Summer 2003

Calibrating Expert Assessments of Advanced Aerospace Technology Adoption Impact

Bruce A. Conway
Old Dominion University

CALIBRATING EXPERT ASSESSMENTS OF ADVANCED AEROSPACE TECHNOLOGY ADOPTION IMPACT

by

Bruce A. Conway
B.S. in AS.E., 1965, Virginia Polytechnic Institute
M.S. in AS.E., 1974, George Washington University

A Dissertation submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

ENGINEERING MANAGEMENT

OLD DOMINION UNIVERSITY
August 2003

Approved by:

Resit Unal (Director)

Charles B. Keating (Member)

Andres Sousa-Poza (Member)

Mary Kae Lockwood (Member)

ABSTRACT

CALIBRATING EXPERT ASSESSMENTS OF ADVANCED AEROSPACE TECHNOLOGY ADOPTION IMPACT

Bruce A. Conway
Old Dominion University
Director: Dr. Resit Unal

This dissertation describes the development of expert judgment calibration methodology as part of elicitation of the expert judgments to assist in the task of quantifying parameter uncertainty for proposed new aerospace vehicles. From previous work, it has been shown that experts in the field of aerospace systems design and development can provide valuable input into the sizing and conceptual design of future space launch vehicles employing advanced technology. In particular (and of specific interest in this case), assessment of operations and support cost implications of adopting proposed new technology is frequently asked of the experts. Often the input consisting of estimates and opinions is imprecise and may be offered with less than a high degree of confidence in its efficacy. Since the sizing and design of advanced space or launch vehicles must ultimately have costs attached to them (for subsequent program advocacy and tradeoff studies), the lack of precision in parameter estimates will be detrimental to the development of viable cost models to support the advocacy and tradeoffs. It is postulated that a system, which could accurately apply a measure of calibration to the imprecise and/or low-confidence estimates of the surveyed experts, would greatly enhance the derived parametric data. The development of such a calibration aid has been the thrust of this effort. Bayesian network methodology, augmented by uncertainty modeling and aggregation techniques, among others, were employed in the tool

construction. Appropriate survey questionnaire instruments were compiled for use in acquiring the experts' input; the responses served as input to a test case for validation of the resulting calibration model. Application of the derived techniques were applied as part of a larger expert assessment elicitation and aggregation study. Results of this research show that calibration of expert judgments, particularly for far-term events, appears to be possible. Suggestions for refinement and extension of the development are presented.

ACKNOWLEDGEMENTS

I express my deepest appreciation to my dissertation committee – Resit Unal, Charles Keating, Andres Sousa-Poza, and Mary Kae Lockwood. My thanks also go to the other faculty , staff, and students in the Engineering Management Department for their support, encouragement, and patience. I am particularly indebted to Trina Chytka of NASA-Langley, a fellow doctoral student, for her part in the collaborative effort around which the present research was carried out. Appreciation is also expressed to Roger Lepsch and Mark McMillin of NASA-Langley who provided invaluable assistance in the development and adjustment of the calibration techniques discussed herein, and to Doug Morris, Nancy White, and Dick Brown of NASA-Langley who offered key support on this and a related precursor study.

Finally, I must acknowledge that I could not have met the challenges of this endeavor without the support and encouragement of my wife, Carol. Her sacrifice and love are endless, and are a continuing source of inspiration to me.

TABLE OF CONTENTS

CHAPTER Page

APPENDICES Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Background

The National Aeronautics and Space Administration (NASA) Langley Research Center has long been responsible for advanced aerospace vehicle conceptual development. In determining attributes for an advanced concept vehicle, NASA utilizes various resources. Among these are current designs and technology, extrapolated to address future requirements and anticipated technology levels. The process of extrapolating current technology requires engineering judgment, and the degree to which the projections will be borne out is dependent upon the expertise of those forecasters performing the extrapolations.

In addition to ascertaining what technologies can or should be included in an advanced vehicle concept, it is important that the cost impact (positive or negative) of incorporating unproven technology be determined. The technologies under consideration can cover many disciplines and affect most, if not all, of the proposed vehicle's systems and subsystems (see, for example, Rowell, Olds, and Unal, 1999). In fact, adoption of a specific technology may impact more than one subsystem, to differing degrees. Because conceptual design specialists do not usually have expertise in every single vehicle system and their technologies, they may not always be able to judge accurately the projected impacts of incorporating future technology. Thus, a methodology to systematically guide the technology forecasting and assess the cost impact of adopting advanced concepts would be very desirable.

The journal model for the references herein is *The American Psychologist*, the journal of the American Psychological Association

In developing a methodology to examine what might happen as a result of some future course of action, forecasters typically rely on such things as past experience, models of behavior or chance, systems models, or several other types of tools. One means that has come into increasing use in the last 20 years is the so-called "expert system." Expert systems come in many levels of sophistication, thoroughness, and ease of use. A common characteristic of these systems is their use of information (facts, preferences, opinions, and other types of knowledge) gleaned from acknowledged experts in the area(s) the expert system is supposed to cover. Studies suggest that expertise results primarily from the ability to perceive, or recognize and differentiate the patterns and invariants in the environment, as opposed to the use of rules and facts. An expert system seeks to capture such expertise (or knowledge) through carefully constructed knowledge acquisition means.

The use of expert systems may cover many scenarios. An expert system may be established to capture "best design practices" as developed over many years by retired (or soon-to-be-retiring) practitioners. Or, an expert system might be constructed to guide a repetitive process for untrained personnel (such as the computer online completion of tests, questionnaires, and the like). Yet another use could involve the design of an expert system to handle some remote operation where a human could not be either physically or virtually present such as in the preparation of robotic planetary exploration craft that must operate out of contact with human controllers for long periods. Finally, an expert system may be designed to capture information needed to establish an architecture or program structure, such as a quality program (see, for example, Kahn and Hafiz, 1999).

The type or class of expert system of interest in this research, however, is one that gathers knowledge and considered opinions from expert practitioners in several related

fields of science and technology and applies that knowledge to the conceptual design and analysis of some future embodiment utilizing current or advanced state of the art in their respective areas of expertise. In the current case, the specific application of an expert system is in the area of developing new concepts (technologies and configurations) for single-stage-to orbit aerospace vehicles designed for efficient access to space. One concern that arises in considering new-technology-solutions to either long-standing or newly developed problems or requirements is the cost of the technology – both its development cost, and its cost impact on operations and support of a particular new vehicle if built and deployed. A knowledge of these cost impacts is crucial in developing tradeoffs when considering more than one potential advanced design. Knowledge of costs is also critical in assessing the life cycle costs of competing vehicle concepts.

The use of expert judgment or opinion to aid in decision-making is reasonably well known. The knowledge of subject matter experts (SMEs) has been "mined" to develop procedures to handle complicated manufacturing or implementation tasks, for example or responses to presented options such as the diagnosis and treatment of medical conditions. More specifically, experts have been tasked with providing estimates for parameters associated with yet-to-be-developed systems (such as advanced space launch vehicles). In utilizing the expertise of an acknowledged expert in a given field, there is often no need to query more than a single expert. The expert's carefully elicited opinion, judgment, or estimate can be accepted as the point estimate for the quantity or parameter under consideration.

There are, however, situations in which it is necessary or desirable to employ the expertise of multiple experts to address a design-related estimating task. The use of multiple experts allows for the coverage of most, if not all, aspects of very sophisticated

design concepts, which may incorporate advances in several disciplines. The potential arises in this scenario for there to be a lack of consensus among the estimators. Such a lack can produce widely divergent estimates of the parameters or quantities being judged, particularly if the estimates are elicited through an anonymous Delphi-type process (where experts independently submit responses to an elicitation of assessments, rather than doing so together in a group). The subsequent utilization of divergent multiple estimates (in projecting associated costs, for example) becomes problematical without a means to adjudicate the disagreement between the estimates.

Complicating the variability of estimates is a degree of uncertainty associated with each expert's judgment. The basis of an individual's uncertainty may arise from an innate (constant) lack of conviction in any projection or estimate, or it may be related to his or her degree of comfort with knowledge in the subject area (which could vary from task to task). Whatever the reason for the uncertainty, the combination of uncertainty with the estimate variability results in a two-dimensional measure that exacerbates the problem of arriving at a single point estimate for the parameter or quantity under consideration.

Elicitation of an expert's degree of certainty about his estimate is crucial to the subsequent use that may be made of that estimate. The level or degree of uncertainty affects the weight that a user of the estimate may assign to the information produced by the assessment. If aggregation of multiple assessments is the use to be made, then weights must be assigned not only to factors associated with the experts producing the assessments (as many aggregation techniques do), but also to the information itself. If decisions are to be made based on the elicited assessments, then the decision maker must perform his own (informal) calibration based on the level of uncertainty in order to make

effective use of the assessments. Uncertainty is thus associated with the assessments of each expert; aggregation of multiple assessments does nothing to reduce that uncertainty, it only masks it.

The elicitation of judgments from subject experts is fraught with several potential pitfalls. Perhaps first among these is identification of an appropriate expert (or experts) whose judgment may be relied upon to produce judgment of acceptable reliability. Ideally, an expert would render consistent judgments in similar environments.

A second hurdle that must be overcome is the development of an appropriate questionnaire or other information elicitation mechanism that is both efficient (time-wise) and effective (content-wise). The time commitment of participating experts must be honored, and an ineffective elicitation instrument would yield information of dubious value.

A third issue with potential ramifications is the decision about the number of experts to be queried. If more than one are used for the same topic, then some means of either "calibrating" each expert's judgments or aggregating the responses of multiple experts in a meaningful way (or both) must be implemented. Otherwise, individual differences in experts' experience, confidence in judgment, and innate baseline from which judgments are based will likely render inconclusive (or at least less precise) results.

It is this last issue that forms the motivation for the present research. While calibration of experts has been studied for some time, the predominance of calibrated judgments have dealt with outcomes which could be verified either at the time of the calibration or within a reasonable time afterwards. In the present instance (advanced launch vehicle concept development), however, the wait for validation of an expert's

judgment could take more than 20 years. In a word, a way must be found to calibrate experts assessing the "unknowable future," to handle information characteristics and content redundancy The subsequent use of calibrated judgments in aggregation techniques using a combination rule that is most compatible with conceptual aerospace vehicle design would thus be made easier.

Problem Statement

Multiple experts can produce widely varied estimates of their tasked judgments; this variability is aggravated by disparate uncertainty in the "calls" made by each of the experts. "Which is the estimate to be used?" Similarly, even a single expert assessing events or state-of-the-art of the future may produce judgments with associated confidence that, without calibration, could be subject to question. In the overall framework of multidisciplinary analysis of advanced conceptual designs, reductions in uncertainty through calibration would be expected to facilitate efficient aggregation and ultimate decision-making that will be based on the expert assessments and analysis.

Synopsis of Report

How then to develop an expert system that can assist in the task of evaluating weight and size estimates, operations and support resource requirements, and multidisciplinary design and optimization questions for proposed new aerospace vehicles? That is the problem that was addressed in this research effort. In particular, a means for effectively calibrating experts whose judgments may not be validated for many years was sought. Both single- and multiple-expert judgment scenarios were examined, and a calibration methodology was successfully developed.

The next section will review predecessor and related research and relevant literature applicable to the topic area. A research methodology is presented which (a) covers continuation of previously begun work by the author in this area, and (b) addresses extension of the effort to define the development of a more robust system, including expert judgment calibration, which efficiently handles a variety of conceptual design-related questions. Next, the application of the methodology as part of a wider study supported by NASA is discussed, followed by a presentation of the results of the investigation pertaining to the expert calibration problem. A discussion of results is followed by conclusions drawn from the application of the methodology and the results of the study. Limitations associated with the current study and recommendations for future effort are presented as the final section of this dissertation.

CHAPTER II

REVIEW OF RELEVANT LITERATURE AND RELATED RESEARCH

Uncertainty

Dealing with uncertainty by those engaged in the physical sciences (including engineering) is almost second nature (Morgan and Henrion, 1990). In engineering, it is common to perform uncertainty modeling and analysis of a system's abstract aspects, with proper consideration of its realized aspects (Ayyub, 2001). Reporting the results of experiments, or designing a new experiment or survey, without attention to uncertainty is deemed risky at best. However, the treatment of the uncertainty associated with an investigation or, as Morgan and Henrion (1990) note, in policy analysis is not uniformly understood; there is no single method of including uncertainty factors in assessing a problem. In addition, risk is usually associated with uncertainty, and means of quantitatively treating the ever-present uncertainty are desirable. Morgan and Henrion include many techniques – philosophical, graphical, and analytic – for handling uncertainty in a given problem or research area.

One aspect of uncertainty is the error associated with making estimates. The credibility of estimates is associated with a variety of fields. Brown (1969) developed a methodology, which, although aimed principally at business-oriented problems, includes tools for a practitioner to use in analyzing (appraising) estimates. His techniques include the use of "credence distributions", which resemble classical probability distributions but which involve personal viewpoints of an investigator in addition to the pure observed research findings. The decomposition of "target variables" which have uncertainty

associated with their values) into several components, each of whose assessment is tractable. Results are combined to yield an overall assessment for the target variable.

There are many methodologies to apply to problems that involve uncertainty. Among them are Bayesian networks (Neil, Fenton, and Nielson, 2000), fuzzy logic (Zadeh 1992, Kosko, 1993), and default reasoning (Antoniou, 1998). Each of these methods provides some means of logically handling the assignment of values (probabilities in the present case). Each method is different, however, and it is problematical as to which (or possibly some other) may be the best to employ in assessing the uncertain probability that specific new technologies will reduce the operations and support costs associated with advanced launch vehicles.

Bayesian networks utilize a directed graph model of causality following Bayes' Rule for influence. Schmitt (1969) and Cyert and DeGroot (1987) provide insight into a variety of applications of Bayesian statistics, which include techniques of modifying probabilities based on accumulating experience. Cyert and DeGroot (1987) focus their work in the field of economics, but develop a concept referred to as "adaptive utility". This concept, analyzing utility functions instead of probabilities, should be directly applicable to the current research problem of cost assessment functions for individual experts. The use of Bayesian networks in expert judgment elicitation for uncertainty has been carried out for many years (see, for example, Renooij, 2001, Neil, Fenton, and Nelson, 2000, and Coupé, van der Gaag and Habbema, 2000). In most cases, the efficacy of the methodology is dependent on the degree of expertise and the comfort of the experts in providing accurate (precise) estimates.

Fuzzy logic methods have been developed over the past 35 years to address uncertainty ("fuzziness") in describing probabilities of certain types of events. Fuzzy

logic treats everything associated with probability as a matter of degree, and thus is not limited to step-function types of probabilities for events (see, for example, Zadeh, 1992 and Kosko, 1993). Hardy (1994) discusses fuzzy logic approaches to multi-objective decision-making in aerospace applications, a work that may be directly applicable to the proposed research. Hardy and Rapp (1994) apply the approach specifically to rocket engine reliability analysis.

In the fields of artificial intelligence and, in particular, knowledge-based systems, researchers have been working for years on the problem of modeling (or quantifying) beliefs, reasoning and opinions, and the uncertainty that, in varying degrees, accompanies those thought processes. For example Mántaras (1990), in his discussion of the modeling of approximate reasoning, focuses on numerical approaches in the framework of rule-based systems. He highlights probabilistic approaches, a fuzzy logic/possibility theory approach, and an approach based on the Dempster-Shafer theory of evidence.

A study related to intuition and analysis cognitive activity, and which produces the more accurate judgments, has found that intuitive and quasi-rational cognition often outperforms analytical cognition in the empirical accuracy of judgments (Hammond, Hamm, Grassia, and Pearson, 1987). The study by Hammond, et al, focused on judgment tasks that ranged on a continuum from purely intuitive to highly analytical; the cognitive tasks would seem to cover a similar range of thought modeling that includes the fuzzy logic and approximate reasoning approaches by Hardy (1994) and Mántaras (1990) have pursued. Their finding would seem to suggest that application of expert judgment elicitation in less-well-defined areas could produce efficacious results.

Expert Judgment Elicitation

The process of obtaining expert opinions, judgments, or assessments with some appellation of confidence (probability of the event being as assessed) must be well structured to minimize or avoid the introduction of bias (Renooij, 2001). The elicitation process would ideally include the selection, motivation and training of experts, proper structuring of the questions to preclude bias, the actual elicitation and documentation phase, and verification of results (Renooij, 2001). There are many forms that the elicitation process can take, as well as a number of different forms of the desired elicitation output.

There have been numerous situations over the years in which it has been desirable to obtain information or assessments from acknowledged experts in a field. Techniques have been developed to "mine" the requisite knowledge or assessments from the experts. Such techniques range from basic one-on-one sessions between an elicitor and expert using specially tailored elicitation aids (see, for example, Keppell, 2001), to methods involving assessments from multiple experts, acting either individually or as a group. Some of the multiple-expert techniques that have been used include brainstorming, the Nominal Group Technique (NGT – see Gustafson, et al, 1973), and Delphi panels (see Dalkey, 1969, Linstone and Turot, 1975, and Rowe, Wright, and Bolger, 1991). In each of these techniques involving groups of experts, the process is designed to develop a consensus among the experts. Decisions would then be made based on the group consensus, provided that one was achieved.

While a consensus approach to eliciting knowledge or judgments from subject matter experts may yield acceptable results, it can be a time consuming process; it may be

hard to assign a degree of certainty to those decisions involving quantitative estimates. In certain fields, notably meteorology, efforts have been underway for several decades to elicit judgments in terms of a probability, sometimes accompanied by an associated degree of confidence in the rendered assessment. Murphy and Winkler (1974) describe experiments whereby a weather forecaster expressed forecasts of (say) a daily maximum temperature in terms of what is referred to as a "credible interval," or interval of values with a probability reflecting the forecasters' degree of belief that the temperature will fall in the given interval. Validation of the forecast methodology is straightforward in this scenario, since the forecasts are for a very short time-horizon (from a few hours to a few days).

Other early work by Beach (1975) found that use of subjective probabilities and Bayes' theorem in real-world decision-making is potentially profitable (or, has economic value). Military decision-making, meteorology, medical diagnosis, and business trend analysis are examples of the classes of problems amenable to probabilistic forecast techniques. Beach also found that, in the majority of the situations she studied, group (or consensus) probability judgments generally yielded results that were superior to those achieved by individuals. The methodology of combining opinions from multiple experts varied and no one technique was found to be better in all cases than another.

