Contributions to Risk-Informed Decision Making
by
Michael A. Elliott

Massachusetts Institute of Technology

Submitted to the Department of Nuclear Science and Engineering in Partial Fulfillment of
the Requirements for the Degree of
Doctor of Philosophy in Nuclear Science and Engineering
at the
Massachusetts Institute of Technology

June 2010

© 2010 Massachusetts Institute of Technology. All rights reserved.

Signature of Author............................................................... Department of Nuclear Science and Engineering
Submitted May 7, 2010

Certified by ................................................................. George E. Apostolakis
KEPCO Professor of Nuclear Science and Engineering
Thesis Supervisor

Certified by ................................................................. Benoit Forget
Norman K. Rasmussen Assistant Professor of Nuclear Science and Engineering
Thesis Reader

Accepted by ............................................................... Jacquelyn C. Yanch
Professor of Nuclear Science and Engineering
Chair, Department Committee on Graduate Students
This page intentionally left blank.
Contributions to Risk-Informed Decision Making

by

Michael A. Elliott

Submitted to the Department of Nuclear Science and Engineering on May 7, 2010 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Nuclear Science and Engineering

ABSTRACT

Risk-informed decision-making (RIDM) is a formal process that assists stakeholders make decisions in the face of uncertainty. At MIT, a tool known as the Analytic Deliberative Decision Making Process (ADP) has been under development for a number of years to provide an efficient framework for implementing RIDM. ADP was initially developed as a tool to be used by a small group of stakeholders but now it has become desirable to extend ADP to an engineering scale that can be used by many individual across large organizations. This dissertation identifies and addresses four challenges in extended the ADP to an engineering scale.

Rigorous preference elicitation using pairwise comparisons is addressed. A new method for judging numerical scales used in these comparisons is presented along with a new type of scale. This theory is tested by an experiment involving 64 individuals and it is found that the optimal scale is a matter of individual choice.

The elicitation of expert opinion is studied and a process that adapts to the complexity of the decision at hand is proposed. This method is tested with a case study involving the choice of a heat removal technology for a new type of fission reactor.

Issues related to the unique informational needs of large organizations are investigated and new tools to handle these needs are developed.

Finally, difficulties with computationally intensive modeling and simulation are identified and a new method of uncertainty propagation using orthogonal polynomials is explored. Using a code designed to investigate the LOCA behavior of a fission reactor, it is demonstrated that this new propagation methods offers superior convergence over existing techniques.

Thesis Supervisor: George E. Apostolakis
Title: KEPCO Professor of Nuclear Science and Engineering
Acknowledgements

I would like to thank Professor George Apostolakis for his support and encouragement throughout my graduate studies at MIT. The guidance he has provided has made my time at MIT both memorable and rewarding.

Dr. Homayoon Dezfuli’s support and guidance were invaluable in focusing this research. The issues he raised formed the basis for this work.

I would like to thank the MIT FCRR design team who participated in this project and provided valuable and timely feedback. Team members included: N.E. Todreas, P. Hejzlar, R. Petroski, A. Nikiforova, J. Whitman, and C.J. Fong. Further thanks are extended to C.J. Fong for advocating on my behalf to the design team.

I would like to acknowledge the department of Nuclear Science and Engineering at MIT for providing me with the intellectual tools to ask the important questions that are the hallmark of academic exploration. The past eight years studying with my colleagues in the department has been a transformative experience.

My readers and members of my thesis committee, Professors Ben Forget, Michael Golay and Jeffery Hoffman, have provided advice and objective review that has greatly improved this research.

This work was supported by the National Aeronautics and Space Administration’s Office of Safety and Mission Assurance in Washington, DC.
1 Dissertation Summary

The design and development of engineered systems involves many decisions that require balancing various competing programmatic and technical considerations in a manner consistent with stakeholder expectations. To assist in making these decisions, a logically consistent, formal decision-making process is required that can by deployed on an engineering scale. Decision making under these conditions has become known as risk-informed decision-making (RIDM) and this dissertation was commissioned by NASA to develop supporting methods.

RIDM is intended to assist a group of multiple stakeholders, each of whom may have different priorities, approach a decision problem in which there are a number of important objectives to achieve and numerous, significant uncertainties surrounding the decision options. It is designed to produce replicable, defensible decisions by adhering to the principles of formal decision theory. It recognizes the importance of subjective and difficult-to-quantify aspects of decision-making and so it pairs a deliberative framework with analysis in all points of the process. RIDM is scalable in that the rigor and complexity of the analysis is commensurate with the significance of the decision that is to be made. RIDM techniques may be applied to a broad range of decision problems, but the focus here is on problems that arise in the development of engineered systems early in their design phase.

One particular subset of RIDM techniques that can be applied to these design phase decision problems is known at the Analytic-Deliberative Decision-Making Process (ADP). ADP is a tool used for making decisions when there is adequate time for analysis and collective discussion. It is not appropriate for real time decision-making, which requires other RIDM techniques. ADP is a methodology that has been under development at MIT for a number of years and has been used to study a number of decision problems including problems in environmental cleanup [38] and water distribution reliability [39]. Here we focus on RIDM techniques as they apply to the ADP.

It will be helpful to present the contributions of this dissertation to the development of ADP based RIDM in terms of a concrete decision problem. Recognizing that ADP may be applied to decision-making in may types of engineered systems; we will proceed with a case study from nuclear reactor design.

As part of the Department of Energy’s Nuclear Energy Research Initiative (NERI), the Department of Nuclear Science and Engineering at MIT studied 2400 MWth lead-cooled and liquid-salt cooled fast reactors known as the Flexible Conversion Ratio (FCR) reactors [1]. The FCR reactors are so named as it is intended that, with minimal modifications, they will be able to operate at both a conversion ratio near zero, to dispose of legacy waste, and near unity to support a closed fuel cycle. Study of these reactors indicates that the most severe transient they are likely to encounter is a station blackout,
i.e., loss of offsite power and emergency onsite ac power. To prevent core damage during such a transient, a passive decay heat removal system is desired.

To remove decay heat, FCR designers developed two options. Both were referred to as Passive Secondary Auxiliary Cooling Systems (PSACS) and they differed primarily in the choice of ultimate heat sink. One system used air and the other water. Details are available in [2]. The choice between these two PSACS systems was made using RIDM techniques and so throughout this summary we will parallel the contributions of this dissertation with addressing the PSACS decision problem.

**RIDM with the Analytic Deliberative Process**

In the previous section, the goals of RIDM and the use of ADP were outlined in general terms. Before continuing, it is necessary to understand the ADP [4] in detail. As shown in Figure 1, the ADP has five steps. It begins with an analyst identifying the stakeholders (SHs) and leading this group in framing a specific decision problem and defining the context in which the decision is to be made. For our purposes, selecting a PSACS system is the problem.

![Figure 1: Steps in the Analytic-Deliberative Process](image)

Once the SHs specify the decision problem, they must identify all of the elements that each individual believes is important to consider in evaluating decision options. Forming an Objectives Hierarchy captures this information and the hierarchy for the PSACS problem is shown in Figure 2 and Table 1. Proceeding from the goal, to the objectives, to attributes, the problem is broken down into a set of elemental issues that express what the decision options must accomplish.

![Figure 2: Objectives Hierarchy and Performance Measures](image)

The final tier in the objectives hierarchy is made up by the quantifiable performance measures (QPMs). QPMs are developed by first examining the attributes and determining a set of appropriate metrics to measure each.
Table 1: Definitions for PSACS Objectives Hierarchy

<table>
<thead>
<tr>
<th>ATTRIBUTES</th>
<th>QPMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_1$ Safe and reliable for normal operations</td>
<td>$v_1$ Results of spurious actuation</td>
</tr>
<tr>
<td>$w_2$ Safe and reliable during SBO</td>
<td>$v_2$ Confidence in PCT</td>
</tr>
<tr>
<td>$w_3$ Accessible</td>
<td>$v_3$ S-CO$_2$ circulation performance</td>
</tr>
<tr>
<td>$w_4$ Simple</td>
<td>$v_4$ Ability to inspect</td>
</tr>
<tr>
<td>$w_5$ Robust</td>
<td>$v_5$ Ability to maintain</td>
</tr>
<tr>
<td>$w_6$ Designable</td>
<td>$v_6$ Number of active SSCs</td>
</tr>
<tr>
<td>$w_7$ Economical</td>
<td>$v_7$ Margin to required energy removal rate</td>
</tr>
<tr>
<td>SSCs: Structures, Systems and Components</td>
<td>$v_8$ Resilience against CCFs</td>
</tr>
<tr>
<td>CCFs: Common Cause Failures</td>
<td>$v_9$ Model confidence</td>
</tr>
<tr>
<td>SBO: Station Blackout</td>
<td>$v_{10}$ Degree of effort required</td>
</tr>
<tr>
<td></td>
<td>$v_{11}$ Component costs</td>
</tr>
</tbody>
</table>

In the context of the QPMs, the SHs must now determine how relatively important each attribute is to achieving the overall goal. To capture these preferences, the analyst conducts an elicitation exercise. Pairwise comparisons between attributes are made that determine which of the pair is more important to achieving the goal and then how much more important. Pairwise comparisons result in a set of person-specific weights, $w_i$, for the attributes that indicate the relative importance of attribute $i$ in the overall context of the decision problem.

With all of this information collected, the objectives hierarchy is fully specified and the ADP process proceeds to its third step in which decision options are identified. In the fourth ADP step, each of the decision options is scored according to the set of QPMs. Appropriate modeling and analysis is conducted and combined with the expert opinion of the participants so that the level of performance of each decision option is understood as well as possible. In general, this step of ADP is the most time consuming and resource intensive as it is the point where external tools are used to study the decision options (see, for example [7]). These may include computer modeling and simulation, physical experiments or extensive literature review.

In the final step of ADP, the DM and the SHs select a decision option using a deliberative process that is led by the analyst. To facilitate deliberation, the analyst presents a preliminary ranking of the decision options. Options are ranked according to a Performance Index (PI). The PI for option $j$ is defined as the sum of the values, $v_{ij}$, associated with the QPMs for attribute $i$ weighted by the attribute weight, $w_i$.

$$PI_j = \sum_{i=1}^{n_{QPM}} w_i v_{ij}$$ (1)

The PSACS selection problem is the first application of ADP to problems in nuclear reactor design. With this problem and others presented by NASA, four areas are
identified in which the current ADP tools and techniques are inadequate for decision making on an engineering scale. These areas are:

1. The pairwise comparison process needed to develop the weights, $w_i$, requires a numerical scale to convert subjective expressions of preference into numbers that may be analyzed. The existing scale is arbitrary and its implications not well understood.

2. The existing process of developing the QPMs requires substantial effort both in modeling and simulation and in training experts in decision theory. This effort makes the ADP too complicated to apply except for the most consequential decisions.

3. The information necessary to make decisions is stored across many individuals. No one person may adequately specify both the performance of a decision option and have knowledge of all the impacts that option will have on the overall system. Yet, the ADP envisions decisions being discussed only across a small group of stakeholders.

4. In cases where extensive modeling and simulation of the QPMs cannot be avoided, long running computer models are often required. These long run times may be prohibitive.

This dissertation addresses each of these four areas and develops the theory, methods and tools necessary to extend ADP to engineering scale decision-making. The contributions made in each area are now presented in turn.

**Selecting numerical scales for pairwise comparisons**

The pairwise comparison process used in the ADP allows an analyst to construct a set of weights, $w_i$, that will indicate how relatively important each of the attributes in a decision problem is to a SH. For example, if a PSACS SH is considering the attributes *simple*, *robust* and *economical*, then it is possible for that SH to respond to questions that will allow us to develop the numerical weights $w_{\text{simple}}$, $w_{\text{robust}}$ and $w_{\text{economical}}$ such that the ratio $w_{\text{simple}}/w_{\text{economical}}$ represents how much more important the simplicity of a PSACS system is to its cost.

To uncover these weights, attributes are presented to a SH two at a time and he is asked to answer two questions about each pair. First, which of the items in the pair is more important and, then, how much more important it is. To answer the second question regarding the degree of preference, the individual is given a list of linguistic phrases to select from. Each phrase used must be assigned a numerical value from the scale in order for a weight vector to be determined.

Up to this point the choice of scale has been somewhat arbitrary with the most common scale simply being the integers 1 – 9. This has led to some confusion in the literature with
authors proposing many scales without any clear benchmarks against which to measure them. Yet the choice of scale is potentially very important to decision making as it can affect not only the relative degree of importance between attributes but also their rank order.

The goal of this work with pairwise comparisons is to remove the arbitrary nature of scale choice by providing a sound theoretical basis for selecting a scale and then to make some assessment of which scale would be best to use with engineering decision problems like the PSACS problem.

Contributions to scales for pairwise comparisons

The arbitrary nature of scale choice is addressed by recognizing a connection between the pairwise comparisons between abstract attributes and the sensations that arise when the human brain processes stimuli perceived by our senses. When our eyes, for example, collect a light stimulus, information is transmitted to our brain and a sensation of that stimulus is produced. Depending on the intensity of the light, a different sensation is produced that we interpret as brightness. Each of our five senses is able to perform this function of converting different magnitudes of physical stimulus into a corresponding level of sensation. These conversions happen in a number of different ways. An exponentially increasing level of sound energy hitting our ears is interpreted as a linear increase in volume, while a linearly increasing level of light energy hitting our eyes is interpreted as a linearly increasing level of brightness.

If we view a pairwise comparison between abstract attributes as producing the stimulus of importance we must ask how the brain interprets the sensations of varying levels of importance. The scale must therefore reflect the manner in which an importance stimulus is converted to sensation. It is found that there are three primacy ways in which the brain treats this conversion from stimulus to sensation. A linear increase in stimulus can produce a linear increase in sensation, an exponential increase in stimulus can produce a linear increase in sensation or a linear increase in stimulus can produce an exponential increase in sensation. This observation leads to three candidate families of scales that correspond to each of the three stimulus-sensation paradigms. Further investigation of the literature demonstrates that several of the existing scales are appropriate for the linear-linear and the exponential-linear paradigms, and this author develops the third family of scale for the linear-exponential paradigm. These families are referred to respectively as the integer, balanced and power families.

With a sound theoretical basis for picking three families of scale, it is recognized that a number of free parameters existed. It is proposed that these parameters be chosen assuming that individuals exhibited rational behavior in selecting between attributes. Here rational behavior is a technical term meaning that an individual acts in accordance with the Von Neumann–Morgenstern axioms of rational behavior [35]. These axioms require completeness, transitivity, continuity, and independence among preferences.
With a cognitive basis for identifying families of scales and an axiomatic basis for picking free parameters, the problem of arbitrary scale selection is solved. It is now necessary to test which of the three families of scales actually represents the manner in which individuals judge importance between attributes. To this end, a study has been conducted of 64 individuals facing a realistic decision problem. What was found is quite interesting. Up until this point, the question being asked in the literature has been: what is the best scale to use for a particular type of decision problem? Our study reveals that this is the wrong question and that in fact the best scale is very much an individual choice. Remarkably, from individual to individual, there are differences in the way in which importance is perceived. It has been found that almost exactly a third of the experimental group preferred each of the three scale families.

As a result of this work, the author has proposed a change in the pairwise comparison process used with the ADP. Individuals are now given a choice of scale to represent their preferences for attributes. In addition, the author hopes that this work will have a significant impact in reducing the number of arbitrary scales that will be proposed in the future without recourse to some sound cognitive basis. The question of best scale should be thought of in terms of individuals rather than in terms of a certain decision problem. This work is published in the journal of Reliability Engineering and System Safety [68] and is the subject of a paper and presentation that will appear in the June 2010 Probabilistic Safety Assessment and Management (PSAM) conference in Seattle, WA [69].

Elicitation of expert opinion

When using the ADP, once decision options have been identified, it is necessary to establish the range of performance each option may achieve relative to each QPM. In the PSACS problem, this required establishing, for example, how easily the air and water options could be inspected or maintained. Objectively establishing ease of inspection and maintenance is not a trivial task. First, some basis or reference system must be established so ease can be assessed relative to a standard. Then, we need to identify and measure a set of criteria or characteristics of the basis system. These would allow us to quantify ease of inspection and maintenance. These characteristics then need to be measured in each of the candidate systems. Once these characteristics are aggregated into a summary measure we would have a QPM.

There are a number of difficulties with this objective approach. First, it is quite complex and requires significant effort. Second, it requires a very detailed understanding of the two candidate decision options. This complexity significantly limits the usefulness of the ADP as it leads decision-makers to only apply the ADP to the most significant decision problems were the consequences are such that the effort required to complete the process is justified. The requirement for detailed understanding of the decision options is even more restrictive for our purposes as we wish to apply RIDM to the design of engineering system, a phase in which, by definition, decision options have not been well characterized. As such, we may wish to avoid the objective evaluation of the QPMs,
when the required level of precision allows, and rely on a subjective analysis in which we
elicit the opinion of experts as to the likely range of performance of the decision options.

The elicitation of expert opinion has been an active area of research and many techniques
have been proposed [10]. Good expert opinion elicitation methods typically focus on
issues related to training the experts to correctly express probabilistic information,
insuring completeness, trading off the opinions of multiple experts, etc. These methods
do an excellent job of allowing us to evaluate QPMs without having access to a fully
design systems, but they do not do much to lessen the complexity of the process. As
such, methods of eliciting expert opinion were studied as part of the PSACS exercise
with the goal of simplifying the process while insuring a level of fidelity adequate to
complete the decision making process.

Contributions to the elicitation of expert opinion

As part of working through the ADP, SHs learn to use pairwise comparisons to rank the
relative importance of attributes. To use pairwise comparisons some training is required.
In the past, after completing pairwise comparisons, participants would be trained on an
entirely new method to collect judgments needed to evaluate the QPMs. This additional
step has been eliminated. Methods have been developed that allow participants to use the
same concepts they learned for pairwise comparisons and apply them to the construction
of the QPMs.

The new process is conducted in a highly deliberative manner between an expert and the
analyst. The expert outlines her believes about the range of reasonable performance any
acceptable system might achieve. This serves as a surrogate for a basis system. The
expert then relates her understanding of the characteristics of the decision options under
design and, using the pairwise comparison methods, compares the decision options to the
range of acceptable performance there by creating a numerical score for the options. This
process is fast while still generating a traceable record of the expert’s thinking and
motivations.

Such a qualitative approach to QPM development can be unsettling to engineers and so it
was important to develop a case study that demonstrated the efficacy of the approach.
This was done using the PSACS decision problem. The experts used the methods
described and were able to reach a logical, defensible decision. This case study has
proved very useful in convincing other groups that such a method of expert opinion
elicitation is viable.

These methods and the results of the PSACS study are the subject of a paper published by
this author in the journal of Nuclear Engineering and Design [70] and a paper and
presentation at the 2008 ANS Probabilistic Safety Analysis (PSA) topical meeting in
Knoxville, TN [71].
ADP with disbursed information

ADP is based on the principles of formal decision theory, which is an axiomatic theory applicable to a single individual. Extensions to multiple individuals, while common due to practical necessity, depart from this axiomatic basis and require special treatment. In a large organization, such as NASA, a large group of individuals is usually involved. Each individual is familiar with a certain piece of knowledge important to the decision problem, but he may be unaware of the overall context of the problem and he may not be an active participant in the decision-making process. In such cases, special tools and techniques are required to implement the ADP. These tools did not exist at the outset of this work and so they have been developed.

Contributions to ADP with disbursed information

As a first step, organizational relationships need to be understood. Instead of having a small number of individuals that can easily come together to discuss a decision problem, as was the case in the PSACS exercise, there are now large portions of the organization whose needs must be represented. Where before there were individual SHs, now each SH represents the aggregate beliefs of entire segments of the organization. These segments have a stake in the outcome of the decision, as it can enable or constrain their future capabilities, but they have little direct authority over the decision. No one segment can be expected to identify all the decision options or be intimately familiar with each. It must rely on experts across the organization to identify and analyze options and relate their relative advantages and disadvantages. In addition, another set of experts is required, who have expertise in the specific areas where consequences of the decision will be experienced, to help specify how desirable particular characteristics of a decision option are.

These relationships lead to a number of separate rolls for individuals participating in the decision problem. For the first time, this work identified these rolls, specified the responsibilities of each and provided guidance as to how individuals in each roll will need to interact over the course of implementing the ADP.

With such a large, varied and potentially geographically distributed group of individuals participating in the ADP, new tools are required to assist in collecting information. These tools have been developed by this author in Microsoft Excel. Excel is a very desirable tool due to its ubiquity, but up until this point, had not been thought to be practical for the ADP due to certain computational requirements. In particular, the ADP requires linear algebra and Monte Carlo simulations that are not standard features in Excel. The necessary algorithms have been developed and implemented in Excel, however, and, as a result, an easy to distribute and intuitive tool is available.

As a result of these innovations, the guidance and tools now exist to apply ADP for large scope decision problems affecting large organizations. This work has been reported on in two white papers by this author distributed internally within NASA.
Uncertainty Propagation using Orthogonal Polynomials

It was previously observed that the quantitative evaluation of the QPMs can require significant effort and so methods have been developed to obtain QPMs from expert opinion when appropriate. In certain cases, however, expert opinion is insufficient as the level of precision needed surpasses what experts are able to provide. In these cases, a certain amount of modeling and simulation may be required. In performing this simulation, a common problem may arise. The simulation might be computationally intensive and require a long runtime for a single realization of the input parameters. Multiple realizations are desired, however, as we wish to understand how input uncertainties effect the distribution of performance a decision option may achieve.

What is required then is a method of obtaining the distribution of performance while running as few realizations of the input parameters as possible. Monte Carlo or Latin Hypercube techniques are not appropriate for this purpose, as they require too many simulations be performed. Many methods have been proposed to do uncertainty propagation with sparse sampling. One of the more popular methods uses response surfaces. Response surfaces are very good at capturing the modal performance of a decision option but not at representing the extremes. Cases like the PSACs, where any failure may lead to catastrophic reactor meltdown, require a faithful representation of the complete range of possible system performance.

Here a new method for uncertainty propagation is proposed based on an expansion in orthogonal polynomials. This method is intended to better represent the full range of possible performance.

Contributions to uncertainty propagation using orthogonal polynomials

The use of orthogonal polynomials in solving integral or differential equations will be familiar to nuclear engineers, where a common example is the application of the $P_n$ method, using Legendre polynomials, to solve the neutron transport equation. Recently, a small number of authors have begun to apply these same principles to the propagation of uncertainties through computational models.

The principle is very simple. A family of polynomials, $\Psi_i$, is selected that is orthogonal to the probability distribution, $w$, being used to represent the input uncertainties, $x_i$, in the computational model. Recall that orthogonally is defined with respect to the inner product where

$$\langle \Psi_k, \Psi_l \rangle_w = \int_{\mathbb{R}} \Psi_k(X)\Psi_l(X)w(X)dX \begin{cases} = 0, & k \neq l, \\ > 0, & k = l. \end{cases}$$

The output of the model, $S(X)$, can then be represented as an expansion in terms of these polynomials.
\[ S(X) = \sum_{i=1}^{M-1} S_i \Psi_i(x_1, \ldots, x_N) \]  

where \( M \) governs the order at which the expansion is stopped for computational convenience. With this representation of \( S(X) \), the distribution of the model output is known analytically. The coefficients \( S_i \) are found in the usual manner

\[ S_i = \frac{\langle S \Psi_i \rangle_x}{\langle \Psi_i \Psi_i \rangle_x} \]  

Up until this point, polynomial expansions could only be performed in cases where the input uncertainties were all represented by the same type of probability distribution, i.e., a lognormal distribution. If the uncertainties were represented my multiple types of distribution, it was not possible to form an orthogonal basis except in certain cases where the complete joint distribution of the \( X \) was known. In most engineering applications, the joint distribution is not known but rather the marginal distributions are estimated along with some assessment of the linear correlation coefficients.

Here, a method has been developed to allow the propagation of arbitrary probability distributions using only the marginal distributions of each of the uncertain parameters and the linear correlation coefficients. This method makes use of an isoprobabilistic transformation that converts the arbitrary probability distributions first to correlated Gaussian distributions and then to uncorrelated standard normal Gaussian distributions.

This method has been tested on the Pagani et al GFR model [64] that simulates the effects of a severe LOCA. In addition to this being the first application of the isoprobabilistic uncertainty propagation approach, it is also the first application of any orthogonal polynomial uncertainty propagation technique to nuclear engineering. The Pagani et al model supposes nine uncertain input parameters, reactor power, pressure, the temperature of certain heat removal surfaces at the time of the LOCA and six other thermohydraulic parameters.

Figure 3 presents the resulting distribution of the maximum hot channel temperature, an important parameters to monitor during the LOCA, obtained from the set of nine differently distributed, correlated uncertainties for three simulations. The first simulation is a Monte Carlo simulation that took 10,000 samples from the underlying model. This number of samples was needed to converge all percentiles to <1%. We will consider this simulation to represent the exact solution. The second simulation is a full factorial response surface requiring 256 samples from the model. The final simulation is a 3rd order isoprobabilistic polynomial expansion requiring 220 samples.
Panel A and B of the figure present the PDFs and CDFs respectively for the three simulations. Panel C presents the error in degrees Celsius for the response surface and isoprobabilistic chaos relative to the Monte Carlo. Note that the response surface initially, at the 0th percentile, underestimates the hot channel temperature by about 25°C then, by the 90th percentile, overestimates it by about 25°C. More disturbing however, is that by the 99th percentile the response surface is underestimating the hot channel temperature by about 150°C. The isoprobabilistic polynomial expansion, on the other hand, very closely tracks the Monte Carlo solution from the 0th until the 97th percentile and then overestimates the hot channel temperature by about 25°C by the 99th percentile. We conclude then that even with 36 fewer samples of the underlying model, the isoprobabilistic chaos produces superior results and may be a promising new technique for uncertainty propagation.

