ABSTRACT

Title of Dissertation: PROJECT SCHEDULING DISPUTES:
                             EXPERT CHARACTERIZATION AND
                             ESTIMATE AGGREGATION

                             Lauren Elizabeth Neely, Doctor of Philosophy,
                             2017

Dissertation directed by: Dr. Gregory Baecher, Civil and Environmental
                             Engineering

Project schedule estimation continues to be a tricky endeavor. Stakeholders bring a

wealth of experience to each project, but also biases which could affect their final

estimates. This research proposes to study differences among stakeholders and

develop a method to aggregate multiple estimates into a single estimate a project

manager can defend. Chapter 1 provides an overview of the problem. Chapter 2

summarizes the literature on historical scheduling issues, scheduling best practices,

decision analysis, and expert aggregation. Chapter 3 describes data

collection/processing, while Chapter 4 provides the results. Chapter 5 provides a

discussion of the results, and Chapter 6 provides a summary and recommendation for

future work.

       The research consists of two major parts. The first part categorizes project

stakeholders by three major demographics: "position", "years of experience", and

"level of formal education". Subjects were asked to answer several questions on risk

aversion, project constraints, and general opinions on scheduling struggles. Using

Design of Experiments (DOE), responses were compared to the different

demographics to determine whether or not certain attitudes concentrated themselves

within certain demographics. Subjects were then asked to provide activity duration

and confidence estimates across several projects, as well as opinions on the activity

list itself. DOE and Bernoulli trials were used to determine whether or not subjects

within different demographics estimated differently from one another. Correlation

coefficients among various responses were then calculated to determine if certain

attitudes affected activity duration estimates.

The second part of this research dealt primarily with aggregation of opinions

on activity durations. The current methodology uses the Program Evaluation and

Review (PERT) technique of calculating the expected value and variance of an

activity duration based on three inputs and assuming the unknown duration follows a

Beta distribution. This research proposes a methodology using Morris' Bayesian

belief-updating methods and unbounded distributions to aggregate multiple expert

opinions. Using the same three baseline estimates, this methodology combines

multiple opinions into one expected value and variance which can then be used in a

network schedule. This aggregated value represents the combined knowledge of the

project stakeholders which helps mitigate biases engrained in a single expert's

opinion.

PROJECT SCHEDULING DISPUTES: EXPERT CHARACTERIZATION AND
ESTIMATE AGGREGATION

by

Lauren Elizabeth Neely

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2017

Advisory Committee:
 Professor Gregory Baecher, Chair
 Dr. Qingbin Cui
 Dr. Mohammad Modarres
 Dr. Allison Reilly
 Dr. Alaa Zeitoun

Preface

The material in this research is based upon work supported by the National

Aeronautics and Space Administration under Contract Number NNG10WA14C and

Contract Number NNG16WA71C.

Any opinions, findings, and conclusions or recommendations expressed in this

material are those of the author and do not necessarily reflect the views of the

National Aeronautics and Space Administration

Throughout this work, to simplify the grammar, the "Decision Maker" is referred to

as a "she" and the "Expert" is referred to as a "he".

Dedication

This dissertation is dedicated to family. To my parents, brother, sister-in-law, and all my extended family who stood by me and encouraged me throughout this endeavor. I can't thank you enough for helping me to keep striving towards my goal. To my Wallops family, without whom this research would not have been possible. The dedication of the men and women of Wallops has contributed to the success of countless missions and I'm forever grateful that they took time to help me succeed in this personal mission.