To avoid some of the problems resulting from group dynamics, Rush and Wallace (1997) developed a technique for eliciting knowledge from multiple experts that were not members of an interacting group. Influence diagrams are used, along with assigned probabilities, to represent an expert's understanding of the problem situation; a multiple expert influence diagram is a composite representation of the multiple experts' knowledge (Rush and Wallace, 1997). This technique appears to be well suited to those

problems where the elicitation is intended to support the making of a decision (go – no go) type of problem.

In another use of experts in judgment or knowledge elicitation, the development of Bayesian belief networks in a medical diagnosis application appears to benefit by sensitivity analysis of a belief-network-in-the-making (Coupé, van der Gaag, and Habbema, 2000). The analysis provides insight into those probabilities requiring a high level of accuracy, and is useful in those problems where there are a large number of probabilities to be assessed by an expert. Subsequent or future elicitations can then be focused in specific areas, based on the sensitivity analysis. It is possible to build rather large Bayesian networks using building blocks (Neil, Fenton and Neilson, 2000). This technique can be applied much like computer-aided design in manufacturing or electronics "assembles" simple elements into larger, more complex structures.

In earlier work related to the present problem, Monroe (1997) developed a methodology for eliciting expert judgment to help overcome uncertainty in decision analysis. The work used as an application example the development of weight estimates for the various major components of advanced single-stage-to-orbit vehicle concepts. A questionnaire was developed to enable the elicitation of expert judgments about weight fractions of the vehicle components. A novel aspect of the technique developed by Monroe was the inclusion of a methodology to allow degrees of uncertainty of the surveyed expert's estimate to be attached to the weight quantities judgments. The methodology consisted of a series of questions to anchor most likely and least likely points and then assign intermediate uncertainty levels. The uncertainty levels were then used to construct probability distributions for the various weight parameters; these probabilities were subsequently used in Monte Carlo simulations to converge to "final"

weight estimates for a vehicle under study. Hampton (2001) adapted Monroe's methodology to the area of integrated risk analysis in a multidisciplinary design environment, aggregating the uncertainty identification and quantification assessments of two experts in a risk analysis problem.

The technique developed by Monroe (1997) was modified and applied to the determination of impact on operations and support costs of proposed vehicle concepts resulting from the adoption of various technologies. The expert judgments of design engineers at three NASA Centers were acquired through a formal questionnaire, which also asked the respondents to provide a confidence level for their estimate, based on a scale developed by the researchers (Unal and Conway, 2000).

The work by Monroe (1997) and Unal and Conway (2000) employed a set of structured guidelines, or rules, by which the experts could apply their knowledge in a consistent manner. In the operations and support costs-related study, the candidate technology spectra were formulated by an advanced vehicles concept group, and refined through pilot surveys of selected experts, who were able to suggest additions or modifications to the technology "menu." The uncertainty levels were applied based on a scale established by the researchers. In an attempt to promote consistent application of the uncertainty levels, this scale provided narrative "anchors" for the expert's use in assigning his rating. In contrast, Monroe (1997) allowed the expert to establish the range and anchors for the uncertainty level (in other words, a subjective probability distribution). The work by Unal and Conway did not include a method to quantitatively apply the confidence levels over multiple experts to enhance the assessment of technology cost savings, nor did the work employ a calibration of experts, a shortcoming that will be discussed in the next section.

A concern that has arisen in some past expert judgment studies is the effect of temporal setting on the elicitation task (specifically, is the assignment of likelihood estimates to past and future outcomes the same or different?). Fischhoff (1976) found that there was no consistent differences in likelihood judgments regarding past and future events which differed solely in their temporal setting. In contrast Wright (1982), in a study involving probability assessment as a function of question type, found that differences did exist: probability assessments for future event questions tended to be less certain than probability assessments for past-event questions. The question thus remains as to the effect of temporal setting on probability or likelihood estimates, and the concomitant effect on the possibility for calibration of expert assessments.

Calibration of Experts

One of the phases of a properly structured expert elicitation process is verification of results (Renooij, 2001). Verification includes ascertaining that the assessments (usually probabilities) are reliable (in a test-retest sense), coherent (obey the laws of probability), and well calibrated (conform to observed frequencies). This last step, calibration, is reasonably straightforward for "knowable" outcomes, but is most difficult for "unknowable" or unobservable events or outcomes, such as those that are future occurrences.

Many authors have discussed calibration of experts in judgment elicitation scenarios. Morgan and Henrion (1990) note that there have been many empirical studies focused on the calibration aspects of people's abilities as probability assessors. Keren (1991) points out that most of the calibration studies he reviewed focused on technical formal issues, presumably because the dominant perspective is that uncertainty is a

reflection of the external world and thus of the events or outcomes being assessed. He proposes that uncertainty is more a characteristic of the assessor, although the two views are often entwined.

Johnson and Bruce (2001) focused on a narrow field (horse wagering) and found that the naturalistic setting of actual wagering facilities at racetracks, as opposed to laboratory settings and experiments, resulted in close correlation between subjective (bettors') probability of winning and objective probability (based on race results).

Perhaps one of the earliest attempts at calibrating expert assessments involved the verification of weather forecasts expressed in terms of probability (Brier, 1950). The methodology was based on a verification score, P, that is a function of outcomes and assessed probabilities associated with the various outcomes, and the actual outcomes themselves (post priori). The P-score would be smaller for "good" forecasting (with zero being perfect) and larger for "bad" forecasting (a score of 2 would be the worst). With appropriate feedback to the forecaster expert, it can be seen that this verification process could also serve a training purpose, because the P-score is minimized by avoiding bias or gamesmanship with the score. In a somewhat similar manner, Schaefer (1976) used a logarithmic form of a proper scoring rule. [Note: a proper scoring rule is one where no strategy by the assessor will produce a better expected score than always reporting one's true beliefs (Lichtenstein and Fischhoff, 1980)]. Schaefer's experiments provided feedback to the subjects for calibration and training purposes. As with the weather forecasts, however, the determination of the actual values of estimated proportions was straightforward.

More recent work on scoring rules and calibration of expert assessments employed an interactive computer-aided graphical means to feed-forward scoring rules

based on assumed subjective probability distributions (SPD) (van Lenthe, 1994). In this technique, assessors could see the effect on their scores of their assumed distribution. In variations of his experiments, van Lenthe presented only the scoring curves or the estimated probability distribution, rather than both. In all experiments where proper scoring rules were used to evaluate subjective probability distributions, the graphical feed-forward technique (both SPD and derived scoring curve) tended to produce better calibrations than those where only the SPD or scoring curve was displayed.

In examining the results of calibration studies, some researchers have addressed differences between actual experts participating in an experiment and non-expert (novice) subjects. Spence (1996) found that experts were less likely to include the actual outcome in their range of probable results than were novices performing the same assessment task, for reasonably straightforward problems. For more complex problems, the experts produced better estimates than did novices. Spence attributes this decline in novice performance to underestimation of the complexity of difficult problems. In contrast, Lichtenstein and Fischhoff (1977) found that calibration is unaffected by differences in expertise or by differences in intelligence or elements of context in the problem setting. In a survey of calibration of probabilities, Lichtenstein, Fischhoff and Phillips (1980) report mixed results: in some studies, experts performed well (were better calibrated), and in other studies they did not. In the cases of poorer performance, difficulty of tasks involving continuous quantities was seen to be a contributing factor.

In the types of applications under consideration in this research, complexity is much more present than is simplicity. The performance of experts in their environment is dependent on their view of the world, or "world view" (Feltovich, Spiro, and Coulson, 1997). Inflexibility in the acquiring or interpretation of knowledge would be expected to

render an expert less effective in his practice or imparting of knowledge to others. While the automation of knowledge-based systems (expert systems) is becoming increasingly more sophisticated, the builders and users of the underlying expert systems must have some assurance that the contributing experts in fact have not been "stuck in a rut" but, rather, have the requisite broader view (and flexibility) demanded by today's applications.

In a study of the relationship between judgmental probability forecasting performance, self-rated expertise, and degree of coherence, it was found that self-rated expertise was found to be a good predictor of subsequent performance (Wright, Rowe, Bolger, and Gammack, 1994). Measures of individual coherence (extent to which a probability assessor's forecasts conform to the axioms of probability theory — probabilities of mutually exclusive and exhaustive events summing to one, for example) were found to be less predictive. The researchers also recomposed the straightforward holistic and marginal probability assessment tasks into more basic intersections, disjunctions and unions, and found that improved performance resulted when compared to the holistic and marginal results. Dawid (1982) makes a case that a well-calibrated assessor will likely not be coherent, because any recalibration required to bring prediction more in line with reality leads to incoherence. He acknowledges, however, that it is theoretically possible that a forecaster's assessed probability distribution could not even potentially miscalibrated by essentially keeping track of and updating past calibration performance when making each succeeding forecast.

Others elicitation practitioners have been working in the field of modeling, to take into account constraints which arise from a given task and applied judgments and which must be propagated through the design process. This can also be aided by the use of

knowledge organization. Kolb (1994) discusses the implementation of a modeling package which combines an approach using object-modeling to organize knowledge about design components and analyses in a modular fashion, and an approach of constraint propagation in analysis and computation. The end result is a flexible schema to undertake the evolutionary design process. Object modeling is a way to organize design knowledge, where the design knowledge is stored in object classes, which are inserted into the design by the user at appropriate times in the design evolution.

A somewhat easy to grasp example of applying expert systems to a problem involving constraints (or at least guidelines) is the design of road bridges that have an esthetic appeal as well as the proper structural strength characteristics (Zuk, 1990). Through the extraction of "rules" from books, articles, or reports (going back more than 100 years), a "comparator bridge" is defined with a given rating (assigned by experts using four separate criteria); target designs would then be compared to the "standard" and assigned ratings higher or lower. All target bridges would first have to meet constraint criteria with respect to design strength. While some of the esthetic guidance involves mathematical criteria, much of it involves criteria that are subject to judgment by either the designer or the evaluator, or both. The knowledge of the experts involved is the key to an objective design or evaluation, but there is no means given to quantify the disparity or provide for consistency among a larger group of evaluators.

In addition to some measure of an expert's level of expertise, a meaningful indication of the expert's tendency toward overconfidence or underconfidence in his or her judgments of probabilities and uncertainties is also needed. Wright, Rowe, Bolger and Gammack (1994) address the over- and underconfidence question, but only in the context of verifiable forecasts. Investigations such as the present one, where judgments

are made about distant-future events and conditions, must necessarily rely on risk-oriented techniques to assess the overconfidence/underconfidence tendency. Utility theory offers tools for such applications. In particular, utility-theory-based techniques that do not depend on the classical monetary wager lottery risk tolerance evaluation would be most helpful. Duarte (2001) has developed a methodology using utility theory that elicits alternative choices in a socio-technical environment that includes qualitative as well as quantitative factors.

For experts whose judgment are being elicited on distant-future concepts and parameters (as in the current research application), it has been noted that calibration (verification of accuracy) will not be possible in the near term. For this situation, then, other means of "verifying" an expert's judgment performance have been sought. One method, developed by James, Demaree, and Wolf (1984) estimates interrater reliability for a group of judges performing the same task. Their methodology seeks to determine the systematic variance among judges participating in an evaluation task. The methodology includes means for addressing influences of response bias that may be common to the multiple judges. Application of the James, Demaree, and Wolf interrater reliability estimation techniques is expected to be useful in assessing the efficacy of calibration of multiple experts on the same judgment tasks.

This section has covered many aspects of expert assessment calibration. Table 1 summarizes some of the more salient literature related to the current problem. The works cited in Table 1 serve to highlight tractable aspects of this research. Incorporation and extension of these concepts are addressed in subsequent sections.

Table 1. Summary of Expert Calibration Literature

Author(s)	Major Points/Findings	Current Research Implications
Brier (1950)	An early "proper scoring rule" for calibration	Use proper scoring rules to avoid gamesmanship
Johnson & Bruce (2001)	"Naturalistic" environments produce better calibration – better performance	Avoid sterile environments or purely academic settings
Lichtenstein & Fischhoff (1977)	Overconfidence increases with knowledge – to a point, then decreases	Take level of knowledge into account during processing of elicitation results
Wright, Rowe, Bolger & Gammack (1994)	Self-rated expertise good predictor of subsequent performance	Include self-designation of expertise in calibration models
Spence (1996)	Calibration varies with expertise for complex problems	Attempt to pinpoint expertise level of assessors
Wright (1982)	Probability estimates for future events less certain than for past events	A key motivator for present study
Keren (1991)	Calibration is a characteristic of assessor, not event	Distinguish between event uncertainty and assessor uncertainty
Dawid (1982)	Calibration and coherence can clash for an assessor	Minimize by updating calibration forward through subsequent assessments
Morgan & Henrion (1990)	Unclear feedback to experts can lead to worse subsequent results	Stress clear feedback in calibration process
Renooij (2001)	Calibration only part of structured elicitation process	Properly structure the assessment elicitation process to include calibration
Duarte (2001)	Non-wager types of alternatives can be used in establishment of utility	Apply to questionnaire construction for calibration
Hammond, et al (1987)	Intuitive and quasi-rational cognition often outperforms analytical in empirical accuracy of judgment	Supports application of expert judgment elicitation and calibration to less-well-defined areas
James, Demaree and Wolf (1984)	Technique for determining interrater reliability addressed	Useful in developing methodology for ascertaining calibration efficacy

Calibration Questionnaire Development

Designing an appropriate instrument to elicit information specific to establishment of a calibration for an expert is important. Elements of this instrument (usually a questionnaire) must ascertain the expert's accuracy (closeness to numerical values of sought responses) as well as the variability of his response. This second quantity (variability) takes the form of a variance of the probability distribution which can be used to models the expert's response. Since variability is typically measured over a number of trials (which would be impractical in the real world of expert elicitation), the associated variance may be obtained through elicitation of a confidence level.

Accuracy of an expert's response is related to his or her level of expertise in the subject area for which judgments are being elicited. As noted previously, self-rated expertise has been found to be a good predictor of performance (Wright, Rowe, Bolger, and Gammack, 1994), suggesting that any elicitation instruments include such a self-rating.

There is at least one other potential indicator of expertise: Crawford and Stankov (1996) and MacCrimmon and Wehring (1986) have found that age and expertise of experts are related. In contrast, studies show that, although there are certain instances of positive correlation between experience and expertise, there is no evidence to support applying this standard universally. It is true that prior conceptions of an expert's level of expertise used the number of years on the job (relevant experience) as a surrogate to expertise. It had been found, however, that while many experts do indeed have significant length in service, time on the job or years in a discipline or field does not

necessarily equate to expertise. Some individuals may work along side experts but never acquire the skills and knowledge to reach true expertise, Shanteau, et al 2002).

A classic means of determining confidence level, which can be equated to as risk tolerance (see, for example, Miller and Byrnes, 1997, and Wang, 2001), has been through the use of utility theory and the determination of an individual's utility function. Typically, monetary wagers are postulated with varying payoffs and associated odds and the "bettor" is asked to indicate a preference between or among alternative wagers. However, some researchers have been able to apply utility theory using non-monetary alternatives to elicit judgments, with confidence being inferred from the choices made. Duarte (2001) has developed a method to solve industrial decision problems using expected utility theory. In his method, Duarte develops technical alternatives with multi-value attributes evaluated for each alternative. While some attributes were measured in terms of monetary value, others, such as image, environmental impact, and flexibility were assigned values by a panel of experts on rating scales established for the attribute. As in traditional (using monetary wagers) applications of utility theory to ascertain choices or risk tolerance, the non-monetary attributes were evaluated along with monetary ones in adjustments to determine indifference to a choice between two alternatives.

MacCrimmon and Wehrung (1986) have conducted studies with executives on the handling of risk in certain business situations. The analysis of risk propensity involved utility-function simulations with monetary and non-monetary situations such as impending threats and opportunities. Their use of situation alternatives to ascertain risk propensity reflects favorably on the concept that risk and confidence level are related. Several researchers, in fields such as finance, entrepreneurship, and psychology have

found that risk-takers tend to be overconfident while risk-averseness is associated with underconfidence (see, for example, Wang, 2001, Simon, Houghton, and Aquino ,1999, and Miller and Byrnes, 1997). Simon, Houghton and Aquino (1999) also found that the risk propensity may not always be conscious but rather may be the result of cognitive biases such as overconfidence. These findings suggest that use of a qualitative form of utility theory application would be appropriate to elicit a risk or confidence-level propensity from a participating expert.

One final aspect of the calibration questionnaire-related research: it has been found by MacCrimmon and Wehrung (1986) that older managers in their studies tended to be more averse to risk than did younger managers. This suggests that there is a relation between age and risk tolerance and also suggests the calibration-related portion of the elicitation instrument include the participant's age.

Other Calibration Considerations

A key consideration in attempting to calibrate expert assessments is training of the assessor. Training is performed in an attempt to improve the quality of an assessor's probability assessments. Lichtenstein and Fischhoff (1980) found that training produced considerable learning, almost all of it after receipt of the first feedback. They found that training could be modestly generalized to some related probability assessment tasks but not to others. Alpert and Raiffa (1982), in an experiment whereby assessors utilized direct fractile assessments in the elicitation process for probability ranges, and found that providing feedback on performance generally improved subsequent performance.

Ayyub (2001) stresses that training should involve experts, observers and facilitators, with one aim of the training being the identification and of sources of

potential bias and their minimization or elimination. In addition, Ayyub notes that experts need to be trained to provide answers or assessments in an acceptable format that facilitates their use in subsequent analysis and application of the elicitation results. For those experts not familiar with probability-related concepts and terminology, additional related instruction might be required.

Another aspect of elicitation methodology and calibration is whether to implement assessor-defined categories (such as the fractiles just mentioned) or allow the subject (expert) to define their own probability categories. Researchers (Browne, Curley, and Benson, 1999) have found that use of subject-defined categories involved a tradeoff in performance: calibration generally became worse as the number of categories increased, but discrimination generally improved.

There are several pitfalls involved with the verification or calibration of probability forecasts, including the key one of calibrating or comparing abilities of assessors ,forecasters) on the basis of assessments that are not comparable to the issue under study (Panofsky and Brier, 1968). Also, interpretation of results can be made difficult through indiscriminate combining of unrelated results to form a single index to be used for comparison purposes. Panofsky and Brier (1968) point out that lack of care in elicitor-provided classes (ranges of probability), such as the use of overlapping classes, can tend to encourage forecasters to hedge by choosing classes with the widest range. Morgan and Henrion (1990) report that unclear feedback to experts regarding their performance can actually lead to worse subsequent results, because of introduced bias that could, for example, result in increased overconfidence. Lichtenstein and Fischhoff (1980) indicate that feedback should include personal discussion of results, since that may be less easy to dismiss than a written numerical summary.