This work has not yet been published but is the subject of manuscript under preparation for the journal *Risk Analysis*. 
# Table of Contents

1 DISSERTATION SUMMARY 5

2 INTRODUCTION 21

2.1 Objectives 22

3 RISK-INFORMED DECISION MAKING TO SELECT THE PSACS ULTIMATE HEAT SINK 23

3.1 Risk-Informed Decision Making and the ADP 24

3.2 Deliberation and Results 29

3.3 Conclusions 31

4 SELECTING NUMERICAL SCALES FOR PAIRWISE COMPARISONS 33

4.1 Pairwise comparisons 35

4.2 The Scale 37

4.2.1 Integer Scales 37

4.2.2 Balanced Scales 39

4.2.3 Power Scale 40

4.3 Choosing a Scale – Normative Factors 41

4.3.1 Distribution of the Weights 41

4.3.2 Consistency of the Pairwise Comparisons 44

4.4 Choosing a Scale - Empirical Arguments 47

4.5 Study I – 64 Individuals, Hypothetical Decision Problem 49

4.5.1 The True Scale 49

4.5.2 Number of Degrees of Preference 50

4.5.3 Calibrating the Participants 52

4.6 Study II – 3 Individuals, Actual Decision Problem 53

4.7 Conclusions 56

5 RISK-INFORMED DECISION MAKING FOR THE ASSESSMENT OF LUNAR SERVICE SYSTEMS PAYLOAD HANDLING 57

5.1 RIDM for LSS 57

5.2 The Analytic Deliberative Process for Large Groups 58

5.2.1 Steps in Implementing the large group ADP 60

5.3 The Role of Experts 64

5.3.1 Formulating Quantifiable Performance Measures 64

5.3.2 Identifying Measurable Consequences 65

5.3.3 Specifying the Range of Consequences 65

5.3.4 Determine the Value of a Level of Consequence 65

5.3.5 Formulating and Scoring Decision Options 67

5.4 Lunar Payload-Handling 68

5.5 Conclusions 79

6 UNCERTAINTY PROPAGATION USING ORTHOGONAL POLYNOMIALS 81

6.1 Orthogonal Polynomials 82

6.1.1 Definition of Orthogonal Polynomials 82

16
Generating Orthogonal Polynomials – Recurrence Relation

Uncertainty Propagation Using Orthogonal Polynomials

Independent Gaussian Random Variables
Independently distributed random variables
Dependent, differently distributed random variables
Calculating the Expansion Coefficients
Applications – A Simple Polynomial Model
Applications – A Simple Non-Polynomial Model
Applications – Nuclear Reactor Transients
Conclusions

REFERENCES

APPENDIX I - PSACS QPMS
APPENDIX II – THE RANDOM INDEX
APPENDIX III – SURVEY QUESTIONS
APPENDIX IV – RIDM PRIMER
## Table of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>Steps in the Analytic-Deliberative Process</td>
</tr>
<tr>
<td>2-2</td>
<td>Objectives Hierarchy and Performance Measures</td>
</tr>
<tr>
<td>2-3</td>
<td>Attribute weights resulting from pairwise comparisons</td>
</tr>
<tr>
<td>2-4</td>
<td>Expected PI for PSACS Decision Options</td>
</tr>
<tr>
<td>2-5</td>
<td>Attribute Contributions to PI for SH 5</td>
</tr>
<tr>
<td>3-1</td>
<td>Distribution of attribute weights obtainable by the three scales for comparisons between 2 and 3 attributes. A simple computer algorithm was written to generate these distributions that itteratively evaluated the weight vectors for all possible sets of judgments.</td>
</tr>
<tr>
<td>3-2</td>
<td>Possible number of consistent judgments and weight vectors for each of the three scales as a function of the permitted percent inconsistency as defined by the consistency ratio CR. Panels A and B are for judgments among three attributes, N=3. Panels C and D are for N = 4. Panels A and C present the total number of consistent judgments. Panel B and D present the number of unique weight vectors that arise from these consistent judgments. A unique weight vector is defined as one with N unique numerical values. The data in this figure was generated by a computer algorithm that generated the weight vectors for all possible sets of judgments and then calculated the consistency ratio corresponding to those judgments.</td>
</tr>
<tr>
<td>3-3</td>
<td>Weight vectors calculated using the three scales. Each weight vector is one possible numerical interpretation of the relative importance of four attributes derived from a single set of judgments obtained from one individual.</td>
</tr>
<tr>
<td>3-4</td>
<td>Distribution of CR for study participants. The CR for each individual was calculated using the RI from the scale that individual preferred. The left panel presents CRs for individuals given 5 degrees of preference to choose from and the right panel presents CRs for 9 degrees. The curves marked as calibrated refer to CR values from the set of individuals who were asked to first complete bounding pairs. The curves marked as random refer to CR values for individual who were presented pairs at random.</td>
</tr>
<tr>
<td>3-5</td>
<td>Absolute percent difference between attribute weights when calculated with the three scales. Results are presented as a function of the attribute weight calculated using the integer scale and are shown separately for the three stakeholders (SH).</td>
</tr>
<tr>
<td>3-6</td>
<td>Performance Index (PI) for the air and water ultimate heat sink options as a function of the three scales for the three stakeholders. The PI has been normalized such that the PI of the water option is unity.</td>
</tr>
<tr>
<td>4-1</td>
<td>Hierarchy of Decision Participants</td>
</tr>
<tr>
<td>4-2</td>
<td>Schematic Objectives Hierarchy [40]</td>
</tr>
<tr>
<td>4-3</td>
<td>Continuous Value Function</td>
</tr>
<tr>
<td>4-4</td>
<td>Lunar Payload Handling Objectives Hierarchy top level</td>
</tr>
<tr>
<td>4-5</td>
<td>Constraints attributes</td>
</tr>
<tr>
<td>4-6</td>
<td>Simplicity attributes</td>
</tr>
<tr>
<td>4-7</td>
<td>Versatility and contingency attributes</td>
</tr>
<tr>
<td>4-8</td>
<td>Optimization attributes</td>
</tr>
<tr>
<td>4-9</td>
<td>Safety attributes</td>
</tr>
</tbody>
</table>
Figure 4-10: Programmatic resources attributes

Figure 5-1: Polynomial expansions with Hermite basis for a simple polynomial with standard normal uncertainties. Panel A presents the PDF of the reference solution as well and 1st and 2nd order expansions (PC). Panel B presents the error relative to the reference solution for the 1st and second order expansions as a function of percentile of the reference solution.

Figure 5-2: Polynomial expansions with a hypergeometric basis for a simple polynomial with uniform uncertainties. Panel A presents the PDF of the reference solution as well and 1st and 2nd order expansions (PC). Panel B presents the error relative to the reference solution for the 1st and second order expansions as a function of percentile of the reference solution.

Figure 5-3: Polynomial expansions with an isoprobabilistic basis for a simple polynomial with differently distributed uncertainties. Panel A presents the PDF of the reference solution as well and 1st and 2nd order expansions (PC). Panel B presents the error relative to the reference solution for the 1st and second order expansions as a function of percentile of the reference solution.

Figure 5-4: Polynomial expansions with Hermite basis for a sin wave. Panel A presents the PDF and panel B the error as a function of percentile.

Figure 5-5: Response surface analysis for a sin wave subject to standard normal uncertainty. Panel A presents the PDF and panel B the error as a function of percentile.

Figure 5-6: Cumulative distribution function of the average outlet temperature from the Pagani et al model for normally distributed uncertainty in reactor power. Results for Monte Carlo, orthogonal polynomial expansion (PC) and response surface simulations are shown.

Figure 5-7: Results of the nine random variable simulation of the Pagani et al model. Panels A, B and C compare the PDFs from a Monte Carlo simulation with those of a full factorial response surface and the 1st through 4th order orthogonal polynomial expansions (PC). Panel D presents the error of the response surface and expansions relative to the Monte Carlo as a function of percentile.
List of Tables

Table 2-1: Definitions for PSACS Objectives Hierarchy ............................................. 25
Table 2-2: Constructed Scale for Number of SSCs ...................................................... 27
Table 3-1: Phrases used in the pairwise comparison process to indicate degree of
importance of item A over item B [11] ......................................................................... 35
Table 3-2: Hypothetical judgments between three items........................................... 36
Table 3-3: Reciprocal matrix and corresponding weight vector obtained from pairwise
comparisons for a hypothetical three-attribute decision problem ................................ 37
Table 3-4: Numerical Values of the Balanced Scale generated using Eq. 5 ................. 40
Table 3-5: Numerical Values of the Power Scale ......................................................... 41
Table 3-6: Percent of possible sets of judgments for N=4 whose corresponding weight
vectors lead to different rankings among the attributes when calculated with the
indicated scales. Percentages are given for rank differences regardless of the
magnitude of the difference between the weights and for when the difference is
greater then 5% and 10%. These results are presented without regard to the CR of
the underlying judgments and using a nine-degree of preference scale .................... 49
Table 3-7: Fraction of study participants believing each of the three scales best captured
their actual attribute weights including the degree of belief over the two other scales.
The degree of belief for each individual was calculated using his preferred numerical
scale. The ranges supplied represent the 75% confidence interval of the responses
........................................................................................................................................ 50
Table 3-8: Percentage of participant given either the 5 or 9 degree of preference
elicitation scheme who believed the number of preference choices was too many,
too few or just right..................................................................................................... 51
Table 4-1: Constructed Scale – Largest mass on a single lander ............................. 62
Table 4-2: The Weighting Process ............................................................................. 63
Table 4-3: Discrete QPM ............................................................................................. 67
Table 5-1: Several hypergeometric polynomials and their corresponding weight functions
that correspond to useful probability distributions. k refers to the degree of the
polynomial, x is the random variable, all other symbols are standard nomenclature
from engineering applications of probability theory .................................................... 87
Table 5-2: Several hypergeometric polynomials and their corresponding recurrence
relations .......................................................................................................................... 88
Table 5-3: Distributions and associated parameters for the four xj ................................ 94
Table 5-4: Input uncertainty distributions and correlations for isoprobabilistic basis
simulation of the Pagani et al model. Correlation coefficients are designated by the
parameter number in the left most column and are listed only once for each pair of
correlated variables ................................................................................................... 99
2 Introduction

The design and development of engineered systems involves many decisions that require balancing various competing programmatic and technical considerations in a manner consistent with stakeholder expectations. To assist in making these decisions, a logically consistent, formal decision-making process is required that can by deployed on an engineering scale. Decision making under these conditions has become known as risk-informed decision-making (RIDM) and this dissertation was commissioned by NASA to develop supporting methods.

RIDM is intended to assist a group of multiple stakeholders, each of whom may have different priorities, approach a decision problem in which there are a number of important objectives to achieve and numerous, significant uncertainties surrounding the decision options. It is designed to produce replicable, defensible decisions by adhering to the principles of formal decision theory. It recognizes the importance of subjective and difficult-to-quantify aspects of decision-making and so it pairs a deliberative framework with analysis in all points of the process. RIDM is scalable in that the rigor and complexity of the analysis is commensurate with the significance of the decision that is to be made. RIDM techniques may be applied to a broad range of decision problems, but the focus here is on problems that arise in the development of engineered systems early in their design phase.

One particular subset of RIDM techniques that can be applied to these design phase decision problems is known at the Analytic-Deliberative Decision-Making Process (ADP). ADP is a tool used for making decisions when there is adequate time for analysis and collective discussion. It is not appropriate for real time decision-making, which requires other RIDM techniques. ADP is a methodology that has been under development at MIT for a number of years and here we focus on RIDM techniques as they apply to the ADP.

The ADP was conceived as a decision-making process to be used by a small number of individuals and so several of its tools and techniques are inadequate for decision making on an engineering scale. In particular, four aspects of the ADP may hinder its application to engineering problems, these are:

5. The existing process of developing performance measures for decision options requires substantial effort both in modeling and simulation and in training experts in decision theory. This effort makes the ADP too complicated to apply except for the most consequential decisions.

6. The process required to rank the relative importance of attributes that contribute to the decision problem requires a numerical scale to convert subjective expressions of preference into numbers that may be analyzed. The existing scale is arbitrary and its implications not well understood.
7. The information necessary to make decisions is stored across many individuals. No one person may adequately specify both the performance of a decision option and have knowledge of all the impacts that option will have on the overall system. Yet, the ADP envisions decisions being discussed only across a small group of stakeholders.

8. In cases where extensive modeling and simulation of the performance measures cannot be avoided, long running computer models are often required. These long run times may be prohibitive.

2.1 Objectives

This dissertation addresses each of these four areas and develops the theory, methods and tools necessary to extend ADP to engineering scale decision-making. Chapter 2 presents a case study application of the ADP to nuclear fission reactor design. Through this case study, the ADP is explained in greater detail and efficient development of performance measures is addressed. It is demonstrated that the usefulness of the ADP can be preserved even when significant effort is not expended in developing performance measures.

The scale needed for ranking attributes is the subject of Chapter 3. This chapter explains the ranking process, develops the theory underlying scale choice, presents a new type of scale and then reports on the results of two experiments that test candidate scales.

Chapter 4 is built around another case study, this one related to a NASA project designing payload handling systems for use on the Lunar surface. This project presented a true engineering scale study and so the chapter explains the use of the ADP in such an environment where decisions require participation from a large group of individuals.

In Chapter 5, the remaining issue of uncertainty propagation in long running computer codes is addressed. A new method based on orthogonal polynomial expansions is proposed. This method is compared to existing techniques in evaluating the performance of a fission reactor system during a severe transient.
3 Risk-Informed Decision Making to Select the PSACS Ultimate Heat Sink

As part of the Department of Energy’s Nuclear Energy Research Initiative (NERI), the Department of Nuclear Science and Engineering at MIT has been studying 2400 MWth lead cooled and liquid-salt cooled fast reactors known as Flexible Conversion Ratio (FCR) reactors. The FCR reactors are so named as it is intended that, with minimal modifications, they will be able to operate at both a conversion ratio near zero to dispose of legacy waste, and near unity to support a closed fuel cycle. Study of these reactors indicates that the most severe transient they are likely to encounter is a station blackout, i.e., complete loss of offsite power and loss of emergency onsite ac power. To prevent core damage during such a transient, a passive decay heat removal system is desired.

To remove decay heat, FCR reactor designers developed an enhanced Reactor Vessel Auxiliary Cooling System (RVACS) similar to General Electric’s S-PRISM design. The RVACS is a passive safety system in that it does not require moving parts, external power, control signals or human actions to accomplish its function. The RVACS removes decay heat through the natural circulation of air over the reactor guard vessel. Further modeling indicated that, although the RVACS was sufficient for most transients, it could not maintain a peak clad temperature (PCT) below appropriate limits during a station blackout. Therefore, a supplementary decay heat removal system would be required. A number of options were considered for this supplementary system. Initially, it was proposed that the power conversion system (PCS) be relied upon to remove decay heat. This was advantageous because the PCS was already in place and would be self-powered, as the removal of decay heat would drive the turbine. This option was abandoned, however, as the design team believed it might have required that PCS and other balance of plant components be categorized as safety related, thereby significantly increasing their cost.

With the rejection of the PCS as a viable supplementary decay heat removal alternative during station blackout, two additional systems were proposed. Both were referred to as Passive Secondary Auxiliary Cooling Systems (PSACS), although they differed in the choice of ultimate heat sink. One system used air and the other water. Details are available in 0. Due to constrained time and resources, the team was only able to design one of these options in detail and so they needed to select an option to proceed without the benefit of prior study. In doing so, the team faced two challenges. They first needed to decide what metrics to use to compare these two options and then they needed to use what little information was available to evaluate the metrics. As with the consideration of the PCS, an informal decision-making process was attempted. In this case, however, the process was unable to choose between the two systems. As a result, the team turned to a formal decision-making process known as the Analytic-Deliberative Process (ADP) originally recommended in [4]. ADP is a process that helps stakeholders make risk-informed decisions. It has been used at MIT in a variety of decision-making problems.
What follows describes the application of ADP to the selection of the PSACS ultimate heat sink.

3.1 Risk-Informed Decision Making and the ADP

Decision analysis and a variety of risk-informed decision making processes have been studied extensively since the mid 1960s. In its most basic formulation, decision analysis provides the basis for making choices between competing alternatives in the face of uncertainty [8]. The type of decision problem that is of concern in this paper has two important characteristics. It is a multi-attribute or multi-criteria problem in which the benefits of any one decision alternative must be judged by its impact on a number of different objectives [9]. It also involves a significant amount of uncertainty that must be characterized by experts whose viewpoints must then be synthesized [10]. As such, the risk-informed decision-making process should bring together the decision maker, those who have a stake in the outcome of the decision (stakeholders), experts who can characterize available options, and an analyst who can gather information and facilitate discussion. It should assist these individuals in systematically identifying the objectives of making a particular decision and defining the performance of options in the context of these objectives. The process must aggregate both objective and subjective information while keeping track of uncertainty. It should combine analytical methods with a deliberation that scrutinizes the analytical results and it should produce a ranking of decision options and a detailed understanding of why certain options outperform others. The ADP is such a process. It consists of two parts [4]:

1. *Analysis*, which uses rigorous, replicable methods, evaluated under the agreed protocols of an expert community - such as those of disciplines in the natural, social, or decision sciences, as well as mathematics, logic, and law - to arrive at answers to factual questions.

2. *Deliberation*, which is any formal or informal process for communication and collective consideration of issues.

As shown in Figure 3-1, the ADP begins with the analyst identifying the decision maker (DM) and stakeholders (SHs) and leading this group in framing a specific decision problem and defining the context in which the decision is to be made. For the purposes of this study, the FCR reactor (FCRR) design team is selecting a supplementary decay heat removal system to maintain PCT below a limiting value during a station blackout. The DM is a senior member of the design team and the SHs are the five remaining members of the team, each of whom also serves as an expert on a different area of the design. The analyst must ensure that these definitions and roles are clearly specified at the beginning of the process and that all individual understand their role and agree to participate in the process as appropriate. This is critical so as to guarantee that all subsequent analysis and risk characterization is done in the context of the specific decision problem at hand and that it all be directed at answering the specific questions that are of interest to the DM and SHs.
Once the DM and SHs specify the decision problem and the context in which it is being addressed, they must identify all of the elements that each individual believes are important to consider in evaluating decision options. Forming an Objectives Hierarchy captures this information and the hierarchy for the PSACS problem is shown in Figure 3-2 and Table 3-1. At the top of the hierarchy is the goal, i.e., a broad statement intended to communicate the overall purpose for making the decision. It reiterates the context in which participants will determine what other elements belong in the objectives hierarchy. The analyst is responsible for making sure that all SHs and the DM agree on the goal.

<table>
<thead>
<tr>
<th>GOAL</th>
<th>PSACS maintains safety during SBO and does not compromise normal operations</th>
</tr>
</thead>
<tbody>
<tr>
<td>OBJECTIVES</td>
<td>PSACS net impact on reactor safety is positive</td>
</tr>
<tr>
<td>ATTRIBUTES</td>
<td>$w_1$</td>
</tr>
<tr>
<td>QPMs</td>
<td>$v_1$</td>
</tr>
</tbody>
</table>

**Figure 3-2: Objectives Hierarchy and Performance Measures**

**Table 3-1: Definitions for PSACS Objectives Hierarchy**

<table>
<thead>
<tr>
<th>ATTRIBUTES</th>
<th>QPMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_1$</td>
<td>Safe and reliable for normal operations</td>
</tr>
<tr>
<td>$w_2$</td>
<td>Safe and reliable during SBO</td>
</tr>
<tr>
<td>$w_3$</td>
<td>Accessible</td>
</tr>
<tr>
<td>$w_4$</td>
<td>Simple</td>
</tr>
<tr>
<td>$w_5$</td>
<td>Robust</td>
</tr>
<tr>
<td>$w_6$</td>
<td>Designable</td>
</tr>
<tr>
<td>$w_7$</td>
<td>Economical</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Objectives are the second tier in the hierarchy. They are the broad categories of elements that the DM and SHs feel must be achieved in order for a decision option to meet the goal. These broad objectives may be further divided into sub-objectives as needed. For this decision problem, the DM and the SHs agreed on two objectives, that the PSACS
must have a net benefit on overall reactor safety and that its inherent design characteristics be favorable. No sub-objectives were needed.

Below the objectives are the attributes. Attributes are the largest set of elements between which a DM or SH is indifferent1 and they describe how to achieve the objectives that lie below. It is helpful to think of attributes as the most detailed level of sub-objectives the DM or SH wishes to consider. Here, seven attributes have been identified, as shown in Figure 3-2.

With input from the DM and the SHs, the analyst will attempt to create a consensus hierarchy and prepare a set of definitions for each objective and attribute. While the DM and SHs need not agree on the structure of the hierarchy, it greatly simplifies the analysis if consensus can be reached. In the case of the PSACS decision problem, the analyst was able to achieve consensus.

It is appropriate to add one qualification. In constructing objective's hierarchies for system design problems, the analyst must expect that the DM and SHs may not yet have sufficient information to specify all attributes that will eventually be determined to be relevant to the decision problem. That is, the initial hierarchy may later be seen to be incomplete. As such, the analyst should stress that results obtained reflect only the current state of knowledge and he should update the hierarchy if new information warrants. Due to the limited scope of this study, no such updating was done for the PSACS decision problem.

The final tier in the objectives hierarchy is made up by the quantifiable performance measures (QPMs). QPMs are developed by first examining the attributes and determining a set of appropriate metrics to measure each. For example, the attribute Simple is measured by one metric, the number of active structures, systems and components (SSCs) required by a particular PSACS design option. Once these metrics are established, the range of performance that any reasonable PSACS option might have is determined and then the relative desirability of different points in this range is assessed. This information is captured in a value function that takes on numbers between zero, for the least desirable performance level, and unity, for the most desirable [11]. The range of performance levels and the corresponding values form a constructed scale, as shown in Table 3-2. The constructed scale can be continuous, with a unique desirability value for every possible performance level, or discrete, as in Table 3-2, with one value corresponding to a range of possible levels. Constructed scales allow any metric to be measured in terms of a common unit and they capture risk aversion to different levels of performance.

1 As an example, if a participant wants to minimize total system cost and does not care if these are capital or operating costs, then we say this individual is indifferent between capital and operating costs. Total cost is therefore an attribute. If, however, a participant thought it was more important to minimize capital costs even though this could lead to higher operating costs, then he or she is not indifferent between the two. Total cost would become an objective or sub-objective and capital and operating costs would each be their own attribute.
Table 3-2: Constructed Scale for Number of SSCs

<table>
<thead>
<tr>
<th>Performance Level</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1-3</td>
<td>0.3</td>
</tr>
<tr>
<td>3 or more</td>
<td>0</td>
</tr>
</tbody>
</table>

A metric and its constructed scale form a QPM. QPMs can be based on quantitative metrics, such as the number of kg, or qualitative ones, such as a subjective understanding of a degree of complexity. They must, however, be metrics for which a constructed scale can be developed. In cases where more than one QPM is used to evaluate a single attribute, QPMs are equally weighted to lead to a single score for the attribute.

Appendix I - PSACS QPM, presents the constructed scales for the eleven QPMs considered in evaluating the two different PSACS ultimate heat sink options. To accelerate the elicitation process for the QPMs, the DM and the SHs were only asked to define the upper and lower bounds of the value function and not to specify a complete set of intermediate values. Intermediate values would be added in step four of the ADP as necessary to score individual decision options. This approach saved considerable time in developing QPMs, but it still allowed for an open discussion and exchange of information between the participants.

In the context of the QPMs, the DM and the SHs must now determine how relatively important each attribute is to achieving the overall goal. To capture these preferences, the analyst conducts an elicitation exercise using a pairwise comparison process based on the Analytic Hierarchy Process (AHP) [12]. AHP requires each SH and DM to make a series of pairwise comparisons between attributes, and then objectives, stating which of the pair is more important to achieving the goal and then how much more important. The constructed scales are critical in providing the necessary context to make these comparisons. As an example, in the absence of context, if an individual is asked to compare safety with a monetary attribute, he or she will likely report that maintaining safety is extremely more important. The constructed scale, however, may reveal that the maximum consequences to safety are minor while the maximum consequences to the monetary attribute are extreme. Within this context, the individual may weigh the two attributes differently.

The pairwise comparison process results in a set of person-specific weights, $w_i$, for the attributes that indicate the relative importance of attribute $i$ in the overall context of the decision problem. The weights across the entire set of attributes sum to unity. Consider Figure 3-3, in which each bar indicates the weight that the DM or a SH assigned to a particular attribute. The figure shows that the weights assigned to each attribute can differ significantly between individuals. As these weights may reveal fundamental differences in the way individuals perceive a decision problem, no attempt is made to reach consensus weights at this stage.
With all of this information collected, the objectives hierarchy is fully specified and the ADP process proceeds to its third step in which decision options are identified. It is important that the participants complete the first two steps before they begin to focus on the decision options. Defining the decision problem, its context and the objectives hierarchy encourages the DM and SHs to consider the problem from all points of view and to see the big picture. If alternatives are proposed a priori, then the participants will begin to focus solely on the characteristics that differentiate these alternatives and not necessarily on the underlying goal [13]. This can lead to an incomplete objectives hierarchy and may impede the formulation of less obvious, but potentially superior alternatives. The analyst might be diligent in encouraging participants not to adopt alternative focused thinking too early in the process. For the case of the PSACS problem the two options, air and water, were already well defined before the start of the ADP process, but the design team found it helpful to begin with an alternative neutral approach that helped ensure they had considered other types of solutions to the decay heat removal problem.

In the fourth ADP step, each of the decision options is scored according to the set of QPMs. Appropriate modeling and analysis is conducted and combined with the expert opinion of the participants so that the level of performance of each decision option is understood as well as possible. The constructed scales are then used to determine the corresponding value of this performance. Uncertainty in performance levels can be tracked rigorously as each decision option may lead to a distribution of possible values, not just a single point value. In general, this step of ADP is the most time consuming and
resource intensive as it is the point where external tools are used to study the decision options (see, for example, [7]). These may include computer modeling and simulation, physical experiments or extensive literature review. Recall, however, that in the case of the PSACS ultimate heat sink problem, time and resources for modeling and analysis were limited, so the QPMs were evaluated as point estimates almost exclusively from the expert opinion of senior members of the design team. The values assigned to the air heat sink and water heat sink options are presented in Appendix I - PSACS QPMs.