Acknowledgements

First and foremost, I thank God for clearing the obstacles I could not and allowing me to pursue this opportunity. I'd also like to thank my advisor for his assistance over the past…well…never mind how long it's been. His recommendations and guidance were instrumental towards focusing my efforts and his suggestions helped shine a light on those efforts when I began to flounder into unknown territory. I would also like to thank Steve Kremer, Nancy Olyha, and Lindsay Robertson for taking time out of their busy schedules to provide a review for this dissertation.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AHP	Analytic Hierarchy Process
AOA	Activity on Arrow
AON	Activity on Node
BC	Best Case
brlt	basic reference lottery ticket
CDF	Cumulative Distribution Function
CDR	Critical Design Review
CI	Consistency Index
COA	Course of Action
CPM	Critical Path Method
DM	Decision Maker
DOE	Design of Experiments
EFT	Early Finish Time
EMV	Expected Monetary Value
EST	Early Start Time
EVM	Earned Value Management
FA	Formulation Agreement
FAD	Formulation Authorization Document
GAO	Government Accountability Office
GEV	Generalized Extreme Value
IG	Inspector General
IRB	Institutional Review Board
ISS	International Space Station
JCL	Joint Cost and Schedule Confidence Level
KDP	Key Decision Point
LFT	Late Finish Time
LoE	Level of Formal Education
LoS	Likelihood of Surprise
LST	Late Start Time
MDR	Mission Design Review
ML	Most Likely
NASA	National Aeronautics and Space Administration
NPR	NASA Procedural Requirement
OMB	Office of Management and Budget
PDF	Probability Distribution Function
PDR	Preliminary Design Review
PERT	Program Evaluation and Review Technique
PMBOK	Project Mangement Body of Knowledge
PRR	Production Readiness Review
SDR	System Design Review

SLS	Space Launch System
SRR	Systems Requirements Review
Te	Total Network Path Duration
WBS	Work Breakdown Structure
WC	Worst Case
WFF	Wallops Flight Facility
YoE	Years of Experience

Chapter 1: Introduction

1.1 The Problem with Scheduling

The Guide to the Project Management Body of Knowledge (PMBOK) tells us that on any given project, several constraints must be managed to achieve project success (PMI 2013, para. 1.3). The schedule constraint, if mismanaged, is one of the more immediate indicators of a problem in the project. On a small scale, if a task does not finish on time, it could drive other tasks in the project also to be late. On a larger scale, when the entire project finishes late, stakeholders begin to question the capabilities of the project manager. How then can a project manager give herself the best chance of success during the planning stages of the project? The quick answer would be to find experts who know the most about the project and ask them for help in putting together the schedule (PMI 2013, para. 6.5.2.1). Herein lies the problem: who exactly is the "expert?" Is it the engineer/technician who does the work? Is it the functional manager who has seen the work over the course of several years? Is it the senior manager who has a better idea of the "bigger picture" across all projects?

The people who actually do the work frequently claim that management does not allow enough time to complete a given project or task (Goldratt 1997, 40). Goldratt, on the other hand, seems to be of the opinion that most time estimates are padded and are larger than they actually need to be (Goldratt 1997, 118). Further compounding the issue is the fact that managers and those who do the work (hereafter referred to as "technicians", to include both engineers and technicians/operators) may

have different views on what defines the success of any activity or project. For example, a technician's key concern may be technical accuracy which could also be interpreted as the project constraint "quality." A manager may be more concerned about the schedule and budget (e.g. it may not be up to full operating specs, but if it meets the requirements, anything further is unnecessary). These different definitions of success could drive different time estimates.

Experience can be another major factor in estimating differences (PMI 2013, para. 6.2.2.3). A senior technician, for example, has seen the worst and will probably make estimates based on those experiences (Kahneman 2011, 236–37; Goldratt 1997, 48). Things do not always turn out badly, however, so when the activity is completed early, it will lead management to believe there was too much padding in the estimate, and they will question the next estimate that is provided (Goldratt 1997, 41). Over time, this back and forth can create tension between management personnel and those they manage. Given the considerations discussed above, how then, should a project manager use the schedule inputs provided by peers and project team members? And if said project manager is questioned on her final schedule estimate, what basis can she use for backing up her decision?