A summary of the key relevant literature applicable specifically to expert judgment calibration questionnaire features and related considerations is given in Table 2.

Table 2. Summary of Calibration Questionnaire-related Literature

Author(s)	Major Points/Findings	Current Research Implications
Lichtenstein & Fischhoff (1980)	Feedback improves learning; should include personal discussion of results	Provide feedback to respondents, stressing personal interaction
Miller & Byrnes (1977)	Links risk-taking philosophy with over- or underconfidence	Use in design of calibration questionnaire
Wang (2001)	Describes risk-taking in non-wager terms	Use concept in design of calibration questionnaire
Ayyub (2001)	Elicitation training can identify sources of bias	Apply to careful design of questions
Panofsky & Brier (1968)	Avoid calibration assessments not related or comparable to issue under study	Tailor calibration questions to reflect appropriate technical "flavor"
Browne, Curley & Benson (1999)	Using subject-defined (response) categories involves tradeoff in performance	Use elicitor-provided categories to eliminate potential disparity among experts

Research Question

Based on the diverse work reported herein, it is seen that tools exist to support the current study. Further, it is also evident that there is a firm basis for moving beyond the immediate effort to the ultimate goal of developing a comprehensive modeling aid for technology cost or impact assessments for advanced launch vehicle operation , support and performance. The question that has been answered by this research is: can a "calibration" function be developed to apply to experts' assessment of technology impacts in order to improve accuracy of those assessments when applied to aerospace

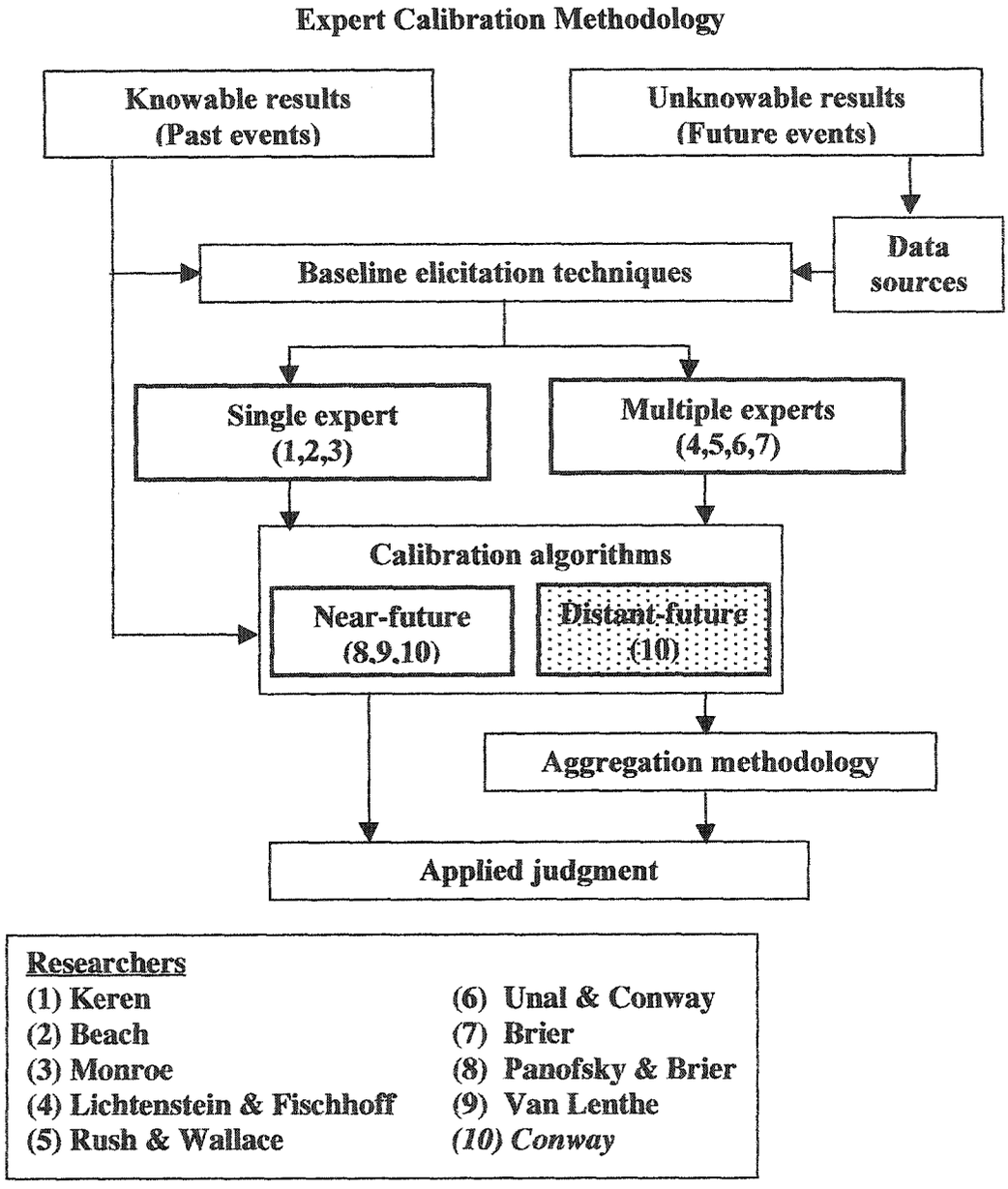
vehicle development? The purpose of the research is to ascertain a means for effectively calibrating experts whose judgments may not be validated for many years.

Significance of the Research

The vast majority of studies on calibration of expert judgments involving probability assessments have dealt with outcomes that could be observed or recorded, either as past events or occurrences or as near-term future events. Little to no applications have been found that address calibration of likelihood estimation of uncertain events in the distant future, where even problem boundary and constraints may be nebulous. Such is the case with the thrust of the proposed effort – the calibration of expert assessments related to operations and support, weights and sizing, and multidisciplinary design considerations in future aerospace vehicle concepts employing many as-yet-unproven technology advances. In particular, the use of multiple experts can exacerbate the assessment problem without some means of calibrating widely divergent raw predictions (Unal and Conway, 2000).

Development of a more robust system, including expert judgment calibration, which can efficiently handle the various disciplines will yield a tool for conceptual designers of advanced-technology systems to assess ultimately critical weight, cost, and multidisciplinary integration impact questions in an upfront (more timely) manner. A properly validated tool can be expected to provide higher levels of confidence in these earlier assessments, resulting in a decision aid for program planners and advocates. Figure 1 places the current research into context with past and current work in this field.

Figure 1. Setting for Current Research



The current effort in development of calibration algorithms which can be applied in either single- or multiple-expert scenarios is thus seen to fill a prominent void in the expert judgment calibration methodology.

CHAPTER III

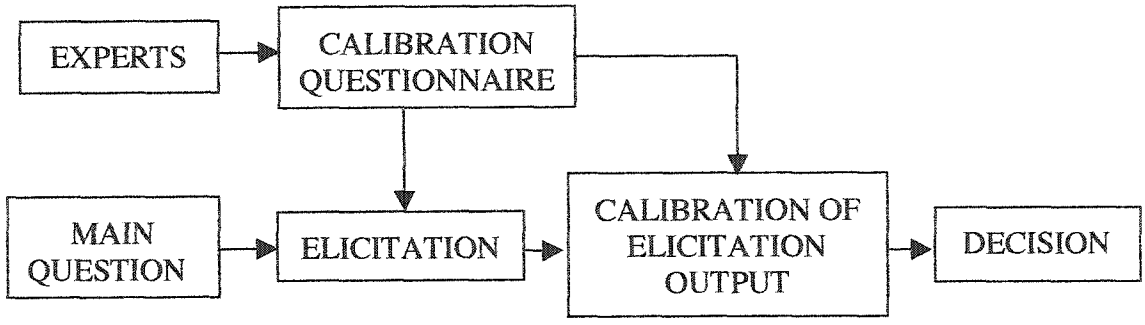
RESEARCH METHODOLOGY

Introduction

The thrust of this work was the development of calibration algorithms to apply to elicited expert judgment information on operation and support, weight and size, and multidisciplinary system requirements for advanced launch vehicles. The purpose is to aid designers in determining a "best" expert estimate of the value of discipline-related parameters in achieving projected performance in the realization of new systems. For example, in the operations and support discipline, improved supportability implies a reduction in the number of failures recorded against a system (measured as a percentage) that then require maintenance actions in order to return the system to flight readiness. It also implies the same reduction in the time and manpower required to maintain and service the system.

Many expert judgment elicitation scenarios involve events whose occurrence can be validated, because they are either past events or near term future events. In such cases, calibration of the expert assessors can include feedback on their performance, which could be expected to improve future performance (self-calibration). In the present research problem application, however, the preponderance of occurrences being assessed are in the distant future – as much as 20 or 30 years. Feedback involving actual results or occurrences is impossible. It is imperative, then, that a calibration technique be found for use by decision makers that does not rely on feedback for credible estimates. Thus, in the present research, an external calibration based on a pre-elicitation calibration questionnaire was sought. Figure 2 presents a schematic of this concept.

Figure 2. Calibration Technique



Approach

Calibration Algorithm Development

Fuzzy logic and Bayesian statistical techniques were employed to develop an Expert Calibration Function (ECF) based on degree (level and time) of past experience and current philosophy. For this study, the ECF has been developed for experts in several technology areas associated with advanced launch vehicles. A simple questionnaire was designed to "pigeon-hole" a responding expert into one of a set of experience classification categories, essentially a self-designation of expertise ,Wright, Rowe, Bolger and Gammack, 1994). A second part of the questionnaire attempts to place the expert in his or her natural confidence level category, such as overconfident (presumes higher success probability than is actually achieved), underconfident (presumes lower), or neutral (places the correct probability). This was achieved through the use of utility theory and the outlining of several "wagers" (or choices of options) related to topics with which the experts are familiar. The questionnaire (and the validation discussed in the next section) were administered to a pilot group of several experts at the Langley Research Center.

From the experience and philosophy responses, calibration factors are determined such that an adjusted probability distribution for the expert's uncertainty for each parameter or analysis tool considered in the elicitation questionnaire can be subsequently constructed. The adjustment takes the following general form, where E is expertise, P is philosophy (confidence level), μ and σ^2 are statistics from the parameter uncertainty distribution, and A_1 and A_2 are arbitrary constants (which will be initially set to 1 for this study) of the adjustment relations:

$$\begin{aligned} \Delta\mu &= f(E,P,\mu)A_1 \\ \Delta\sigma^2 &= f(P,\sigma^2)A_2 \end{aligned} \tag{1}$$

It should be noted that the adjustment factors so determined will only be placeholder estimates until validated (and possibly modified) through a validation procedure as outlined next.

Expert Calibration Function Validation

Validation of the calibration methodology and resulting calibration functions is accomplished through an interrater comparison between initial and calibrated results from multiple experts participating in trial testing of an overall expert judgment elicitation, calibration, and aggregation methodology. One of the principal motivators for the current research was the (sometimes wide) disparity among multiple experts addressing the same uncertainty-related questions. A successful reduction in disparity among expert respondents' results would suggest at least partial validation of the calibration methodology. Further support would be provided by "movement" of the results from

respondents with a somewhat lower level of expertise toward results from respondents possessing a higher level of expertise.

Calibration Function Reliability

Consistency of the expert calibration function's performance, or reliability, is expected to be high, given the mathematical nature of its form. This assumes that interpretation by experts of questions in the Background section of the expert elicitation instrument is consistent. Care was taken in the phrasing of the questions, and formulation followed guidance from the literature on similar constructions (see, for example, Monroe, 1997 and Duarte, 2001). Administration of the questionnaires was such that each participating expert responded without consultation or influence (bias) from other experts.

CHAPTER IV

APPLICATION OF THE METHODOLOGY

Application of Technique – Overview

Working with program officials from NASA Langley Research Center, three aerospace vehicle disciplines and an example conceptual design case were selected to apply the calibration technique in conjunction with other ongoing expert judgment elicitation and aggregation research. In this larger scale research effort, a survey for eliciting expert judgment for the selected disciplines has been developed. Included in the questionnaire development is elicitation of background data on experts for the purpose of calibration. The satisfaction of Institutional Review Board requirements for protection of experimental subjects was achieved through careful design and handling of the questionnaire instruments. The final survey design is capable of being administered to selected experts via the World Wide Web.

The administration of the overall expert judgment elicitation questionnaire, including calibration-specific questions, was accomplished by querying discipline experts at NASA Langley Research Center. The questions were administered using a Microsoft Excel® spreadsheet, on which responses were entered for subsequent data collection and analysis. Because expert participants will likely be in geographically dispersed locations, and responding to the expert elicitation questionnaire by the several experts involved will be asynchronous, use of web-based tools is deemed crucial to the efficient collection of information. Accordingly, automated web-based survey software, Inquisite® (Catapult Systems, Austin TX), has been used to develop a web-based version of the expert

judgment elicitation questionnaire for future application. The Inquisite® web version was also used in parts of the current study.

Calibration Questionnaire Design and Implementation

Specific questions were designed to be used in a Background section of the expert judgment elicitation questionnaire, based on previously noted findings from the literature review. To ascertain level of expertise, questions to ascertain the participating expert's self-assessment of his own expertise were posed, per Wright, Rowe, Bolger and Gammack (1994). Another background question asked the responding expert to compare his degree of expertise in the discipline being addressed with those of his peers in the discipline. This was intended to provide a second indicator of the expert's self-designated level of expertise related to a more absolute scale. Also, age was included as a requested background response, in accordance with findings by MacCrimmon and Wehring (1986) and Crawford and Stankov (1996) that expertise can be related to the age of an elicitee.

Several Background questions attempted to place an expert on a continuum with respect to his confidence level, or comfort with expert judgments rendered in response to an elicitation. The first of these was included as the second part of a question designed to gauge an expert's knowledge of the discipline by asking a discipline-specific question (with a numerical answer) that practitioner experts would be able to answer. Another set of questions was designed to help ascertain the expert's assessment of his attitude or philosophy with respect to manifesting confidence in judgments made in his specific field (discipline). These questions' purpose was to develop a baseline to help gauge response to the final set of Background entries.

The last Background questions consisted of a series of options for which the participating expert was asked to make a choice. The available choices for each set of options were designed to reflect (1) a more risky situation whose choice would imply a tendency toward overconfidence, or (2) a less risky situation whose choice would signal a tendency away from overconfidence and toward neutrality or underconfidence.

The relatively brief Background section to the expert judgment elicitation questionnaire provided the necessary input to develop a calibration function for the responding expert. The complete Background section is shown in Figure 3.

Figure 3. Background Question Section of Questionnaire (Weight and Sizing Example)

BACKGROUND (Weight and Sizing Specific)

1. Name or USERID: _____

2. Your age _____

3. In this subject area, rate your own level of expertise on a scale of 1 (least) to 5 (most) _____

4. Think of others with similar experience working in this discipline. On a scale of 1 (much less than peers), to 3 (about the same), to 5 (much more than peers), how would you compare yourself to your peers with respect to expertise? _____

5. Payload mass fraction for the Space Shuttle is _____. Your assessment of the probability that your estimate is correct: _____. [Discipline-specific]

6. Think about predicting weights of hardware system elements; do you usually predict more than actually occurs (5), less than actually occurs (1), or about the amount/number of times that actually occur (3)? ___

7. In estimating in your subject area in areas that have associated uncertainty, do you think it is better to be (a) close to the actual value without a lot of confidence in the estimate, or (b) not very close to the actual value, but with a high degree of confidence in your estimate? ___

8. In making estimates related to weight and sizing model input parameters, would you say you were, (a) usually right-on with a high degree of confidence, (b) right-on without a high degree of confidence, or (c) not very close but with a high degree of confidence, or (d) not very close, and with not much confidence _____

For the following pairs of choices, please select the one in each pair that is most comfortable or appealing to you:

Figure 3. Background Question Section of Questionnaire (concluded)

9. (a) Setting, in advance, the completion date for a multi-year project OR (b) Establishing, in advance, technical milestones for a multi-year project 10. (a) Estimating, in advance, total cost outlays for a multi-year project OR (b) Identifying, in advance, cost elements for a multi-year project 11. (a) Identifying, at conceptual design review, utilization scenarios for the successful project OR (b) Predicting, at conceptual design review, technical performance characteristics of the completed hardware

Principal Expert Judgment Questionnaire Design and Implementation

Introduction

In order to evaluate a conceptual launch vehicle, the vehicle must be defined in terms of performance characteristics (length, width, height, thrust level, payload delivery capability, etc.) These performance characteristics are the direct result of the vehicle configuration and mission requirements. For each discipline of interest in conceptual launch vehicle design, the disciplinary analysis tools have estimating relationships (ER) associated with them. Each estimating relationship may be comprised of a set of parameters, which define the ER. For example, in the weights and sizing discipline, a subsystem may be the wing, the ER may be the wing surface area and the parameters

would be a set of configurable inputs which are used to compute the estimated wing surface area value.

For each of the discipline questionnaires used in this study, a list of input parameter variables (with associated nominal values) for a vehicle concept were compiled by subject matter experts associated with the conceptual design team. A form of the classical Nominal Group Technique was employed to identify the most highly uncertain input parameters from the list, using a Pareto principle approach.

Questionnaire Flow Process

The questionnaires included instructions for the respondent as well as the calibration-related Background section discussed previously. The questionnaire methodology follows that of Monroe (1997). This methodology assumes a default symmetrical triangular probability distribution associated with expert's assessed uncertainty about parameter values, where the distribution variance is proportional to the level of uncertainty. However, to avoid possibilities of respondents contradicting themselves with respect to the shape (skewness) and variance of assumed uncertainty distributions, an instruction was included early in the parameter uncertainty quantification sequence (rather than later as in Monroe) to permit the specification of a skewed (triangular) distribution. Figure 4 presents the set of instruction for the experts responding to the questionnaire; instruction 3 reflects the key modification made to Monroe's methodology.

Figure 4. Expert Elicitation Questionnaire Instructions

A list of (discipline specific model) input parameters whose values are potentially uncertain will be provided on a subsequent screen. You will be asked to evaluate these parameters using the following guidelines.