In the final step of ADP, the DM and the SHs select a decision option using a deliberative process that is led by the analyst. To facilitate deliberation, the analyst presents a preliminary ranking of the decision options. Options are ranked according to a Performance Index (PI). The PI for option $j$ is defined as the sum of the values, $v_{ij}$, associated with the QPMs for attribute $i$ weighted by the weight for that attribute, $w_i$.

$$PI_j = \sum_{i=1}^{N_{qpm}} w_i v_{ij}$$

The distribution of the PI is calculated separately for the DM and each SH. The decision options can then be ranked according to their expected PIs and the effect of performance uncertainty can be shown.

### 3.2 Deliberation and Results

With the calculation of the PIs, the analysis portion of the ADP ends and deliberation begins. The DM and the SHs each review their individual PIs to understand how the current state of knowledge about the decision options and their individual preferences for the attributes affect the decision problem. Individuals meet in a forum moderated by the analyst to discuss the similarities and differences between their rankings in order to reach a collective decision.

Figure 4 shows the expected PIs for each participant for the two PSACS decision options, air and water. First, note that the PI is a number between zero and unity. An option has a PI of unity if it achieves the highest possible level of performance in each QPM. The absolute magnitude of the PI therefore gives a general indication of the "goodness" of the options. For the PSACS problem, the air option’s characteristics put it at a level of about 40% of what the participants would have considered a best possible decision alternative and water at a level of about 60%. While neither of these options achieves top ranking relative to every attribute, comparing the PI values between the two options does indicate that all participants preferred the water option to the air one. Notice that this is a holistic comparison based collectively on all the attributes that pertain to the problem.

Recall that Figure 3-3 showed a large variation in the weights each participant assigned to the attributes. Figure 3-4 reveals that these differences do not affect the final ranking of the decision options. This information is valuable in guiding the deliberation. If the
differences in the weights did affect the ranking, then a large portion of the deliberation would focus on understanding how each of the participants arrived at their weights. In the initial informal decision-making process the design team attempted, their discussion focused around these issues and very little consensus was reached. As the variation in the weights still led to a consistent ranking, however, the deliberation was able to skip over these issues and focus on confirming that the participants were confident in the postulated performance levels assigned to each option.

ADP results can be further unpacked to help the DM and the SHs understand exactly why the water option was preferred to the air. Figure 3-5 shows the contribution of the eleven QPMs to the PI of the two options for SH 5. Water significantly outperformed air with respect to component cost, potential for common cause failures, supercritical carbon dioxide circulation performance and confidence in PCT. Air outperformed water in only one category, potential for spurious actuation.
Upon completion of the ADP process, the design team felt confident in selecting the PSACS option with the water ultimate heat sink for the FCR reactor design and they continued to develop it to the exclusion of the air option. ADP provided a transparent means to evaluate the two options, placing each of their performance characteristics in the broader context of the overall system. After completing the ADP, several members of the design team remarked that the water option was obviously the better choice and they were surprised they had not come to this conclusion using their informal decision making process.

In addition to resolving the question about proceeding with either the air or water option, the ADP also gave the design team valuable insights to help them produce a well-balanced design. ADP revealed, as indicated in Figure 3-5, that the water option would likely lead to a design that would be hard to inspect and maintain and that would be difficult and time consuming to characterize with the available modeling technology. The team was able to focus their efforts on these attributes that received a very low PI as opposed to ones receiving a higher PI.

### 3.3 Conclusions

In the very early phases of the design of complex systems there is reluctance by design teams to employ formal analysis and decision-making processes as it is often felt that information is too scarce or uncertainties too large for formal methods to yield any better
results then their less formal counterparts. Resistance also comes from a belief that the
time and effort required to implement a formal method is excessive and would be better
spent working on the actual design. This study suggests both of these assumptions may
be incorrect. Formal methods can still be applied in very early design phases and teams
can benefit from the ability of these methods to organize and prioritize information.

To be useful, however, formal methods must be scalable. They must be able to
incorporate information of varying fidelity from expert supposition to first-principle
simulations and they must require an effort to implement that is commensurate with the
significance of the decision problem at hand. ADP is such a method. Its use of
constructed scales allows users to incorporate whatever types of information they have
access to. This study has shown that, even if the information is almost exclusively
derived from expert opinion, useful results can still be obtained. Additionally, ADP is
not overly burdensome to participants. The process of constructing the objectives
hierarchy, developing constructed scales and ranking decision options took only about 20
person-hours or a little under three hours for each member of the design team. This is on
par with the time the team’s informal decision process required.
4 Selecting Numerical Scales for Pairwise Comparisons

A single-attribute decision problem is one in which the decision makers believe that the value of any particular decision alternative can be fully described by a single characteristic of that alternative. A common metric for this type of problem is monetary value. Each decision alternative can be assigned one number, in terms of dollars, that represents its value. Having this single value is critical as it allows alternatives to be ranked or selected using decision trees or other common techniques. A multiattribute decision problem is one where multiple attributes are required to fully describe the value of any decision alternative. These attributes are typically not commensurate and so the critical single value that should be associated with each alternative is not obvious. Examples of such attributes are costs, environmental impacts, and fatalities due to accidents.

One method that can be applied to obtain the needed value number is Multiattribute Utility Theory (MAUT) [11]. MAUT follows from a self-consistent set of axioms based on an individual’s indifference between various lotteries. It allows the value that a decision alternative derives from each attribute to be measured on a common scale and it provides a scheme for aggregating these into a single value [14]. Applying MAUT without significant simplifications is a complex process requiring the assistance of a skilled analyst and well-trained, patient decision makers. This complexity is prohibitive and a number of methodologies and simplifying assumptions have been proposed to obtain the necessary MAUT inputs while reducing the complexity of the elicitation process.

One such assumption that has become quite popular is that of additive independence, which holds that the value, $V_j$, of the jth decision alternative is simply a sum of the values derived from each attribute, $v_{ij}$, weighted by an attribute specific weight, $w_i$.

\[
V_j = \sum_{i} w_i v_{ij}
\]

where \[
\sum_{i} w_i = 1
\]

Regarding this assumption, Clemen [11] states: “Even if used only as an approximation, the additive utility function takes us a long way toward understanding our preferences and resolving difficult situations.” Clemen also states that “In extremely complicated situations with many attributes, the additive model may be a useful rough-cut approximation.” The additive independence assumption is found to yield good approximations for $V_j$ in a number of practical cases [15] but users should be cautious in problems that involve large uncertainties. The popularity of this assumption is partly due to the very simple and intuitive pairwise comparison method it permits in eliciting the $w_i$ from decision makers. Some recent applications of additive independence MAUT include a power grid vulnerability and risk assessment by Koonce [16] and a community hazard and risk analysis by Li [17].
Another popular method for addressing multiattribute decision problems is the Analytic Hierarchy Process (AHP). The AHP was intended as a method that would allow individuals or groups of decision makers to easily formulate multiattribute decision problems in a manner that would be intuitive, encourage compromise and build consensus while not requiring inordinate assistance from a skilled analyst [18]. The AHP is based on a different set of axioms from additive independence MAUT but it also leads one to calculate the value, \( V_j \), of the jth decision alternative using equation 1. Some practitioners find AHP simpler than additive independence MAUT as it permits both the \( w_i \) and the \( v_{ij} \) to be elicited using pairwise comparisons. Recent applications of the AHP include a study of road safety by Hermans, Van den Bossche and Wets [19], a method for estimating human error probabilities by Park and Lee [20], a method for prioritizing component maintenance by Bertolini and Bevilacqua [21] and a method for risk-informed safety categorization by Ha and Seong [22].

It should be noted that, over the years, experts have criticized various parts of the AHP, with the most significant criticisms relating to the issue of rank reversal. Rank reversal occurs when attributes or decision alternatives that are added or removed from a decision problem cause the rank of other decision alternatives to change relative to one another. The March 1990 issue of Management Science provides an excellent overview of this and other issues with perspectives both from the critics and proponents of the AHP [23],[24],[25].

The shortcomings of the AHP are, however, mostly restricted to the elicitation of the \( v_{ij} \), and it remains a very powerful tool for developing the \( w_i \). As such, a large number of authors continue to use the AHP. We point out that, despite the similarities between additive independence MAUT and the AHP, the AHP is not a replacement for additive independence MAUT.

In this paper we focus on the pairwise comparison elicitation method that is common to both additive independence MAUT and to the AHP. We will examine the numerical scale that is used when making paired comparisons as the choice of scale reflects the cognitive context in which an individual approaches the problem. We show that the choice of scale is not obvious, nor does there appear to be one best choice. We also examine how best to present comparisons to an individual and how this presentation can affect the response.

With the wide and continued usage of the pairwise comparison process, it is appropriate to reexamine the assumptions made in structuring the elicitation. In what follows, we present a brief overview of pairwise comparisons and then examine the numerical scale, presenting two alternatives to the traditional scale type. We examine the results of a study of 64 individuals responding to a hypothetical pairwise comparison exercise and draw conclusions about the appropriateness of the different scales and different variations on the elicitation process. Finally, we consider a pairwise comparison study involving an engineering design team in which three individuals worked on an actual decision problem. The design team was tasked with developing a technology for decay heat
removal from a nuclear fission reactor under accident conditions. This study highlights the effect of the choice of scale on the preferred decision.

4.1 Pairwise comparisons

A fundamental assumption of the pairwise comparison process is that, for any group of attributes, it is possible for an individual to supply information that allows an analyst to construct a set of weights, \( w_i \), that will indicate how relatively important each of the attributes is to the individual in a certain predefined context. For example, if an individual is considering a decision problem in which the cost, quality and manufacture time of some widget are attributes, then it is possible for the individual to respond to questions that will allow us to develop the numerical weights \( w_{\text{cost}} \), \( w_{\text{quality}} \) and \( w_{\text{time}} \) such that the ratio \( w_{\text{cost}} / w_{\text{quality}} \) represents how much more important cost is to quality in the context of this particular decision problem.

To uncover these weights, items in a group are presented to an individual two at a time and he is asked to answer two questions about each pair. First, which of the items in the pair is more important in the predefined context and, then, how much more important it is. The individual is, of course, permitted to respond that the items are equally important. To answer the second question regarding the degree of preference, the individual is given a list of linguistic phrases, shown in Table 4-1, to select from. The number of phrases that should be presented is taken up later in this paper but it is traditional to present either five or nine. If presenting five, the odd numbered phrases are used. If presenting nine, both the odd and even are used. In any particular application, the analyst may modify these phrases to better reflect the specific context. What is important is that the individual responding understands that each successive phrase represents an additional increment of importance. Presenting one pair to an individual and collecting the response to the two questions is referred to as eliciting a judgment.

<table>
<thead>
<tr>
<th></th>
<th>Phrases used in the pairwise comparison process to indicate degree of importance of item A over item B [12]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A is equally important to B</td>
</tr>
<tr>
<td>2</td>
<td>A is weakly or slightly more important than B</td>
</tr>
<tr>
<td>3</td>
<td>A is moderately more important than B</td>
</tr>
<tr>
<td>4</td>
<td>A is moderately plus more important than B</td>
</tr>
<tr>
<td>5</td>
<td>A is strongly more important than B</td>
</tr>
<tr>
<td>6</td>
<td>A is strongly plus more important than B</td>
</tr>
<tr>
<td>7</td>
<td>A is very strongly more important than B</td>
</tr>
<tr>
<td>8</td>
<td>A is very, very strongly more important than B</td>
</tr>
<tr>
<td>9</td>
<td>A is extremely more important than B</td>
</tr>
</tbody>
</table>

Determining the weights implied by a set of judgments requires two components. Each phrase used in the judgments must be assigned a numerical value from the scale and then these values must be manipulated to determine the weight vector. One scale we might
consider is the simple set of integers between 1 and 9 shown in Table 4-1. We will use this scale for the moment to illustrate how the weight vector is determined.

Returning to the example of the three attributes cost, quality and time in the manufacturing of a widget, we recognize that there are three possible judgments we could collect. In general, if we wish to determine weights for the relative importance of \( n \) items then there are "\( n \) choose 2" judgments that may be collected:

\[
\binom{n}{2} = \frac{n!}{2!(n-2)!} = \frac{n(n-1)}{2}
\]  

(3)

The reader may notice that some of these judgments are redundant, in that they can be calculated from other judgments. At a minimum, \( n-1 \) judgments must be collected, but redundant judgments are helpful in determining whether the individual is responding consistently to the comparisons. Let us assume that an individual supplies the judgments shown in Table 4-2 for the attributes cost, quality and time.

**Table 4-2: Hypothetical judgments between three items.**

<table>
<thead>
<tr>
<th>Cost is <strong>moderately</strong> more important than Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost is <strong>strongly</strong> more important than Time</td>
</tr>
<tr>
<td>Quality is <strong>weakly</strong> more important than Time</td>
</tr>
</tbody>
</table>

Using the scale in Table 4-1, we may form the following ratios to express these three judgments:

\[
\frac{\text{Cost}}{\text{Quality}} = \frac{3}{1} ; \quad \frac{\text{Cost}}{\text{Time}} = \frac{5}{1} ; \quad \frac{\text{Quality}}{\text{Time}} = \frac{2}{1}
\]  

(4)

It is now understood that the individual believes that cost is three times as important at quality, cost is five times as important at time and quality is twice as important as time. We may reorganize this information into the following reciprocal matrix, where it is understood that the entry in row \( i \), column \( j \) corresponds to the ratio of attribute \( i \) over attribute \( j \). Unities are always present on the main diagonal as these represent ratios comparing an attribute to itself. Saaty [12] has shown that the vector of weights, \( w_i \), which best represents these judgments is found by calculating the normalized eigenvector corresponding to the maximum eigenvalue, \( \lambda_{max} \), of this reciprocal matrix.
Table 4-3: Reciprocal matrix and corresponding weight vector obtained from pairwise comparisons for a hypothetical three-attribute decision problem.

<table>
<thead>
<tr>
<th>Matrix</th>
<th>Eigenvector</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C$</td>
<td>$Q$</td>
</tr>
<tr>
<td>$C$</td>
<td>1</td>
</tr>
<tr>
<td>$Q$</td>
<td>1/3</td>
</tr>
<tr>
<td>$T$</td>
<td>1/5</td>
</tr>
<tr>
<td>$\lambda_{max} = 3.0037$</td>
<td></td>
</tr>
</tbody>
</table>

4.2 The Scale

In selecting an appropriate scale to use when converting judgments into weights, it is necessary to consider normative as well as empirical factors. In the area of normative factors, we must examine the cognitive implications of a scale, that is, how it converts stimuli into sensations. Here, the stimuli are generated when one considers a pair of attributes that have a certain difference in preference between them, and the sensation is generated when one reports that difference in preference [26]. We must examine the distribution of possible attribute weights that the scale allows us to specify. Using a scale that permits only a finite number of responses means there are a finite number of weight values that can be calculated. This restricted set of weights may or may not be sufficient to capture the possible preferences of an individual and we wish to know if these values are distributed in a logical fashion. We also must examine the consistency the scale permits between the judgments. Noting that some judgments will be redundant; we wish to know whether the scale allows these redundant judgments to be consistent with the others. In the area of empirical factors, we must consider, in actual practice, what type of scale seems most natural to users and which they believe most accurately captures their preferences.

We will now introduce three candidate scales: an integer scale, a balanced scale and a power scale and look at the motivation for each. We will then examine the normative and empirical implication of each.

4.2.1 Integer Scales

The integer scales were the ones originally proposed by Saaty [12]. Though many such scales exist [27], the most popular, and the one we will focus on for this discussion, relates each of the phrases in Table 4-1 to the integers 1-9. This scale is motivated by the work of Ernest Weber and Gustav Fechner. In 1834, Weber [28], [29] found that the ability of a person to distinguish the magnitude of two stimuli is governed not by the absolute difference between the stimuli but rather by the ratio of the difference between the two stimuli, and the smaller of the two stimuli. Consider a person exposed to a stimulus of magnitude $s$, for our purposes, a paired comparison where the degree of...
preference for one item over the other is of magnitude \( s \). Now we increase \( s \) by an amount \( \Delta s \) such that the person can just distinguish between \( s \) and \( s + \Delta s \). \( \Delta s \) is referred to as the just noticeable difference and it is found that it is not constant but depends on the magnitude of the original stimuli \( s \). Further, it is found that the ratio, \( r = \Delta s / s \), is constant. That is, if a person is asked to hold two objects of different mass in opposite hands, the stimuli, his ability to distinguish that they are different, depends on the magnitude of their masses. As the objects get heavier, the difference between their masses must get larger in order for the person to be able to distinguish between them.

In 1860, Fechner [30] considered a sequence of just noticeable stimuli and found that, when considering just distinguishable masses, the magnitude of such stimuli would follow a geometric progression, i.e.,

\[
\begin{align*}
\text{s}_0 & \\
\text{s}_1 &= \text{s}_0 + \Delta \text{s}_0 = \text{s}_0 + \frac{\Delta \text{s}_0}{\text{s}_0} \text{s}_0 = \text{s}_0 + rs_0 = \text{s}_0(1 + r) \\
\text{s}_2 &= \text{s}_1 + \Delta \text{s}_1 = \text{s}_1(1 + r) = \text{s}_0(1 + r)^2 \\
\text{s}_n &= \text{s}_{n-1}(1 + r) = \text{s}_0(1 + r)^n
\end{align*}
\]

where \( r = \frac{\Delta s_i}{s_i} \) is a constant independent of \( i \).

Further, Fechner found that an individual’s report of the corresponding sensation, that is, what one perceived to be the relation between the just distinguishable masses, \( p \), would follow an arithmetic progression related to the logarithm of the stimuli:

\[
p_n = a \ln s_n + b
\]

Saaty postulated that if the stimuli arose from making pairwise comparisons between attributes, the sensation, or judgment, derived from the pair would also be of the form shown in Eq. (4). Saaty further conjectured that if we require the stimulus \( p_0 \) to be undetectable, that is, to result in no sensation of preference, and if we define the first next noticeable stimulus, \( p_1 \), as the sensation obtained from comparing one attribute to itself, that is, the sensation of equal preference, then we may conclude that

\[
\begin{align*}
p_0 &= a \ln s_0 + b = 0 \\
p_1 &= a \ln s_1 + b = 1
\end{align*}
\]

\[
\Rightarrow a = \frac{1}{\ln[s_0(1 + r)]} ; b = 0
\]

\[
\Rightarrow p_n = \frac{1}{\ln[s_0(1 + r)]} \ln s_n = \frac{1}{\ln[s_0(1 + r)]} \ln[s_0(1 + r)^n] = n
\]
and thereby obtain the integer sequence 1,2,3... corresponding to the next noticeable judgments resulting from pairwise comparisons. This is the derivation of the integer scale shown in Table 1.

To be clear, the implication of assuming a logarithmic relationship between stimuli and judgment in a pairwise comparison is that when an individual judges that attribute A is one increment more important than item B, i.e., the individual reports \( p_2 \) corresponding to the phrase “weakly more important”, then A must be two times as important as B. When \( p_3 \) is reported, then A is three times as important as B, etc.

Although an integer scale might be the right choice for making pairwise comparisons between objects of different mass, it is not clear that it is psychologically appropriate for making comparisons between the sometimes abstract attributes in a decision problem. There are other physical stimuli that are not interpreted in this way. Schoner and Wedley [31], for example, found that the integer scale produced poor results for judging optical stimuli. It should also be said that the popularity of the integer scale may, in part, be attributed to its simplicity rather than its physiological roots in the Weber-Fechner law.

4.2.2 Balanced Scales

It has been noted by several authors that the integer scales produce an uneven distribution of attribute weights. That is, if we consider a problem involving only two attributes, and if the judgment of the importance of A relative to B is increased from very, very strongly (8) to extremely (9), the increase in the weight of A is only about 1/15 as much as the weight would increase if its importance is increased from equally (1) to weakly (2). Salo and Hamalainen [32] proposed a class of scale where the increment in the weights is constant for all increments in judgment. That is, they proposed a scale where, for two attribute decision problems, \( \Delta s \) is constant and independent of \( s \). These so called balanced scales postulate that the weights, \( w \), deduced from the pairwise comparison between any two attributes with associated numerical judgments from the scale, \( x \), should have evenly distributed numerical values, such as 0.1, 0.15, 0.2, ..., 0.9 depending on the judgment rendered. By solving the eigenvalue problem associated with such a two-attribute reciprocal matrix,

\[
\begin{bmatrix}
1 & x \\
1/x & 1
\end{bmatrix}
\begin{bmatrix}
w \\
1-w
\end{bmatrix}
= 2
\begin{bmatrix}
w \\
1-w
\end{bmatrix}
\Rightarrow x = \frac{w}{1-w}
\]

we can determine the scale, \( x \), which would yield these weights, \( w \). Numerical values for the 0.1 - 0.9 based scale, which is the most popular and will be used for this discussion, are shown in Table 4-4.
Table 4-4: Numerical Values of the Balanced Scale generated using Eq. 5

<table>
<thead>
<tr>
<th>1</th>
<th>A is <strong>equally</strong> important to B</th>
</tr>
</thead>
<tbody>
<tr>
<td>11/9</td>
<td>A is <strong>weakly or slightly</strong> more important than B</td>
</tr>
<tr>
<td>4/3</td>
<td>A is <strong>moderately</strong> more important than B</td>
</tr>
<tr>
<td>13/7</td>
<td>A is <strong>moderately plus</strong> more important than B</td>
</tr>
<tr>
<td>( \frac{0.7}{1 - 0.7} = \frac{7}{3} )</td>
<td>A is <strong>strongly</strong> more important than B</td>
</tr>
<tr>
<td>3</td>
<td>A is <strong>strongly plus</strong> more important than B</td>
</tr>
<tr>
<td>4</td>
<td>A is <strong>very strongly</strong> more important than B</td>
</tr>
<tr>
<td>17/3</td>
<td>A is <strong>very, very strongly</strong> more important than B</td>
</tr>
<tr>
<td>1</td>
<td>A is <strong>extremely</strong> more important than B</td>
</tr>
</tbody>
</table>

It is noted that while the balanced scale produces evenly distributed weights for two-attribute decision problems, the weights will not be evenly distributed as the number of attributes increases. This will be taken up further in later sections.

### 4.2.3 Power Scale

A third possible scale we will consider is a power scale that is of the general form \( \alpha^x \), where \( x + 1 \) are the integers corresponding to the phrases in Table 4-1. Power scales have been proposed by a number of authors including Lootsma [33] who cites work by Stevens [34] in suggesting that a geometric series of stimuli should also produce a geometric series of sensations

\[
\frac{p_n}{p_{n-1}} = \left( \frac{s_n}{s_{n-1}} \right)^\beta = \alpha
\]  

for some positive constant \( \beta \). Lootsma points out that this is known to be an accurate representation of how auditory stimuli are perceived for which there is an experimentally determined value of \( \beta \). As with the integer scale, it is unknown whether pairwise comparisons of abstract attributes lead to stimuli that produce sensations similar to the sensations produced by auditory stimuli. Saaty has criticized power scales because of what he deems to be the arbitrary nature of choosing \( \alpha \) [12].

In this paper, we propose a new, non-arbitrary, power scale based on two principles. First, we adopt the motivation behind the balanced scale, that \( \Delta s \) should be constant and independent of \( s \). Unlike the balanced scale however, the intent is to maintain constant \( \Delta s \) for a decision problem involving any number of attributes, not just two. This will lead to a distribution of the weights implied by the judgments that is uniform for any number of attributes. Second, we wish to permit consistency between the judgments. By consistency we mean that redundant judgments can be calculated exactly from other judgments.

As such we propose,

\[
\alpha = r^{-1/\sqrt{9}}
\]  

40
where \( \gamma \) is the number of increments of judgment that is deemed appropriate for comparing any set of attributes. As mentioned before, this is usually either five or nine increments. Table 4-5 provides the numerical values implied by the power scale for each of the possible preference phrases.

### Table 4-5: Numerical Values of the Power Scale

<table>
<thead>
<tr>
<th>((\sqrt{9})^0)</th>
<th>A is equally important to B</th>
</tr>
</thead>
<tbody>
<tr>
<td>((\sqrt{9})^1) = 1.32</td>
<td>A is weakly or slightly more important than B</td>
</tr>
<tr>
<td>((\sqrt{9})^2) = 1.73</td>
<td>A is moderately more important than B</td>
</tr>
<tr>
<td>((\sqrt{9})^3) = 2.28</td>
<td>A is moderately plus more important than B</td>
</tr>
<tr>
<td>((\sqrt{9})^4) = 3</td>
<td>A is strongly more important than B</td>
</tr>
<tr>
<td>((\sqrt{9})^5) = 3.95</td>
<td>A is strongly plus more important than B</td>
</tr>
<tr>
<td>((\sqrt{9})^6) = 5.20</td>
<td>A is very strongly more important than B</td>
</tr>
<tr>
<td>((\sqrt{9})^7) = 6.84</td>
<td>A is very, very strongly more important than B</td>
</tr>
<tr>
<td>((\sqrt{9})^8) = 9</td>
<td>A is extremely more important than B</td>
</tr>
</tbody>
</table>

In the next sections, we will discuss further how this choice of \( \alpha \) leads to a more uniform distribution of weights and the benefits a power scale has on the consistency of judgments.

### 4.3 Choosing a Scale – Normative Factors

Now that we have identified three candidate scales, we wish to ask which scale is most attractive from a normative point of view. In other words, which scale leads to weights with properties that seem appropriate based on simple logical arguments. There are no well-established axioms to draw from in discriminating between scales, but we can still develop simple hypotheses regarding what properties should make one scale better than another. We will consider two such hypotheses regarding how weights should be distributed and how judgments should be consistent.