1.2 Goals and Objectives

The goals of this dissertation are broken down into two parts. The first goal is to develop an understanding of differing perspectives of project stakeholders and how project stakeholders estimate differently from one another. The second goal is to provide project managers a method to incorporate multiple opinions when developing inputs for a network schedule.

Through work experience, it was noted that in an effort to develop project schedules, there appeared to be disagreements among certain groups of stakeholders regarding how long activities should take. Based on this observation, the first objective of this research is to analyze the differences in stakeholder opinions about various project constraints and practices based on three major demographics: Position (manager vs technician), Years of Experience (YoE) , and Level of Formal Education (LoE) . Using these same demographic categories, the next objective is to study how project stakeholders differed from one another when asked to provide duration estimates on project activities.

Based on the results noted in the scheduling estimation study, the final objective is to develop a procedure to allow a project manager to use Bayesian methods to update her own beliefs about activity durations based on stakeholder estimates. This updating model is not tied to the results of the first part of the study in that the updating method only considers the estimate provided by the decisions maker and experts, without consideration of their demographic or scheduling trends.

1.3 Potential Implications

Whether it exists or not, there is a perception of a divide between those who manage the work and those who perform the work. If this research can find where the differences hide or if, in fact, project stakeholders are not actually so different from one another, then perhaps the two groups can open a better dialog. Management's perception seems to be that the technicians inflate their estimates when asked "how long will this take." The technicians, on the other hand, seem to be of the opinion that the schedules are not realistic. If this research can expose and

document these underlying beliefs, then perhaps the dialog between the two groups can be improved.

Current scheduling methodology focuses on creating network schedules based on three point estimates (PMI 2013, para. 6.5.2.4). Personal biases (known and unknown) and gaps in information can affect these estimates and ultimately provide bad inputs to the network schedule (Regnier 2005b, 8). By incorporating the estimates of multiple stakeholders, biases can be more readily filtered out. This method could also increase the level of stakeholder engagement in the project by allowing everyone to have their say in the schedule (Surowiecki 2005, 212, 227; PMI 2013, para. 6.5.2.5). The final estimate may not match any one stakeholder's estimate, but it does reflect the collective assessment of the team.

Creating this aggregate estimate represents a departure from the current methodology both by incorporating multiple estimates and by requiring a new distribution model for the three point estimates required by PERT.

1.4 Background of Wallops Flight Facility

The data gathered for this research was obtained by analyzing several active projects at Wallops Flight Facility (WFF), a launch range and test facility located on the Eastern Shore of Virginia. Like Cape Canaveral, WFF provides a spacelift capability (although on a smaller scale than Cape Canaveral), as well as providing a launch area for smaller rockets whose primary mission is atmospheric study or vehicle validation. WFF is owned and operated by the National Aeronautics and Space Administration (NASA) and its primary mission has been to support smaller test and scientific launches as opposed to major spacelift operations, although it has

started to expand its spacelift capabilities (Kremer 2013b, 8–10). "Spacelift" is the ability to use a rocket to launch a payload. Rockets typically consist of two parts: the booster and the payload. The booster comprises most of what one typically thinks of when one hears the term "rocket." It provides the thrust required to allow the payload to travel along its intended trajectory. That trajectory can either be orbital (the payload will orbit the earth) or suborbital (the payload will fly in a parabolic shape and return to the earth without ever reaching orbit). The payload is, in most cases, the booster's raison d'être. It can be anything from a space shuttle to a simple bank of instruments and transmitters (Jenner 2015).

WFF supports a unique subset of the spacelift mission known as the "sounding rocket." In the context of the rocket world, a sounding rocket typically carries a scientific payload on a sub-orbital voyage to gather atmospheric data or data on the geomagnetic fields that create the stunning auroras that can be seen in the extreme northern and southern parts of Earth. These smaller rockets are also used to demonstrate vehicle capability. In these cases, the intent of the mission is not to gather data about our atmosphere, but to gather data about the booster itself ("NASA Sounding Rockets Annual Report 2013" 2013, 4, 20).