1. Rate each INPUT parameter uncertainty QUALITATIVELY using a 5-point rating scale (Low, Low/Moderate, Moderate, Moderate/High, High). Focus only on those INPUT parameters that you feel should be evaluated in this manner.
2. If you feel a parameter's default value should be modified, you may provide a new point estimate for the nominal value.
3. If you feel the range of possible values (due to uncertainty, physical limitations, design constraints, etc.) around the nominal value is not symmetrical, please provide your own estimates of minimum and maximum values.
4. Describe the reason for the uncertainty and the reasoning behind the parameter value ranges for the UNCERTAIN INPUTS that you rated. Include a rationale for those parameters to which you have assigned new nominal values. Do this simultaneously while rating each INPUT parameter to document your thinking.
5. Think of any other cues (or reasons that you have not documented) and record that information at this time.
6. Once the INPUT parameters provided have been rated for uncertainty, you may add parameters not shown which you assess to have a level of uncertainty associated with their value. Use the OTHER option listed at the bottom of the INPUT parameter listing for this purpose.
7. After rating all INPUT parameters, next anchor your Low, Moderate, and High QUALITATIVE measures of uncertainty to QUANTITATIVE measures on the 5-point scales (provided).
8. Describe any scenarios that may change INPUT parameter values. Provide the alternate INPUT parameter values that in your judgment would be appropriate for the scenario

Each expert, working alone, was asked in the questionnaire to consider each of the input parameters in turn for what he believed to be the degree of uncertainty associated with it. For those which he believed to have no uncertainty he simply accepted the provided nominal value. For those which were deemed to have a degree of uncertainty associated with them, additional questions were asked of the expert to ascertain his estimate of the amount of uncertainty. He was asked to rate each input parameter uncertainty qualitatively using a 5-point rating scale (Low, Low/Moderate, Moderate, Moderate/High, High). If the expert believed a parameter's default value

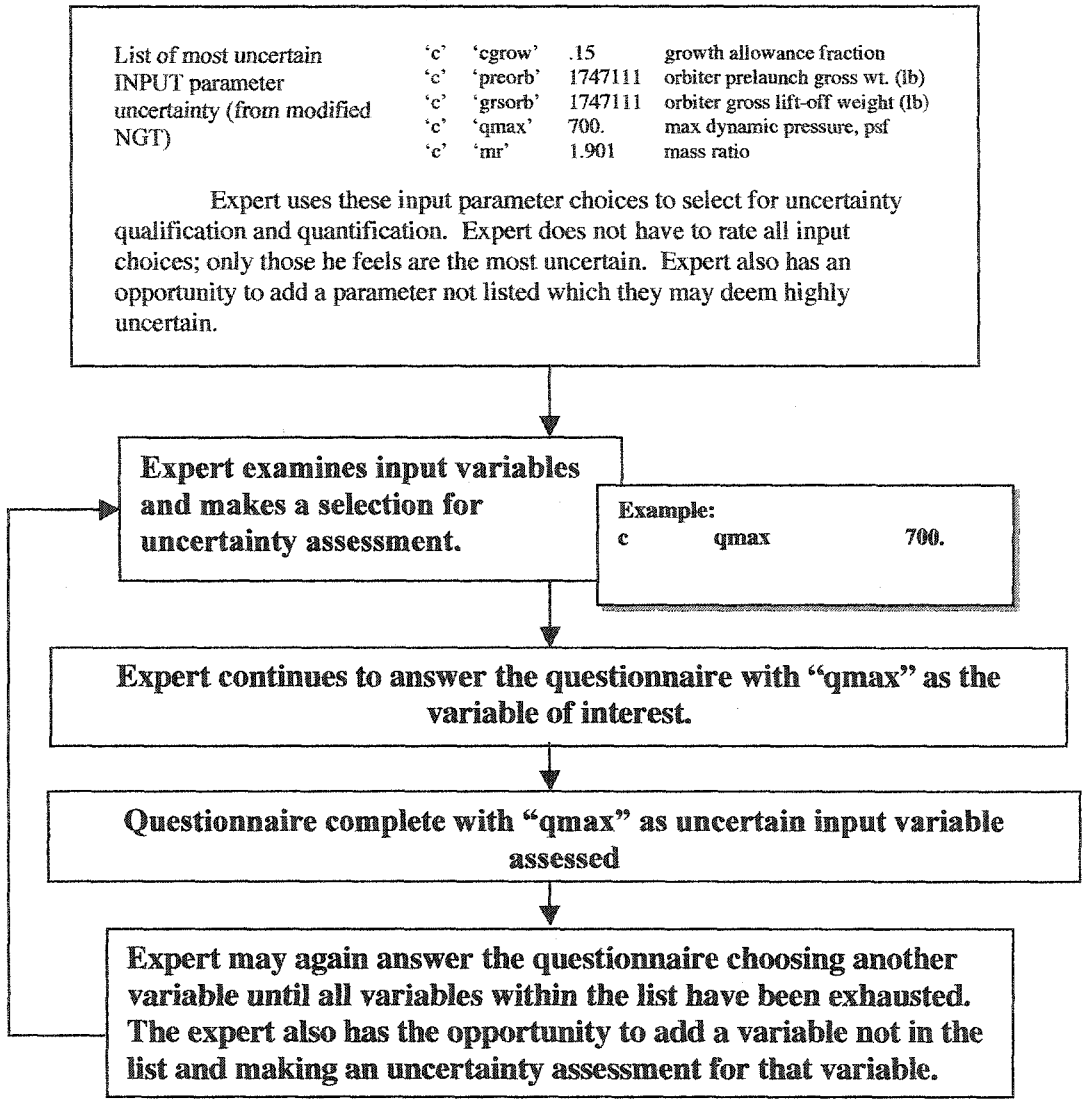
should be modified, he was asked to provide a new point estimate for the nominal value. He was also allowed to establish a nonsymmetrical distribution of he deemed appropriate.

The expert was next asked to describe the reason for the uncertainty and the reasoning behind the resultant parameter value ranges for the uncertain inputs that he rated. He was asked to include a rationale for those parameters to which he assigned new nominal values.

The expert subject was then asked to think of any other cues (or reasons that he had not documented) and record that information. Once the input parameters provided had been rated for uncertainty, the expert was given a chance to add parameters not shown which he believed to have a level of uncertainty associated with their value. After rating all input parameters, the expert anchored his Low, Moderate, and High qualitative measures of uncertainty to quantitative measures on 5-point scales provided. He was asked to describe any scenarios that could change input parameter values, and to provide the alternate input parameter values that in his judgment were appropriate for the scenario.

Figure 5 illustrates the questionnaire response process, using Weights and Sizing as an example discipline. A complete questionnaire for the Weight and Sizing discipline is presented in Appendix A as an example.

Figure 5. Questionnaire response flow schematic



Institutional Review Board Considerations

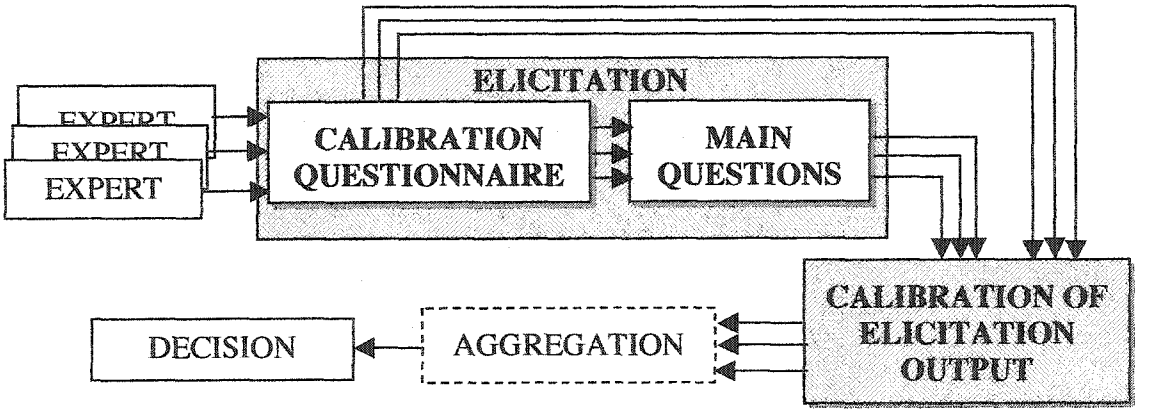
Questionnaires were developed for two disciplines for this study: weights and sizing, and operations and support. In addition, for each discipline separate questions were developed to address analysis tool uncertainty for the discipline. Examples of complete questionnaires for these disciplines and uncertainty type are given in Appendix A. The questionnaires and the questionnaire application process were reported to

Institutional Review Board (IRB) representatives at Old Dominion University, and copies of the questionnaires were furnished. It was concluded that this research would qualify for an exemption from full IRB procedures for human subject research based on the questionnaire output NOT being damaging in any way (civil or criminal liability, employability, or financial) to subject participants, and NOT dealing with sensitive aspects of any subject's behavior.

Data Collection and Handling

For this study, the experts provided input via Microsoft Excel® spreadsheet questionnaires. The information from the Background portion of each expert's questionnaire was maintained in a separate file for subsequent analysis and application to that expert's responses to the main uncertainty elicitation questionnaire sections. The responses to the uncertainty portions of the questionnaires were also maintained in separate files for the calibration application, and will be subsequently processed in trials of aggregation schemes that are being studied in a related investigation. Figure 6 is a schematic of the entire expert elicitation, calibration, and aggregation process.

Figure 6. Expert Data Collection and Handling Process



Data Analysis

Expertise

Data analysis was carried out using analysis routines in Microsoft Excel®. The first analysis performed was the determination of each participating expert's level of expertise, based on his or her response to Questions 2 – 5 of the Background portion of the questionnaire. The following relation was used in the determination.

$$\text{Expertise, } E = \frac{5\left[\frac{AGE}{60}\right] + SEL + ECP + 5\left[\frac{1}{1+\left|\frac{DSK - ACT}{ACT}\right|}\right]}{4} \quad (2)$$

The variables AGE, SEL, ECP, DSK, and ACT are described in Table 3.

Table 3. Definitions of Expertise Relation Variables

Variable in Expertise Relation	Definition
AGE	Expert's age, years
SEL	Expert's self-designation of expertise in the discipline area being elicited, scale of 1 (least) – 5 (most)
ECP	Expert's perception of expertise compared to peers in the discipline area being elicited, scale of 1 (least) – 5 (most)
DSK	Expert's numerical response to question involving discipline-specific knowledge
ACT	Actual (true) value of the response to the question involving discipline-specific knowledge

Since none of the research cited in the literature presented any rationale for differentiating weighting factors affecting expertise, equal weights were assigned to the four factors age, self-assessment of expertise, perceived expertise compared to peers, and discipline-specific knowledge. These initial weights could be allowed to vary in order to ascertain if different values would yield more consistent results in the main questionnaire elicitation responses. However, if responses appear consistent, no adjustments would be made.

Confidence/Risk Philosophy

Background responses to questions 6 – 11 were compiled in a confidence/utility (risk) philosophy profile, as shown in Figure 7 with an example set of responses. Table 4 maps questionnaire response options for questions 6-11 to the scale used in Figure 7.

Figure 7. Confidence/Utility Philosophy Profile, from Background Questionnaire

		PDO Q6	PVvC Q7	PerfVvC Q8	CDvTM Q9	TCvCE Q10	USvTPC Q11
RISK TOLERANT (OVERCONFIDENT)	5		X	X		X	X
	2.5	X					
RISK NEUTRAL	0						
	-2.5						
RISK AVERSE UNDERCONFIDENT	-5				X		

Table 4. Questionnaire Response (Questions 6-11) vs. Philosophy Profile Scale

Question	Assigned Philosophy Profile Response				
	5	2.5	0	-2.5	-5
6	5		3		1
7	(b)				(a)
8	(c)	(a)		(b)	(d)
9	(a)				(b)
10	(a)				(b)
11	(b)				(a)

Although not all questions from numbers 6-11 permitted responses that covered every philosophy "level," the mapping shown in Table 3 proved useful in defining an overall confidence or risk philosophy factor for an expert.

From the confidence/risk profile, a confidence/risk philosophy was defined as

$$P = \sum \frac{[\text{Responses, Questions 6-11})]}{6} \tag{3}$$

Here the responses could take on values from –5 to 5, according to the criteria in Table 3.

Determination of Adjustment Factors

The expertise and confidence philosophy determined from the Background responses are utilized to adjust the mode, c (most likely value), of that expert's uncertainty distribution for a parameter according to the following:

$$\Delta c = c\left\{-\text{sign}(P)\left[1-\frac{E}{5}\right]\right\} A_1 \tag{4}$$

Similarly, the confidence philosophy was utilized in determining a variance adjustment factor as shown in the following:

$$\Delta\sigma^2 = \sigma^2\left[\frac{P}{5}\right]A_2 \tag{5}$$

For symmetrical distributions, the mean and mode are equivalent; they are different for unsymmetrical distributions. For a symmetrical triangular distribution, the adjusted distribution's parameters are given by the following:

$$\begin{aligned} \mu_2 &= \mu_1 + \Delta c \\ \sigma_2^{\,2} &= \sigma_1^{\,2} + \Delta\sigma \\ c_2 &= c_1 + \Delta c \end{aligned} \tag{6}$$

The new distribution's endpoints are calculated using the above by solving the simultaneous equations for mean and variance of a triangular distribution:

$$\begin{aligned} \mu_2 &= \frac{a_2 + b_2 + c_2}{3} \\ \sigma_2^{\,2} &= \frac{a_2^{\,2} + b_2^{\,2} + c_2^{\,2} - a_2b_2 - a_2c_2 - b_2c_2}{18} \end{aligned} \tag{7}$$

When an expert judges a parameter's distribution to be unsymmetrical (skewed), it will be assumed that this judgment reflects (possibly) some physical or other constraints that must be obeyed. Thus, for the skewed distribution cases, the end points will be taken as fixed, and

$$c_2 = c_1 + \Delta c \tag{8}$$

The mean and variance of the new distribution will then be calculated from equations (7) above.

As suggested in the previous chapter, arbitrary constants denoted A_1 and A_2 may be used in equations (4) and (5) and allowed to vary, ultimately resulting in a "best fit" calibration that maximizes the interrater reliability among multiple experts responding to the same elicitation questionnaire (after James, Demaree, and Wolf, 1984). This type of

parametric or optimization study was not performed as part of the present research, but is a strong candidate for future expansion of the current methodology development.

Interrater reliability is defined as the degree to which judges "agree" on a set of judgments (James, Demaree, and Wolf, 1984). Interrater reliability can be defined as a proportion, or ratio, of systematic variance to total variance for the set of judgments. In the present case, this will be given as follows:

$$r = \frac{J[1-(\bar{\sigma}_2^2/\sigma_{EU}^2)]}{J\left[1-(\bar{\sigma}_2^2/\sigma_{EU}^2)\right]+(\bar{\sigma}_2^2/\sigma_{EU}^2)} \tag{9}$$

In equation (9), r is interrater reliability, J is the number of experts participating on a given set of questions, $\bar{\sigma}_2^2$ is the mean variance among the J experts for the set of questions, and σ_{EU}^2 is the expected variance if every choice of parameter value by the experts was equally likely (from a continuous uniform distribution).

While the data analysis methodology presented thus far could be extended to types of distributions other than the triangular distribution used as an example for the parameter uncertainty judgments elicited from each expert, the questionnaire methodology does result in a triangular distribution being the default function. Accordingly, for purposes of the present research, calibration adjustments are assumed made to a triangular distribution to yield another (adjusted) triangular distribution. The adjusted distributions from multiple experts then becomes the input for a follow-on aggregation process, which might be expected to take into account such factors as level of expertise or other weighting factors associated with the experts and their elicitation.

Methodology Application Example

To demonstrate the application of the calibration methodology presented in this report, the following example will be used. It is assumed that an expert has completed the background section of the expert elicitation questionnaire in the manner shown in Table 5:

Table 5. Example Background Input

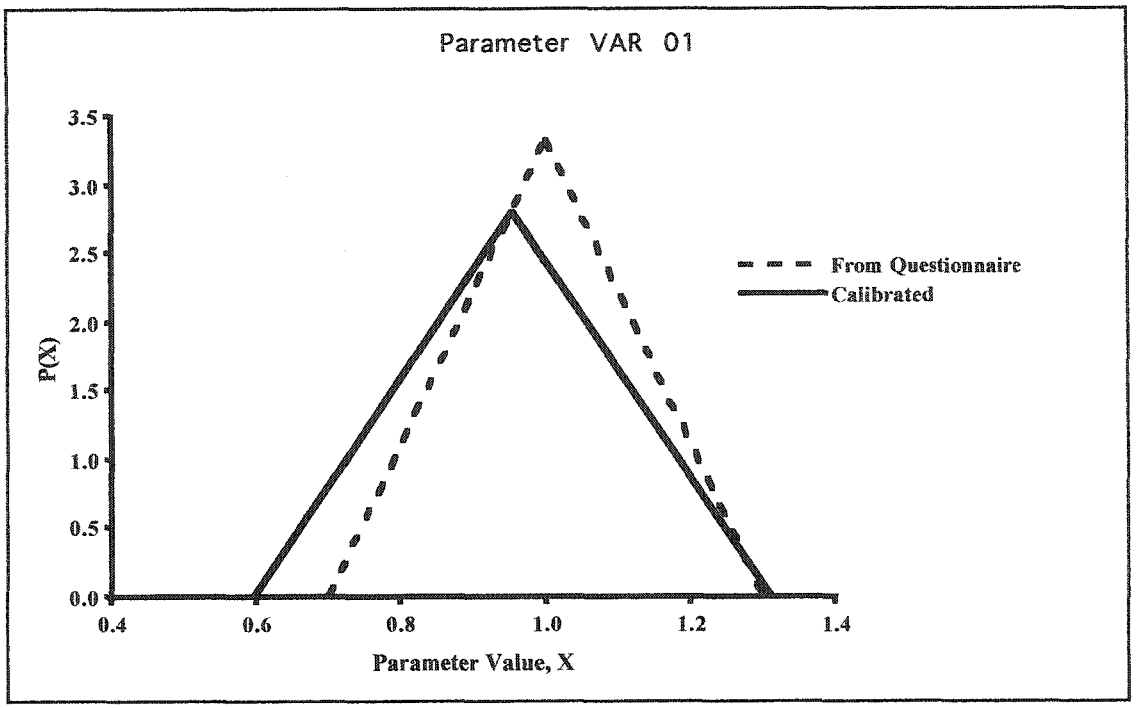
Question No.	Topic	Response	Point Conversion
EXPERTISE			
1	Name or USERID	5551212	-
2	Age (years)	50	50
3	Self-rated expertise	Most	5
4	Expertise compared to others in field	Much more	5
5	Discipline-specific knowledge question	300	
CONFIDENCE/RISK PHIOSOPHY			**(Table 4)**
6	Predicting discipline-related quantities	(3)	0
7	Estimating uncertainty preference	(a)	-5
8	Estimating trend in discipline	(a)	2.5
9	Completion vs. milestones	(a)	5
10	Total outlays vs. cost elements	(a)	5
11	Utilization scenarios vs. performance	(b)	5

From the responses in Table 5, given that the actual value of the discipline knowledge question parameter (Question 5) is 295, the expertise of the subject is calculated from equation (2) as 4.448. This value places the expert high in the expertise category. The expert's confidence/risk philosophy determined from equation (3), is 2.083. This indicates a mild tendency to overconfidence, or risk tolerance.