#### 4.3.1 Distribution of the Weights

Let us postulate that the optimal distribution of weights is uniform, where the possible weights occur at evenly spaced intervals between 0.1 and 0.9 for any attribute. We pick the 0.1 and 0.9 bounds somewhat arbitrarily, with the logic being that any attribute not capable of contributing at least 10% to the total value of a decision alternative should be
screened from the process. We require a uniform distribution from concluding that, a
priori, before anything is known about the attributes, there is no reason that certain
weight values in the 0.1-0.9 range should be more preferable relative to others. All that is
left to consider is the number of evenly spaced weight values within the range that should
be permitted. The number of degrees of preferences presented to an individual controls
this. As mentioned previously, either five or nine degrees are most common. With five
degrees of preference there are eight possible judgments between two attributes. A is
equally preferable to B, A is moderately, strongly, very strongly, or extremely more
important than B, or B is moderately, strongly, very strongly, or extremely more
important than A. In the 0.1-0.9 weight range, these eight possible judgments lead us to
the weights, 0.1, 0.2, 0.3, ..., 0.9 with an increment of 0.1 between each. If we permit
nine degrees of preference there are 17 possible judgments between any two attributes
leading to weights in the 0.1-0.9 range separate by an increment of 0.05.

Now let us examine the impact of the three scales on the distribution of attribute weights
in the context of these optimal requirements. Consider Figure 4-1, which shows all
possible attribute weights that can be obtained using a 5-degree of preference scheme for
the comparisons between two and three attributes. Each point in the figure defines a
possible combination of weights for two of the attributes using the specified scale. For
the case of N = 2, two attributes are being compared and so each point in the figure
defines the complete weight vector. For the case of N=3, three attributes are being
compared and the figure shows only two of the three weights but the third can easily be
inferred by the reader recalling that the sum of the weight vector must be unity. The
reader will also recognize that the distribution of weights for N>3 will follow the same
pattern as that for N=3 with the range of the distribution decreasing as N increases.

To be clear on the interpretation of Figure 4-1, consider again the three-attribute example
of Table 4-2 whose weight vector was calculated using the integer scale in Table 4-3. Let
the attribute labeled A in Figure 4-1 correspond to Cost and that labeled B correspond to
Quality. The weight vector corresponding to this example is located on the Integer Scale
panel of Figure 4-1 at A = 0.65, B = 0.23. The weight of the third attribute, Time, is
found by taking 1 - 0.65 - 0.23 = 0.12. In the Balanced Scale panel, these same
judgments lead to a weight vector with A = 0.47, B = 0.31. In the Power Scale panel, A
= 0.48, B = 0.28.
Figure 4-1: Distribution of attribute weights obtainable by the three scales for comparisons between 2 and 3 attributes. A simple computer algorithm was written to generate these distributions that iteratively evaluated the weight vectors for all possible sets of judgments.
Considering the \( N=2 \) case of comparisons between two attributes, we find that the balanced scale exactly meets the optimal criteria. This is hardly surprising as the scale was specifically designed to do so. The power scale comes close to achieving a uniform distribution, though it is slightly skewed to more extreme values of the weights. What is more interesting is the behavior of the integer scale. This scale is highly skewed to extreme values of the weights. If an individual indicates that attribute A is moderately more important than attribute B, the first possible increment of preference over equally important, then the integer scale implies a weight vector of 0.75 to 0.25, a factor of 3 increment in preference.

Considering the \( N=3 \) case, we find that the balanced scale no longer implies a uniform distribution of the weights and the figure shows evidence of sparse regions where no weight values lie. It would, of course, be possible to develop a new balanced scale specific to \( N=3 \), but it would require 28 degrees of preference (to achieve a weight increment of 0.1) and be prohibitively complex for the user. If one considers extending this logic to even higher values of \( N \), this is clearly a prohibitive approach. The integer scale fairs even worse for \( N=3 \); there are very large sparse regions, and significant clustering of the weights. While not shown, this behavior becomes even more pronounced for higher values of \( N \). The power scale, on the other hand, produces fairly reasonable results. There is again some skewness to more extreme weights but there is still fairly complete coverage of values around 0.5 with a much more uniform distribution.

Based on these findings, we conclude that the balanced scale provides an optimal distribution of the weights for \( N=2 \) and that, while none of the scales are optimal for \( N>2 \), the power scale provides the best approximation.

### 4.3.2 Consistency of the Pairwise Comparisons

If an individual provides responses to pairwise comparisons that lead to the numerical values \( c_{ij} \) from the scale, then the comparisons are said to be consistent if the \( c_{ij} \) are equal to the ratios of the weights \( w_i/w_j \). That is,

\[
\begin{bmatrix}
1 & c_{i2} & \cdots & c_{ij} \\
1/c_{i2} & 1 & c_{2j-1} & c_{2j} \\
\vdots & 1/c_{2j-1} & \ddots & \vdots \\
1/c_{ij} & 1/c_{2j} & \cdots & 1
\end{bmatrix}
= 
\begin{bmatrix}
1 & w_i/w_2 & \cdots & w_i/w_j \\
w_2/w_i & 1 & w_2/w_{j-1} & w_2/w_j \\
\vdots & w_{j-1}/w_2 & \ddots & \vdots \\
w_j/w_i & w_j/w_2 & \cdots & 1
\end{bmatrix}
\]

To be more specific, if we consider the first two judgments of the pairwise comparison example of Table 4-2, we can infer the third. Using the power scale, we have:
Then if the individual supplies the third judgment, "Quality is weakly more important than time", which has the numerical value of 
\[
\left( \frac{\sqrt{3}}{\sqrt{3}} \right)^2 = 1.73
\]
we can say this judgment is consistent with the first two. Let us postulate that, at optimum, individual’s preferences should exhibit this mathematical consistency.

It is expected, however, that an individual will naturally make some inconsistent judgments when doing comparisons and Saaty [12] has proposed a framework to determine whether the level of inconsistency warranted revising the judgments. He proposed that two quantities, called the consistency index (CI) and consistency ratio (CR), be calculated:

\[
CI = \frac{\lambda_{\text{max}} - n}{(n - 1)}; \quad CR = \frac{CI}{RI}
\]

where \(\lambda_{\text{max}}\) is the maximum eigenvalue of the matrix of the judgments \(c_{ij}\), \(n\) is the order of the matrix, or the number of attributes being compared, and \(RI\) is a quantity called the random index. CI is a measure of the inconsistency of the judgments \(c_{ij}\). When the matrix contains completely consistent judgments, then \(\lambda_{\text{max}} = n\) and \(CI = 0\). As the judgments become more inconsistent, \(\lambda_{\text{max}}\) increases along with \(CI\). CR is a measure of the level of inconsistency of the judgments relative to the level of inconsistency of random judgments. The random index, \(RI\), is the average CI for a large number of random realizations of the reciprocal matrix. \(RI\) depends on \(n\), the scale and the number of degrees of preference being used. \(RI\) for \(n\) from 3 to 10, for the three scales and for 5 and 9 degrees of preference are provided in Appendix II – The Random Index. Note that all judgments for \(n=2\) must be consistent and so it is not necessary to consider CR in this case.

Saaty proposed that, if the CR value is less that 5% for \(n = 3\), 8% for \(n = 4\) and 10% for \(n > 5\), the judgments should be deemed to have acceptable consistency and not require further refinement. There are two reasons Saaty puts forth for allowing some level of inconsistency. First, small levels of inconsistency will not have a large impact on the weights and so excessive effort should not be expended to resolve it. Second, small inconsistencies might not be errors on the part of the individual supplying the judgments.
but rather might carry some information that makes the judgments more consistent with
the individual’s experience [18].

Figure 4-2: Possible number of consistent judgments and weight vectors for each of the three scales
as a function of the permitted percent inconsistency as defined by the consistency ratio CR. Panels A
and B are for judgments among three attributes, N=3. Panels C and D are for N = 4. Panels A and C
present the total number of consistent judgments. Panel B and D present the number of unique
weight vectors that arise from these consistent judgments. A unique weight vector is defined as one
with N unique numerical values. The data in this figure was generated by a computer algorithm
that generated the weight vectors for all possible sets of judgments and then calculated the consistency
ratio corresponding to those judgments.

These two reasons aside, we point out a more fundamental motive for permitting some
inconsistency. Consider Figure 4-2, which shows the number of possible weight vectors
that may be obtained for n=3 and n=4 using the three scales and permitting different
levels of inconsistency. Panels A and C in the figure show the total number of possible
sets of judgments that are possible for a given CR level. Note that multiple sets of
judgments may lead to the same weight vectors and so Panels B and D were generated to
show the number of unique weight vectors that can be achieved. Here, a unique weight
vector is defined as one that has n unique numerical values. From the figure we see that,
if we permit 5% (or less) inconsistency using the integer scale for n=3, then there are 103
possible sets of judgments and 78 possible unique weight vectors. This is a sufficiently
large number of unique weight vectors as to allow us to elicit an acceptable range of preferences subject to the restrictions mentioned in the previous section. Consider however, what happens if we require actual consistency (CR = 0) in the judgments. Then, for n=3 and integer scale, there are only 31 possible sets of judgments and only 6 possible unique weight vectors. This is clearly unacceptable as surely there are more than 6 possible variations of weight vector that individuals may wish to express when comparing three attributes. If we consider n=4 then there are no possible unique weight vectors permitted by the integer scale for CR=0. While not shown in the figure, this remains true for all n>4 as there must be at least n choose 2 unique combinations of two numbers in the scale whose ratio is itself a number in the scale in order for a unique weight vector to exist. So we may conclude that the most fundamental reason to permit inconsistency when using the integer scale is that if one does not, no judgments leading to unique weight vectors would be deemed acceptable for n≥4. If we consider the balanced scale, the situation is even worse with no such acceptable judgments being possible for n≥3.

In contrast, the power scale, gives rise to a comparatively large number of acceptable sets of judgments and unique weight vectors at CR=0. For n=3, the power scale permits 61 sets of judgments and 36 unique weight vectors. For n=4, it permits 359 sets of judgments and 96 unique weight vectors. Unlike the balanced and integer scales the number of permitted judgments and unique vectors increases with n as opposed to rapidly dropping to zero.

We postulate that a desirable scale is one that permits a large number of possible judgments at CR = 0. Notwithstanding Saaty’s two arguments, individuals should strive to be consistent in their responses to pairwise comparisons so as to maintain transitivity among their preferences, a cornerstone of rational behavior [35]. A scale should be used that allows them to do to and, among the three we are considering here, the power scale is clearly best.

4.4 Choosing a Scale - Empirical Arguments

In the previous sections, we presented three possible scales that can be used to convert phrases expressing different degrees of preference into numerical values. We explored the implications each scale has on the distribution of possible weight vectors and on making consistent judgments. We must now recognize that, apart from these normative concerns, the choice of scale has a real and significant affect on the actual numerical weights that are calculated from a given set of judgments. Figure 4-3 illustrates the range of different weight vectors that arise from a single, hypothetical set of judgments (with CR < 5%) depending on which of the three scales is applied. Notice that the integer scale suggests attribute 1 is about twice as important as attribute 2 while the balanced scale would suggest that the individual is essentially indifferent between the two attributes.
Perhaps of more concern than the difference in the magnitude of the weights is a reversal of rank between one or more attributes based on the scale that is chosen. Table 4-6 shows that, if all possible sets of judgments for comparing four attributes are considered², there is a significant probability of such rank reversal occurring. To explain, the table indicates there is an 89.6% chance that a set of random judgments comparing 4 attributes will result in weight vectors that imply a different ranking of those attributes if one weight vector is calculated with the integer scale and the other with the balanced scale. Notice that the table implies that the integer and balanced scales are the least consistent with one another while the power and balanced scales are most consistent.

---

² All possible set of judgments are generated by iterating thought all possible responses to the six paired comparisons necessary for four attributes. For the nine-degree of preference scale used, there are 531,441 such sets.
Table 4-6: Percent of possible sets of judgments for N=4 whose corresponding weight vectors lead to different rankings among the attributes when calculated with the indicated scales. Percentages are given for rank differences regardless of the magnitude of the difference between the weights and for when the difference is greater than 5% and 10%. These results are presented without regard to the CR of the underlying judgments and using a nine-degree of preference scale.

<table>
<thead>
<tr>
<th></th>
<th>Δw &gt; 0</th>
<th>Δw &gt; 5%</th>
<th>Δw &gt; 10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integer vs. Balanced</td>
<td>89.6</td>
<td>23.2</td>
<td>4.02</td>
</tr>
<tr>
<td>Integer vs. Power</td>
<td>58.6</td>
<td>5.91</td>
<td>0.43</td>
</tr>
<tr>
<td>Balanced vs. Power</td>
<td>37.3</td>
<td>0.91</td>
<td>0.00</td>
</tr>
</tbody>
</table>

4.5 Study I – 64 Individuals, Hypothetical Decision Problem

Given the magnitude and prevalence of the differences between the weight vectors that result from the different scales, the author wished to determine whether the vectors produced by one of the scales was a better match to the actual preferences of individuals. To this end, 64 individuals (undergraduate college students) were presented with a set of four attributes to compare using the pairwise comparison process described in this paper. The attributes were chosen as part of a hypothetical decision problem. Participants where presented with the three weight vectors that could be calculated using the three scales. Each individual was asked to compare these vectors relative to how well they represented that individual’s actual preference for the attributes. Appendix III – Survey Questions provides a more detailed description of the questions posed to the participants. We will now examine the results from this study.

4.5.1 The True Scale

Table 4-7 summarizes the extent to which study participants believed each of the three scales captured their true weight vector. To better understand the table, let us consider the first row. It indicates that 32% of participants believed the integer scale best captured their actual preference for the attributes. It further indicates that participants who chose the integer scale believed its resulting weight vector was closer to their actual weights by an average degree of 5.2 relative to the balanced scale. Recalling Table 4-1, for the integer scale, the numerical value 5.2 most closely corresponds to the linguistic phrase “strongly more important” implying that, on average, these respondents replied that the weight vector produced by the integer scale was strongly closer to their actual weights than that produced by the balanced scale. The uncertainty range given represents the 75% confidence interval of the responses. So 75% of those individuals believing the integer scale indicated it was superior to the balanced scale to a degree of 5.2 - 2.6 = 2.6 and 5.2 + 2.1 = 7.3. In other words, the integer scale was roughly between moderately superior and very strongly superior to the balanced scale. The other rows in the table can be interpreted in a similar fashion, recalling that degree of belief numbers are drawn from the individuals’ preferred scale. To determine the linguistic phases corresponding to preferences for individuals believing the balanced scale, the reader should refer to Table 4-4 and to Table 4-5 for those believing the power scale.
Table 4-7: Fraction of study participants believing each of the three scales best captured their actual attribute weights including the degree of belief over the two other scales. The degree of belief for each individual was calculated using his preferred numerical scale. The ranges supplied represent the 75% confidence interval of the responses.

<table>
<thead>
<tr>
<th>Believed Scale</th>
<th>% Of Participants</th>
<th>Degree of Belief Over Scale X</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Integer</td>
</tr>
<tr>
<td>Integer</td>
<td>31</td>
<td>1</td>
</tr>
<tr>
<td>Balanced</td>
<td>28</td>
<td>6.3 (+2.2, -1.3)</td>
</tr>
<tr>
<td>Power</td>
<td>41</td>
<td>5.0 (+2.3, -1.7)</td>
</tr>
</tbody>
</table>

Notice that, while slightly more participants believed the power scale, none of the scales was believed by an overwhelming majority of people. Therefore, it cannot be concluded that one scale is clearly better at representing the preferences of all individuals. The data indicating the degree to which an individual believed his chosen scale is more insightful however. Notice that the individuals who believed the Power scale preferred it weakly to moderately over the Balanced scale but preferred it strongly to very strongly over the Integer scale. Similarly, those who believed the Balanced scale preferred it equally to weakly over the Power scale but preferred it strongly to very, very strongly over the Integer Scale. That is, participants who believed either the Balanced scale or the Power scale were essentially indifferent between these two scales but thought they produced significantly better results than the Integer scale. The same is true for those who believed the Integer scale. These individuals strongly preferred the Integer scale to either the Balanced or the Power scale.

We can conclude that participants fall into one of two categories, those whose judgments were best represented by the Integer scale and those whose judgments were best represented by either the Balanced or Power scales. This would seem to suggest that when individuals are comparing attributes where the degree of the preference, the stimulus, is judged by a psychological construct rather then from direct sensory input, not all individuals perceive degrees of preference in the same way. Some individuals perceive degree of preference as increasing in multiples, \( \Delta s/s = constant \), and so they believe the Integer scale. We can say that they judge preference between attributes in a similar manner as they would judge the physical stimulus, weight of an object. Other individuals perceive preference as increasing in equal increments, \( \Delta s = constant \), and so they prefer either the Balanced or Power scales. We can say they judge preference in the same manner they would judge the physical stimulus of sound.

4.5.2 Number of Degrees of Preference

Consider the number of degrees of preference that should be presented to individuals during the judgment elicitation process. Standard practice is to present either 5 or 9 degrees. Saaty proposed first using 5 degrees but allowing the full set of 9 to be used if
individuals indicate they need more granularity in making judgments or if compromise is needed between individuals [18].

To test what effect this number might have, half of all study participants were asked to complete judgments using the 5-degree scheme and the other half where asked to use the 9-degree one. After completing all judgments, the participants were asked whether they believed the number of choices they were given was too many, too few or just right. Responses, presented in Table 4-8, indicate that the overwhelming majority of participants (84.4%) believed the 9-degree scheme gave too many choices as opposed to 15.6% who felt the number of choices were just right. Participants were about evenly split as to whether the 5-degree scheme gave too many choices or was just right (43.8% versus 53.1%).

<table>
<thead>
<tr>
<th></th>
<th>Too Many</th>
<th>Just Right</th>
<th>Too Few</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 Degree</td>
<td>43.8</td>
<td>53.1</td>
<td>3.1</td>
</tr>
<tr>
<td>9 Degree</td>
<td>84.4</td>
<td>15.6</td>
<td>0</td>
</tr>
</tbody>
</table>

In addition to how participants felt about the number of choices, we can also look at the consistency of the judgments obtained when using the two schemes. The distribution of the consistency ratio, CR, for all study participants was tabulated and it was found that the average CR was 11.9% for the 5-degree scheme and 17.6% for the 9-degree scheme. Participants who used the 5-degree scheme provided noticeably more consistent judgments relative to those using the 9-degree scheme. That a smaller number of choices provides more consistent judgments should not be surprising. The fewer the choices, the easier it is for individuals to remember the degree of preference expressed by each choice and so they will apply their own internal definitions of each choice more consistently. This phenomenon is well understood in psychology. Miller [36] argued that people could maintain, on average, seven plus or minus two items in their working memory at any one time. If more items were presented some of the original group would be forgotten. For simple items, such as integers between 1 and 10, people could remember closer to nine items and for more complex and abstract concepts that number would drop toward five.

In this context, the 5-degree scheme would seem to offer benefits over the 9-degree scheme. Participant felt more comfortable using it and it produced lower levels of inconsistency between judgments. It should be noted, however, that this study only considered comparisons among four attributes and it is not clear how these results would change if more attributes were compared or if the relative importance of the attributes differed greatly from those in this study.
4.5.3 Calibrating the Participants

An open issue in the elicitation of judgments is the order in which pairwise comparisons should be presented. Comparisons can be presented sequentially in random order. They can also be presented concurrently with the participant first being asked to identify the pairs exhibiting the minimum and maximum degrees of preferences, render judgments for those two pairs, and then complete the remaining pairs in random order. The motivation for the first elicitation order is to eliminate biases that may result from any particular ordering of the judgments. The motivation for the second is to calibrate the participants so that they have a better understanding for the degree of preference implied by each of the phrases in the scale and thereby improve the consistency of the judgments.

With this study, the author wished to see if the calibrated order does in fact improve consistency. To this end, half of the study participants were presented with pairs in random order and the other half in calibrated order. Figure 4-4, presents the distribution of CR obtained from participants separated both by the number of degrees of preference used and by order of elicitation. For those participants given 5 degrees of preference, average CR improved from 13.7% for random order to 10.1% for calibrated while the range of CR remained the same. For participants given 9 degrees of preference, average CR improved from 21.7% to 13.6% and the range fell from 50% to 40%.

It would seem, then, that the calibrated order does in fact improve consistency of the comparisons and that the amount of improvement increases as the number of degrees of preference used increases.

![Figure 4-4: Distribution of CR for study participants. The CR for each individual was calculated using the RI from the scale that individual preferred. The left panel presents CRs for individuals given 5 degrees of preference to choose from and the right panel presents CRs for 9 degrees. The curves marked as calibrated refer to CR values from the set of individuals who were asked to first complete bounding pairs. The curves marked as random refer to CR values for individual who were presented pairs at random.](image-url)
4.6 Study II – 3 Individuals, Actual Decision Problem

The results presented in the previous section were obtained under controlled conditions by posing a set of pairwise comparisons, which described a hypothetical decision problem, to a large group of random individuals. Not, to a group of actual decision makers and stakeholders who were attempting to make a real decision. As such, it is reasonable to question if and when the sensitivities to numerical scale affect the preferred decision alternatives in an actual decision problem. The drawback to studying actual decision problems is that they tend to involve only a small number of individuals, making it hard to draw the kind of statistically significant observations as was possible in the previous section. Notwithstanding this limitation, we will now consider again the actual PSACS decision problem from chapter 3.

Let us consider results from three members of the design team, whom we will refer to as stakeholders (SHs) 1, 2 and 3. We begin by looking at the differences in the attribute weight vectors that result from each stakeholder’s pairwise comparisons when processed using each of the three scales. Figure 4-5 shows the absolute percent difference between attribute weights for attributes of the same rank order for each pair of scales. These are presented as a function of the attribute weight calculated using the integer scale. To explain, consider the panel for stakeholder 2 and the three data points furthest to the right on the horizontal axes. These points show that the attribute that this stakeholder ranked most important was given a weight of about 0.23 by the integer scale. Further, they show that the weight value for this highest ranked attribute varies about 10% if calculated instead by the balanced scale, about 5% if instead calculated by the power scale and that the difference between the weights calculated with the balanced and power scales is about 4%. Though not shown in the figure, it is the case that the rank order of the attributes was not affected by the choice of scale for stakeholders 2 or 3, but there is rank reversal between the two lowest ranked attributes for stakeholder 1. For those two attributes the integer scale produced one ranking and both the power and balances scales produced the reverse.
Several interesting observations can be made from Figure 4-5. First, the magnitude of the difference in weights produced by the three scales is significant and varies by stakeholder. The average difference in weight between attributes of the same rank order is 10%, 10% and 21% for stakeholders 1, 2 and 3 respectively and the maximum difference is 26%, 42% and 110%. Recalling from previous discussions that it is common to control consistency of the pairwise comparison process so that weights are “accurate” to within 10%, clearly, the variations produced by choice of scale is also worthy of attention.

Second, note the correlation between the range of the underlying attribute weights and the magnitude of the weight differences produced by the scales. The difference in weight between the maximum and minimum ranked attributes for stakeholder 3 is almost twice that of either stakeholders 1 or 2 and the average difference in weights produced by varying scale for stakeholder 3 is twice that for stakeholders 1 and 2. This suggests that controlling the choice of scale becomes more important as the range of the weight vector increases.

Third, note that the magnitude of the absolute percent differences in weight is inversely correlated with the magnitude of the underlying weight. That is, the largest percent differences correspond to the lowest ranked attributes. For all three of the stakeholders, the figure shows, that the choice of scale has the greatest effect on the weights of the lower ranked attributes and that the effect diminishes progressively for each higher ranked attribute, except for the highest ranked attribute. This can be understood, as, the
effect of changing scale is to “take” weight from the highest ranked attribute and to redistribute it preferentially to the lower ranked attributes. This suggests that in decision problems where the options’ variation in performance is dominated by characteristics of the lower ranked attributes, it is more important to select an appropriate scale than is cases where the options vary by characteristics of the higher ranked attributes.

With this understanding of the variations in weight vector produced by the three scales, we examine the effect these variations have on the preferred decision of air versus water ultimate heat sink. Figure 4-6 presents the performance index (PI) for the air and water options for each of the three stakeholders using each of the three scales. The PI has been normalized so that the PI of the water option is unity. Note that for stakeholders 1 and 2 the choice of scale has no effect on the preferred decision, the water option is always preferred. Also note that the PI of the air option is mostly insensitive to the choice of scale. For stakeholder 3, however, the choice of scale is more significant. If the integer scale is chosen, the water option is preferred, if the power scale is chosen the water option is only very slightly preferred, and if the balanced scale is chosen the two options have essentially equal PI.

![Figure 4-6: Performance Index (PI) for the air and water ultimate heat sink options as a function of the three scales for the three stakeholders. The PI has been normalized such that the PI of the water option is unity.](image)

This behavior of the PI between the three stakeholders is consistent with the previous observations about the degree of variation in their weight vectors. Stakeholder 3’s preferences were most sensitive to the choice of scale and so it is logical that his PIs would similarly be most affected. It can be seen, that if the three stakeholders were only
presented with results from the integer scale, they would have little difficulty agreeing that the water option is preferred. Introducing the other two scales, however, would lead stakeholder 3 to carefully consider which scale most accurately represented his preferences before making a decision. In this particular decision problem all three stakeholders eventually used the integer scale to model their beliefs and they did chose the water option for their design.

4.7 Conclusions

It has been demonstrated that, in applications of the pairwise comparison process, the choice of scale has a significant effect on the attribute weights that are calculated from the judgments supplied by an individual and potentially on the preferred decisions these judgments imply. It is important therefore to examine the choice of scale. From a theoretical perspective, the traditional integer scale has two significant shortcomings. It results in a distribution of weight vectors that is highly skewed to extreme weight values and produces clustering of the weights for all n>2, a behavior that does not appear to be logically justifiable. It also precludes consistent, unique weight vectors for all n>4. On these grounds, we were motivated to explored two additional scales, the balanced and power scales, and it was found that the power scale produced a superior distribution of weights and allowed for more consistent, unique weight vectors then either of the other scales.

When these three scales were presented to individuals in a judgment elicitation exercise involving a hypothetical decision problem, it was found that individuals appeared to think about the preference between attributes in two distinct manners, one consistent with the integer scale and the other consistent with either the balanced or power scales. When these three scales were considered in a judgment elicitation exercise involving an actual decision problem, it was found that the choice of scale has the greatest potential to affect the outcome of a decision problem when there is a large range in the attribute weight vector and when differences in the decision options pertain mostly to the lower ranked attributes.