Just as the rocket has two parts, a launch campaign also has two parts: the vehicle (described above) and the ground support. The vehicle gathers data and transmits it back to systems waiting on the ground. In order to receive and process these signals, an extensive network of equipment is required. Typically, this ground equipment can be divided into three parts: radar, telemetry, and command (Kremer 2013a, 6). Radars are used to track the flight of the vehicle, which not only tells the

5

scientists/engineers where the vehicle is headed, but also helps determine how well the vehicle is performing (Kremer 2013b, 50). Telemetry assets can also be used to track the vehicle during fly-out, but typically telemetry assets are more concerned with receiving the data transmitted back from the vehicle during its flight (Kremer 2013b, 43–44). Command assets protect public safety by ensuring that an errant vehicle can be destroyed before it violates federal safety criteria (Kremer 2013b, 47). Beyond these major categories, several other systems tie together to provide the required support infrastructure, including communications and networking, data processing, weather measurements, and photo/optical products. Together, all of these systems provide the ground support required to ensure that the data provided by the vehicle during fly-out gets back to the appropriate stakeholders (Kremer 2013b, 42).

1.5 Background of Project Types

  This research deals with three major types of activities at WFF: operations, maintenance, and engineering. Although all three project types accomplish different tasks, they all ultimately point to the same end goal and are necessary to accomplish WFF's mission.

  Operations projects involve supporting the preparation, launch, and post-flight data collection of the vehicles that launch from WFF or one of its deployed ranges such as Poker Flat Research Range in Alaska or the Andøya Space Center in Norway (Kremer 2013b, 7). These projects involve reviewing the requirements of the various range customers and supporting pre-launch testing to ensure that the range instrumentation (telemetry, radar, command, etc.) is interacting correctly with the

vehicle and with the other range instrumentation. When all of these pieces are in place, the range supports a launch by tracking the vehicle and recording the data sent back from the vehicle during flight. After the flight, that data is processed and provided to the customer for further analysis. When supporting at one of its deployed ranges, operations projects involve not only supporting the actual mission along with its pre-launch tests, but in some cases, also bringing up a site that has not been used in several months and ensuring it is still in good working order. This usually requires a team of people to travel to the location prior to the actual operation to get ready for the mission before the customer first requires support.

Beyond operations activities, personnel at WFF are also responsible for maintenance projects which entail maintaining the instrumentation and systems that support launch operations. When personnel are not actively supporting a launch, they must perform scheduled maintenance activities on the instrumentation. This applies to both WFF and deployed sites that have a more permanent set up (i.e. the instrumentation stays in place although the site is not actively manned the entire year by WFF personnel). For the truly deployed sites, the instrumentation is returned to WFF where it undergoes its standard maintenance. Maintenance activities vary in complexity and frequency depending on the type of instrumentation or system on which the maintenance is being performed. Typically, there are two types of maintenance performed on the instrumentation/systems: preventative maintenance and corrective maintenance. The former is scheduled and known. These are specific activities to check out the system/instrumentation and ensure it is in good working order (e.g. clearing dust out, checking connections, greasing gears, etc.). The latter is

unscheduled and unknown. This type of maintenance is performed when something

breaks or does not perform as expected. This type is harder to estimate with respect

to completion time (Kremer 2015, 36–37).

Engineering projects at WFF can be extremely varied in their scope and type.

For this research project, the engineering projects could be described in one of two

ways: system upgrades and system acquisitions. Projects of the "system upgrade"

type typically involve upgrading an already-existing system with a new part,

capability, or software. These projects take systems that already exist and make

changes using locally (at WFF) developed products or "Commercial-Off-The-Shelf"

products which are then tested and integrated into the already-existing infrastructure.

Projects dealing with system acquisition occur when WFF purchases an already-

developed system and integrates it into the WFF infrastructure. These projects

typically involve finding a physical location for the system, assembling and testing

the system, integrating the system with the existing infrastructure at WFF, and finally,

certifying the system for operational use (Kremer 2013b, 37–45).