It is next assumed that the expert evaluates a sample parameter whose nominal value is given as 1. Accepting the symmetrical nature of his uncertainty about this parameter, the example expert rates the uncertainty as "moderate"; he subsequently assigns a quantitative value of 30 per cent as his interpretation of "moderate" uncertainty. Interpreting the expert's assessment of uncertainty as a triangular probability distribution, the parameters of the distribution (lowest value most likely value, highest value) are given as (0.7, 1.0, 1.3).

Applying the adjustments to mean and variance given by equations (4) and (5), parameters for the calibrated distribution are found to be (0.53, 0.89, 1.25). The adjusted distribution has lowered the most likely value (also the mean of the distribution) and has increased the variance, commensurate with the expertise and philosophy of the expert ,following, among others, Wright, Rowe, Bolger, and Gammack, 1994, MacCrimmon and Wehring, 1986, Crawford and Stankov, 1996, and Duarte, 2001. Figure 8 compares the two distributions: uncalibrated (from the questionnaire response) and calibrated.

Figure 8. Results of Sample Methodology Application



The methodology described in this chapter was applied to cases utilizing advanced conceptual vehicle parameters provided by NASA-Langley. The results of these cases are presented in the next chapter.

CHAPTER V

RESULTS

Case Descriptions

Weights and Sizing

Two separate cases were considered to exercise the calibration methodology developed and described in the preceding chapter. The first case involved weight and sizing study input parameter uncertainty for a two-stage-to-orbit (TSTO) reference configuration (orbiter and booster), which features staging at Mach 3. The initial list of user-defined parameters for this case is given in Appendix B (booster and orbiter parameters are listed separately). Additional parameters are also provided as "pass-through" parameters from another application. The parameters are utilized in a NASA configuration sizing program (CONSIZ) developed at the Langley Research Center to size a vehicle and determine the weights of its components. CONSIZ provides the capability of sizing and estimating weights for a variety of aerospace vehicles using weight estimating relationships based on historical regression, finite element analysis, and technology readiness or maturity level. Within CONSIZ, the vehicle is modeled as a collection of components representing structure, subsystem, and propulsion elements (Monroe, Lepsch, and Unal, 2002).

A modified Nominal Group Technique (NGT) evaluation of the initial user-defined parameter list was conducted by NASA project team personnel to identify which of the parameters would be expected to have the most impact on vehicle performance. The modified NGT yielded parameters rated by each team member from 1 (most impact)

to 5 (least impact) for the 112 user input parameters in the list. Ratings for each parameter were totaled over all team members.

From the list of rated parameters, assuming adherence to the Pareto principle, the 20 per cent of parameters deemed to have the most impact were selected for inclusion in the uncertainty judgment elicitation questionnaire. The reduced parameter list for the booster and the orbiter is given as Appendix C.

For this case, two experts participated in the elicitation. They received questionnaires via e-mail, and were given a week to provide responses. They worked at their own pace, in their normal work setting (office). Background questionnaire section results for these experts are presented in Appendix D, along with interpretation of the responses in terms of the expertise and confidence or risk philosophy ratings calculated.

Operations and Support

The second case evaluated expert judgments of uncertainty involving input parameters associated with a NASA reliability and maintainability analysis tool (RMAT), used at NASA-Langley. The target vehicle was the same one described in the weight and sizing section above – a TSTO vehicle with Mach 3 staging. RMAT is based on evaluating comparability between support requirements for current operational aircraft and launch vehicles and proposed future vehicle concepts. Using RMAT, operational characteristics such as mission completion reliability, maintenance actions per mission, manpower and support requirements can be estimated for a particular vehicle concept and mission scenario. RMAT is a complex, stand-alone, operational analysis code requiring expert user inputs (Unal, 2002). The reduced RMAT input variable list for the booster

and the orbiter, after applying the modified NGT discussed above, is given as Appendix E.

Three experts participated in the operations and support case elicitation, working in an environment and time frame similar to that of the weights and sizing experts. Background questionnaire section results for these operations and support experts are presented in Appendix F, along with interpretation of the responses in terms of the expertise and confidence or risk philosophy ratings calculated.

Case Study Output – Uncertainty Distributions

Weights and Sizing

The reduced parameter lists for both the booster and orbiter in the weights and sizing case contained variables that either were fixed, such as a physical constant that could be maintained during a launch mission (liquid oxygen density is an example) or that would be held constant in any analyses (such as the number of common booster stages). These parameters had virtually no uncertainty associated with them and were thus not considered in further probability distribution calibration analysis. In addition, there were other variables that one expert felt could not be addressed without knowledge of an assumed technology and for which the other expert established a triangular distribution with tight tolerances, thus essentially fixing the final distribution. These also were omitted from the calibration analysis. The remaining variables were addressed consistently by both experts and were included in the remainder of the analyses. Appendix G presents the data for the weights and sizing case, giving both the expert's

input-derived uncertainty distributions and the distributions obtained by applying the calibration algorithms derived in this study.

Operations and Support

Experts participating in the operations and support case elicitation provided consistent responses on most variables in the reduced parameter list. As in the weights and sizing case, there were a few variables for which uncertainty was either not present (defined constants for a study) or for which application of the continuous-distribution-oriented methodology made little sense (such as in determining a number of work shifts to assign. These parameters were excluded from the summary analysis. Appendix H provides uncertainty distribution data for the operations and support case in a similar format as that of Appendix G.

Case Study Output – Interrater Reliability

A key measure of the efficacy of the present calibration methodology development is the degree to which interrater reliability (equation (9)) for the group of experts responding to a given elicitation questionnaire improves following calibration. Thus, interrater reliability was determined for both the initial set of uncertainty distributions and the calibrated set, for each of the two cases studied. Tables 6 and 7 present in tabular form the interrater reliability results for the initial and calibrated distribution. Note that the interrater reliability, r, is determined for each variable addressed in the output.

Following Tables 6 and 7, Figures 9 – 12 present the interrater reliability comparisons in graphical form for the weights and sizing and operations and support cases. Discussion of the results presented will follow in the next chapter.

Table 6. Interrater Reliability for Weights and Sizing

Interrater Reliability, Booster						
	Initial IRR			Calibrated IRR		
	$\overline{\sigma_1^2}$	$\sigma^2_{EU_1}$	r_{init}	$\overline{\sigma_2^2}$	$\sigma^2_{EU_2}$	r_{cal}
VAR 02	0.0011	0.0051	0.8790	0.0005	0.0014	0.7888
VAR 03	0.0004	0.0033	0.9333	0.0004	0.0166	0.9883
VAR 08	0.0015	0.0053	0.8293	0.0015	0.0052	0.8365
VAR 09	0.0247	0.0824	0.8240	0.0106	0.0516	0.8854
VAR 11	18.9444	56.3333	0.7979	13.2917	108.7180	0.9349
Interrater Reliability, Orbiter						
	Initial IRR			Calibrated IRR		
	$\overline{\sigma_1^2}$	$\sigma^2_{EU_1}$	r_{init}	$\overline{\sigma_2^2}$	$\sigma^2_{EU_2}$	r_{cal}
VAR 03	0.0015	0.0165	0.9517	0.0007	0.0119	0.9714
VAR 12	252.3000	908.2800	0.8387	107.2275	413.0948	0.8509
VAR 13	52.0833	133.3333	0.7573	24.4792	79.2510	0.8174
VAR 14	425.5894	1004.6700	0.7313	191.8641	727.8784	0.8482

Table 7. Interrater Reliability for Operations and Support

Interrater Reliability, Booster						
	Initial IRR			Calibrated IRR		
	$\overline{\sigma_1^2}$	$\sigma^2_{EU_1}$	r_{init}	$\overline{\sigma_2^2}$	$\sigma^2_{EU_2}$	r_{cal}
VAR 01	159886574	833333333	0.9267	175100492	833333333	0.9186
VAR 04	0.0590	0.1302	0.7834	0.0610	0.1302	0.7730
VAR 05	0.0005	0.0012	0.8067	0.0004	0.0012	0.8470
VAR 06	50616.667	100833.333	0.7485	50554.912	100833.333	0.7490
VAR 07	0.9063	3.0000	0.8739	0.7500	8.7579	0.9697
VAR 08	38.7272	75.0000	0.7375	38.1279	75.0000	0.7437
VAR 09	0.2222	0.3333	0.6000	0.2143	0.3333	0.6248
VAR 10	0.0178	0.0533	0.8565	0.0176	0.0533	0.8591
VAR 12	0.2901	1.3669	0.9176	0.3037	1.9436	0.9419
VAR 14	0.0000	0.0000	0.9109	0.0000	0.0000	0.9646
VAR 15	0.0003	0.0010	0.8540	0.0003	0.0015	0.9253

Table 7. Interrater Reliability for Operations and Support (concluded)

Interrater Reliability, Orbiter						
	Initial IRR			Calibrated IRR		
	$\overline{\sigma_1^2}$	$\sigma^2_{EU_1}$	r_{init}	$\overline{\sigma_2^2}$	$\sigma^2_{EU_2}$	r_{cal}
VAR 01	332274478	1752083333	0.9276	355529122	1752083333	0.9218
VAR 04	0.0590	0.1302	0.7834	0.0610	0.1302	0.7730
VAR 05	0.0002	0.0014	0.9472	0.0002	0.0014	0.9462
VAR 06	50272.815	100833.333	0.7511	50251.907	100833.333	0.7512
VAR 07	0.9063	3.0000	0.8739	0.7500	8.7579	0.9697
VAR 08	38.7272	75.0000	0.7375	38.1279	75.0000	0.7437
VAR 09	0.0749	0.3675	0.9213	0.0730	0.3675	0.9236
VAR 10	0.0139	0.0300	0.7776	0.0137	0.0300	0.7814
VAR 11	63.9450	211.6800	0.8739	52.9200	617.9593	0.9697
VAR 12	0.2901	1.3669	0.9176	0.3037	1.9436	0.9419
VAR 14	0.0000	0.0000	0.9109	0.0000	0.0000	0.9652
VAR 15	0.0003	0.0008	0.8485	0.0002	0.0012	0.9224