As such, it is concluded that applications of the pairwise comparison process would benefit from allowing individuals a choice between the scales that will be used to process their judgments. An obvious criticism of allowing such a choice is that it would complicate the elicitation process and require tutoring individuals on each scale. This need not be the case, however. Recall from Appendix II, that the elicitation can be accomplished independent of the scale, as the verbal phases remain the same for the three scales. Individuals supplying judgments need never be told that there is uncertainty in the choice of scale and never need to be tutored in the underlying numerical values. After the elicitation, the analyst can calculate the weight vectors corresponding to the three scales and then present these three vectors to the individual so that she may select that which she feels best represents her preferences. Given that the weight vectors, the ranking of the attributes, and even the preferred decision option can be affected dramatically by the choice of scale, this minor additional task is more than justified.
5 Risk-Informed Decision Making for the Assessment of Lunar Service Systems Payload Handling

The following presents background and a partial case study for the application of risk-informed decision-making (RIDM) using the Analytic Deliberative Decision Making Process (ADP) to an engineering decision problem at NASA. The decision problem of interest is to iteratively refine and optimize payload-handling concepts for the lunar surface. Each concept is designed to meet the lifting, handling and mobility requirements as defined by evolving architecture and scenario definitions [37] and the intent of applying ADP to these concepts is to better focus research efforts.

The lunar cargo-handling problem arises from the constellation program’s Lunar Surface Systems (LSS) architecture effort that is developing technologies to allow extended human exploration of the lunar surface. In this analysis, expectations related to technical performance attributes, such as safety, reliability, and mission performance, as well as programmatic performance attributes, such as cost and schedule, are captured. Though the constellation program has been cancelled at the time of this writing and this case study could not be completed, a number of improvements in the ADP process have been possible.

We will begin by presenting background information on risk-informed decision-making and the ADP as it pertains to NASA and lunar surface systems. This will be somewhat repetitive of the concepts in Chapter 3 but will focus on ADP as it applies to a large organization. We will then address improvements in the specification of the role of experts. Finally we will present the objectives hierarchy that was developed to address lunar payload handling.

5.1 RIDM for LSS

The Office of Safety and Mission Assurance at NASA headquarters is interested in developing formal methods to assist NASA programs and projects in making risk-informed decisions. The intent is not to develop an expert system, or black box, to replace existing decision-making practices, but rather to allow decision makers to organize information in a way that clearly reveals the benefits and risks associated with candidate decision options. The development of lunar payload handling systems is an example of such a decision problem.

Risk-informed decision-making brings together the decision maker, those who have a stake in the outcome of a decision (stakeholders), and experts who can characterize available options. It assists these individuals in systematically identifying the objectives of making a particular decision and defining the performance of options in the context of these objectives. The process aggregates both objective and subjective information while keeping track of uncertainty. It combines analytical methods with a deliberation that
scanty: the analytical results. It produces a ranking of decision options and a detailed understanding of why certain options outperform others.

To facilitate risk-informed decision-making of the lunar payload handling system, the Analytic Deliberative Process (ADP) is used. ADP is a method of implementing multiattribute decision theory that is an axiomatic theory for the behavior of one individual. Extensions of this single individual theory to a small group of individuals are quite common, where it is assumed that all individuals involved may meet in a group setting to deliberate and reach consensus. In a large organization like NASA, decisions may involve many individuals across the organization making full group deliberation impractical and so ADP must be extended to this large group model. Let us explore the steps in the ADP in the context of large group decision-making.

5.2 The Analytic Deliberative Process for Large Groups

In most large organizations, such as NASA, decisions are made in the context of the organization’s hierarchal structure. One person, or a small group of individuals, is the decision maker (DM). The seniority of the DM is commensurate with the scope and complexity of the problem at hand. Applying a formal methodology like the ADP can help organize information when the scope and complexity are sufficiently large such that the DM must gather information from a number of sources.

A decision with a large scope has consequences that affect portions of the organization for which the DM is not directly accountable. These portions are led by individuals near the level of the DM in the organization’s hierarchy. These individuals have a stake in the outcome of the decision, as it can enable or constrain their future capabilities, but they have little direct authority over the decision. If the DM is to make a decision that best benefits the organization as a whole, he or she must capture the elements that are important to these stakeholders (SHs) in addition to elements he or she feels are important.

In a complex decision problem, the DM cannot be expected to identify the decision options or be intimately familiar with each. He or she must rely on experts further down in the organization’s hierarchy to identify and analyze options and relate their relative advantages and disadvantages. Similarly, the DM and SHs require experts with expertise in the specific areas where consequences will be experienced to help them understand how desirable particular characteristics of a decision option are. These organizational relationships between these large groups of individuals are shown in Figure 5-1.
The choice of lunar payload handling system is a decision problem that is both sufficient in scope and complexity to require participation from all the groups in Figure 5-1 and it may benefit from using the ADP. As an example of this dynamic, consider one element important to the decision, the mass of the handling system. From experience, the DM knows that mass is an important element to consider, but some of the most important consequences of the system’s mass are the constraints it places on the lunar lander. The lander is a portion of the Constellation architecture not directly under LSS and therefore not directly under the control of the DM. A suitably senior member of the lander team must therefore be a SH in this decision to help the DM understand the relative importance of mass in the context of all other elements he must consider.

While the lander SH can provide this broad context, he or she is likely unable to say exactly how much more desirable a handling system that weights $x$ kg is than one that weights $2x$ kg. For this information, an expert intimately involved in lander mass allocation is required; this is the role of the consequence expert. Finally, the DM must know the mass of each of the handling options and any associated uncertainty. As the DM will not perform these calculations himself, option experts are required who have studied various systems.

To appropriately synthesize this information for all of the elements relevant to payload handling, a structured approach is required. The ADP methodology does just this. In the analysis portion of the ADP, information is collected and aggregated using the principles of multi-attribute utility theory to provide a preliminary ranking of options. During the deliberation, SHs and the DM review the analysis to insure it is consistent and has correctly captured their preferences. The DM then makes a decision with the collective input of the SHs.
The ADP is facilitated by an analyst who either has some experience in decision analysis or who has learned to use the ADP over the course of its application to a prior decision problem. The ADP has been implemented to address a number of decision problems similar to that of payload handling including problems in environmental cleanup [38], water distribution reliability [39] and NASA project planning [40]. ADP is designed to be simple to use and intuitive to participants so that an expert in decision theory is not required for each application.

5.2.1 Steps in Implementing the large group ADP

Phase One
The first step of the ADP is to clearly identify the DM, define the decision problem and establish the context in which the decision is to be made. Here, the DM is the manager of Lunar Surface Systems. The decision problem is to select a payload handling system or systems to be used on the lunar surface. The context is provided by a representative campaign plan [37].

The DM now selects SHs based on his or her preliminary assessment of the elements important to the decision. As the process evolves, more stakeholders may be added as required. In principle, the number of SHs need not be limited, but in practice, holding to approximately six or fewer SHs helps to keep the analysis tractable for a single analyst. As a result, each SH may not represent just his or her own views, but those of an entire segment of the organization.

The analyst then meets with the DM and each SH, either individually or as a group, to identify all of the elements that each individual believes are important to consider in evaluating the decision options. This information is captured by forming an Objectives Hierarchy, which is shown schematically in Figure 5-2. At the top of the hierarchy is the goal, a broad statement intended to communicate the overall purpose for making the decision. It reiterates the context in which SHs will determine what other elements belong in the objectives hierarchy. All SHs and the DM must agree on the goal.

Objectives are the second tier in the hierarchy. They are the broad categories of elements that the SH feels must be achieved in order for a decision option to meet the goal. These broad objectives may be further divided into sub-objectives as needed. Objectives for payload handling might include maintaining safety and maximizing the simplicity of using the system.
Below objectives are attributes. Attributes are the largest set of elements a SH is indifferent between and describe how to achieve the objective they lie below. It is helpful to think of attributes as the most detailed level of sub-objective the DM or SH wishes to consider. As an example, a SH may determine that minimizing the dimensional constraints on payload is important. If the SH is indifferent between constraints on the height, length and width then dimensional constraints is an attribute. If not, then dimensional constraints is an objective or sub-objective and the individual constraints on each dimension are attributes.

With this input from the DM and the SHs, the analyst will attempt to create a consensus hierarchy and prepare a set of definitions for each objective and attribute. While the DM and SHs need not agree on the structure of the hierarchy, it greatly simplifies analysis if consensus can be reached.

**Phase Two**
The analyst now takes the list of attributes identified by the DM and the SHs and passes them to the experts. These experts have two tasks. First to develop quantifiable performance measures (QPMs) so that the extent to which an option satisfies an attribute can be specified and second to develop decision options and report the level of performance of each with associated uncertainty. QPMs make up the very last tier of the Objectives Hierarchy in Figure 5-2.

Developing QPMs starts with the consequence experts. These individuals examine the attributes and determine a set of appropriate metrics to measure each. For example, an attribute important to payload handling systems might be to minimize the mass of the system. A consequence expert determines how mass should be measured. He or she might determine that the appropriate metrics are the largest single mass on a lunar lander and the cumulative mass brought to the lunar surface, measured in kilograms. These metrics will define two QPMs.
Option experts take metrics and determine the range of consequences any reasonable decision option might have, and the performance level of particular decision options. They formulate decision options using their experience or by conducting trade space searches. For each option, they will have a set of models to determine its level of performance. These models are quantitative but it is expected that engineering judgment will play a key role. Continuing the example, the largest mass of a payload handling system on a single lander might range between 0 kg and 5,000 kg with a crane being 700 ± 100 kg and a davit being 200 ± 50 kg. In determining the range of all reasonable options it should be the case that, if a new option is proposed in the future, its performance falls within the established range.

Consequence experts now examine the range of performance levels for each metric and determine the relative desirability of different values in the range. This information is captured in a value function that takes on values between zero, for the least desirable consequence, and one, for the most desirable. The range of performance levels and the corresponding values form a constructed scale, as shown in Table 5-1. The constructed scale can be continuous, with a unique desirability value for every possible performance level, or discrete, as in Table 5-1, with one value corresponding to a range of possible levels. Constructed scales allow any metric to be measured in terms of a common unit and they capture risk aversion to different levels of performance.

<table>
<thead>
<tr>
<th>Performance Level</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 49 kg</td>
<td>1</td>
</tr>
<tr>
<td>50 - 499 kg</td>
<td>0.9</td>
</tr>
<tr>
<td>500 - 4,999 kg</td>
<td>0.5</td>
</tr>
<tr>
<td>5,000 kg</td>
<td>0</td>
</tr>
</tbody>
</table>

A metric and its constructed scale form a QPM. QPMs can be based on quantitative metrics, such as a number of kg, or qualitative ones, such as a subjective understanding of a degree of complexity. They must, however, be metrics for which a constructed scale can be developed. In cases where more than one QPM is used to evaluate a single attribute, the consequence expert determines how the QPMs should be weighted to lead to a single score for the attribute. Depending on the particular attribute being considered, a single individual may be qualified to act as both consequence and option expert.

**Phase Three**

Once the experts have provided QPMs for each attribute, the DM and the SHs must determine how relatively important each attribute is to achieving the overall goal. To capture these preferences, a pairwise comparison process is used. This process requires each individual to make a series of comparisons between pairs of attributes, and then objectives, saying which of the pair is more important to achieving the goal and then how much more important.

The constructed scales are critical in providing the necessary context to make these comparisons. In the absence of context, if an individual is asked to compare maintaining
human safety with reducing cost, he or she will likely report that maintaining human safety is extremely more important. The constructed scale, however, may reveal that the maximum consequences to human safety are cuts and scrapes, while the maximum consequences to cost are a many millions of dollars budget overrun. With this context the individual may weigh the two attributes more equally.

Results from pairwise comparisons lead to a series of person-specific weights for the attributes. Consider two stakeholders and the objectives Ensure Affordability and Ensure Technical Success. In context, SH 1 believes affordably and technical success are equally important while SH 2 believes that technical success is twice as important as affordability. The pairwise comparison process would result is the weights shown in Table 5-2. As these weights reveal fundamental differences in the way individuals perceive a decision problem, no attempt it made to reach consensus weights at this stage.

Table 5-2: The Weighting Process

<table>
<thead>
<tr>
<th></th>
<th>SH 1</th>
<th>SH 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affordability</td>
<td>0.5</td>
<td>0.25</td>
</tr>
<tr>
<td>Technical Success</td>
<td>0.5</td>
<td>0.75</td>
</tr>
</tbody>
</table>

With all of this information collected, a Performance Index (PI) can be determined for each of the decision options. Exactly as in Section 3.1, the PI for alternative \( j \) is defined as the sum of the values, \( v_{ij} \), associated with the QPMs for attribute \( i \) weighted by the weight for that attribute, \( w_i \).

\[
PI_j = \sum_{i=1}^{N_{QPM}} w_i v_{ij}
\]

Again, the distribution of the PI is calculated separately for the DM and each SH. The decision options can then be ranked according to their expected PIs and the effect of performance uncertainty can be shown.

With the calculation of the PIs, the analysis portion of the ADP ends and deliberation begins. The DM and the SHs each review their individual PIs to understand how the current state of knowledge about the decision options and their individual preferences for the attributes affect the decision problem. Individuals can now discuss the similarities and differences between their rankings in order to reach a collective decision.

Results of ADP
It is important to reiterate that the ADP does not produce one best decision, rather it is designed to separate out the components of the decision making process so that the DM and SHs can reach consensus. The ADP clearly separates the issue of uncertainty in the performance of a decision alternative from variation in the preferences of individuals. The ADP is typically used to show each stakeholder how his or her rankings of alternatives change if preferences are changed or if postulated option performance
changes. The DM and stakeholders can then focus their efforts around only those issues that have high impact on the decision. They might decide to conduct additional modeling to better understand option performance and reduce uncertainty, they might reconcile their preferences or they might find an obvious optimal decision.

5.3 The Role of Experts

In the ADP, experts are individuals who have intimate knowledge of a particular piece of the overall decision problem. In contrast to stakeholders, their primary focus is to understand this piece as well as possible, not to place it in the broader context of all other elements important to the problem. Two types of experts are required: decision option experts (OEs) and decision consequence experts (CEs).

OEs are responsible for identifying decision options and assessing their performance based on provided metrics. CEs understand the repercussions that a decision might have and are responsible for developing metrics to measure the level of consequences associated with different decision options.

In theory, at least one individual must be designated as an OE for each decision option and at least one individual must be designated as a CE for each attribute specified by the stakeholders (SHs). In practice, a single individual may take on multiple roles, or a group of people may combine their knowledge to fulfill a single role.

Together these experts develop quantifiable performance measures (QPMs) and identify and assess the performance of options.

5.3.1 Formulating Quantifiable Performance Measures

Recall that in the ADP, it is the responsibility of the SHs to develop the objectives hierarchy down to the level of attributes as shown in Figure 5-2. Attributes are the most detailed level of elements the SHs wish to consider. It is the responsibility of the experts to develop the final tier, the QPMs, that will be used to determine the effects that a given decision options has on the attributes. Each attribute must be associated with one or more QPMs and a single QPM can be associated with more than one attribute when appropriate.

The process of formulating the QPMs requires three steps. First, measurable consequences must be identified. These are characteristics of the performance of the decision options. Second, the range of these consequences, or performance levels, must be specified. Third, the value of achieving any level of performance must be determined.
5.3.2 Identifying Measurable Consequences

It is the responsibility of the CEs to consider each attribute and to determine what the measurable consequences of designing, developing and implementing payload-handling systems are. These will determine the extent to which a decision option affects an attribute. As an example, the attribute, maintain human safety, might be measured by a QPM based on the consequence number of injuries to astronauts.

It is important for the CEs to communicate with the OEs when defining the measurable consequences, as the OEs have an understanding of the type and amount of information that is available to characterize decision options. Determining the number of injuries to astronauts would require a detailed study of the operations of a payload handling system. If this information is not available, a surrogate metric is needed. For this case, one might be, the amount of time astronauts interact with the handling system.

5.3.3 Specifying the Range of Consequences

Once the measurable consequences are defined, the OEs must determine their range. The range should be large enough to encompass the performance levels of all current decision options, as well as all options that might be considered in the future. The intent is that the range is large enough so that it will not need to be redefined to accommodate future options. At the same time, the range should not be so large such that all decision options end up clustered in one small part making them hard to distinguish from one another.

The OE may specify the range in two ways. One way is as a set of continuous numbers between two values. This is appropriate when a consequence is easily measured on a continuous quantitative scale, such as a mass between, say, 0 and 1,000 kg. The second way is as a set of discrete descriptive categories. This is used when the consequence is more subjective, and not easily paired with a quantitative scale. It is up to the OE to choose the number of discrete categories to use, but there must be at least three, corresponding to minimum, intermediate and maximum consequence.

5.3.4 Determine the Value of a Level of Consequence

Once the range of consequences has been specified, the CEs must determine how important or how desirable a particular level is. This is done by creating a value function, a numerical scale between zero and one, where zero corresponds to the least desired consequence and one to the most desired.

If the range of consequences is continuous, then the value function is determined using the Bisection method. To begin, the CE examines the range provided by the OE to determine the level of consequences, $C_0$ and $C_1$, which should correspond to the values zero and one. $C_0$ and $C_1$ may correspond to the extremes provided by the OE, or they might be interior to these extremes. If the CE finds a portion of the range completely unacceptable then $C_0$ is the consequence at the interior bound of the unacceptable region.
Similarly, if the CE finds a point in the range beyond which no additional value is obtained, then $C_1$ corresponds to this point.

The CE next determines a midpoint, $m_1$, such that he or she is indifferent between the value difference $v(m_1) - v(C_0)$ and $v(C_1) - v(m_1)$. The consequence level $m_1$ now corresponds to a value of $\frac{1}{2}$. Midpoints $m_2$ and $m_3$ are placed such that $v(m_2) = \frac{1}{4}$ and $v(m_3) = \frac{3}{4}$. Additional midpoints are added until the CE’s value function is specified to the desired accuracy. A continuous value function is shown schematically in Figure 5-3.

![Continuous Value Function](image)

If the range of consequences is discrete, then the CE uses a Ratio method to determine the value function. The CE begins again by determining $C_0$ and $C_1$, the categories to which values of 1 and 0 should be assigned. Then each category is assigned a number of points between 0 and 100 to indicate how desirable it is. Using these points, a value is derived for each category as shown in Table 5-3.
Table 5-3: Discrete QPM

<table>
<thead>
<tr>
<th>Level of crew interaction required</th>
<th>Performance Level</th>
<th>Points</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No crew interaction required</td>
<td>100</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Minor crew interaction required</td>
<td>30</td>
<td>0.222</td>
<td></td>
</tr>
<tr>
<td>Significant crew interaction</td>
<td>10</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

The value of the interior category is calculated using linear interpolation as follows for points, \( p \), value, \( v \), and constants \( a \) and \( b \): 

\[
ap + b = v
\]

\[
100a + b = 1 \
10a + b = 0 
\Rightarrow 
\]

\[
a = \frac{1}{90} 
\]

\[
b = -\frac{1}{9} 
\]

\[
\frac{1}{90}(30) - \frac{1}{9} = \frac{2}{9} = 0.222
\]

While the methods described here for obtaining value functions are quite simple, the process requires vigilance on the part of the CE. Each value judgment must be justified as rigorously as available information will permit.

5.3.5 Formulating and Scoring Decision Options

In the development of the objectives hierarchy for this application of ADP, it has been assumed that only decision options that can accomplish a set of minimum payload handling functions will be considered. That is, the hierarchy is not set up to evaluate options that don’t actually “work”.

In this context, a decision option is a collection of devices and/or operations, as well as any supporting materials, that can be used to accomplish all the primary payload-handling tasks described in the decision scenario document. As an example, a ramp, by itself, does not define a complete decision option because it cannot accomplish payload handling without assistance. A complete decision option that includes a ramp would need to specify a concept of operations and all other equipment that would be needed to successfully move payload.
Formulating these complete decision options is the responsibility of the OEs. Once formulated, the OEs must score them against each QPM. To do so, an OE must determine the level of performance the option can achieve. As the performance is uncertain, the OE should provide as much information about the distribution of possible performance as possible. At a minimum, this distribution should be a statement of the lowest, highest and most likely level of performance.

5.4 Lunar Payload-Handling

To execute the ADP process described in the previous sections, stakeholders and experts were engaged from the NASA Goddard Space Flight Center (GSFC), the NASA Johnson Space Center (JSC), the NASA Langley Research Center (LRC) and the NASA Jet Propulsion Laboratory (JPL) to address the decision problem of selecting systems and technologies to manipulate payloads on the lunar surface to support human exploration.

To familiarize these participants with the RIDM process they were briefed by the analyst, this author, and provided with the document included in Appendix IV - RIDM Primer. In addition to describing the ADP, this process provided background information and discussion questions for the participants to consider to help them frame the decision problem and it described to tools that are used to collect the information that the participants provide.

Iteratively over about a half dozen interactions with the analyst, this group formed a consensus objectives hierarchy to describe the decision problem. The top level of this hierarchy is shown in Figure 5-4.

The group decided that the goal of the decision-making process should be to optimize the net capabilities of the payload handling system as an integrated part of the Lunar Surface Systems (LSS) architecture over the course of the lunar campaign as defined by the SARD. Net capabilities are defined as the total of functions available for lifting, handling and mobility considering the resources, complexities and hazards the system involves to design, deploy and operate. Programmatic issues such as cost and impact on schedule would be considered to the extent determined by upper level management not directly participating in the decision process.
To meet this goal, six primary objectives were identified. These were:

1. Minimize Constraints on Payload and Lander due to Characteristics of the Handling System - The manner in which payload handling is accomplished may impose constraints (mass, volume, location, center of gravity, lifting fixtures etc.) on a given payload or lander. These constraints should be minimized.

2. Optimize Simplicity of Handling System Use - Handling functions and operations should be as simple and as easy as possible. This objective refers to all of the constituent processes, functions, and operations necessary to accomplish payload handling. It does not refer to any contingency use of the system or its ability to perform secondary or alternative functions. It also does not refer to the simplicity of designing, building or testing the handling system.

3. Maximize System Versatility and its ability to conduct Contingency Operations - The payload handling system should have two types of versatility, the versatility to complete handling functions using redundant or alternative means, and the versatility to perform other functions. The first type of versatility refers to the system's ability to respond to a fault or failure (fail-operational) or an unexpected operating environment. The system should have adequate fault tolerance and
ability to conduct contingency workarounds. This versatility does not refer to the system’s likelihood of incurring a fault or defect. The second versatility refers to the adaptability of the system and its ability to perform functions not directly related to the handling of payload around a lander. Of these functions, providing mobility to payloads is likely to be of particular interest.

4. Optimize Intrinsic Characteristics of Handling System - The design of the handling system should optimize required resources both for individual Landers/flights and cumulatively over the course of the lunar campaign. This objective refers to optimizing the intrinsic properties of the system such as its mass, and reliability and also refers to possible external resources required for operation such as power, commanding, and control.

5. Maintain Safety during Payload Handling Operations - An acceptable level of safety during operations needs to be maintained. In keeping with the standard NASA interpretation, safety is defined broadly to include human, environmental and equipment safety. Hazard control features that may be inherent in the design (automated safety monitoring/ “watch-dog”, safety inhibits, or fail-safe) as well as procedural protections, for remote or crew-assisted operations, are considered here. Safety during the design and development of the system are not considered.

6. Minimize Programmatic Resources - The design development and deployment of the handling system should utilize as few programmatic resources as possible. In addition the system design should have the versatility to respond to changing programmatic requirements.

Preparing definitions for these six objectives required care. Not only is it important to specify which characteristics of payload handling are covered by each objective but it is also necessary to clearly indicate which are not covered. This allows the participants to focus their thinking into six distinct categories which helps simplify later discussion about candidate options, but also allows numerical ranking of the system against an orthogonal set of measures.

Next each of the six objectives is further broken down into sub-objectives and attributes as necessary. The constraints objective (1) is divided into three attributes as shown in Figure 5-5.
1.1 Minimize Constraints on Payload Configuration on Lander due to Characteristics of the Handling System - The configuration of the payload on the lander is defined as its horizontal or vertical orientation as well as its physical location on the lander. The capabilities of the handling system may constrain how payload may be oriented, where it may be placed or how it may be stacked on the lander. Such constraints are undesirable to the extent that they preclude the optimal packing of payload on the lander. A handling becomes more desirable as it reduces these constraints.
1.2 Minimize Constraints on Payload Dimensions due to Characteristics of the Handling System - Payload dimensions are the physical length, width and height of the payload. It is assumed that payload is roughly rectangular, spherical or cylindrical in shape and either solid or placed in a solid container. The maximum dimensions of the payload may be limited by the capabilities of the handling system. This is undesirable, as components of the lunar architecture may have to be constrained to a size other than that which would otherwise be optimal. A handling system becomes more desirable as it can accommodate payloads with a broader range of dimensions.

1.3 Minimize Constraints on Payload due to Devices required to attach Payload to Handling System - In order to complete handling, the system will first need to be attached to the payload in some manner. Payload may need to contain certain equipment or be designed to allow certain operations to ensure payload retention/release, and attachment to handling system. Equipment might include lifting fixtures and operations might include manual attachment and/or crew assist. Additional equipment adds mass to the payload and the need for operations adds complexity to payload design. A handling system becomes more desirable as it requires the payload to contain less equipment and permit fewer operations to attach to the system.

Notice that below the three attributes in the figure there are a number of un-numbered boxes. These are potential QPMs that may be used to measure the performance of candidate handling options against the attributes. The participants found it helpful to think about what observable characteristics of an option would be obtainable when they were considering appropriate attributes. While technically the ADP process call for the specification of the attributes before the QPMs, considering both at the same time was beneficial, particularly to those participating in the ADP as experts. At this stage, these are not complete QPMs with constructed scales but rather just indication of the metrics that might later be used to rank the systems.

