace. This explains why we use an alternative descriptor.

[22] Forrester (1961) originally elaborated systems dynamics models. His colleagues and students at MIT and elsewhere have improved the techniques and applied them to many areas. The essence of the approach consists of chains of actions or events that both influence downstream events and have feedback loops that can either dampen or amplify upstream events. A range of software exists to for the implementation of systems dynamics models, such as Stella®, Powersim® and Vensim®.

[23] Optimization procedures generally fall into one of three types: linear programming and its derivatives; dynamic programming, and genetic algorithms. Linear programming methods exploit mathematical features of the problem – specifically the convexity of the feasible region if it exists – to develop an optimal solution. Dynamic programming exploits a different possible property of the system, the independence of the contributions of different elements to the result, to find an optimal solution. These two approaches are very powerful for the specific cases for which they are applicable. The third type is a form of patterned search that attempts to find optimal spots and then to examine them closely to arrive at solution which is superior, but which may not be optimum.

[24] Manufacturing experience indicates when a group performs a task again and again, it learns to do so more rapidly with less effort – and thus less expensively. The empirical observation is that the effect is strongest at the beginning, and decreases exponentially thereafter. A typical way to express this is to say that the labor cost decreases by X percent each time production doubles, when X can range from a few percent up to 30%, depending on the complexity of the task. This equation describes the "learning curve". This effect can be dramatic when applied to major, rarely produced systems – such as satellites, deep-water oil platforms, nuclear power plants and other unique engineering systems.

[25] Suh (2005) provides a detailed example of flexibility in automobile production based on platform design.

[26] See Brushlinsky et al (2006).

[27] See for example Routledge (1980); Sneeuwjagt (1998); Thompson et al (2000); Tolhurst et al (2006); Brillinger et al. (2004); and González et al, 2005).

[28] Wang (2005) describes this case in detail.

[29] A convenient discussion of economies of scale in practice, as in particular of the "cost function," is in Chapter 4 of de Neufville (1990). In thinking about the cost function, we need to keep two points in mind. It refers to the minimum cost of the output capacity; not directly to the inputs used. Second, it assumes that the entire capacity is used. It does not represent the actual cost if the plant is operating below capacity. However, whenever a large plant is not operating below capacity, its average cost per unit

produced can easily be much higher than in a smaller plant. This is because the production of the larger plant then has to bear the cost of the idle capacity. In short, large plants with "economies of scale" often actually deliver product at a higher average cost than smaller plants.

[30] Lin (2009) discusses this case in detail.

[31] See Steel (2008).

[32] See Maseda (2008) for further discussion of this case.

[33] Lin et al (2009) provides details on this case and the great increase in value.

[34] Crystal Ball®, @Risk®, and RiskSolver® are widely used commercial software packages. Savage's XLSim® software, used to produce graphics in this book, is a good entry-level package, considerably less expensive but sufficiently powerful for many practical applications. As standard in the field, each allows the user to choose the kind of distribution they want to use, such as uniform, normal, triangular, lognormal and many others. Each also enables the analyst to prepare attractive graphical representations. Other simple add-ins to Excel® spreadsheets are also freely available from the web. As expected, the free add-in modules are not as sophisticated or user-friendly as the commercial products.

[35] The question of how many evaluations a Monte Carlo simulation should run has no obvious answer. If we could be confident in the description of our input uncertainties, we could develop a test for how many runs we need to develop a specified level of confidence in accuracy of the simulation. However, we are not likely to be confident in the long-range forecasts of important factors. Therefore, any test we might devise is unlikely to be of much use. Insofar as we can run each simulation quickly, it is good to use a generous number of samples. This reasoning explains the practice of using 1000 or more simulations. If the model is complex and computational resources are limited, however, fewer runs may be appropriate.

[36] A skew in the distribution refers to the fact that one tail of its shape is longer than the other. The tail that is longer, say the lower 50% of the distribution, pulls the average toward that side. The median however will stay put so long as there is no change in the other tail, say the upper 50%. Thus, the difference between the median and the mean is an indication of skewness.

[37] In financial markets, an option is a contract between traders that gives the owner the "right, but not the obligation" to take some future action at agreed terms. A "call" option gives the owner the right to buy something valuable at an agreed price if the right opportunity presents (such as a higher market price for the asset). A "put" option gives the owner the right avoid losses, as by getting rid of a property at an agreed price if the market price of the property dips sufficiently. An insurance contract is a form of option. For example, if you total your car and your insurer pays you for it, you have essentially sold this now worthless property for a good price.

[38] Kulatilaka (1993) is responsible for the seminal description of the case of dual-fuel burners.

[39] The utility function embodies overall expression of value for all dimensions of choice. Theoretically, we could apply utility functions to translate multiple dimensions of choice into a single metric, and we could thus avoid the problem of choosing between multiple dimensions. However, this is not practical. Reliable, accurate measurement of the utility function for any individual requires careful extended laboratory

procedures. In practice, we cannot apply these methods to decision-makers, who have neither the time nor the appetite to be experimental subjects. Despite extended efforts to apply utility theory to the evaluation and selection of major projects, this approach has not worked out to be practical.