Figure 9. Interrater Reliability for Weights and Sizing, Booster

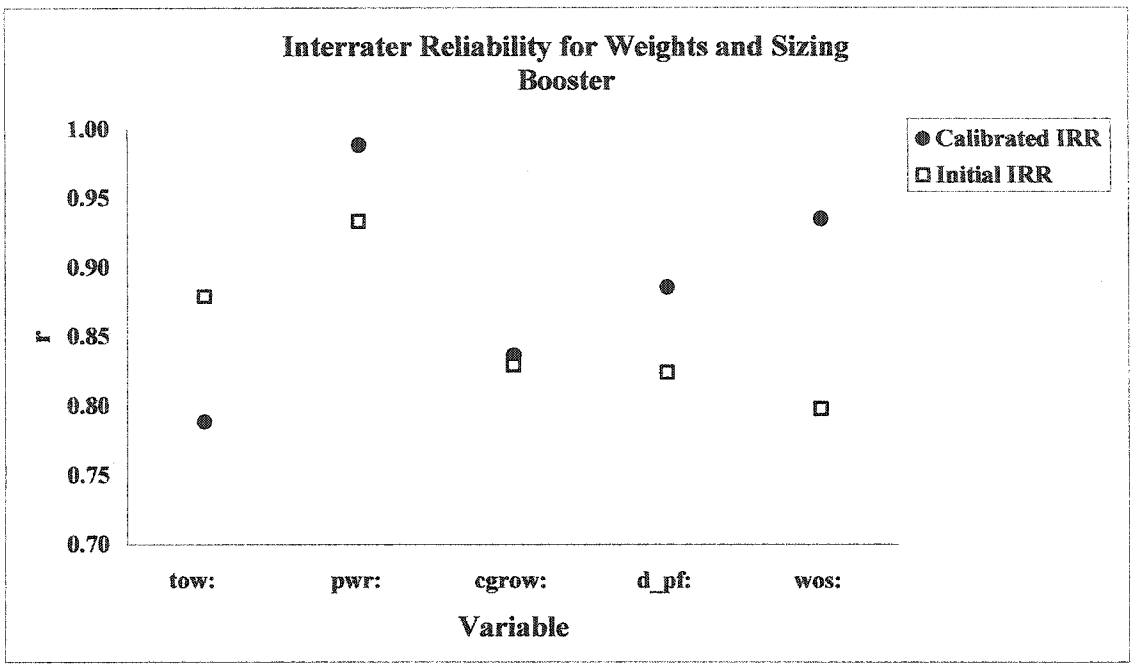


Figure 10. Interrater Reliability for Weights and Sizing, Orbiter

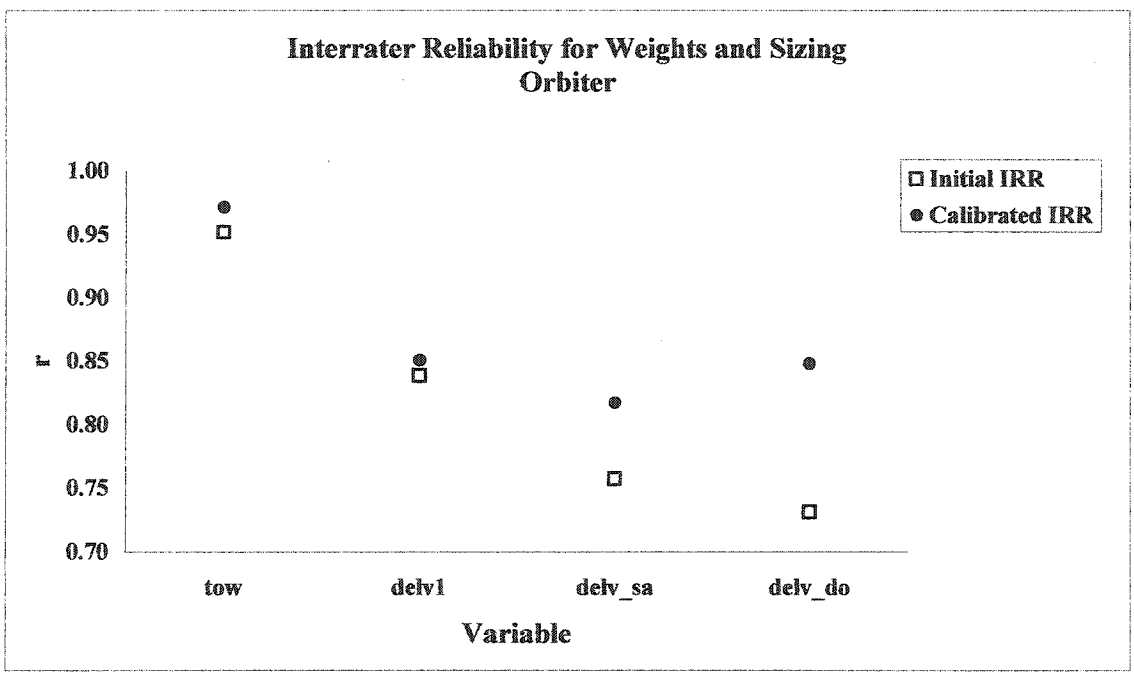


Figure 11. Interrater Reliability for Operations and Support, Booster

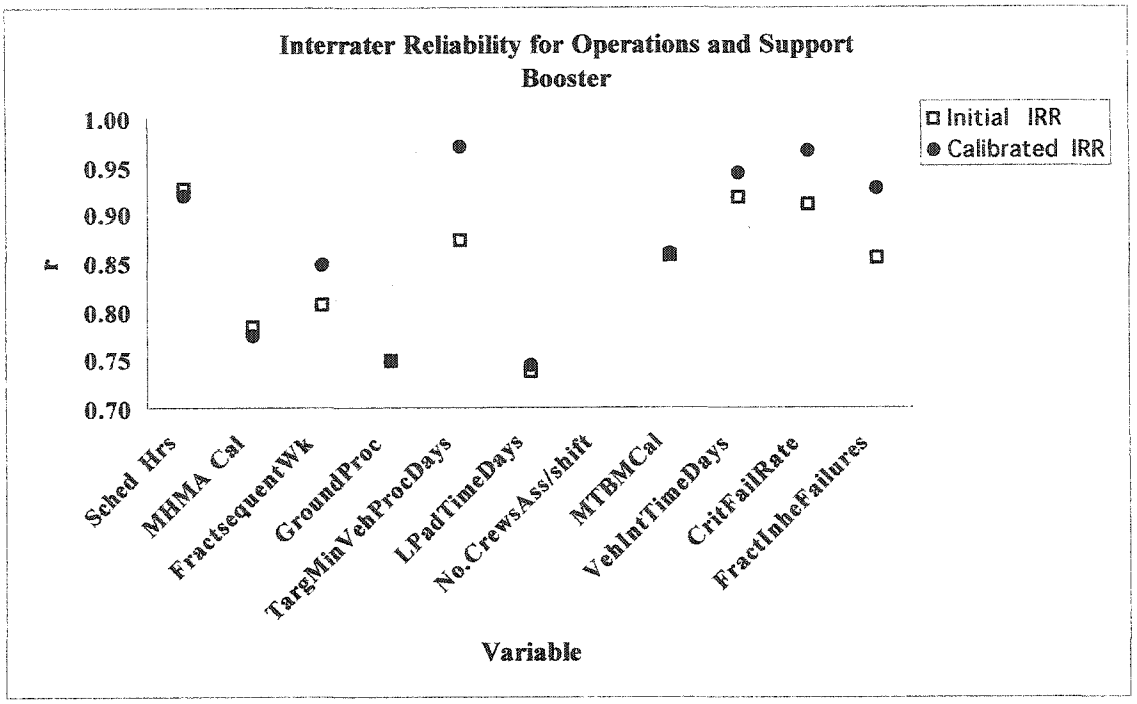
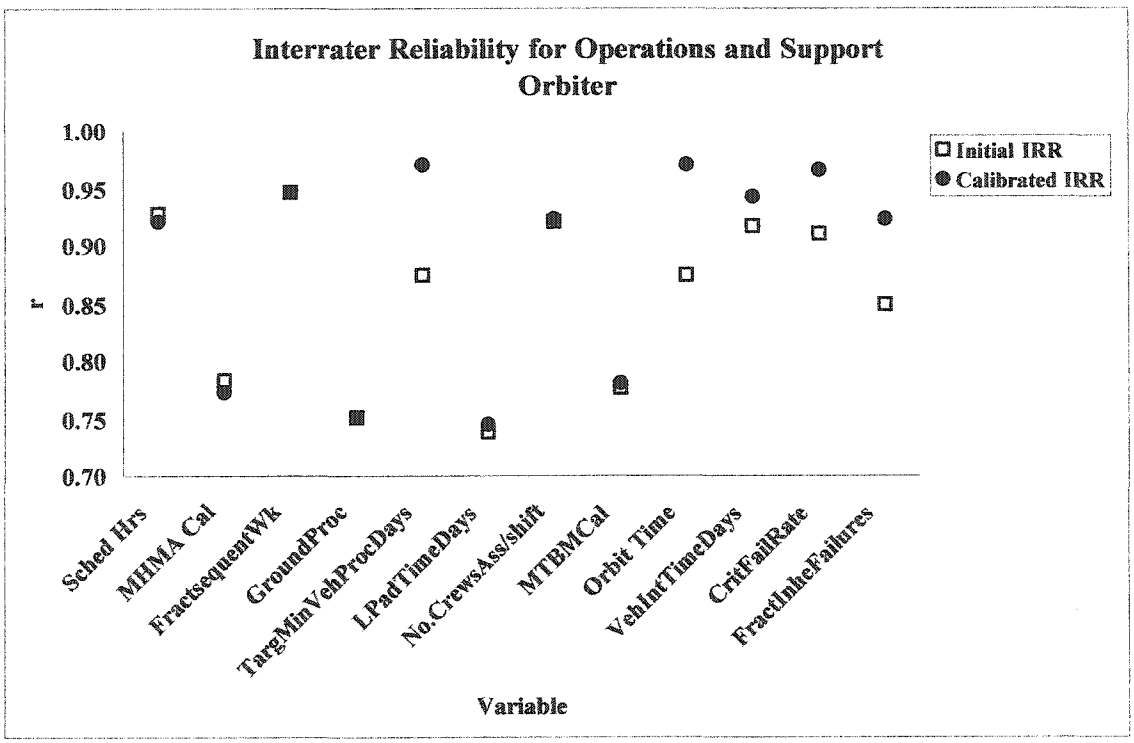


Figure 12. Interrater Reliability for Operations and Support, Orbiter



CHAPTER VI

DISCUSSION

Methodology Development and Application

While several references were consulted in the development of the expert elicitation calibration methodology reported herein, the decisions about how to weight the various factors that appear to be involved in such concepts as expertise and confidence or risk philosophy ultimately rest with the researcher. As was discussed earlier in the chapter about applying this methodology there is no indication, let alone consensus, as to weights for individual factors in the two quantities. Thus, the equal weights chosen for this study were felt to be a reasonable starting set. There were some adjustments in arriving at the final form of the expertise and philosophy relationships. In particular, the influence of age on expertise, which had originally been taken as multiplicative, was allowed to be additive to factors such as expertise self-designation. Also, the age chosen to serve as the ultimate benchmark -- originally set at 70 years -- was finally set at 60, in recognition of the typical retirement age of the experts participating in this study (and expected follow-on applications involving advanced aerospace technology impact judgments). It is suggested that perhaps the particular field of knowledge involved in comparable elicitations might better determine the benchmark age.

Related to the arbitrary nature of assigning weights to factors affecting expertise and philosophy is determining the means for ascertaining pertinent information from an expert to allow his correct (or at least reasonable) placement at appropriate locations on the expertise and philosophy scale. As noted in the literature review and in the discussion about applying the methodology, the classic means for determination of someone's

attitude about risk or confidence involves the proposing of monetary wagers to ascertain a point at which the subject would be indifferent in the choice of two or more options. Here a somewhat similar technique (presenting options) was used wherein the expert respondent to the background section of the elicitation questionnaire was asked to choose between two non-monetary preferences for three separate scenarios. While it is believed that the responses taken together reasonably assigned participating experts to appropriate broad philosophy categories based on knowledge of the subjects' professional backgrounds), there is likely room for improvement of the granularity of the philosophy classification.

Two experts responding to the operations and support case questionnaires omitted answers to two of the philosophy-related background questions, explaining that they did not understand what the questions were looking for. Because the default point assignments in this case offset each other (see Appendix F), the end result was neutral with respect to causing an undue shift in perceived philosophy (this result helped to demonstrate the robustness of the background questionnaire and its scoring). The other operations and support participant and the two weights and sizing participants completed the background portion with no comments or omissions.

Perhaps the biggest challenge in the development of the present methodology was determining a means for validating the calibration technique. The distant-future nature of the aerospace technology impact render classic judgment- or prediction-verification techniques moot. A concept explored early in the current research involved the identification of an already-accomplished aerospace development program whose performance characteristics would be unknown to a selected set of aerospace discipline experts (who would be questioned about relevant technology impacts and

accomplishments). One potential source of such a program might be a foreign country-based development. The difficulties associated with establishing credible baseline performance characteristics in the selected experts' discipline for an unfamiliar benchmark development program included access to the appropriate program documentation and the small likelihood that a discipline expert would not have at least some knowledge about other development programs in his field. It was determined that an alternate means for validating the current calibration approach would be more appropriate and efficient to carry out. Interrater reliability was identified as a viable alternative. This concept has been discussed in an earlier chapter (Chapter IV), and will be discussed further later in this chapter.

Finally, the discipline questionnaire development and implementation should be addressed. The basic techniques employed followed earlier work related to the present effort in both discipline-related and expert elicitation areas. The questionnaire was developed as part of a larger related effort that will include aggregation of expert responses. The questionnaire used in this work seems to have been generally efficient in eliciting the appropriate judgments. There were some cases of inconsistency among experts in a case that rendered responses with respect to the affected variables questionable for use in the calibration portion of the data analysis activity. The inconsistency suggests that vigilance in questionnaire development and testing be maintained and emphasized. There was m ore consistency among the operations and support experts, compared to the weights and sizing experts.

Methodology Application Results

Calibration

Applying the calibration methodology to the experts' variable-related uncertainty responses was straightforward. The expert's initial uncertainty distribution was easily determined from his responses and displayed for subsequent comparison to the distribution obtained after applying a calibration adjustment, determined from responses to the background questionnaire for each expert.

The calibrated distributions follow the trends (compared to the initial distributions) suggested by the present research. For those experts whose confidence/risk philosophy tended toward risk-averse (denoted by negative philosophy scores, the calibrated distributions reflected a lower variance. There were no participating experts in either case whose philosophy score was positive, so observation of an expanded distribution was not possible. For the one expert with a zero philosophy score (in the operations and support case), the expected constancy of variance was observed in the case results.

The calibrated distributions also reflected the expected response to difference in expertise. Those experts with a higher expertise score displayed less adjustment in "most likely" values than did those with lower expertise scores. The shifts in modes also occurred in the direction established by the adjustment algorithm (which in turn is based on the expertise and confidence philosophy of each expert). There were responses that resulted in calibrated distributions that did not completely follow precisely those suggested by expertise and philosophy scores; these distributions were the result of an experts assignment of either an alternate "most likely" value to a parameter, or to the

assignment of minimum and/or maximum values that resulted in a nonsymmetrical distribution. As noted in an earlier discussion, endpoints so assigned to a distribution were honored in the subsequent analysis.

It is observed that applying the variance adjustments (based on the philosophy ratings of the experts) will tend to equalize variances among participants. This tendency reflects that associated with more traditional calibrations (that of measuring instruments, for example). A principal reason for applying calibrations is the reduction or elimination of measurement errors resulting from bias, thus rendering measurements more consistent from trial to trial. The use of multiple instruments to improve system redundancy often necessitates the aggregation of the output of these instruments to provide a "best" value for system operation. The removal of biases renders the aggregation task more easy, whether it be simple averaging or the use of more sophisticated weighting functions. In the present case, using the experts' calibrated distributions for the subsequent aggregation process should likewise result in more consistent output for use in the disciplinary or multidisciplinary analysis upon which programmatic decisions may ultimately be based.

Feedback of calibration results to participating experts is considered an important part of the methodology. Such feedback can perhaps indicate to an expert tendencies that could prove helpful in other analysis situations. The feedback will also allow the expert to contribute to the optimization and efficiency of the calibration methodology. Feedback and subsequent adjustment of the calibration techniques would be expected to occur over a period of time, since adjustments based on only a few respondents may not be representative those resulting from a larger pool of expert participants.

Validation

Interrater reliability was chosen as a reasonable means of validating the efficacy of the current calibration methodology. The results of applying the technique developed by James, Demaree, and Wolf (1984) indicate that the calibration techniques are successful in increasing the interrater reliability for the groups of experts addressing each case. Because the nature of the parameters involved in the evaluation was not uniform (different physical quantities and different ranges of possible values), each variable was considered separately. To evaluate the overall effectiveness of the calibrations, the individual interrater reliabilities (r's) were tabulated and displayed. It can be seen (see Tables 6 and 7, and Figures 9 – 12) that the preponderance of calibrated r's were higher than those for the uncalibrated distribution for the variable. This general result, even though attained with a small number of participating experts, indicates that the calibration has a positive effect on moving expert responses toward agreement (this was one of the key motivators for the present research, as discussed in Chapter I). It should be pointed out, however, that interrater reliability should not be interpreted as an aggregation technique. An effective aggregation technique must (usually) employ knowledge about an expert, with weighting factors assigned to experience along with other factors not considered in the present study. Much research has been undertaken in the aggregation techniques area, including a companion study to the present one.

Further validation of these techniques may be expected in the future with feedback from participating experts, as noted in the previous section.

CHAPTER VII

CONCLUSIONS

General

The research and application study reported herein has resulted in a tool that permits calibration of experts' judgments for future-occurring events or developments. While the tool may permit calibrated predictions that ultimately turn out to be inaccurate (given the time horizon of their fulfillment), it has been shown that the methodology employed yield uncertainty levels around estimates that are more consistent with an estimator's (or expert judge's) experience and risk philosophy than other attempts thus far. Since it is a somewhat new area of research (calibration for distant-future events) being initially applied to aerospace disciplines that are dynamic and continuously evolving, the methodology and algorithms developed here should prove an effective aid to decision making associated with aerospace development. In particular, the techniques will facilitate the investigations associated with examining impact on aerospace vehicle performance of adopting new technology for future concepts.

The methodology that includes the calibration technique developed here, along with ongoing work on aggregation of uncertainty distributions from multiple experts, is already being utilized in multidisciplinary conceptual design and analysis programs. The calibration methodology itself possesses rigor that is expected to lead to more informed decisions than those made without the benefits of this new technique that reduces the effects of uncertainty in expert assessments.

Based on the results of this research, then, the answer to the research question posed earlier is, "yes": a calibration function has been developed that can be applied to experts' assessment of technology concepts and impacts.

Expertise and Philosophy Characterization

It appears that the use of factors such as self-identified expertise and age in determining an expertise adjustment to an elicited uncertainty distributions is a reasonable aspect of a calibration methodology. As noted in the Discussion (Chapter VI), weighting factors for the various elements of the expertise determination may be subject to future adjustments after experience with these tools.

Risk or confidence philosophy characterization also seems to be feasible to include in an expert judgment calibration scheme. Elicitation of appropriate background to assign a philosophy score may represent a continuing challenge.

Limitations and Suggestions for Future Work

The present study has been subject to some limitations that may reduce its efficacy in dealing with large-scale applications. The pool of experts for the cases included here was small, with judges for each case having similar work and organizational settings. Ascertainment of interrater reliability (for example) would be more prone to what James, Demaree, and Wolf (1984) refer to as "response bias." While response bias can include other factors such as psychological or social attitudes, the current effort did not attempt to qualify these. It would be desirable to reduce any potential bias resulting from similarity in organizational or work setting by employing experts from multiple organizations. The experts in such a broader group would also be

expected to possess similar backgrounds in their discipline, and approximately comparable levels of expertise.

This work has utilized triangular distributions for its elicitation and subsequent analysis and calibration. This follows previous work in expert judgment elicitation, specifically in related aerospace conceptual design applications. Other distributions may provide better representations of uncertainty; these other models are amenable to the same calibration methodology development steps reported here, including adjustments for expertise and confidence/risk philosophy. Broadening the set of uncertainty distribution calibration tools would be expected to facilitate application to other disciplines or different expert elicitation protocols than those reported herein.

Yet another limitation of the current research may be an inability to ascertain whether a participating expert whose initial assessment may be correct, but whose expertise and philosophy scores are such that his calibrated assessment uncertainty distribution would not include the ultimate actual value (of a parameter, say). This scenario is equivalent to committing a Type I error in statistical hypothesis testing (rejecting a true null hypothesis). The question thus arises, what is the danger of losing information that may be of value? While the likelihood of this occurrence is thought to be low in the present research because of the care taken in selection of expert participants, it would not be zero. For the target application of the current methodology, conceptual design and analysis of conceptual aerospace vehicles, the assessment of a single expert in a single discipline would be expected to have relatively little influence on the entire set of analyses that use input from the expert elicitations. The possibility of occurrence, however, suggests that sensitivity studies would be a profitable area of study in future extensions of this methodology.

A fourth and to some perhaps key limitation is the lack of actual performance, cost, or other results by which to gauge the effectiveness of the calibration methods developed. The interrater reliability determination discussed above provides some measure of validation for the methodology, but can never be as effective as real-world results. Because these would not be available for many years, if ever, the testing of the methodology in an aerospace application would be desirable. It has been suggested that evaluation involving a past program with documented performance could be used with experts uninvolved in the program unknowledgeable about the outcomes (performance, cost, or final geometry. Another alternative might be involvement of discipline experts in related applications areas (such as automotive or hydrodynamics) in a similar evaluation as the foregoing suggestion. Such an experiment could help provide a more robust validation of the developed methodology.

Two additional areas for possible future research in extension or application of the methodology are noted here. First, the addition of experts from the fields of construction and operation of aerospace vehicles as similar as possible to the concepts be studied could bring fresh perspectives to some of the analyses, particularly in the assessments of impacts on performance and the lessening of uncertainty about operational- or performance-related parameters. Second, the potential exists for utilizing the current approach as a means of developing expertise in younger practitioners in a field or discipline. The questionnaire and calibration process, with feedback, might be useful in a training syllabus to help compress the learning experience of future experts.

Summary

This work began through motivation from expert elicitation applications that resulted in widely disparate evaluations of advanced technology impact on future aerospace vehicle design, performance and operations. Through an extensive review of literature on uncertainty, expert judgment elicitation, and calibration, a methodology has been developed that utilizes characteristics of an expert elicitee to adjust the rendered judgments. These adjusted, or calibrated judgments lead to uncertainty distribution that provided a more consistent response among multiple experts in analytical modeling of aerospace vehicle concepts, performance and cost.

REFERENCES

Antoniou, G. (1998). *A tutorial on default reasoning*. Cambridge University Press.

Ayyub, B.M. (2001). *Elicitation of Expert Opinions for Uncertainty and Risks*. Boca Raton, FL: CRC Press.

Beach, B. H. (1975). Expert Judgment About Uncertainty: Bayesian Decision Making in Realistic Settings. *Organizational Behavior and Human Performance*, 14, 10-59.

Brier, G. (1950). Verification of Forecasts Expressed in Terms of Probability. Monthly Weather Review, Vol. 78, No. 1, pp 1-3.

Brown, R.V. (1969). *Research and the Credibility of Estimates*. Boston: Harvard University.

Browne, G.J., S.P. Curley, and P.G. Benson (1999). The Effects of Subject-Defined Categories on Judgmental Accuracy in Confidence Assessment Tasks. *Organizational Behavior and Human Decision Processes*, 80, No. 2, 134-154.

Coupé, V.M.H., L.C. van der Gaag, and J.D.F. Habbema (2000). Sensitivity analysis: an aid for belief-network quantification. *The Knowledge Engineering Review*, Vol. 15, 3, pp 215-232.

Crawford, J. D. and L. Stankov (1996). Age Differences in the Realism of Confidence Judgements: A Calibration Study Using Tests of Fluid and Crystallized Intelligence. *Learning and Individual Differences*, Vol. 8, No. 2, 83-103.

Cyert, R.M. and M.H. DeGroot (1987). *Bayesian Analysis and Uncertainty in Economic Theory*. Totowa, NJ: Rowman & Littlefield.

Dalkey, N.C. (1969). The Use of Self-Ratings to Improve Group Estimates. *Technology Forecasting* 12, 283-291.

Dawid, A.P. (1982). The Well-Calibrated Bayesian. *Journal of the American Statistical Association*, Vol. 77, Issue 379, 605-610.

de Mántaras, R.L. (1990). *Approximate Reasoning Models*. Chichester, England: Halstead Press: a division of John Wiley & Sons.

Duarte, B. P. M. (2001). The expected utility theory applied to an individual decision problem -- what technological alternatives to implement to treat industrial solid residuals. *Computers and Operations Research* 28, 357-380.

Fischhoff, B. (1976). The Effect of Temporal Setting on Likelihood Estimates. *Organizational Behavior and Human Performance*, 15, 180-194.

Feltovich, P.J., R.J. Spiro, and R.L. Coulson (1997). Issues of Expert Flexibility in Contexts Characterized by Complexity and Change. In P.J. Feltovich, K.M. Ford and R.R. Hoffman (Eds.), *Expertise in Context*. Menlo Park: AAAI Press/The MIT Press.

Gustafson, D.H., R.E. Shukla, A. Delbecq, , and G.W. Walster (1973). A Comparative Study of Differences in Subjective Likelihood Estimates Made by Individuals, Interacting Groups, Delphi Groups, and Nominal Groups. Organizational Behavior and Human Performance, 9, 200-291.