Similarly, for the remaining five objectives the following sub-objectives and attributes were identified.
2.1 Maximize Ease of Initial Handling System Deployment - When the handling system first arrives on the lunar surface it will need to be deployed. Deployment might include un-stowing, assembling and connecting. With better handling systems, this process is faster and requires little or no human remote or manual operations or monitoring. This attribute is limited to the initial deployment of the handling system and refers only to deployment necessary to accomplish handling functions. All considerations of reusability are captured elsewhere.

2.2 Maximize Ease of Handling System Operations - The operations of the handling system required to accomplish primary functions should not be time consuming and require as little human interaction as possible. This attribute is limited to handling functions. All considerations of reusability and versatility are captured elsewhere.
2.3 Minimize the additional items required to accomplish handling functions - The handling system may not be able to accomplish all handling functions without help from other sources. This might include equipment or function required to detach payload from the lander or to move payload to an acceptable location after it is off the lander. A payload handling system becomes more desirable as it reduces the need for these items.

3. Maximize System Versatility and Contingency Operations

3a. Maximize ability to conduct contingency operations

3a.1 Maximize number and ease of operating modes

3a.2 Maximize range of operational environments

- Fully autonomous operations
- Partially autonomous operations
- Manual operations
- Number of alternative means to accomplish primary functions (#)

3b. Maximize functional versatility

3b.3 Maximize secondary functions

- Secondary function rating scale required
- Time needed to redeploy system for alternative use (time)

3b.4 Maximize reusability

- Availability
- Time needed to redeploy system

Figure 5-7: Versatility and contingency attributes

3a.1 Optimize range of operating modes available to handling system - A handling system will have a primary operating mode, but it could be capable of operating
in any number of modes from completely automatic to unpowered manual in the event of system faults or failures. The system becomes more desirable as it can operate in the more advantageous of these modes in response to failures.

3a.2 Maximize Range of lunar Environments Handling System can operate in - Any handling system will be designed to operate in a range of nominal lunar environments. As a lunar lander may not land in its intended landing zone or the scope of the lunar campaign may change over time, additional versatility is obtained if the system has contingency modes or capabilities that allow it to operate in unexpected environments.

3b.3 Maximize Secondary Functions of Handling System - While a handling system is primarily intended to move payload to and from a lander, many other functions can be envisioned, from outpost construction to logistical support. Similarly, a transportation system that is primarily intended to provide mobility may be used for handling as a secondary function. The ability of a system to perform other functions reduces the cumulative amount of equipment needed on the lunar surface and increases flexibility to respond to contingencies and changes to campaign implementation over time. A system becomes more desirable as it is more versatile and can perform more functions.

3b.4 Maximize Reusability of Handling System - After the system completes its primary handling activities it may be useful in completing other tasks later on in the lunar campaign. In order for the system to be reusable for these additional tasks, it must either remain functional until needed or be able to be restored to a functional state when needed. In addition, the time needed to redeploy the system so that it can perform its new functions must be reasonable. Systems that have a higher availability, can be operated and stowed (kept in cold stand-by mode) longer without maintenance, and require less time to redeploy are more desirable. This attribute is distinguished from “Maximize Secondary Functions of Handling System” as it does not answer the question can the as designed system perform a particular additional task, but rather, by the time an additional task is required will the system still be in working order to accomplish the task and how hard will it be to set up the system to accomplish that task.
4.1 Minimize the Mass and Volume of the Handling System - As the masses and volumes that can be transported to the lunar surface by a lander are likely to be highly constrained, it is desirable to minimize the mass and volume of the handling system. Both the largest mass and volume on a single lander as well as the cumulative mass and volume sent to the surface over the lunar campaign should be minimized. More desirable systems are lighter and more compact.

4.2 Maximize Ease of Maintenance and Repair of the Handling System - The handling system will likely be needed for operations on the lunar surface for a period of a number of years. It is expected that over this time the system will require some form of maintenance and repair. It is desirable for the system to be easy to inspect to identify failed components or insure it is in functional condition and for it to be easy to repair or replace failed components.

4.3 Maximize the Robustness of the Handling System - It is desirable for the handling system to be as reliable as possible. A more desirable system is resistant to faults.
and failures, avoids undue complexity in moving parts or control systems and can be completely tested prior to launch. This attribute refers only to robustness in performing primary functions not to subsequent availability to conduct secondary functions.

4.4 Minimize the Energy Requirements of the Handling System - The energy that the handling system requires should be minimized. The available power at any lunar outpost is likely to be limited with constraints being particularly severe at the beginning of a campaign. The handling system should both minimize the power it draws and the total energy it consumes.

5.1 Maintain Human Safety - Maintain EVA crew safety by eliminating or mitigating potential hazards resulting from the operations of or proximity to offloading. The offloading approach should minimize unique hazards and prioritize safety controls applying the hazard reduction precedence.

5.2 Maintain Environmental Safety - Maintain environmental safety by eliminating or mitigating potential hazards materials or fluid releases resulting from the operations of or proximity to offloading. The offloading approach should
minimize unique environmental hazards and prioritize safety controls (for example: spill cleanup) applying the hazard reduction precedence.

5.3 Maintain Equipment Safety - Maintain equipment safety by eliminating or mitigating potential faults or failures leading to inadvertent, unintended, or uncontrolled operations resulting in collisions or impacts. The offloading approach should minimize unique equipment hazards and prioritize safety controls applying the hazard reduction precedence and maximizing approaches that are “must-work safety critical”, “fail-operational” or “fail-safe”.

Figure 5-10: Programmatic resources attributes

6.1 Minimize Costs - Ensure design and development costs for offloading within constraints established by the use of medium/high technology maturity. Maximize system operational life/reliability and optimize system availability/maintainability.
6.2 Minimize Impact on Schedule - Ensure design and development schedule for offloading within constraints established by the use of medium/high technology maturity with good fabrication/produclibility characteristics.

6.3 Versatility of the System Design to Adapt to changes in Program Requirements - It is reasonable to expect there will be some changes to the lunar program after initial design of the handling system is completed. These changes may affect the required capabilities of the handling system. A handling system is more desirable as its design permits easy accommodation of the requirements most likely to be changed.

5.5 Conclusions

At the time of this writing, after the completion of the objectives hierarchy, the constellation program has been cancelled and so to further development of the QPMs, rankings of the attributes or options is possible. Nonetheless, the lessons learned in the process of developing the hierarchy are valuable.

Establishment of clear rolls and responsibilities for expert and stakeholders was found to be very beneficial. When performing ADP with a small group of individuals there is substantial focus on the deliberative nature of the process. Each aspect of the decision problem is discussed with all participants so that consensus, or at least a clear understand of opposing viewpoints, is established at every stage. This type of deliberation is not possible when dealing with a large group as the group is too difficult to physically assemble and not all individuals understand the overall context of the decision problem. Information must be compartmentalized and summarized. Deliberations around a single attribute or QPM will occur between a small number of stakeholders and experts but top-level stakeholders only deliberate the full body of information. If the small group formulation of ADP is presented to a large group of participants, as was originally attempted, the process seems overly confusing, time consuming and as a result individuals are unwilling to participate. Being able to clearly delineate the roll and scope of each person’s involvement in the process makes him more willing to participate and take ownership for his piece.

Aside from understanding their respective roll in the ADP process, participants must appreciate the context in which the decision is being made. For a problem such a lunar payload handling the context in which handling will take place is extremely important in thinking about how the capabilities of a handling system should be judged. Lunar payload handling can be envisioned in a very wide range of contexts from a one-time use system that must move one payload from a lander to the surface of the moon and then never be used again. To semiautonomous robotic systems that must travel around a lunar base assisting in base construction and maintenance, lander unloading, etc. When the development of payload handling system begins the context is not yet clear, as the lunar campaigns have not yet been designed.

For stakeholders and experts thinking about payload handling systems, the question then is whether to develop a single ADP framework to judge systems whatever the context
ends up being or to develop a framework for several candidate contexts. The advantage of the former being reduced effort and the latter being better specificity. What has been discovered during the course of this case study is that a single objectives hierarchy can be developed that is inclusive of all foreseeable contexts. Then the pairwise comparison process would be used in light of the specific context to weight attributes accordingly. As an example, the attributes related to the payload handling objective versatility (3) would have little weight if the final context called for single use systems but would have much greater weight if the context called for a multifunctional reusable system.
6 Uncertainty Propagation using Orthogonal Polynomials

The use of mechanistic models in the study of physical phenomena is ubiquitous in many fields of science and engineering. It is often the case that the input parameters to these models are uncertain and that this uncertainty may affect the decisions one may wish to base on the model output. Not surprisingly then, methods to treat input uncertainties have received substantial attention. These methods are broadly referred to as methods of uncertainty propagation.

The method one chooses to use in conducting uncertainty propagation depends on the information required to make any applicable decisions. In general there are three types or grades of information that may be relevant. In order of increasing complexity these are:

1. Knowledge of the mean and variance of the output. This provides an understanding of the central tendency of the output distribution and therefore the likely behavior of the system. Methods used to determine these are referred to as output variability methods and include the perturbation method [42].

2. Knowledge of the probability of exceeding a critical threshold. This provides an understanding of the likelihood the system will fail or operate beyond prescribed bounds. Methods in this area are referred to as reliability methods and include the well-known FORM/SORM [43] techniques from structural reliability analysis.

3. Knowledge of the complete output probability density function (PDF). This, of course, provides an understanding of the entire range of possible output and is most useful when decisions involve comparing the relative goodness of alternatives rather than acceptance relative to hard constraints. Monte Carlo simulation is the most common method used here.

Focusing on the third category, where the entire PDF is of interest, we find that, while Monte Carlo simulation is a very powerful technique, it requires many samples from the mechanistic model and so is not a practical tool in cases where the model is computationally intensive. Alternate techniques have been proposed involving response surfaces[44] and other such meta-models in combination with Monte Carlo simulation. In these methods, the computationally intensive physical model is replaced with a computationally inexpensive surrogate on which Monte Carlo simulation may be performed. These methods are primarily limited by the quality of the meta-model. Langewisch [45] provides an excellent review of the relevant issues.

More recently, Ghanem and Spanos [46], working in the area of stochastic finite element analysis, have proposed methods based on a spectral representation of the output. In this
technique the response is expanded onto an orthogonal polynomial basis of the input probability space. It is these expansions\(^3\) that will be the focus of this paper.

We begin by briefly reviewing the theory of orthogonal polynomials. We then present the theory of uncertainty propagation via orthogonal polynomials focusing first on the original presentations that limited propagations to very specific types of input uncertainty. We then present a computationally simple method for extending the original methods to arbitrary input uncertainties. We examine the performance of these techniques in propagating the uncertainty in a thermohydraulics model of a nuclear reactor system and compare these results to response surface meta-models. We conclude with several general observations.

### 6.1 Orthogonal Polynomials

Over the years there has been significant study of orthogonal polynomials and there are now a number of texts that detail their properties [47][48][49]. Let us begin by reviewing a few of these that will be germane to our discussion.

#### 6.1.1 Definition of Orthogonal Polynomials

Orthogonally is defined with respect to the inner product which, in turn, requires a weight function, \(w\). A continuous weight function is defined as,

\[
 w(x) \text{ on } [a,b], \quad -\infty \leq a \leq b \leq \infty
\]  

(16)

Similarly, a discrete weight function is defined as,

\[
 w_N(x) = \sum_{i=1}^{N} w_i \delta(x - x_i), \quad x_1 < x_2 < \ldots < x_N
\]  

(17)

where \(\delta\) is the Dirac delta function and the points \(x_i\) form the support of \(w\).

The weight function, which is usually positive over an interval \((a,b)\), will be of particular interest to us, as it will be shown later that, for certain polynomials, the weight function is identical to particular probability distributions.

Let us make the usual assumption that for continuous \(w\) all moments,

\[
 \mu_k = \int_a^b x^k w(x) dx, \quad k = 0, 1, 2, \ldots
\]  

(18)

---

\(^3\) The use of orthogonal polynomials in this manner is sometimes referred to as a polynomial chaos [52]. The usage of chaos in polynomial chaos should not be confused with the more common usage in chaos theory. The term polynomial chaos was coined before the advent of chaos theory when “chaos” referred to any representation of stochastic phenomena.
exist and are finite. Then, the inner product of any two polynomials \( p \) and \( q \) with respect to a weight function \( w(x) \) is well defined by

\[
(p,q)_w = \int_{\mathbb{R}} p(x)q(x)w(x)\,dx
\]

Orthogonal polynomials relative to the weight function \( w(x) \) are defined as follows:

\[
(p,q)_w = \begin{cases} 
0, & k \neq l, \\
>0, & k = l.
\end{cases}
\]

(19)

(20)

Where \( \Psi_k(x) \) is an orthogonal polynomial of degree \( k, k=0,1,2,... \)

For continuous \( w(x) \) there are an infinite number of such polynomials while for discrete \( w_N(x) \) there are exactly \( N \) orthogonal polynomials of order \( k = 0,1,...,N-1 \). These polynomials are unique up to the value of their leading coefficient.

6.1.2 Generating Orthogonal Polynomials – Recurrence Relation

A set of orthogonal polynomials can be generated relative to an arbitrary weight function \( w(x) \) using a simple three-term recurrence relation

\[
\begin{align*}
\Psi_{k+1}(x) &= (x - \alpha_k)\Psi_k(x) - \beta_k \Psi_{k-1}(x), \quad k = 0,1,...,n-1, \\
\Psi_{-1}(x) &= 0, \quad \Psi_0(x) = 1
\end{align*}
\]

(21)

where \( \alpha_k = \alpha_k(w(x)) \) and \( \beta_k = \beta_k(w(x)) \) are the recurrence coefficients. It can be shown by using Darboux’s formula [51] that

\[
\begin{align*}
\alpha_k(w) &= \frac{(x\psi_k,\psi_k)_w}{(\psi_k,\psi_k)_w}, \quad k = 0,1,2,..., \\
\beta_k(w) &= \frac{(\psi_k,\psi_k)_w}{(\psi_{k-1},\psi_k-1)_w}, \quad k = 1,2,...
\end{align*}
\]

(22)

as \( \beta_0 \) multiplies \( \Psi_{-1}(x) = 0 \), its value is arbitrary, but it is convenient to define

\[
\beta_0(w) = \int_{\mathbb{R}} w(x)\,dx
\]

(23)

6.2 Uncertainty Propagation Using Orthogonal Polynomials

We will now introduce the theory needed to propagate uncertainties using orthogonal polynomials. Consider any second order random process, \( S(X) \), which is any process with finite variance. As most all physical systems we may wish to model have finite
variance, we may think of $S(X)$ as representing the output of some model with respect to the input random variables $X = (x_1, x_2, ..., x_N)$. We may represent $S$ by the expansion

$$S(X) = \sum_{i=1}^{\infty} S_i \phi_i(x_1, ..., x_N),$$  

(24)

where the set of $\phi_i$ are an appropriate Hilbertian basis of the response space of $S$. Here we wish to find an orthogonal polynomial basis, $\Psi_i$, for the $\phi_i$ so as to permit the efficient computation of the coefficients $S_i$. Noting, of course, that for computational convenience we will truncate this expansion. For an expansion of $N$-dimensional orthogonal polynomials not exceeding degree $k$, the expansion becomes

$$S(X) \approx \sum_{i=1}^{M-1} S_i \Psi_i(x_1, ..., x_N),$$  

(25)

where

$$M = \binom{k + N}{k} = \frac{(k + N)!}{k!N!}$$  

(26)

The appropriate polynomial basis to be used depends on the nature of the vector of the input random variables $X$. The basis is simplest if $x_n$ are independent and distributed according to the same probability distribution and becomes more complex for correlated, differently distributed $x_n$. Let us consider several cases.

### 6.2.1 Independent Gaussian Random Variables

The original formulation of polynomial expansions for stochastic systems was proposed by Wiener [52] in 1938 primarily as a method to solve stochastic differential equations. It assumes that the $x_n$ are identically distributed standard Gaussian random variables with zero mean and unit variance. The appropriate polynomial basis for this $X$ is one that is orthogonal to the weight function

$$w(X)dx = \frac{1}{\sqrt{(2\pi)^n}} e^{-\frac{1}{2}x^T x} dx$$  

(27)

where the weight function is the standard multivariate Gaussian distribution such that a complete orthogonal basis is formed in the space of the standard Gaussian random variable. The polynomials that are orthogonal to this weight function are the well-known Hermite polynomials, $H$. In one dimension, the first five of these are
\[ H_0 = 1 \]
\[ H_1 = x \]
\[ H_2 = x^2 - 1 \]
\[ H_3 = x^3 - 3x \]
\[ H_4 = x^4 - 6x^2 + 3 \]

and then in \( N \) dimensions

\[ H_{\gamma}(X) = \prod_{i=1}^{N} H_{\gamma_i}(x_i), \gamma = \{\gamma_i, i = 1, ..., N\} \] (29)

or in notation more standard in the orthogonal polynomial literature,

\[ H_{\gamma}(x_1, ..., x_{N}) = (-1)^\gamma e^{\frac{1}{2}x^T x} \frac{\partial^\gamma}{\partial x_1 ... \partial x_\gamma} e^{\frac{-1}{2}x^T x} \] (30)

Our second order random process, \( S \), may then be expressed as

\[ S(X) = a_0 H_0 + \sum_{i=1}^{M-1} a_i H_i(x_i) + \sum_{i=1}^{M-1} \sum_{i=1}^{M-1} a_{i,i} H_2(x_i, x_i) + \sum_{i=1}^{M-1} \sum_{i=1}^{M-1} \sum_{j=1}^{M-1} a_{i,j} H_3(x_i, x_i, x_j) + ... \] (31)

According to the Cameron-Martin theorem [53], as the Hermite polynomials form a complete orthogonal basis in the Hilbert space of the Gaussian distribution, this expansion will converge in the \( L_2 \) sense\(^4\) as long as the \( x_i \) are all Gaussian. Once the coefficients \( a \) have been calculated, which we will discuss later, we are able to approximate any physical model whose input uncertainties are represented by standard Gaussian random variables.

While the theory underlying polynomial expansions using the Hermite basis is useful, the method is practically limited as it constraints the input uncertainties to independent standard Gaussian random variables. To propagate the uncertainty in models where the input is not independent standard Gaussian, we propose two alternatives. In cases where the input random variables are independent and distributed according to the same non-Gaussian probability distribution, we may search for an alternative polynomial basis around which to perform the expansion. In cases where the input random variables are dependent and/or differently distributed we may use an isoprobabilistic transform to convert to a common distribution. We will consider each of these in turn.

\(^4\) Recall \( L_2 \) convergence implies \( \int_{\mathbb{R}} (S_M - S)^2 dX \to 0 \) where \( S_M \) is the approximation of the true \( S \), generated by \( M \) samples of the underlying model.
6.2.2 Independently distributed random variables

As discussed in Section 6.1.1, it is in general possible to develop a family of orthogonal polynomials for any weight function, \( w(x) \), and thereby to create an orthogonal polynomial expansion for any physical model with independently distributed uncertainties in exactly the same manner as with the Hermite basis. In practice, computation of the recursion coefficients maybe nontrivial for an arbitrary probability distribution. There is, however, a well understood scheme of orthogonal polynomials, known as the hypergeometric polynomials, whose weight functions are identical to certain probability distributions that are of practical interest in the modeling of physical uncertainties. When presented with a physical model whose input uncertainties match one of the weight functions from the hypergeometric scheme it is possible to easily develop an appropriate orthogonal basis. To demonstrate, we will briefly discuss the hypergeometric polynomials and then describe the construction of the orthogonal basis.

The Hypergeometric Orthogonal Polynomials

The hypergeometric scheme of orthogonal polynomials is based on the hypergeometric series and can be defined as follows [54]:

Let \( (\cdot)_n \), referred to as the rising factorial or Pochhammer symbol, be defined as

\[
(\cdot)_n = \begin{cases} 
1 & n = 0 \\
\frac{\Gamma(\cdot + n)}{\Gamma(\cdot)} & n > 0 
\end{cases}
\] (32)

Then the hypergeometric series \( \,_{p}F_{q} \) is defined by

\[
_{p}F_{q}(a_1, \ldots, a_p; b_1, \ldots, b_q; x) = \sum_{n=0}^{\infty} \frac{(a_1)_n \cdots (a_p)_n x^n}{(b_1)_n \cdots (b_q)_n n!} 
\] (33)

Common hypergeometric series that the reader will recognize include the \( _0F_0 \) exponential series and the \( _1F_0 \) binomial series.

If one of the \( a_i \) is a negative integer, i.e. \( a_i = -k \), then the series ends at the \( k \)th term and becomes a \( k \) order polynomial in \( x \).

\[
_{p}F_{q}(-k, \ldots, a_p; b_1, \ldots, b_q; x) = \sum_{n=0}^{k} \frac{(-k)_n \cdots (a_p)_n x^n}{(b_1)_n \cdots (b_q)_n n!} 
\] (34)

It is these polynomials that are of interest in forming an orthogonal. Four families of hypergeometric polynomials have weight functions that are identical, to within a normalization constant, to nine probability distributions that are of practical use in the
modeling of physical systems [55]. Table 6-1 presents these polynomials with their weight functions, their corresponding probability distributions as well as supports. Table 6-2 shows the recurrence relations for these polynomials using the notation of Table 6-1.

Table 6-1: Several hypergeometric polynomials and their corresponding weight functions that correspond to useful probability distributions. \( k \) refers to the degree of the polynomial, \( x \) is the random variable, all other symbols are standard nomenclature from engineering applications of probability theory.

<table>
<thead>
<tr>
<th>Definition</th>
<th>Weight Function</th>
<th>Probability Distributions</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Discrete Functions</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charlier Polynomials</td>
<td>( F_0(-k,-x;\frac{1}{\lambda};x) ) ( \frac{(\lambda t)^x}{x!}e^{-\lambda t} )</td>
<td>Poisson</td>
<td>{0,1,2,...}</td>
</tr>
<tr>
<td>Krawtchouk Polynomials</td>
<td>( F_1(-k,-x;\frac{1}{p};-N) ) ( N \choose x p^x(1-p)^{N-x} )</td>
<td>Binomial</td>
<td>{0,1,2,...,N}</td>
</tr>
<tr>
<td>Meixner Polynomials</td>
<td>( F_2(-k,-x;\frac{1}{p};-m,\frac{1}{N};-N) ) ( (x+r-1) \choose (r-1) p^r(1-p)^{r-1} )</td>
<td>Negative Binomial</td>
<td>{0,1,2,...}</td>
</tr>
<tr>
<td>Hahn Polynomials</td>
<td>( F_2(-k,-x;\frac{1}{p};-m,\frac{1}{N};-N) ) ( m \choose N-x \frac{N}{x} \frac{T}{T} )</td>
<td>Hypergeometric</td>
<td>{0,1,...,N}</td>
</tr>
<tr>
<td><strong>Continuous Functions</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hermite Polynomials</td>
<td>( F_0(-k;\frac{1}{2},\frac{k-1}{2};\frac{1}{x^2})x^k ) ( \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} )</td>
<td>Normal / Gaussian</td>
<td>((-\infty,\infty))</td>
</tr>
<tr>
<td>Laguerre Polynomials</td>
<td>( F_1(-k;\alpha;x) ) ( \frac{1}{k!} x^{-1} e^{\frac{x}{\beta}} )</td>
<td>Gamma</td>
<td>([0,\infty))</td>
</tr>
<tr>
<td>Jacobi Polynomials</td>
<td>( F_1(-k;\alpha+\beta;\frac{1-x}{2}) ) ( \frac{1}{k!} (1-x)^{\alpha-1}(1+x)^{\beta} )</td>
<td>Exponential ((\alpha=1))</td>
<td>([0,\infty))</td>
</tr>
<tr>
<td>Legendre Polynomials</td>
<td>( F_1(-k;\frac{1-x}{2}) ) ( \frac{1}{2} )</td>
<td>Beta</td>
<td>((0,1))</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Uniform</td>
<td>((0,1))</td>
</tr>
</tbody>
</table>
Table 6-2: Several hypergeometric polynomials and their corresponding recurrence relations.

<table>
<thead>
<tr>
<th>Polynomial</th>
<th>Recurrence Relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Charlier</td>
<td>( C_k(x; \lambda t) = \frac{\lambda t C_{k+1}(x; \lambda t) + kC_{k-1}(x; \lambda t)}{k + \lambda t - x} )</td>
</tr>
<tr>
<td>Krawtchouk</td>
<td>( K_k(x; N, p) = \frac{p(N-k)K_{k+1}(x; N, p) + k(1-p)K_{k-1}(x; N, p)}{p(N-k) + k(1-p) - x} )</td>
</tr>
<tr>
<td>Meixner</td>
<td>( M_k(x; r, p) = \frac{p(r+k)M_{k+1}(x; r, p) + kM_{k-1}(x; r, p)}{k + (k + r + x)p - x} )</td>
</tr>
<tr>
<td>Hahn</td>
<td>( Q_k(x) = \frac{\alpha_k Q_{k+1}(x) + \beta_k Q_{k-1}(x)}{\alpha_k + \beta_k - x} ) ; ( \alpha_k = \frac{(k-T-1)(k-m)(N-k)}{(2k-T-1)(2k-T)} ) ; ( \beta_k = \frac{k(k-T+1+N)(k+m-T-1)}{(2k-T-1)(2k-T-2)} )</td>
</tr>
<tr>
<td>Hermite</td>
<td>( H_k(x) = \frac{H_{k+1}(x) + kH_{k-1}(x)}{x} )</td>
</tr>
<tr>
<td>Laguerre</td>
<td>( L_k(x) = \frac{(k+1)L_{k+1}(x) + (k+\alpha)L_{k-1}(x)}{(2k+\alpha+1-x)} )</td>
</tr>
<tr>
<td>Jacobi</td>
<td>( P_k(x) = \frac{2(k+1)(k+\gamma+1)(2k+\gamma)P_{k+1}(x) + 2(k+\alpha)(k+\beta)(2k+\gamma+2)P_{k-1}(x)}{(2k+\gamma+1)(\alpha^2 + \beta^2) + (2k+\gamma)x} ) ; ( \gamma = \alpha + \beta )</td>
</tr>
<tr>
<td>Legendre</td>
<td>( P_k(x) = \frac{(k+1)P_{k+1} + kP_{k-1}}{(2k+1)x} )</td>
</tr>
</tbody>
</table>

Hypergeometric Expansion Algorithm
Let us now introduce the hypergeometric generalization to the original Hermite basis. If the output of any physical model, \( S(X) \), containing input uncertainties, \( X \), has independent \( X \) quantified by one of the probability distributions in Table 6-1, then it is possible to approximate \( S(X) \) in terms of an orthogonal set of hypergeometric polynomials, \( \xi(X) \),

\[
S(X) = a_0 \xi_0 + \sum_{k=1}^{M-1} a_k \xi_k(x) + \sum_{i=1}^{M-1} \sum_{j=1}^{M-1} a_{ij} \xi_i(x, x_j) + \sum_{i=1}^{M-1} \sum_{j=1}^{M-1} \sum_{k=1}^{M-1} a_{ijk} \xi_{i,j,k}(x, x_i, x_j) + \ldots \quad (35)
\]

As each family of these polynomials forms a complete orthogonal basis in the Hilbert space of its corresponding probability distribution, this expansion will converge in the \( L_2 \) sense as a generalized consequence of the Cameron-Martin theorem [56].