Simple textbook descriptions of how to measure utility in rapid interviews lead to inconsistent results. de Neufville and McCord (1986) describe the kind of controlled analyses required to develop meaningful results. Starting with the seminal analysis of plans for a new airport for Mexico City (Keeney and de Neufville, 1973) there have been many attempts to promote utility analysis as a means to deal with multi-dimensional decision-making, for example Keeney and Raiffa (1993) but these have not proven practical for the evaluation and selection of major projects. Note finally that it is impossible, from a theoretical perspective, to define the utility function of the group of decision-makers without specifying the exact voting rules that will prevail among them, as Arrow (1950) demonstrated. In short, attempts to use utility functions to define the best choices for a system are futile from a practical perspective.

In this context, analysts should particularly be careful about the simplistic approach that defines the utility function as the linear sum of the metric for each dimension times a relative value or weight for each metric. This approach is appealing, insofar as it would dodge any complicated judgments. However, this linear approach is fundamentally misguided. Theory indicates that the utility functions are non-linear, according to the principle of diminishing marginal utility. That is, the relative weight given to any factor depends on one's level of satisfaction. For example, if you have no food, you might pay a lot for a meal; but if you already have a lot of food, you might not value extra food highly at all. Experiment confirms this general rule. Thus, any scheme for defining the overall best design based on relative fixed weights is highly suspect.

[40] Concern about regret can be thought of as a form of risk aversion, of wanting to avoid decisions that might look wrong in retrospect, of preferring "to be safe than sorry". Theoretical discussions define the measure of regret as the difference, for any set of circumstances, between what actually results from a choice, and what the chooser could have obtained from the alternative most suitable to those circumstances.

[41] In design, it is conventional to define "robustness" in terms of the ratio of the standard deviation of the performance compared to the average value. A design with the lowest ratio is said to be the most robust. Robustness is a desirable characteristic of design in many cases, particularly when it is possible to define a single most desirable value – such as when one tunes into a broadcast at a known frequency.

Robustness is not however desirable when the goal is to maximize (or minimize) a value, for example when we want to maximize profit. In cases where we want to maximize value, we would like to minimize the downside, and maximize the possible upside opportunities. In terms of flexible designs, we want solutions with great upside potential, the greater the better, that is, a design with higher standard deviation is more desirable than one that limits upside gains. This point is counter-intuitive to those who think that robustness is always better; it certainly is not.

[42] Petroleum engineers distinguish between the total oil in a reservoir, and the amount they can recover economically. The former is known as the STOIIP, the Stock Tank Oil Initially In Place. The reserves, that is, the amount that it is economically worthwhile to extract, are some fraction of the STOIIP, maybe between 1/3 and 1/2.

[43] It is common in extractive industries (mining of all forms in addition to oil and gas) for management to define a long-term price of the product that designers should use in valuing possible projects. They do this in order to create a level playing field among the design teams working on different projects across the company. Remarkably, management frequently changes these statements about long-term prices, often within a few years. The prices actually used by a company are often deeply confidential, as they reveal how aggressive a company might be in bidding for reserves.

[44] For a good example of how it is possible to explore tradeoffs in system design, see the work by Cohon and Marks (1973) on the development of river flows for agricultural, industrial and other purposes, available also in de Neufville and Marks (1974).

[45] Large Canadian airports, such as Calgary, Edmonton, Toronto and Vancouver, have designed their terminals with moveable partitions to enable them to direct passengers to the segregated international facilities or not, depending on the need. Other leading airports do likewise as de Neufville and Odoni (2003) describe. Progressive manufacturers of major equipment, such as cars or tractors, can similarly switch their production rapidly by changing the jigs or redirecting the robots on their assembly lines.

[46] For details, see Hassan et al (2005) for satellite fleets and Lin et al (2009) for oil platforms.

[47] This is a common issue. From the point of view of final design, it is desirable to have the climate control plant as close as possible to the main buildings, to reduce the cost of conduits and cut energy losses. However, it is almost impossible to move such plants since they are vital to the operation of the buildings and have to keep operating. The Dartmouth-Hitchcock hospital offers a good example of how the placement of their heating and chilling plant enhanced the possibility of future expansion (Lee, 20XX).

[48] See Guma et al (2009) and Wittels and Pearson (2008) for full descriptions.

[49] See Barker (2006) and American Institute of Architects (2008). Different versions of the process exist, and some people refer to integrated project management.

[50] This refers to the so-called South Suburban Airport near Peatone, Illinois. Its future is currently doubtful, as the City of Chicago has focused attention on rebuilding its O'Hare airport. The US Federal Aviation Administration has primarily worked with the State of Illinois. Airport authorities elsewhere have also acquired land for possible future airports in advance of a possible decision to build, such as Sydney, Toronto, and Bangkok – which eventually did build its Suvarnabhumi airport on such a site.

[51] Wittels and Pearson (2008) discuss the HCSC and the Tufts cases in detail.

[52] Walker (2001) has discussed the concept of signposts.

[53] de Neufville et al (2008) provide detailed discussions of hospital developments.

[54] McConnell (2007) gives interesting accounts of some of the politics surrounding Houston Metro.