Hammond, K. R., R. M. Hamm, J. Grassia, and T. Pearson (1987). Direct Comparison of the Efficacy of Intuitive and Analytical Cognition in Expert Judgment. *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-17, No. 5, September/October.

Hampton, K.R. (2001). An Integrated Risk Analysis Methodology in a Multidisciplinary Design Environment. Ph.D. Dissertation. Old Dominion University, Norfolk, VA.

Hardy, T.L. (1994). Fuzzy Logic Approaches to Multi-Objective Decision-Making in Aerospace Applications. *AIAA Paper 94-3163, presented at the AIAA 30th Joint Propulsion Conference*, Colorado Springs, CO, 1994.

Hardy, T.L. and D.C. Rapp (1994). Rocket Engine System Reliability Analyses Using Probabilistic and Fuzzy Logic Techniques. *AIAA Paper 94-2750, presented at the AIAA/ASME/SAE/ASEE 30th Joint Propulsion Conference*, Indianapolis, IN, 1994.

James, L. R., R. G. Demaree, and G. Wolf (1984). Estimating Within-Group Interrater Reliability With and Without Response Bias. *Journal of Applied Psychology*, Vol. 69, No. 1, 85-98.

Johnson, J.E.V. and A.C. Bruce (2001). Calibration of Subjective Probability Judgments in a Naturalistic Setting. Organizational Behavior and Human Decision Processes, 85, No. 2, 265-290.

Kahn, M. H. and N. Hafiz (1999). Development of an Expert System for implementation of ISO 9000 quality systems. *Total Quality Management*, Vol. 10, Issue 1, 47-59.

Keppell, M. (2001). Optimizing Instructional Designer-Subject Matter Expert Communications in the Design and Development of Multimedia Projects. *Journal of Interactive Learning Research*, 12 (2/3), 209-27.

Keren, G. (1991). Calibration and probability judgments: Conceptual and methodological issues. *Acta Psychologica*, 77, 217-273.

Kolb, M.A. (1994). Object-modeling with constraints for aerospace conceptual/preliminary design. *Proceedings of SPIE, v. 2244*. Bellingham, WA.

Kosko, B. (1993). *Fuzzy Thinking: The New Science of Fuzzy Logic*. New York: Hyperion.

Lichtenstein, S. and B. Fischhoff (1977). Do Those Who Know More Also Know More About What They Know? Organizational Behavior and Human Performance, 20, 159-183.

Lichtenstein, S., B. Fischhoff, and L.D. Phillips (1982). Calibration of probabilities: The state of the art to 1980. In Kahneman, Slovic, and Tversky (Eds.), *Judgment Under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press.

Linstone, H. and M. Turot (1975). *The Delphi Method: Techniques and Applications*. Reading, MA: Addison Wesley.

MacCrimmon, K. R. and D. A. Wehrung (1986). *Taking Risks: The Management of Uncertainty*. New York. The Free Press.

Miller, D. C. and J. M. Byrnes (1997). The Role of Contextual and Personal Factors in Children's Risk Taking. *Developmental Psychology*, Vol. 33, No. 3, 814-823.

Monroe, R.W. (1997). *A Synthesized Methodology for Eliciting Expert Judgment for Addressing Uncertainty in Decision Analysis*. Ph.D. Dissertation. Old Dominion University, Norfolk, VA. August 1997.

Monroe, R.W., R. A. Lepsch, and R. Unal (2002). Using Expert Judgment Methodology to Address Uncertainty in Launch Vehicle Weight Estimates. *AIAA Paper 2002-5183 presented at 9th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, Atlanta, GA, 2002.

Morgan, M.G. and M. Henrion (1990). *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. London: Cambridge University Press.

Murphy, A. H. and R. L. Winkler (1974). Credible Interval Temperature Forecasting: Some Experimental Results. *Monthly Weather Review*, 102, 784-794

Neil, M., N. Fenton and L. Nielson. (2000). Building large-scale Bayesian networks. *The Knowledge Engineering Review*, Vol. 15, 3, pp 257-284.

Panofsky, H.A. and G.W. Brier (1968). *Some Applications of Statistics to Meteorology*. University Park, Pennsylvania: The Pennsylvania State University,.

Renooij, S. (2001). Probability elicitation for belief networks: issues to consider. *The Knowledge Engineering Review*, Vol. 16, 3, 255–269.

Rowe, G., G. Wright, and F. Bolger (1991). Delphi: A reevaluation of research and theory. *Technological Forecasting and Social Change*, Vol. 39, Issue 3, 235-251.

Rowell, L. F., Braun, R. D., J. R. Olds, and R. Unal (1999). Multidisciplinary Conceptual Design Optimization of Space Transportation Systems. *AIAA Journal of Aircraft*, Vol. 26, No. 1, pp 218-226.

Rush, R. and W.A. Wallace. (1997). Elicitation of Knowledge from Multiple Experts Using Network Inference. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 9, No. 5. September/October 1997.

Schaefer, R.E. (1976). The Evaluation of Individual and Aggregated Subjective Probability Distributions. *Organizational Behavior and Human Performance*, 17, 199-210.

Schmitt, S.A. (1969). *Measuring Uncertainty: An Elementary Introduction to Bayesian Statistics*. Reading, MA: Addison-Wesley.

Shanteau, J.; D. Weiss; R. Thomas and J. Pounds (2002). Performance-based assessment of expertise: How to decide if someone is an expert or not. *European Journal of Operational Research*, 136, 253-263.

Simon, M., S. M. Houghton, and K. Aquino (1999). Cognitive Biases, Risk Perception, and Venture Formation: How Individuals Decide to Start Companies. *Journal of Business Venturing* 15, 113–134.

Spence, M. T. (1996). Problem - Problem Solver Characteristics Affecting The Calibration of Judgments. *Organizational Behavior and Human Decision Processes*, Vol. 67, No. 3, September, 271-278.

Unal, R. (2002). Development Of Response Surface Models And Technology Support Levels. Final Report, NASA PO # H-30938D (Old Dominion University Research Foundation Project No: 11720). February 2002.

Unal, R. and B.A. Conway (2000). A Survey to Determine Influence of Design Parameters on Operations & Support Complexity and Cost for Launch Vehicles. Final Report, NASA PO# L-12288 (Old Dominion University Research Foundation Project No: 104881). December 2000.

Van Lenthe, J. (1994). Scoring-Rule Feedforward and the Elicitation of Subjective Probability Distributions. *Organizational Behavior and Human Design Processes*, 59, 188-209.

Wang, F. A. (2001). Overconfidence, Investor Sentiment, and Evolution. *Journal of Financial Intermediation*, Vol. 10, No. 2, 138-170.

Wright, G. (1982). Changes In The Realism And Distribution Of Probability Assessments As A Function Of Question Type. Acta Psychologica, 52: 165-174.

Wright, G., G. Rowe, F. Bolger, and J. Gammack (1994). Coherence, Calibration, and Expertise in Judgmental Probability Forecasting. *Organizational Behavior and Human Decision Processes*, 57, 1-25.

Zadeh, L.A. (1992). Knowledge Representation in Fuzzy Logic. In R.R. Yager and L.S. Zadeh (Eds.), *An Introduction to Fuzzy Logic Applications in Intelligent Systems.* Boston: Kluwer Academic Publishers.

Zaff, B. S., M. D. McNeese, and D. E. Snyder (1993). Capturing multiple perspectives: a user-centered approach to knowledge and design acquisition. *Knowledge Acquisition*, 5, 79-116.

Zuk, W. (1990). An Expert System for the Esthetic Rating of Bridges (Report No. VTRC 90-R13). Charlottesville, VA: Virginia Transportation Research Council.

APPENDIX A

EXPERT JUDGMENT ELICITATION QUESTIONNAIRE
(WEIGHTS AND SIZING EXAMPLE)

INPUT PARAMETER UNCERTAINTY QUESTIONNAIRE

Input Parameter Uncertainty: Weights & Sizing (Mach 5 booster)

Select the INPUT parameters from the following list that you want to evaluate for uncertainty. If you wish to add a parameter not listed, select "OTHER".

INPUT parameters

'nbst'	1.	number of common booster stages
'cballast'	.0000	ballast weight fraction of empty weight
'cgrow'	.15	growth allowance fraction
'preorb'	1747111	orbiter prelaunch gross weight (lb)
'grsorb'	1747111	orbiter gross lift-off weight (lb)
	.	
	.	
'qmax'	700.	max dynamic pressure, psf
'dcruise'	150.	cruise distance (nmi)
	.	
	.	
towe'	61.3	main engine thrust/weight (vacuum) at 100% power
ispvac'	452.5	main engine I_{sp} (vacuum)
ispvac2'	452.5	main engine I_{sp} (vacuum), orbiter
	.	
	.	
'stf'	374.481	tip fin planform area (ft^2)
'sflap'	260.376	body flap planform area (ft^2)
'mr'	1.9011	mass ratio
OTHER	_____	_____

From the **WEIGHT and SIZING** INPUT parameters you have selected:

'__' [name] [value]

Rate the degree of uncertainty that you associate with this parameter:

Low	Low/moderate	Moderate	Moderate/high	High

If you feel this INPUT parameter's default value should be modified, you may provide a new point estimate for the INPUT parameter's nominal value. [ ]

If you feel the range of possible values around the nominal value is not symmetrical, please provide your own estimates of minimum and maximum values.

Min [ ] Max [ ]

Now that you have rated the uncertainty for this INPUT parameter, please provide a reason or reasons for your rating. Include a rationale for any change you made to the parameter's nominal value.

To further document your thinking, please provide any cues (or triggers) that influence your thinking about this parameter.

After completing the preceding steps for all parameters you have rated as uncertain, please provide a quantitative explanation of your understanding of Low, Moderate and High uncertainty, using the 5-point scales provided.

The amount of uncertainty or variation that I associate with Low Uncertainty is:

Less	**5%**	**7.5%**	**10%**	**12.5%**	**15%**	**More**

The amount of uncertainty or variation that I associate with Moderate Uncertainty is:

Less	**10%**	**15%**	**20%**	**25%**	**30%**	**More**

The amount of uncertainty or variation that I associate with High Uncertainty is:

Less	**20%**	**30%**	**40%**	**50%**	**60%**	**More**

APPENDIX B

INITIAL PARAMETER LIST (WEIGHTS AND SIZING)
(TWO –STAGE-TO-ORBIT LAUNCH VEHICLE WITH MACH 3 STAGING)

INITIAL PARAMETER LIST – TSTO LAUNCH VEHICLE

Stage: Orbiter			
Input Variable Description	Variable Name	Value	Data Source
ballast weight fraction of empty wt	cballast	0	user input
growth allowance fraction	cgrow	0.15	user input
payload weight (lb)	payld	35000	user input
additional down-payload (lb.)	adpay	25000	user input
space radiator area (ft2)	srad	700	user input
mission duration (days), design	tday	10.5	user input
mission duration (days), reserve	tmar	2	user input
number of crew	ncrew	0	user input
maximum man-day capability	tmday	0	user input
nominal fuel cell power (kw)	pfcnom	14	user input
oms delta v for tank sizing (ft./sec.)	delvt	900	user input
oms delta v (ft./sec.) - burn 1	delv1	348	user input
oms delta v (ft./sec.) - burn 2	delv2	0	user input
oms delta v (ft./sec.) - burn 3	delv3	0	user input
oms delta v (ft./sec.) - station appr.	delv_sa	100	user input
oms delta v (ft./sec.) - deorbit	delv_do	366	user input
max dynamic pressure, psf	qmax	700	user input
cruise distance (nmi)	dcruise	0	user input
number of main engines	neng	9	user input
total number of fly-back jet engines	njeng	0	user input
initial t/w, orbiter	tow	1.3113	user input
lift-off t/w, 2-stage vehicle	towi	1.3369	user input
engine power level fraction	pwr	1.04	user input
design max engine power level fraction	pwrmax	1.04	user input
oxidizer-to-fuel ratio	rmix	6	user input
propellant bulk density, o/f=6.0	dbulk	22.54	user input
fuel density (lb./cu. ft.)	d_pfl	4.42	user input
lox density (lb./cu. ft.)	d_lox	71.14	user input
ullage volume fraction	ull	0.015	user input
ullage volume fraction, wing	wull	0.03	user input
wing loading (psf)	wos	65	user input
technology factor - wing str	fwstr	1	user input
technology factor - vertical fin str	fvstr	1	user input
technology factor - body dry str	fbstr	1	user input
technology factor - fuel tank	fpf1tnk	1	user input
technology factor - LO2 tank	flo2tnk	1	user input
technology factor - fuselage TPS	fbtps	1	user input
technology factor - wing & fin TPS	fwtps	1	user input
technology factor - body flap TPS	fbftps	1	user input
technology factor - landing gear	fgear	1	user input

Stage: Orbiter (continued)

Input Variable Description	Variable Name	Value	Data Source
technology factor - main engines	fmeng	1	user input
technology factor - propellant feed sys	fpfs	1	user input
technology factor - gimbal actuation	fgim	1	user input
technology factor - main engine ht shld	fhtsld	1	user input
technology factor - he pneumatic sys	fhesys	1	user input
technology factor - RCS	frcs	1	user input
technology factor - OMS	foms	1	user input
technology factor - APU	fapu	1	user input
technology factor - fuel cell sys	ffcell	1	user input
technology factor - ECD	fecd	1	user input
technology factor - hydr conv & distr	fhcd	1	user input
technology factor - control surface act.	fcs	1	user input
technology factor - avionics	fav	1	user input
technology factor - environmental contrl	fec	1	user input
technology factor - internal insulation	finsl	1	user input
technology factor - purge, vent, & drn	fpvd	1	user input
technology factor - range safety	frng	1	user input
technology factor - payload container	fplcon	1	user input

Stage: Booster

Input Variable Description	Variable Name	Value	Data Source
number of common booster stages	nbst	1	user input
ballast weight fraction of empty wt	cballast	0	user input
growth allowance fraction	cgrow	0.15	user input
ascent time (min)	t_asc	2	user input
descent time (min)	t_desc	20	user input
operating time margin (min)	t_mar	5	user input
number of crew	ncrew	0	user input
maximum man-day capability	tmday	0	user input
electrical power req. (kw), ascent	p_asc	11.3	user input
electrical power req. (kw), descent	p_desc	7.7	user input
nominal electrical power (kw)	pfcnom	11.3	user input
max dynamic pressure, psf	qmax	700	user input
cruise distance (nmi)	dcruise	0	user input
number of main engines	neng	8	user input
total number of fly-back jet engines	njeng	0	user input

Stage: Booster (continued)			
Input Variable Description	Variable Name	Value	Data Source
lift-off t/w, 2-stage vehicle	tow	1.3372	user input
initial t/w, orbiter	toworb	1.3113	user input
engine power level fraction	pwr	1.04	user input
design max engine power level fraction	pwrmax	1.04	user input
oxidizer-to-fuel ratio	rmix	6	user input
propellant bulk density, o/f=6.0	dbulk	22.54	user input
propellant bulk density, o/f=6.0 (orb.)	dbulk2	22.54	user input
fuel density (lb./cu. ft.)	d_pfl	4.42	user input
lox density (lb./cu. ft.)	d_lox	71.14	user input
ullage volume fraction	ull	0.015	user input
ullage volume fraction, wing	wull	0.03	user input
wing loading (psf)	wos	65	user input
technology factor - wing str	fwstr	1	user input
technology factor - vertical fin str	fvstr	1	user input
technology factor - body dry str	fbstr	1	user input
technology factor - fuel tank	fpf1tnk	1	user input
technology factor - LO2 tank	flo2tnk	1	user input
technology factor - fuselage TPS	fbtps	1	user input
technology factor - wing & fin TPS	fwtps	1	user input
technology factor - body flap TPS	fbftps	1	user input
technology factor - landing gear	fgear	1	user input
technology factor - main engines	fmeng	1	user input
technology factor - propellant feed sys	fpfs	1	user input
technology factor - gimbal actuation	fgim	1	user input
technology factor - main engine ht shld	fhtsld	1	user input
technology factor - he pneumatic sys	fhesys	1	user input
technology factor - RCS	frcs	1	user input
technology factor - OMS	foms	0	user input
technology factor - APU	fapu	1	user input
technology factor - fuel cell sys	ffcell	1	user input
technology factor - ECD	fecd	1	user input
technology factor - hydr conv & distr	fhcd	1	user input
technology factor - control surface act.	fcs	1	user input
technology factor - avionics	fav	1	user input
technology factor - environmental contrl	fec	1	user input
technology factor - internal insulation	finsl	1	user input
technology factor - purge, vent, & drn	fpvd	1	user input
technology factor - range safety	frng	1	user input
technology factor - payload container	fplcon	1	user input

APPENDIX C

REDUCED PARAMETER LIST (MOST IMPACT ON PERFORMANCE)
TWO –STAGE-TO-ORBIT LAUNCH VEHICLE WITH MACH 3 STAGING
AFTER MODIFIED NGT - ASSUMING PARETO DISTRIBUTION
WEIGHTS AND SIZING

REDUCED PARAMER LIST AFTER APPLICATION OF MODIFIED NGT TSTO LAUNCH VEHICLE – WEIGHTS AND SIZING

Vehicle: ISAT Ref TSTO - M3 Staging
Model: CONSIZ
Stage: Orbiter

Variable Name	Description	Nominal Value
cgrow	growth allowance fraction	0.15
payld	payload weight (lb)	35000
tow	initial t/w, orbiter	1.3113
towi	lift-off t/w, 2-stage vehicle	1.3369
pwr	engine power level fraction	1.04
fmeng	technology factor - main engines	1
fbstr	technology factor - body dry str	1
fpf1tnk	technology factor - fuel tank	1
flo2tnk	technology factor - LO2 tank	1
fbtps	technology factor - fuselage TPS	1
delvt	oms delta v for tank sizing (ft./sec.)	900
delv1	oms delta v (ft./sec.) - burn 1	348
delv_sa	oms delta v (ft./sec.) - station appr.	100
delv_do	oms delta v (ft./sec.) - deorbit	366
d_pf1	fuel density (lb./cu. ft.)	4.42
d_lox	lox density (lb./cu. ft.)	71.14
wos	wing loading (psf)	65
fwstr	technology factor - wing str	1