Notice that the scheme of hypergeometric polynomials contains the family of Hermite polynomials. As such, we may view this hypergeometric basis expansion as a generalization of the Hermite basis expansion, extending the theory to additional probability distributions.
6.2.3 Dependent, differently distributed random variables

To this point we have shown how to construct an orthogonal basis in cases where the input uncertainties, \( X \), that must be propagated to determine the output of our physical model, \( S(X) \), are independently distributed according to one of the probability distributions in Table 6-1. Now we expand the orthogonal basis concept to dependent input uncertainties represented by any well-formed probability distribution.

First let us consider what we typically start with when presented with a physical model that has uncertainties of this type. We rarely know the joint distribution, \( f(X) \), but rather we are presented with a population of data, expert opinion, etc. from which we must approximate \( f(X) \) by assigning an analytical distribution to each of the marginal distributions \( f_{x_i}(x_i) \) and then make some assessment of the linear correlation coefficients, \( r \), between them. As such, we require a method of forming an orthogonal basis using only the marginal distributions and linear correlation coefficients. By adapting the Nataf isoprobabilistic transform [57], we will show that it is possible to do so by using this information to express \( f(X) \) so that each of the \( x_i \) is represented by an independent standard Gaussian distribution with zero mean and unit variance. Such a representation allows us to apply the principles of the simple Hermite basis expansions of Section 6.2.1.

An Appropriate Isoprobabilistic Transformation

For this discussion, we will adopt the notation of Lebrun and Dutfoy [58]. Let us assume that we have been able to construct the marginal distributions, \( f_{x_i}(x_i) \), for the input uncertainties \( X \). These marginal distributions have corresponding cumulative distributions \( F_i \), and for the correlated components of \( X \) we have the positive-definite \(^5\) correlation matrix \( R=(r_{ij}) \) such that

\[
r_{ij} = E \left( \frac{x_i - \mu_i}{\sigma_i} \right) \left( \frac{x_j - \mu_j}{\sigma_j} \right)
\]

where \( \mu_i \) is the mean and \( \sigma_i \) the standard deviation of \( x_i \).

Our isoprobabilistic transformation \( T \) is accomplished by the composition of two functions, \( T_1 \) and \( T_2 \). In the first, we transform the input vector \( X \) into a vector \( Y \) by means of the cumulative distribution function of the standard normal variable, \( \Phi \).

\[
T_1 : X \mapsto Y = \begin{pmatrix}
\Phi^{-1}(F_1(x_1)) \\
\Phi^{-1}(F_2(x_2)) \\
\vdots \\
\Phi^{-1}(F_n(x_n))
\end{pmatrix}
\]

\(^5\) Recall positive-definite implies \( x^T A x > 0 \) for a real matrix \( A \)
\( Y \) is now a Gaussian vector with standard normal marginal distributions. \( Y \) has correlation matrix \( R_0 \) where \( r_{0ij} \) is found by solving the integral equation

\[
\begin{align*}
\rho_{ij} &= E \left[ \left( \frac{F_i^{-1}(\Phi(y_i)) - \mu_i}{\sigma_i} \right) \left( \frac{F_j^{-1}(\Phi(y_j)) - \mu_j}{\sigma_j} \right) \right] \\
&= \frac{1}{\sigma_i \sigma_j} \int_{\mathbb{R}} \left[ \frac{F_i^{-1}(\Phi(y_i)) - \mu_i}{\sigma_i} \right] \left[ \frac{F_j^{-1}(\Phi(y_j)) - \mu_j}{\sigma_j} \right] \omega(y_i, y_j, r_{0ij}) \, dy_i \, dy_j
\end{align*}
\] (38)

where \( \omega \) is the bivariate standard normal PDF

\[
\omega(y_i, y_j, r_{0ij}) = \frac{1}{2\pi \sqrt{1 - r_{0ij}^2}} e^{-\frac{y_i^2 + y_j^2 - 2r_{0ij} y_i y_j}{2(1 - r_{0ij}^2)}}
\] (39)

The second part of the transform involves transforming \( Y \) into a vector \( U \) that is a Gaussian vector with the same marginal distributions as \( Y \) but independent components.

\[
T_2 : Y \mapsto U = \Gamma Y
\] (40)

where \( \Gamma \) is the Cholesky factor of \( R_0^{-1} \). The Cholesky factor [59] is simply a square root operation satisfying

\[
\Gamma^T \Gamma = R_0^{-1}
\] (41)

This transform results in the following relationship between the components of \( X, Y \) and \( U \)

\[
\begin{align*}
y_i &= 0 + L_{R_0}^T u_i \\
x_i &= \mu_i + L_{R}^T y_i
\end{align*}
\] (42)

where \( L_{R_0}^T \) and \( L_{R}^T \) are the \( i \)-th rows of the matrices forming the Cholesky factors of \( R_0 \) and \( R \) respectively. Note in particular the one-to-one correspondence between the values of \( X, Y \) and \( U \).

Before moving on we must caution the reader on the application of this method. First, the computation of \( R_0 \) may require significant care. The integral equation needed to find \( r_{0ij} \) may not converge easily if \( r_{ij} \) is too close to -1 or 1. In fact, a solution will not exist if \( r_{ij} \) is equal to \( \pm 1 \). Second, the matrix \( R_0^{-1} \) may be ill-conditioned leading to numerical instability in the calculation of \( \Gamma \). For these two reasons, it is difficult to develop one simple set of numerical algorithms that will successfully execute the isoprobabilistic...
transformation in all cases. It is often necessary to scrutinize the numerical output at each step of the transformation and adjust numerical methods as needed.

**Isoprobabilistic basis expansion Algorithm**

With the transformation of the input uncertainties vector \( X \) into the independent identically distributed vector \( U \), we may express \( S(X) \) using the Hermite basis expansion

\[
S(X) = a_0 H_0 + \sum_{i=1}^{M-1} a_i H_i(u_i) + \sum_{i=1}^{M-1} \sum_{j=1}^{M-1} a_{ij} H_{ij}(u_i,u_j) + \sum_{i=1}^{M-1} \sum_{j=1}^{M-1} \sum_{k=1}^{M-1} a_{ijk} H_{ijk}(u_i,u_j,u_k) + \ldots
\]  

(43)

As before, this expansion will converge in the \( L_2 \) sense.

**6.3 Calculating the Expansion Coefficients**

Now that we have outlined the three forms of orthogonal polynomial expansion that may be used depending on the nature of the input uncertainties we must take up the very important topic of how to calculate the expansion coefficients. We will focus on the Galerkin methods that have historically been used in computing response surface coefficients [60]. These methods can be divided into the intrusive and the non-intrusive.

The intrusive methods are so named as they typically call for careful modification of the underlying physical model before expansion coefficients may be calculated. They typically offer improved convergence at the cost of additional computational overhead as compared to the non-intrusive methods [61]. While the possibility of improved convergence is attractive, the need to modify each physical model in order to proceed with uncertainty propagation is not practical. Many of the models of interest would make modification prohibitively complex or, in some cases, licensing issues limit access to the model source code making modification impossible. As such we shall avoid further development of the intrusive methods.

Of the non-intrusive methods, the projection method [62] is the most popular and the one preferred by this author. The projection method takes advantage of the orthogonality of the polynomial basis. Recall the general form of the expansion in Eq. (24). By pre-multiplying by \( \Psi_j \) and taking the inner product we obtain

\[
\left( \Psi_j S \right)_w = \left( \Psi_j \left[ \sum_{i=1}^{M} S_i \Psi_i \right] \right)_w
\]

(44)

But orthogonally guarantees \( \left( \Psi_i \Psi_j \right) = 0 \) for all \( i \neq j \) and so we obtain

\[
S_j = \frac{\left( \Psi_j S \right)_w}{\left( \Psi_j \Psi_j \right)_w}
\]

(45)
The denominator in Eq. (45) is known analytically and the numerator is the multidimensional integral

\[
    \left( \Psi_j S \right)_w = \int_\mathbb{R} S(X) \Psi_j(X) w(X) dX
\]  

All that is left is to choose a method to calculate this integral. If the model that must be run to produce the response \( S \) is computationally inexpensive then standard Monte Carlo or Latin hypercube methods may be used. If the model is computationally inexpensive however, then uncertainties could be propagated using these methods and the polynomial expansion would not be required in the first place. In the more likely case that the model is extremely expensive to run, then an effort must be made to evaluate the integral while requiring as few realizations of \( S \) as possible.

This problem of numerically evaluating an integral, called quadrature, is well known in the field of numerical methods and we will not attempt to review the entirety of that work here. We will briefly introduce one of the more popular quadrature methods, Gaussian quadrature, as well as point the reader to a well-contained discussion on the topic by Gautschi [50]. In Gaussian quadrature the integral in Eq. (46) is computed by taking a weighted sum of the integrand evaluated at particular points. In general then, the \( n \) point Gaussian quadrature formula is given by

\[
    \int_\mathbb{R} S(X) w(X) dX = \sum_{i=1}^{M} w_i S(\tilde{t}_i)
\]  

The weight factors, \( w_i \) are the same as in the sense of Eq. (17), they are the weight function for a particular set of orthogonal polynomials, chosen to be orthogonal to \( X \), evaluated at the points \( \tilde{t}_i \). These points, \( \tilde{t}_i \), called the integration points, are chosen as the roots of the \( k \)th order orthogonal function.

To be specific, if the \( X \) in Eq. (46) are standard Gaussian, then the \( \Psi \) are the Hermite polynomials. To evaluate this integral according to Eq. (47), three steps are needed. The \( M \) roots, \( \tilde{t}_1, \ldots, \tilde{t}_M \), of the \( k \)th order \( N \) dimensional Hermite polynomials must be calculated. The corresponding values, \( w_1, \ldots, w_M \), of the \( N \) dimensional Gaussian weight function at those roots must be determined such that all lower order polynomials integrate exactly. The response \( S \) must be calculated at the points \( \tilde{t}_1, \ldots, \tilde{t}_M \). This procedure is said to have maximum algebraic degree of exactitude in the sense that, if \( S \) were a polynomial of degree \( k \) or less, the procedure would give the true value of the integral.

### 6.4 Applications – A Simple Polynomial Model

Let us begin by considering the simple polynomial model

\[
    S(X) = \sum_{i=1}^{N} x_i^2
\]
where the $x_i$ are independent identically distributed random variables. Assume first that the $x_i$ are standard Gaussian, then the appropriate orthogonal basis to apply is the original Hermite basis proposed by Wiener. Let us confirm that the polynomial expansion with the Hermite basis is able to exactly determine the PDF of $S$. Consider Figure 6-1 that compares the reference $S(X)$ for $N=4$ to the $1^{st}$ and $2^{nd}$ order expansions. We note that the reference solution was generated using $10^7$ random samples, not from an analytic expression, but this was adequate to converge all percentiles to $<0.1\%$. Panel A shows that the $1^{st}$ order expansion captures only the mode of $S(X)$ but that the second order expansion reproduces the entire distribution. This is confirmed in Panel B, which presents the frequency error of the 2 expansions relative to the reference solution as a function of percentile. Note that the error for the $2^{nd}$ order expansion is below the $0.1\%$ bounds of the reference solution for all percentiles. This should not be surprising as Eq. (48) is second order. We conclude that the polynomial expansion with the Hermite basis behaves as expected for this simple case.

![Figure 6-1: Polynomial expansions with Hermite basis for a simple polynomial with standard normal uncertainties. Panel A presents the PDF of the reference solution as well and 1st and 2nd order expansions (PC). Panel B presents the error relative to the reference solution for the 1st and second order expansions as a function of percentile of the reference solution.](image)

Again consider the polynomial of Eq. (48) for $N=4$, but now assume that the $x_i$ are uniformly distributed on $[0,1]$. For uniform random variables we apply the Legendre polynomial basis from the hypergeometric scheme. Figure 6-2 contains similar results as before. The expansion correctly reproduces the distribution of $S$ at second order within the error bounds of the reference solution and we conclude that the hypergeometric polynomial bases behaves as expected.
Finally, consider the polynomial of Eq. (48) for $N=4$ for independent but differently distributed $x_i$ as shown in Table 6-3.

Table 6-3: Distributions and associated parameters for the four $x_i$

<table>
<thead>
<tr>
<th>$x_i$</th>
<th>Distribution</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>Beta</td>
<td>$(\alpha = 1; \beta = 1)$</td>
</tr>
<tr>
<td>$x_2$</td>
<td>Gamma</td>
<td>$(\alpha = 5; \beta = \frac{1}{2})$</td>
</tr>
<tr>
<td>$x_3$</td>
<td>Exponential</td>
<td>$(\beta = \frac{1}{2})$</td>
</tr>
<tr>
<td>$x_4$</td>
<td>Lognormal</td>
<td>$(\mu = \frac{8}{10}; \sigma = \frac{1}{10})$</td>
</tr>
</tbody>
</table>

For this $X$, the appropriate isoprobabilistic orthogonal basis, described previously, must be applied. This was done and results appear in Figure 6-3.
As before we notice that the second order expansion matches within the error bounds on the reference solution and this confirms that the expansions based on the isoprobabilistic orthogonal basis behave as expected.

In the three cases of the Hermite basis, the hypergeometric basis, and the isoprobabilistic basis, we have shown that the output uncertainty can be represented exactly then the underlying model is polynomial. While not particularly profound, this confirms that the three formulations are accurate implementations of an orthogonal polynomial expansion. What is more interesting is the computational inexpensive nature of the expansions. While it required $10^7$ Monte Carlo random samples to construct the output distributions for $>0.1\%$ error, (and $10^4$ for $>3\%$ error) it takes only 15 samples to construct the second order polynomial expansion for $N=4$.

6.5 Applications – A Simple Non-Polynomial Model

It is now appropriate to ask how well the polynomial expansions are able to produce the output distribution from a model that is not polynomial. To this end we consider the model of Eq. (49), a simple one dimensional sin wave.

$$S(X) = \sin(x)$$ (49)

If we require $x$ be standard normal then $f(S)$ may be calculated analytically,

$$f(S) = e^{-\frac{\left(\sin^{-1}(s)\right)^2}{2}} \frac{1}{\sqrt{2\pi|1-S^2|}}$$ (50)
For this $x$ we apply the Hermite orthogonal basis and obtain the results in Figure 6-4. The figure shows both the PDF and the error as a function of percentile for the 1st, 3rd and 7th order expansions compared to the exact solution of Eq. (50). The even order expansions are not shown, as they do not provide any additional information for the odd function $\sin$. That is, the 2nd order expansion is the same as the 1st; the 4th is the same as the 3rd, etc. We stop at 7th order as very little improvement in convergence is observed at higher orders. We don’t show the 5th order expansion simply to avoid cluttering the figure, but it lies between the 3rd and 7th order expansions.

![Figure 6-4: Polynomial expansions with Hermite basis for a sin wave. Panel A presents the PDF and panel B the error as a function of percentile.](image_url)

Notice that none of the expansions are able to represent the true $f(S)$ accurately. The reader may wonder if these are actually the closest approximations of $f(S)$ that may be obtained by an orthogonal polynomial expansion. They are not. The limitations in the representation shown are due not to fundamental limitations of the expansion but rather to the limited number of quadrature points used in calculating the expansion coefficients. For example, a 7th order expansion in one dimension requires a minimum of 8 quadrature points to compute the eight coefficients. For each order, this minimum number of points is what was used to generate the results in the figure. It is possible to obtain a better representation of $f(S)$ if we permit a large number of quadrature points and for this $S(X)$ it would be computationally inexpensive to do so. Recall, however, that we are attempting to develop tools to propagate uncertainties for computationally intensive models and so it would be quite misleading to perform a low order expansion of this simple $S(X)$ with a large number of quadrature points and claim the results are indicative of those for a more complex $S(X)$ where such a large number of points would be impossible. If the reader is interested in the results obtained by a large number of quadrature points he may consult the work of Prange et al [63].

Aside from the limitations of the few point expansions, Figure 6-4 also indicates that the representation of $f(S)$ improves very little in going from the 3rd to 7th order expansion. In other words, the expansion converges very quickly. It is interesting to compare this behavior to that of typical response surface methodologies. Figure 6-5 presents results from a 1st, 3rd and 7th order response surface of the sin wave $S(X)$ with standard normal $x$. 

96
First note that the 7th order response surface produces a very accurate representation of \( f(S) \) which is superior to the 7th order polynomial expansion. Now notice that the 3rd order response surface bares little resemblance to \( f(S) \). We conclude that the response surface may converge much more slowly then the polynomial expansion but may produce better representations at high order.

### 6.6 Applications – Nuclear Reactor Transients

Let us now consider a more complex \( S(X) \) as found in a model by Pagani et al [64]. The model in question is intended to study the performance of passive safety systems under design for a gas-cooled nuclear fission fast reactor. The model assesses the performance of hypothetical safety systems in removing decay heat from the reactor after a severe loss of cooling accident (LOCA). The model is of interest to us as it is one where nominal conditions are not particularly informative. Nominally, the safety systems function as expected and the reactor performs correctly. Rather, what is informative is how the totality of the input uncertainties affect the probability that the safety systems will fail to provide their designed function and the reactor will become damaged.

Details on the phenomenology of this model can be found in [65]. For our purposes, it suffices to say that the model is primarily thermohydraulic in nature and contains nine potential input uncertainties. Six of these are parameters that relate to relevant thermohydraulic correlations, one relates to the power level of the reactor immediately after the start of the LOCA, one to the pressure of the system and the last to the temperature of certain heat removal surfaces. The output of the model is the maximum average and hot channel temperature observed during the course of the transient. This model is computationally efficient for its kind and it is possible to run a single deterministic sample in only several seconds on a modern desktop computer. This speed allows us to perform Monte Carlo simulations to create a reference output distribution.

To examine how useful the polynomial expansion is in propagating uncertainties in this model, let us consider two cases. First we will consider a simple situation where there is
uncertainty only in the reactor power. We will assume that reactor power is normally distributed with a mean value of 17.9 MW\textsubscript{th} and a standard deviation 1\% of the mean.

We apply the polynomial expansion using the isoprobabilistic transform and the Hermite basis, take three samples from the underlying model and obtain the results in Figure 6-6.

![Cumulative distribution function of the average outlet temperature from the Pagani et al model for normally distributed uncertainty in reactor power. Results for Monte Carlo, orthogonal polynomial expansion (PC) and response surface simulations are shown.](image)

The figure shows the cumulative distribution functions of average outlet temperature for a 1000 sample Monte Carlo simulation, a three-sample response surface simulation and the orthogonal polynomial expansion with isoprobabilistic basis simulation. Both the polynomial expansion and response surface methodologies under predict the average outlet temperature at all but the extreme tail of the distribution but the error associated with the polynomial expansion is less than that of the response surface. This is consistent with the results in Section 6.5 where it was shown that the polynomial expansion tends to converge more quickly than the response surface. Whether or not this implies that the polynomial expansion results are acceptable while the response surface results are not is not a determination we try to make here. It is possible that the level or error in both simulations is either acceptable or unacceptable.

Now let us examine a more interesting case where we assume there is uncertainty in all nine of the input variables. These uncertainties are presented in Table 6-4. The mean and variance of each parameter is consistent with those used by Pagani et al [64]. The correlations between the parameters are chosen by this author based on the phenomenological relationships described by Incropera and DeWitt [65], and Todreas and Kazimi [67]. Notice that because the uncertainties are differently distributed and several are correlated, the isoprobabilistic polynomial basis is needed.
Table 6-4: Input uncertainty distributions and correlations for isoprobabilistic basis simulation of the Pagani et al model. Correlation coefficients are designated by the parameter number in the left most column and are listed only once for each pair of correlated variables.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Distribution</th>
<th>Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Reactor Power (MWth)</td>
<td>Lognormal ($\mu = 2.884; \sigma = 0.01$)</td>
<td>$r_{12} = 0.25$</td>
</tr>
<tr>
<td>2 System Pressure (kPa)</td>
<td>Lognormal ($\mu = 7.406; \sigma = 0.075$)</td>
<td></td>
</tr>
<tr>
<td>3 Cooler Wall Temperature (°C)</td>
<td>Gamma ($\alpha = 400; \beta = 0.225$)</td>
<td></td>
</tr>
<tr>
<td>Thermohydraulic Error Factors (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Nusselt – forced convection</td>
<td>Normal ($\mu = 1; \sigma = 5%$)</td>
<td>$r_{46} = 0.2$ $r_{47} = r_{48} = 0.1$ $r_{49} = 0.1$</td>
</tr>
<tr>
<td>5 Nusselt – mixed convection</td>
<td>Normal ($\mu = 1; \sigma = 15%$)</td>
<td>$r_{54} = r_{56} = 0.58$ $r_{57} = r_{58} = 0.1$ $r_{59} = 0.1$</td>
</tr>
<tr>
<td>6 Nusselt – free convection</td>
<td>Normal ($\mu = 1; \sigma = 7.5%$)</td>
<td>$r_{67} = r_{68} = 0.1$ $r_{69} = 0.1$</td>
</tr>
<tr>
<td>7 Friction factor – forced convection</td>
<td>Normal ($\mu = 1; \sigma = 1%$)</td>
<td>$r_{79} = 0.2$</td>
</tr>
<tr>
<td>8 Friction factor – mixed convection</td>
<td>Normal ($\mu = 1; \sigma = 10%$)</td>
<td>$r_{78} = r_{89} = 0.71$</td>
</tr>
<tr>
<td>9 Friction factor – free convection</td>
<td>Normal ($\mu = 1; \sigma = 1.5%$)</td>
<td></td>
</tr>
</tbody>
</table>

Figure 6-7 presents the distribution of the hot channel temperature obtained from these differently distributed uncertainties for six simulations. The first is a Monte Carlo simulation that took 10,000 samples from the underlying model. This number of samples was needed to converge all percentiles to <1%. We will consider this simulation to represent the reference solution. The second simulation is a full factorial response surface for the 9 uncertain factors requiring 256 samples from the model. To generate the probabilistic results in the figure the response surface is sampled using Monte Carlo methods. For this 10,000 samples were taken but note that sampling the response surface is computationally inconsequential as compared to the underlying model. The final four simulations are the 1st through 4th order orthogonal polynomial expansions with the isoprobabilistic basis requiring 10, 55, 220 and 715 samples respectively.
Panels A, B and C of the figure present the PDFs for the five test simulations compared to the Monte Carlo. Notice in panel B, that the 1st and 2nd order polynomial expansions do not adequately capture the hot channel distribution. The 1st order expansion captures only the range, and the second order only the mode. The response surface and 3rd and 4th order expansions in panels A and C, however, do have PDFs that resemble the Monte Carlo results. To further compare these three simulations, panel D presents the error in degrees Celsius relative to the Monte Carlo results. Note that the response surface’s full range of error is not shown so that it is easier to compare the performance of the simulations. Initially, at the 1st percentile, the response surface underestimates the hot channel temperature by about 65°C then, by the 99th percentile the response surface underestimates the hot channel temperature by about 153°C.

The polynomial expansions, on the other hand, more closely track the Monte Carlo solution. The 3rd order expansion has a maximum error of 16°C at the 1st percentile and the 4th order expansion has a maximum error of 3°C at the 93rd percentile. We note then that even with 36 fewer samples of the underlying model, the 3rd order polynomial...
expansion produces superior results to the response surface. This is consistent with our earlier observations that the polynomial expansions tend to converge faster than the response surface. The 4th order expansion does offer some additional improvement but requires 3.25 times more samples.

It is appropriate at this point to bring two items to the reader’s attention. First, in general it is not easy to determine a priori the order of the expansion that will be required to adequately produce the output distribution. What should be done is to calculate the output distribution based on sequentially higher order expansions and stop when little change is observed from one to the next. For the Pagani et al. model, in the absence of the Monte Carlo results, it would not have been known that the 3rd order expansion produced a reasonable output distribution until the 4th order expansion had been calculated. Accordingly, when trying to propagate uncertainties through a new model the reader should expect to have to run one \( k+1 \) order expansion to confirm that the \( k \)th order provides adequate results. This \( k+1 \) expansion can be quite computationally intensive, but this difficulty is not unique to orthogonal polynomial expansions. A similar procedure is required with response surfaces in determining the number of layers to run.

Second, models of the Pagani et al. type are often used in assessing the probability of system failure. With this model, a failure is assumed to occur for hot channel temperatures in excess of 1200°C. That being the case, we would estimate a 0.55% chance of failure from the Monte Carlo results, a 0.63% chance from the 3rd order expansion results, a 0.56% chance from the 4th order expansion but less than a 0.001% chance from the response surface results. If the reader were interested in system safety, then he may believe the polynomial expansion results look particularly attractive, as they are conservative. This is not a valid conclusion. It is purely incidental to this model that the polynomial expansion results are conservative. All we can conclude is that the polynomial expansion converges more quickly than the response surface, not that the error associated with it will always overestimate the extreme percentiles of the output distribution.