1.6 Research Summary

Given the dynamic nature of projects and specifically projects at WFF,

developing an accurate schedule can be a challenge. Some believe too much time is

given on a project while others believe not enough time is allotted. Unexpected

challenges during project execution frustrate the technicians who execute the tasks,

leaving them with a desire for more time for the next similar project. When the next

project goes smoothly and does not require the full amount of allotted time,

management is left feeling like the project could have been completed more quickly. As time progresses, these mindsets become engrained while the project manager is left trying to find the "right" answer (Kahneman 2011, 80–81; Goldratt 1997, 40–41).

In order to determine trends in estimating practices, subjects from a variety of different backgrounds were asked to provide activity duration estimates on several projects of the types described above. Subjects were provided several surveys, the first of which was a survey that captured basic demographic and project-constraint preference information. Later, subjects were provided different project surveys with lists of activities required to complete each project. These surveys were designed to capture estimates on how long activities should take and determine whether or not subjects believed the provided list was accurate. A second survey was provided to those engaged in executing the projects to record how long the activities actually took along with any other changes or challenges that took place during the project.

These survey responses were compiled and analyzed using Design of Experiments (DOE) to determine if there was any correlation between the demographics of the subjects and the results of the other surveys (Montgomery 2008, 208–10). A new estimating method was then developed which used Bayesian updating to combine the inputs of multiple experts (Morris 1977).

Because the human element plays a heavy role in project planning and execution, responses obtained during the period of study were also analyzed to determine if project stakeholders think differently from one another and if those opinions are part of the disconnect that seems to occur when determining how long a

project or activity should take. These observations were then compared to the scheduling data to determine if the stated opinions of different stakeholders matched their scheduling estimates in the hope of revealing some of the underlying reasons for why different stakeholders estimate the way they do.

Ultimately, this research seeks to provide insight into the mindsets of a diverse group of project stakeholders and provide a method to combine these diverse opinions into one estimate that can be used in the development of a network schedule. By having a better understanding of the thought process behind the estimates and by including estimates from multiple experts, a project manager can not only create a better project schedule, but can also better defend one should it go awry. By gathering real world data, it is hoped that this will be reflective of what a project manager will actually encounter when asked to develop a schedule, making the results of this research a useful tool to help accurately assess how long it should take to successfully complete a project.

Chapter 2: Literature Review

The process of scheduling a project can be very complicated. Politics, budgets, past experiences, and present "unknowns" are just some of the challenges faced by a project manager trying to determine a likely completion date for a given project. Several scheduling "best practices" exist and are available for use by a project manager, but those best practices are entirely dependent on the input provided to them (Malcolm et al. 1959, 650–51; Grubbs 1962, 914; Pickard 2004, 1569). The inputs to these scheduling best practices should come from the "experts" (PMI 2013, para. 6.5.2.1), but how do those experts decide on what their inputs should be? Are scheduling challenges seen at Wallops Flight Facility unique or has NASA as an organization encountered similar problems? This chapter will provide an overview of the scheduling challenges faced by NASA over the past several decades to see if there are any trends that can be applied to the scheduling challenges at WFF. The chapter will then go on to discuss best practices for scheduling and some caveats that accompany those best practices. It will then move on to the current literature on decision analysis and how it can affect scheduling estimates. It will conclude with a discussion of the Bayesian aggregation method used in this project.

2.1 Scheduling in NASA – GAO Reports

While Wallops Flight Facility may have a unique mission within the constructs of NASA, the project management (and specifically scheduling) challenges

experienced by the project teams at WFF are not unique to the facility. According to

its website, the Government Accountability Office (GAO) is responsible for

monitoring government spending of American tax dollars. Within this role, they

provide reports on how well certain programs are being managed along with any

concerns about the ability of the project to be successful. These reports document

challenges encountered and often provide recommendations for overcoming these

challenges and how to proceed. ("About GAO" 2015) A word search on "Schedule"

was conducted on the GAO website, with those results being further narrowed down

to those reports related to NASA. This search returned nearly 800 results, and of

those approximately 75 were chosen and reviewed based on the apparent applicability

provided in the report's abstract. These reports spanned a variety of projects and

several decades, but many seemed to have several common themes that played out

over and over again. The information below is a summary of the issues identified in

those reports which seem to be contributing factors to schedule challenges. One

interesting thing to note throughout this section are the years shown in the references.