Stage: Booster

Variable Name	Description	Nominal Value
nbst:	number of common booster stages	1
tow:	lift-off t/w, 2-stage vehicle	1.3372
pwr:	engine power level fraction	1.04
fmeng:	technology factor - main engines	1
fbstr:	technology factor - body dry str	1
fpf1tnk:	technology factor - fuel tank	1
flo2tnk:	technology factor - LO2 tank	1
cgrow:	growth allowance fraction	0.15
d_pf:	fuel density (lb./cu. ft.)	4.42
d_lox:	lox density (lb./cu.ft.)	71.14
wos:	wing loading (psf)	65
fwstr:	technology factor - wing str	1

APPENDIX D

EXPERT BACKGROUND RESPONSE RESULTS
WEIGHTS AND SIZING
TWO –STAGE-TO-ORBIT LAUNCH VEHICLE WITH MACH 3 STAGING

EXPERT BACKGROUND RESPONSE RESULTS
WEIGHTS AND SIZING

Question No.	Topic	Expert 1 Response	Point Conversion	Expert 2 Response	Point Conversion
EXPERTISE					
1	Name or USERID	8646275	-	8647643	-
2	Age (years)	42	3.5	41	3.4
3	Self-rated expertise	Less	2	More	4
4	Expertise compared to others in field	More	4	Same	3
5	Discipline-specific knowledge question	.01	3.9	.012	4.3
	Expertise =	**3.347**		**3.675**	
PHILOSOPHY					
6	Predicting discipline-related quantities	(1)	-5	(3)	0
7	Estimating uncertainty preference	(b)	5	(a)	-5
8	Estimating trend in discipline	(b)	-2.5	(d)	-5
9	Completion vs. milestones	(b)	-5	(b)	-5
10	Total outlays vs. cost elements	(b)	-5	(b)	-5
11	Utilization scenarios vs. performance	(a)	-5	(b)	5
	Philosophy =	**-2.917**		**-2.5**	

APPENDIX E

REDUCED PARAMETER LIST (MOST IMPACT ON PERFORMANCE)
TWO –STAGE-TO-ORBIT LAUNCH VEHICLE WITH MACH 3 STAGING
AFTER MODIFIED NGT - ASSUMING PARETO DISTRIBUTION
OPERATIONS AND SUPPORT

REDUCED PARAMER LIST AFTER APPLICATION OF MODIFIED NGT

TSTO LAUNCH VEHICLE – OPERATIONS AND SUPPORT

Vehicle: ISAT Ref TSTO - M3 Staging Booster **Model: RMAT**	
Variable Name	**Nominal Value**
Scheduled Hours	114750
Shifts per Day	2
Missions per Year	8
MHMA Calibration	1
Fraction of sequential (independent) work	0.05
Ground Processing	7689
Target minimum vehicle processing days	30
Launch Pad Time in Days	25.3
Number of Crews Assigned per shift	1
MTBM Calibration	0.833
Orbit Time	0
Vehicle Integration Time (days)	5.5
Technology Growth	0
Critical Failure Rate	0.0006052
Fraction Inherent Failures	0.1836
Vehicle: ISAT Ref TSTO - M3 Staging Orbiter **Model: RMAT**	
Variable Name	**Nominal Value**
Scheduled Hours	159897
Shifts per Day	2
Missions per Year	8
MHMA Calibration	1
Fraction of sequential (independent) work	0.05
Ground Processing	7771
Target minimum vehicle processing days	30
Launch Pad Time in Days	25.3
Number of Crews Assigned per shift	1
MTBM Calibration	0.804
Orbit Time	252
Vehicle Integration Time (days)	5.5
Technology Growth	0
Critical Failure Rate	0.0005745
Fraction Inherent Failures	0.1645

APPENDIX F

EXPERT BACKGROUND RESPONSE RESULTS
OPERATIONS AND SUPPORT
TWO –STAGE-TO-ORBIT LAUNCH VEHICLE WITH MACH 3 STAGING

EXPERT BACKGROUND RESPONSE RESULTS
WEIGHTS AND SIZING

Quest No.	Topic	Exp't 1 Resp	Point Conv	Exp't 2 Resp	Point Conv	Exp't 3 Resp	Point Conv
EXPERTISE							
1	Name or USERID	8643684	-	8646262	-	8643425	
2	Age (years)	60	5	59	4.9	51	4.3
3	Self-rated expertise	More	4	Average	3	More	4
4	Expertise compared to others in field	Most	5	More	4	Same	3
5	Discipline-specific knowledge question	$85	5	$75	4.5	$75	4.5
	Expertise =	**4.750**		**4.098**		**3.931**	
PHILOSOPHY							
6	Predicting discipline-related	(3)	0	(5)	5	5	5
7	Estimating uncertainty	-	5	-	5	(b)	5
8	Estimating trend in discipline	-	-5	-	-5	(c)	5
9	Completion vs. milestones	(b)	-5	(b)	-5	(b)	-5
10	Total outlays vs. cost elements	(b)	-5	(b)	-5	(b)	-5
11	Utilization scenarios vs. performance	(a)	-5	(b)	5	(a)	-5
	Philosophy =	**-2.500**		**-1.667**		**0.000**	

APPENDIX G

UNCERTAINTY DISTRIBUTION RESULTS
WEIGHTS AND SIZING
TWO –STAGE-TO-ORBIT LAUNCH VEHICLE WITH MACH 3 STAGING

WEIGHT AND SIZING CASE RESULTS

Summary Uncertainty Distributions, Booster								
	Expert 8646275 Initial Distribution				Expert 8646275 Calibrated Distribution			
	a_1	c_1	b_1	σ_1^2	a_2	c_2	b_2	σ_2^2
VAR 02	1.1563	1.2500	1.3438	0.0015	1.6027	1.6632	1.7237	0.0006
VAR 03	0.8500	0.9000	0.9500	0.0004	0.8500	0.9500	0.9500	0.0006
VAR 08	0.1000	0.2000	0.3500	0.0026	0.1000	0.2661	0.3500	0.0027
VAR 09	3.9228	4.4200	4.9173	0.0412	5.5601	5.8811	6.2020	0.0172
VAR 11	55.0000	65.0000	70.0000	9.7222	55.0000	70.0000	70.0000	12.5000
	Expert 8647643 Initial Distribution				Expert 8647643 Calibrated Distribution			
	a_1	c_1	b_1	σ_1^2	a_2	c_2	b_2	σ_2^2
VAR 02	1.2703	1.3372	1.4041	0.0007	1.6382	1.6854	1.7327	0.0004
VAR 03	0.9500	1.0000	1.0500	0.0004	1.2251	1.2604	1.2958	0.0002
VAR 08	0.0975	0.1500	0.2025	0.0005	0.1519	0.1891	0.2262	0.0002
VAR 09	4.1990	4.4200	4.6410	0.0081	5.4148	5.5710	5.7273	0.0041
VAR 11	52.0000	65.0000	78.0000	28.1667	72.7347	81.9271	91.1195	14.0833

WEIGHT AND SIZING CASE RESULTS
(CONTINUED)

Summary Uncertainty Distributions, Orbiter								
	Expert 8646275 Initial Distribution				Expert 8646275 Calibrated Distribution			
	a_1	c_1	b_1	σ_1^2	a_2	c_2	b_2	σ_2^2
VAR 03	0.9319	1.0500	1.1681	0.0023	1.3208	1.3971	1.4733	0.0010
VAR 12	295.8000	348.0000	400.2000	454.1400	429.3384	463.0333	496.7283	189.2250
VAR 13	85.0000	100.0000	115.0000	37.5000	123.3731	133.0556	142.7380	15.6250
VAR 14	311.1000	366.0000	420.9000	502.3350	451.5455	486.9833	522.4211	209.3063
	Expert 8647643 Initial Distribution				Expert 8647643 Calibrated Distribution			
	a_1	c_1	b_1	σ_1^2	a_2	c_2	b_2	σ_2^2
VAR 03	1.245735	1.3113	1.376865	0.0007	1.6064	1.6528	1.6991	0.0004
VAR 12	330.600	348.000	365.400	50.4600	426.3213	438.6250	450.9287	25.2300
VAR 13	80.000	100.000	120.000	66.6667	111.8995	126.0417	140.1838	33.3333
VAR 14	320.250	366.000	411.750	348.8438	428.9624	461.3125	493.6626	174.4219

APPENDIX H

UNCERTAINTY DISTRIBUTION RESULTS
OPERATIONS AND SUPPORT
TWO –STAGE-TO-ORBIT LAUNCH VEHICLE WITH MACH 3 STAGING

OPERATIONS AND SUPPORT CASE RESULTS

Summary Uncertainty Distributions, Booster								
	Expert 8643684 Initial Distribution				Expert 8643684 Calibrated Distribution			
	a_1	c_1	b_1	σ_1^2	a_2	c_2	b_2	σ_2^2
VAR 01	40000	114750	125000	358822917	40000	120488	125000	381211120
VAR 04	0.2500	1.0000	1.2500	0.0451	0.2500	1.0500	1.2500	0.0467
VAR 05	0.2000	0.0500	0.0800	0.0011	0.2000	0.2000	0.0800	0.0008
VAR 06	150.0000	640.0000	1250.0000	50616.667	150.0000	672.0000	1250.0000	50460.222
VAR 07	28.5000	30.0000	31.5000	0.3750	30.4393	31.5000	32.5607	0.1875
VAR 08	15.0000	25.3000	45.0000	38.7272	15.0000	26.5650	45.0000	38.1555
VAR 09	1.0000	1.0000	3.0000	0.2222	1.0000	1.0500	3.0000	0.2168
VAR 10	0.7000	0.8330	1.4000	0.0230	0.7000	0.8747	1.4000	0.0221
VAR 12	2.0000	5.5000	6.0000	0.7917	2.0000	5.7750	6.0000	0.8417
VAR 14	0.0006	0.0006	0.0006	0.0000	0.0006	0.0006	0.0007	0.0000
VAR 15	0.1700	0.1836	0.1900	0.0000	0.1700	0.1900	0.1900	0.0000
	Expert 8646262 Initial Distribution				Expert 8646262 Calibrated Distribution			
	a_1	c_1	b_1	σ_1^2	a_2	c_2	b_2	σ_2^2
VAR 01	90000	114750	140000	104170139	90000	135460	140000	127423689
VAR 04	0.2500	1.0000	1.5000	0.0660	0.2500	1.1805	1.5000	0.0703
VAR 05	0.0300	0.0500	0.0600	0.0000	0.0300	0.0590	0.0600	0.0000
VAR 06	150.0000	640.0000	1250.0000	50616.667	150.0000	755.5088	1250.0000	50587.845
VAR 07	27.7500	30.0000	32.2500	0.8438	33.5774	35.4145	37.2516	0.5625
VAR 08	15.0000	25.3000	45.0000	38.7272	15.0000	29.8662	45.0000	37.5010
VAR 09	1.0000	1.0000	3.0000	0.2222	1.0000	1.1805	3.0000	0.2040
VAR 10	0.6000	0.8330	1.2000	0.0152	0.6000	0.9833	1.2000	0.0154
VAR 12	5.0875	5.5000	5.9125	0.0284	6.1558	6.4927	6.8295	0.0189
VAR 14	0.0006	0.0006	0.0007	0.0000	0.0007	0.0007	0.0008	0.0000
VAR 15	0.1285	0.1836	0.2387	0.0005	0.1718	0.2167	0.2617	0.0003

OPERATIONS AND SUPPORT CASE RESULTS

(CONTINUED)

Summary Uncertainty Distributions, Booster								
	Expert 8643425 Initial Distribution				Expert 8643425 Calibrated Distribution			
	a_1	c_1	b_1	σ_1^2	a_2	c_2	b_2	σ_2^2
VAR 01	90000	100000	110000	16666667	90000	100000	110000	16666667
VAR 04	0.2500	1.0000	1.5000	0.0660	0.2500	1.0000	1.5000	0.0660
VAR 05	0.0500	0.1000	0.1500	0.0004	0.0500	0.1000	0.1500	0.0004
VAR 06	150.0000	640.000	1250.000	50616.667	150.0000	640.000	1250.000	50616.667
VAR 07	27.0000	30.0000	33.0000	1.5000	27.0000	30.0000	33.0000	1.5000
VAR 08	15.0000	25.3000	45.0000	38.7272	15.0000	25.3000	45.0000	38.7272
VAR 09	1.0000	1.0000	3.0000	0.2222	1.0000	1.0000	3.0000	0.2222
VAR 10	0.6000	0.8330	1.2000	0.0152	0.6000	0.8330	1.2000	0.0152
VAR 12	4.9500	5.5000	6.0500	0.0504	4.9500	5.5000	6.0500	0.0504
VAR 14	0.0005	0.0006	0.0007	0.0000	0.0005	0.0006	0.0007	0.0000
VAR 15	0.1285	0.1836	0.2387	0.0005	0.1285	0.1836	0.2387	0.0005

OPERATIONS AND SUPPORT CASE RESULTS

(CONTINUED)

Summary Uncertainty Distributions, Orbiter								
	Expert 8643684 Initial Distribution				Expert 8643684 Calibrated Distribution			
	a_1	c_1	b_1	σ_1^2	a_2	c_2	b_2	σ_2^2
VAR 01	50000	159897	170000	738317256	50000	167892	170000	786192572
VAR 04	0.2500	1.0000	1.2500	0.0451	0.2500	1.0500	1.2500	0.0467
VAR 05	0.0200	0.0500	0.0800	0.0002	0.0200	0.0525	0.0800	0.0002
VAR 06	160.0000	648.0000	1250.0000	49684.666	160.0000	680.4000	1250.0000	49537.786
VAR 07	28.5000	30.0000	31.5000	0.3750	30.4393	31.5000	32.5607	0.1875
VAR 08	15.0000	25.3000	45.0000	38.7272	15.0000	26.5650	45.0000	38.1555
VAR 09	1.0000	1.0000	3.0000	0.2222	1.0000	1.0500	3.0000	0.2168
VAR 10	0.6000	0.8040	1.1000	0.0105	0.6000	0.8442	1.1000	0.0104
VAR 11	239.400	252.000	264.600	26.460	255.691	264.600	273.5095	13.230
VAR 12	2.0000	5.5000	6.0000	0.7917	2.0000	5.7750	6.0000	0.8417
VAR 14	0.0005	0.0006	0.0006	0.0000	0.0006	0.0006	0.0006	0.0000
VAR 15	0.1500	0.1645	0.1800	0.0000	0.1500	0.1727	0.1800	0.0000

OPERATIONS AND SUPPORT CASE RESULTS

(CONTINUED)

Summary Uncertainty Distributions, Orbiter								
	Expert 8646262 Initial Distribution				Expert 8646262 Calibrated Distribution			
	a_1	c_1	b_1	σ_1^2	a_2	c_2	b_2	σ_2^2
VAR 01	140000	159897	195000	129253089	140000	188756	195000	151141704
VAR 04	0.2500	1.0000	1.5000	0.0660	0.2500	1.1805	1.5000	0.0703
VAR 05	0.0300	0.0500	0.0600	0.0000	0.0300	0.0590	0.0600	0.0000
VAR 06	150.000	648.000	1250.000	50566.889	150.000	764.953	1250.000	50651.047
VAR 07	27.7500	30.0000	32.2500	0.8438	33.5774	35.4145	37.2516	0.5625
VAR 08	15.0000	25.3000	45.0000	38.7272	15.0000	29.8662	45.0000	37.5010
VAR 09	0.9250	1.0000	1.0750	0.0009	1.1192	1.1805	1.2417	0.0006
VAR 10	0.6000	0.8040	1.2000	0.0155	0.6000	0.9491	1.2000	0.0151
VAR 11	233.1000	252.0000	270.9000	59.5350	282.0498	297.4816	312.9134	39.6900
VAR 12	5.0875	5.5000	5.9125	0.0284	6.1558	6.4927	6.8295	0.0189
VAR 14	0.0005	0.0006	0.0006	0.0000	0.0006	0.0007	0.0007	0.0000
VAR 15	0.1152	0.1645	0.2139	0.0004	0.1539	0.1942	0.2345	0.0003

OPERATIONS AND SUPPORT CASE RESULTS

(CONCLUDED)

Summary Uncertainty Distributions, Orbiter								
	Expert 86434252 Initial Distribution				Expert 8643425 Calibrated Distribution			
	a_1	c_1	b_1	σ_1^2	a_2	c_2	b_2	σ_2^2
VAR 01	140000	159897	195000	129253089	140000	159897	195000	129253089
VAR 04	0.2500	1.0000	1.5000	0.0660	0.2500	1.0000	1.5000	0.0660
VAR 05	0.0500	0.1000	0.1500	0.0004	0.0500	0.1000	0.1500	0.0004
VAR 06	150.000	648.000	1250.000	50566.889	150.000	648.000	1250.000	50566.889
VAR 07	27.0000	30.0000	33.0000	1.5000	27.0000	30.0000	33.0000	1.5000
VAR 08	15.0000	25.3000	45.0000	38.7272	15.0000	25.3000	45.0000	38.7272
VAR 09	0.9000	1.0000	1.1000	0.0017	0.9000	1.0000	1.1000	0.0017
VAR 10	0.6000	0.8040	1.2000	0.0155	0.6000	0.8040	1.2000	0.0155
VAR 11	226.8000	252.0000	277.2000	105.8400	226.8000	252.0000	277.2000	105.8400
VAR 12	4.9500	5.5000	6.0500	0.0504	4.9500	5.5000	6.0500	0.0504
VAR 14	0.0005	0.0006	0.0007	0.0000	0.0005	0.0006	0.0007	0.0000
VAR 15	0.1152	0.1645	0.2139	0.0004	0.1151	0.1645	0.2139	0.0004

VITA

Bruce A. Conway
Department of Engineering Management and Systems Engineering
Old Dominion University
Norfolk, VA 23529

Bruce A. Conway earned his Bachelor of Science in Aerospace Engineering from Virginia Polytechnic Institute in Blacksburg, Virginia in 1965. He received a Master of Science degree in aerospace engineering from The George Washington University, Washington, D.C. in 1974. His career includes 37 years with the National Aeronautics and Space Administration, from which he retired as a Senior Executive in 1998, and more than 25 years as an adjunct and full-time faculty member of Embry-Riddle Aeronautical University's Extended Campus (currently holding the rank of Associate Professor of Aeronautical Science). While pursuing his doctorate, he has conducted research in expert judgment elicitation and calibration through the Old Dominion University Research Foundation. He is a member of Tau Beta Pi engineering honor society, Sigma Gamma Tau aerospace engineering honor society, and Phi Kappa Phi academic honor society.