### 6.7 Conclusions

Propagating uncertainties through mechanistic computer models is most complex when a complete construction of the output distribution is required. We have shown that this task may be accomplished by use of the orthogonal polynomial expansion procedure. Initial applications of the procedure were limited to uncertainties that could be described by certain independent distributions. While practically limited, these forms of expansion are useful in scoping studies and in understanding general system behavior. We presented two such formulations, the Hermite and hypergeometric basis expansions.

To be practically useful to those making decisions based on the results of mechanistic models, however, it is necessary to be able to propagate whatever arbitrary uncertainties are found to exist in a system. We have presented a computationally efficient method to do so that makes use of an isoprobabilistic transformation and an orthogonal polynomial
expansion using the original Hermite basis. This method is made even more useful as it requires only the marginal distribution of each input random variable and their corresponding linear correlation coefficients and not the complete joint PDF describing the system.

The polynomial expansion using the isoprobabilistic transformation was shown to be theoretically consistent by its ability to exactly produce the output distribution of random variables propagated through a polynomial model. For a simple second order, four-dimensional polynomial, it was shown that the output distribution produced by the orthogonal polynomial expansion was indistinguishable from the Monte Carlo reference solution.

In addition, the procedure was shown to have advantages over the response surface uncertainty propagation method as it offers superior convergence at low order, requiring fewer samples from the underlying model and reducing computation time. At high order, however, the benefits of the procedure are less clear if the expansion coefficients must be determined by a minimal number of quadrature points. This was demonstrated by applying the polynomial expansion procedure to a simple but fundamentally non-polynomial oscillatory model.

Finally, the procedure was tested on a model of a nuclear reactor transient where it was found to again produce superior results to the response surface. The procedure requires slightly fewer samples from the model as compared to the response surface but nearly two orders of magnitude less than the Monte Carlo method.

We conclude that the isoprobabilistic transformation and Hermite basis expansion procedure is a potentially useful method for uncertainty propagation that deserves additional study. This study should focus on numerical methods for determining the expansion coefficients accurately using a minimum number of realizations of the underlying model and for solving the integral equation needed to resolve the correlation matrix in our isoprobabilistic transform.
7 References


[34] Stevens, SS. On the psychophysical law, Psychol. Rev. 1957; 64:153-181.


[37] Lunar Surface Architecture Reference Document. NASA. S12_SARD_DRAFT_0.2.


Appendix I - PSACS QPMs

Below is a listing of the eleven QPMs used to evaluate the two ultimate heat sink alternatives for the PSACS system. Each table gives a summary of the conditions that the DM and the SHs believe bound the value function. A value of zero corresponds to the worst possible outcome any PSACS system might lead to and a value of one corresponds to the best possible outcome. With these bounds in mind, the DM and the SHs assigned specific values to the air and water ultimate heat sink options. These values are listed in the middle two columns of each table.

<table>
<thead>
<tr>
<th>v₁ – Results of spurious actuation of PSACS system</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (worst possible outcome)</td>
</tr>
<tr>
<td>Spurious actuation leads to reactor transients that challenge other SSCs including other safety systems and the power conversion system</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>v₂ – Confidence in ability to model the peak clad temperature (PCT) with available tools and information</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (worst possible outcome)</td>
</tr>
<tr>
<td>Uncertainties related to heat transfer are large and/or PCT is very sensitive to these uncertainties.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>v₃ – S-CO₂ circulation performance after actuation of PSACS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (worst possible outcome)</td>
</tr>
<tr>
<td>S-CO₂ flow will not initiate or will reverse after being initiated.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>v₄ – Ability to inspect the PSACS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (worst possible outcome)</td>
</tr>
<tr>
<td>Demonstration of operability is very difficult. Latent system failures are identified only upon system actuation.</td>
</tr>
</tbody>
</table>
### $v_5$ - Ability to maintain the PSACS

<table>
<thead>
<tr>
<th>0 (worst possible outcome)</th>
<th>Air</th>
<th>Water</th>
<th>1 (best possible outcome)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No online maintenance is possible</td>
<td>0.35</td>
<td>0.75</td>
<td>Entire system can be accessed and maintained online.</td>
</tr>
</tbody>
</table>

### $v_6$ - Number of active SSCs required by PSCAS

<table>
<thead>
<tr>
<th>0 (worst possible outcome)</th>
<th>Air</th>
<th>Water</th>
<th>1 (best possible outcome)</th>
</tr>
</thead>
<tbody>
<tr>
<td>System requires multiple active components or support systems to function.</td>
<td>0.7</td>
<td>0.7</td>
<td>System is entirely passive and requires no active components or support systems.</td>
</tr>
</tbody>
</table>

### $v_7$ - Margin to required energy removal rate obtainable by PSCAS

<table>
<thead>
<tr>
<th>0 (worst possible outcome)</th>
<th>Air</th>
<th>Water</th>
<th>1 (best possible outcome)</th>
</tr>
</thead>
<tbody>
<tr>
<td>System removes just enough heat to meet peak clad temperature limit with no margin.</td>
<td>0.5</td>
<td>0.8</td>
<td>System provides at least 20% margin.</td>
</tr>
</tbody>
</table>

### $v_8$ - Resilience of PSACS against common cause failures (CCF)

<table>
<thead>
<tr>
<th>0 (worst possible outcome)</th>
<th>Air</th>
<th>Water</th>
<th>1 (best possible outcome)</th>
</tr>
</thead>
<tbody>
<tr>
<td>System is susceptible to CCF with the RVACS, i.e. single event could disable both systems.</td>
<td>0.2</td>
<td>0.9</td>
<td>Common cause events do not challenge both PSACS and RVACS.</td>
</tr>
</tbody>
</table>

### $v_9$ - Confidence in models needed to assess each PSACS system

<table>
<thead>
<tr>
<th>0 (worst possible outcome)</th>
<th>Air</th>
<th>Water</th>
<th>1 (best possible outcome)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heat transfer correlations needed to assess system performance are not benchmarked and/or exhibit large (&gt;20%) uncertainties.</td>
<td>0.6</td>
<td>0.8</td>
<td>There is complete confidence in the heat transfer correlations used to assess the system.</td>
</tr>
</tbody>
</table>

### $v_{10}$ - Degree of effort required by design team to design and model PSACS

<table>
<thead>
<tr>
<th>0 (worst possible outcome)</th>
<th>Air</th>
<th>Water</th>
<th>1 (best possible outcome)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Considerable design effort is required. Other systems may be affected by PSACS design decisions.</td>
<td>0.6</td>
<td>0.7</td>
<td>No additional design work is required.</td>
</tr>
</tbody>
</table>
\( v_{11} \) – Component cost of PSACS

<table>
<thead>
<tr>
<th></th>
<th>Air</th>
<th>Water</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (worst possible outcome)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>System represents a substantial part of plant capital costs.</td>
<td>0.4</td>
<td>0.6</td>
</tr>
<tr>
<td>1 (best possible outcome)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

System’s share of plant capital cost is negligible.


9 Appendix II – The Random Index

In the following two tables, the random index, RI, has been calculated for $n$ between three and ten for the integer scale, the balanced scale and the power scale proposed in this paper. The first table gives RI values for pairwise comparisons using five degrees of preference corresponding to the odd integers between 1 and 9. The second table gives RI values for nine degrees of preference corresponding to all integers between 1 and 9. Each RI value is based on 250,000 random realizations of the reciprocal preference matrix.

<table>
<thead>
<tr>
<th>$n$</th>
<th>Integer</th>
<th>Balanced</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.555</td>
<td>0.317</td>
<td>0.346</td>
</tr>
<tr>
<td>4</td>
<td>0.938</td>
<td>0.527</td>
<td>0.569</td>
</tr>
<tr>
<td>5</td>
<td>1.181</td>
<td>0.662</td>
<td>0.714</td>
</tr>
<tr>
<td>6</td>
<td>1.329</td>
<td>0.755</td>
<td>0.809</td>
</tr>
<tr>
<td>7</td>
<td>1.426</td>
<td>0.820</td>
<td>0.873</td>
</tr>
<tr>
<td>8</td>
<td>1.495</td>
<td>0.866</td>
<td>0.920</td>
</tr>
<tr>
<td>9</td>
<td>1.544</td>
<td>0.902</td>
<td>0.954</td>
</tr>
<tr>
<td>10</td>
<td>1.582</td>
<td>0.930</td>
<td>0.981</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n$</th>
<th>Integer</th>
<th>Integer</th>
<th>Balanced</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.525</td>
<td>0.52</td>
<td>0.267</td>
<td>0.387</td>
</tr>
<tr>
<td>4</td>
<td>0.883</td>
<td>0.89</td>
<td>0.437</td>
<td>0.646</td>
</tr>
<tr>
<td>5</td>
<td>1.108</td>
<td>1.11</td>
<td>0.547</td>
<td>0.813</td>
</tr>
<tr>
<td>6</td>
<td>1.251</td>
<td>1.25</td>
<td>0.620</td>
<td>0.922</td>
</tr>
<tr>
<td>7</td>
<td>1.341</td>
<td>1.35</td>
<td>0.671</td>
<td>0.997</td>
</tr>
<tr>
<td>8</td>
<td>1.404</td>
<td>1.40</td>
<td>0.710</td>
<td>1.051</td>
</tr>
<tr>
<td>9</td>
<td>1.450</td>
<td>1.45</td>
<td>0.739</td>
<td>1.091</td>
</tr>
<tr>
<td>10</td>
<td>1.486</td>
<td>1.49</td>
<td>0.761</td>
<td>1.122</td>
</tr>
</tbody>
</table>

* As calculated by Forman [72], for comparison.

The author wishes to caution the reader about the often-overlooked dependence of RI on the number of degrees of preference used. Some authors use a 5-degree of preference elicitation scheme while applying RI values from Saaty’s original papers that are based on 9 degrees of preference. If a 10% threshold is being used to judge acceptable inconsistency, then the reader will observe that the variation in these RI values is quite significant.
10 Appendix III – Survey Questions

What follows is a summary of the instructions provided to survey participants. This is intended to give the reader a flavor of what was asked. The actual survey was conducted electronically, on a computer, with each question appearing on a separate screen. Participants were placed in a distraction-reduced environment and were not allowed to interact with anyone while completing the survey. The following omits all preparatory exercises, examples and actual pairwise comparisons. In addition to the responses to each question, the time taken to complete each was recorded. 64 undergraduate college students were surveyed the afternoon before a major exam.

The following problem was posed:

Consider tomorrow’s [subject] exam. Imagine you may do some combination of the following four things in order to best prepare yourself for exam.

1.) Review homework problems
2.) Review exams from previous years
3.) Review lecture notes
4.) Get a good night’s sleep.

We wish to determine how useful you believe each of these activities is in helping you prepare.

TASK 1:
A brief orientation to determining priority using pairwise comparisons was presented to the participants and then they were asked to complete comparisons to rank these four activities using one of the four following variations. 16 participants responded to each variation.

Variations A and (B): Nine (5) degrees of preference, Random presentation order
On the next screens you will be asked to compare each of the activities to one another. For each pair, please decide which activity you believe is more useful in helping you prepare for the exam and check the box to its left. Then select one of the nine (five) verbal descriptions to indicate how much more useful this activity is than the other one in the pair. If the two activities are equally useful, check neither box, then select “Equally Useful” from the list of verbal descriptions.

Variations C and (D): Nine (five) degrees of preference, Extremes presented first
On the next screen you will be shown all possible pairings of these four activities. Please select the pair that contains the two activities that have the greatest disparity in their level of usefulness.
Now, please decide which activity in this pair you believe is more useful in helping you prepare for the exam and check the box to its left. Then select one of the nine (five) verbal descriptions to indicate how much more important it is.

On the next screen you will be shown all remaining pairings of these four activities. Please select the pair that contains the two activities that have the least disparity in their level of usefulness.

Now, please decide which activity in this pair you believe is more useful in helping you prepare for the exam and check the box to its left. Then select one of the nine (five) verbal descriptions to indicate how much more important it is. If the two activities are equally useful, check neither box, then select “Equally Useful” from the list of verbal descriptions.

On the following screens you will be shown the remaining pairs. For each pair please decide which activity you believe is more useful in helping you prepare for the exam and check the box to its left. Then select one of the nine (five) verbal descriptions to indicate how much more useful this activity is than the other one in the pair. If the two activities are equally useful, check neither box, then select “Equally Useful” from the list of verbal descriptions.

**TASK 2:**
After completing the pairwise comparisons, participants were asked to rank the usefulness of the activities using a simple numerical scale.

By placing the numbers 1 through 4 in the boxes next to each activity, please rank their usefulness. Use 1 for the most useful activity and 4 for the least useful.

**TASK 3:**
Participants were asked to choose between the three possible weight vectors calculated with the integer, balanced and power scales by completing pairs to rank them according to how representative they were of their actual preferences.

Below are three plots, lettered A, B and C. Each represents the results of one method of turning your previous responses into a weighting of the usefulness of the four activities. Please complete the following pairs to indicate how well each plot reflects your actual beliefs of the usefulness of the four activities.

**TASK 4:**
Participants were asked if the number of degrees of preference they were allowed to express was too many, too few or sufficient.

Consider the scale, shown below, you used to complete each of the pairwise comparisons. Did this scale contain too many choices, too few choices or an acceptable number? Please select the appropriate response at the bottom of this page.
RISK-INFORMED DECISION-MAKING
Assessment for the Lunar Service Systems (LSS)
Payload Offloading

OBJECTIVES

We are making use of the risk-informed decision-making framework being developed by NASA Headquarters and Massachusetts Institute of Technology to systematically evaluate the competing priorities that Lunar Surface Systems decision makers face. In this analysis, decision makers’ expectations related to technical performance attributes, such as safety, reliability, and mission performance, as well as programmatic performance attributes, such as cost and schedule, are defined in terms of quantitative performance measures that can be assessed and monitored during the entire life-cycle of the program.

Risk Informed Decision Making. This process requires collaboration among analysts, decision makers, and engineers to develop appropriate models to provide a valid representation of the decision-space or trade-space.

Risk-informed decision making is the formal process of analyzing various decision alternatives with respect to their impact on the measures of performance, of assessing uncertainty associated with their degree of impact, and of selecting the optimal decision alternative using formal decision theory and taking into consideration program constraints, stakeholder expectations, and uncertainties. The assessment is based on an analytic deliberative decision-making methodology applied to the trade study process that consists of three major steps:

1. formulation and selection of decision alternatives,
2. analysis and ranking of decision alternatives, and
3. selection of the best decision alternative.

These steps are linked and are supported by a deliberative process in which the decision-makers scrutinize the initial set of decision alternatives and results of the analysis of decision alternatives to ensure that they are meaningful and that important concerns have been captured.

The process aggregates both objective and subjective information while keeping track of uncertainty. It combines analytical methods with a deliberation that scrutinizes the analytical results. It produces a ranking of decision options and a detailed understanding of why certain options out perform others.
Decision Elicitation

Scenario Definition & Assessment. As a baseline, lunar surface architecture(s) known as Scenario 12, will be used as a baseline for this assessment. (Surface Architecture Reference Document (SARD) Revision S12_SARD_DRAFT_0.2)

Assumptions & Screen Criteria.
- All handling approaches that have the capability for off loading cargo/payload in the “heavy” and “moderately heavy” ranges should be considered.
- It should be assumed that “crew” and possibly “EVA” are there at the time of offloading.
- In general, it is assumed that handling will be accomplished with some level of autonomous or remote operations though manual operations may not necessarily disqualify an approach.
- For evaluation purposes, campaign sorties defined in the baseline should be used as additional scenario context description for handling. For example, these sortie definitions provide additional information about whether or not handling devices, components, or systems can be “pre-positioned” prior to initial use or if there is a strong need for lander flight specific handling capabilities.
- If common or standardized methods for accomplishing a particular function or functions are identified then they shall either be scored consistently across all alternatives or removed from the evaluation.
- If more than one function can be accomplished or if intermediate steps can be avoided (for example, handling to service then to transporter vs. handling directly to transporter) by a particular handling alternative then that will be grounds for elevating the simplicity/ease of use score.

Handling Definition.
"Heavy" Payloads/Cargo include Habitable Modules and Transporters. "Moderately Heavy" Payloads/Cargo include: Pressurized Rovers and Unpressurized Rovers and logistics that can be packaged/containerized for those volumes & weight ranges.

All handling approaches or combinations must be able to accomplish a minimum set of cargo/payload handling functions including:
- cargo/payload release;
- cargo/payload reconfiguration from lander’s flight configuration to handling configuration;
- handling setup (for example, attachment of lifting or handling fixtures);
- handling from lander to the lunar surface; and,
- loading onto mode of conveyance or conveyance.
**Phase I – Options, Goals & Objectives Concurrence (Stakeholders)**

**Handling Options.**
The following options should be evaluated for “heavy or moderately heavy” payload/cargo offloading:
1) Tri-ATHLETE;
2) LSMS;

*As an identified stakeholder, consider the handling options listed above. For each make a determination if each handling approach is an acceptable option for cargo/payload handling describe above. Indicate whether the option is Acceptable or Unacceptable. If unacceptable, please provide rationale in terms of not meeting handling functions listed above.*

<table>
<thead>
<tr>
<th>Handling Option</th>
<th>Acceptable(A) or Unacceptable (U)</th>
<th>Remarks/Justification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Tri-Athlete;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2) Crane;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3) Davit;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4) Athlete a) Pre-Integrated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4) Athlete b) EVA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4) Athlete c) Sliding (Slide to other Side, Attached Directly)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*If there are other handling options not listed, please provide description of those options.*

**Other Handling Options:**

**Description:**

**Goals & Objectives Concurrence**
A notional Objectives Hierarchy is provided in Figure 1 showing the relationship with Goals, Attributes, and Decisions. At the top of the hierarchy is the goal, a broad statement intended to communicate the overall purpose for making the decision. It reiterates the context in which stakeholders (SHs) will determine what other elements belong in the objectives hierarchy.

Objectives are the second tier in the hierarchy. They are the broad categories of elements that the SH feels must be achieved in order for a decision option to meet the goal. These broad objectives may be further divided into sub-objectives as needed.

As an identified stakeholder, consider the handling options hierarchy developed for payload/cargo handling (Summarized Above). Consider each of the objectives listed. Additional detailed descriptions of the objectives and attributes are provided in Attachment ( ).
Are the objectives, as listed, consistent with meeting the top level goal?
Do the objectives correctly capture the important characteristics for handling approaches?
Are the objectives, to the extent that you can determine, independent of each other?
If yes, please describe why you agree with the objective and any additional information needed.
If no, please provide clarification. Should the objective be deleted, modified, or replaced?

<table>
<thead>
<tr>
<th>Objectives</th>
<th>Consistent (Y/N)</th>
<th>Important Charact. (Y/N)</th>
<th>Indep. (Y/N)</th>
<th>Concurrence/Clarification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min. Constraints on Payload</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Opt. Simplicity of Use</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max. Syst. Flex. &amp; Contingency Ops.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Opt. Design</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max. Safety</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Goals & Objectives Consensus Development & Iteration**
With feedback provided from the stakeholders, the RIDM analyst will create a consensus hierarchy and prepare a set of definitions for each objective and attribute. While the stakeholders do not need to agree on the structure of the hierarchy, it greatly simplifies analysis if consensus can be reached.

*The consensus hierarchy will be made available to stakeholders for review and concurrence and the previous steps will be iterated.*
Phase II – Pair-wise Comparison & Weighting Factors (Stakeholders)

Objectives & Attributes – Pairwise Comparison & Weighting Factors. The objectives for the consensus hierarchy will have been further developed to include detailed attributes as shown in Figure 3.

To capture these preferences, the Analytic Hierarchy Process (AHP) is used to capture how relatively important each attribute is to achieving the overall goal. AHP requires each individual to make a series of pairwise comparisons between attributes, and then objectives, saying which of the pair is more important to achieving the goal and then how much more important.

As an identified stakeholder, consider the handling objectives and attributes. Starting with the first Objective, use the pair-wise comparison to determine the relative importance of each Attribute to achieving the higher level Objective. Please note that the Quantifiable Performance Measures (QPM) are notional at this time and will be further developed in the future. They provided for information only to more fully describe the intended context for each attribute.

The ADP AHP spreadsheet is provided as a tool to capture these pair-wise comparisons (See Figure 4)
Results from pairwise comparisons lead to a series of stakeholder-specific weights for the attributes. Consider two stakeholders and the objectives ‘Ensure Affordability’ and ‘Ensure Technical Success’. In context, SH 1 believes affordability and technical success are equally important while SH 2 believes that technical success is twice as important as affordability. The AHP process would result in the weights shown in Table 2. As these weights reveal fundamental differences in the way individuals perceive a decision problem, no attempt is made to reach consensus weights at this stage.

<table>
<thead>
<tr>
<th></th>
<th>SH 1</th>
<th>SH 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affordability</td>
<td>0.5</td>
<td>0.25</td>
</tr>
<tr>
<td>Technical Success</td>
<td>0.5</td>
<td>0.75</td>
</tr>
</tbody>
</table>

Table 2: The AHP Weighting Process
As an identified stakeholder, once you have completed a pair-wise comparison of all the attributes for each objective, then you will compare each Objective to the Goal. (See Figure 5.)

Figure 5 – Objectives Pairwise Comparison
Phase III – Performance Measures & Ranking (Experts)

The next phase involves taking the list of attributes identified and passing them to the experts who develop quantifiable performance measures (QPMs) so that the extent to which a cargo handling option satisfies an attribute can be specified. Along with the expected level of performance, the associated uncertainty is captured. QPMs make up the very last tier of the Objectives Hierarchy in Figure 1.

Since the Lunar Surface Systems (LSS) payload/cargo handling options are currently at the concept phase of development most, if not all, of the performance measure are not be completely quantifiable. It will therefore be necessary to develop more qualitative scales based engineering judgment and expert opinion. As the design matures, more quantitative methods can be incorporated into the decision process.

Options/Consequence Ranking – (Experts). Experts are individuals who have intimate knowledge of the cargo/payload handling options and devices. In contrast to stakeholders, their primary focus is on understanding how each option may perform, individual QPMs, not placing that performance into the broader context of all other elements important to the problem. Two types of experts are required: option experts and consequence experts. For LSS payload/cargo handling, crane designer(s)/engineering would be considered option experts and systems safety engineering would be considered consequence experts.

Option Experts, such as payload/cargo handling systems engineers, are responsible for identifying handling options and assessing their performance based on provided metrics. Consequence Experts understand the repercussions that a decision might have and are responsible for developing metrics to measure the level of consequences associated with different decision options. Together these experts develop quantifiable performance measures (QPMs) and identify and assess the performance of options.

Since all of the payload/cargo off-load options are in the concept phase most of the QPMs will need to be qualitative or notional at this phase of the evaluation. Scales such as Very High, High, Medium, Low, Very Low will need to be used. With descriptive context/justification provided to clarify ranking.
Phase IV – Performance Index Calculation

With all of this information collected, a Performance Index (PI) can be determined for each of the decision options. The PI for alternative \( j \) is defined as the sum of the values, \( v_{ij} \), associated with the QPMs for attribute \( i \) weighted by the APH determined weight for that attribute, \( w_i \).

\[
PI_j = \sum_{i=1}^{N_{PM}} w_i v_{ij}
\]

The distribution of the PI is calculated separately for the DM and each SH. The decision options can then be ranked according to their expected PIs and the effect of performance uncertainty can be shown.

With the calculation of the PIs, the analysis portion of the ADP ends and deliberation begins. The decision maker and the stakeholders each review their individual PIs to understand how the current state of knowledge about the decision options and their individual preferences for the attributes affect the decision problem. Individuals then have an opportunity discuss the similarities and differences between their rankings in order to reach a collective decision.

Review of Important Terms

Objectives Hierarchy

An objectives hierarchy is a communication tool used to show how a stakeholder thinks about a particular decision problem. It captures all of the elements important to a stakeholder in discriminating between different decision options and it organizes these elements hierarchically. The hierarchy contains at least four levels. A goal, objectives, attributes and quantifiable performance measures. Objectives may be divided into sub-objectives to the extent necessary.

Stakeholder

A stakeholder is an individual who has an interest in the outcome of the decision problem. When making risk-informed decisions it is expected that there will be multiple stakeholders and each may have a slightly different objectives hierarchy. At a minimum, the individual with ultimate authority over the decision must be a stakeholder. This individual typically decides who else will be considered a stakeholder. A group of individuals may be considered a single stakeholder if their preferences can be represented by a single objectives hierarchy and they agree on the relative importance of the elements in the hierarchy.
Goal
The goal is the top level of the objectives hierarchy. It is a broad statement intended to communicate the overall purpose for making the decision. It provides the context in which stakeholders will determine what other elements belong in the objectives hierarchy and how relatively important they are to one another. All stakeholders must agree on the goal.

Objectives
Objectives are the second tier in the hierarchy. They are the categories of elements that the stakeholder feels must be achieved in order for a decision option to meet the goal. Objectives may be further divided into sub-objectives as needed.
As an example, maintain safety may be an objective. This objective may be further broken down into the sub-objectives maintain human safety and maintain environmental safety.

Attributes
Attributes fall below objectives in the hierarchy. They are the largest set of elements a stakeholder is indifferent between and often answer the question how an objective will be achieved. It is helpful to think of attributes as the most detailed level of sub-objective the stakeholder wishes to consider.
Continuing the example from above, the attributes associated with the sub-objective maintain human safety may include minimize death and injury to the public and minimize death and injury to the workforce. If the stakeholder were indifferent between death and injury to the public and to the workforce then the sub-objective maintain human safety could become minimize death and injury to humans and be an attribute under the objective maintain safety.

Quantifiable Performance Measures (QPMs)
QPMs are the final tier in the objectives hierarchy. They are measureable characteristics of every decision option that are used to determine how well an option meets an attribute. Each attribute must have one or more QPMs associated with it and a single QPM can be associated with multiple attributes.
Continuing the example, the QPM associated with the attribute minimize death and injury to the workforce might be number of deaths and injuries to workforce.