The first two-digit number in each of the references describes the year the report was

written. In several cases, the same issue is described years (and even decades) apart.

2.1.1 Lack of Resources/Inadequate funding

One recurring theme seen throughout several of the reports was that of

schedule delays being caused by a lack of resources and/or inadequate funding. In

the movie Apollo 13, there is a scene where engineers are working to develop a

procedure to turn the Command Module back on after it had been shut down for

several days. The required systems are determined, but those systems will overreach

the available power budget. At one point, one of the engineers states that the

command module thrusters must be warmed up due to the extreme cold of space and

the other engineer replies that he will have to trade off the parachutes or something to

make that happen. The first engineer responds that if the parachutes do not open,

then there is no point to continue trying. The second engineer then replies with a

statement that has stuck with this author as applicable to nearly all resources

constraints: "You're telling me what you need. I'm telling you what we have to

work with at this point. I'm not making this stuff up." (Howard 1995). The same

principle can be seen with nearly any resource required for a project. Although

funding is the resource that comes most readily to mind, there are several which must

be considered, including: time, money, technology, personnel, and knowledge (GAO

2011, 7, 2009b, 6). In one example involving the Space Launch System (SLS), the

report stated that the program's budget was $400 million short of what it needed.

Without the required funds in place (among other issues) officials at NASA were not

able complete the contracts needed to proceed with development. This in turn

increased the risk to both the cost and the schedule to the program. (GAO 2014, 10–

11). NASA told the government what it needed and the government replied with

what NASA had to work with. This is just one example, but it can be seen over and

over again across multiple projects spanning nearly forty years. Without the

resources required to execute the tasks in the schedule, whether it be people, money,

or equipment, it does not matter how well one estimates how long something should

take. Without the capability to get started, the duration will remain "indefinite".

Returning to the example of the SLS, NASA realizes its need to operate within a constrained budget. While it is doing its best to keep within the prescribed funding limits, the program has consistently struggled to ensure technical and programmatic requirements of the system are met within the constraints of available funding. The program has listed this as its number one risk and stated that it does not believe its current planned budget will cover the current design, which does not even account for changes and challenges during development and testing. This lack of funding is predicted to delay the launch date by six months which, in turn, increases the overall cost (GAO 2014, 11). Even forty-five years later on a project designed to once again carry humans into space, one group is telling the other what it needs, the other responds with what it has to work with. Based on a recommended "best practice called the Joint Cost and Schedule Confidence Level (JCL), NASA requires its launch programs to have a 70% probability of meeting its cost and schedule baselines. The JCL looks at the proposed requirements, cost, and schedule goals of a given project and analyzes the probability that the project can meet those goals (GAO 2014, 5–6). Given the problems already encountered by the SLS system, NASA must decide what it will sacrifice in order to keep the project moving forward: increased cost, increased schedule, or pressing forward with a JCL rating of less than 70% (GAO 2014, 10–11).

A mismatch of resources and requirements is not necessarily always the fault of the project team, especially in the case of research and development. In some cases, the teams knew what was required to successfully complete the project, but the resources simply were not available (GAO 1991a, 29; Martin 2012, 27). A recurring

theme throughout several of these reports seems to be delays in receipt of funding

from Congress (GAO 1988a, 1,2,5, 12, 14-15, 1991a, 4, 1977d, 3). In some cases,

this was due to governmental constraints that were out of NASA's hands. One report

released in 2012 states that, since its inception in 1959, NASA has started the fiscal

year with its allocated funding only seven times. Without the funding in hand,

managers had to restructure the project plan in order to conform to the available

resources (usually in the form of some type of continuation) (Martin 2012, vii). In

other cases, if Congress does not believe that a project can meet its stated cost and

schedule estimates, it can delay funding until NASA can provide such assurances

(GAO 1991a, 31, 2008, 10, 1997, 6). If designs and plans lag behind early in the

project, Congress may delay funding until it has some assurance that the project can

succeed. If the perception is that the project is mired in problems, then it is less likely

that Congress will authorize funding, even if the program is already in work (GAO

1991a, 30–31, 1991b, 5). In other cases, funds were simply not approved, causing

delays in start dates which propagates through the project (GAO 1980a, 44–45). In

one report, a response from NASA criticizes the author for failing to acknowledge

that funding constraints were a major contributor to projects running behind schedule

and that these funding constraints were externally driven (GAO 1980a, 65).

In yet another budgeting challenge with Congress, project managers must

contend with increased scope and stagnant budgets (Martin 2012, 29). This is

another example of the *Apollo 13* phenomenon of funding: NASA tells Congress

what it needs, Congress responds with what NASA has to work with. As mentioned

before, in the SLS program, NASA is striving to remain within the budget profiles set

by Congress. Despite efforts to remain within this profile, the number one risk is that it will run out of funding prior to the first launch. Which will push the launch date out. Which will cause an increase in required funding. Which will push the launch date out… (GAO 2014, 11). Given the vast portfolio that must be managed, NASA works to create levels of prioritization among its projects. The theory is that NASA will rank its projects such that the approved projects will fall within the funding profile allocated by Congress and ensure that the most important projects get the funding they need. The problem, though, is that even with this prioritization, NASA was exceeding the likely allocation it would be provided by Congress. When the allocated funding is not received, sacrifices must be made to other project constraints (GAO 1994a, 1–2).

Another major recurring theme was that of NASA officials having to manage and estimate project costs based on annual budgets as opposed to life-cycle costs (GAO 1988a, 19, 2002a, 2). Because NASA is required to manage projects based on annual funding requirements, funding may not necessarily be available in accordance with the planned schedule (GAO 2002b, 10–11). In cases such as these, the funding seems to be driving the schedule as opposed to matching funding to scheduled milestones as would be recommended in an Earned Value Management (EVM) construct (GAO 1994a, 1; Mantel Jr. et al. 2004, 237–44). When that funding is not available, adjustments must be made to the project in order to remain within the budget constraints (GAO 1988a, 5, 19). Even high priority projects such as the space shuttle fall victim to managing by annual budget. One report stated that aspects of the program experienced schedule extensions of 13-15 months with the primary driver

being the need to remain within the annual budget (GAO 1977d, 3). A report issued

that same year described a space telescope project that was delayed from the

beginning by at least one year due to requests for funds being denied by the Office of

Management and Budget (OMB) (GAO 1977c, iii, 4). In another example, from a

later date, even the International Space Station (ISS) experienced schedule delays that

resulted from trying to make the project plan fit the annual funding schedule. In this

same report, NASA admitted that in this instance the funding delays were not a major

issue, but that the uncertainty caused by unstable funding profiles did negatively

affect the stability of the project. It also stated later in the report that trying to match

the project plan to allocated funding forced a schedule delay of 18 months, although it

did provide some improved stability in the plan (GAO 1991a, 4, 29, 34). In an

interview conducted by the Inspector General (IG), personnel across NASA were

asked about different challenges facing their projects. In this survey, funding

instability was cited as a major challenge (nearly 75% of respondents listed this).

When the budget was changed, the teams had to adjust their projects accordingly

which often affected the overall schedule (Martin 2012, 25). Because of the lifespan

of several of these development projects, NASA also faces the challenge of keeping

funding in the face of changing government officials in both the executive and

legislative branch. An effort that was a priority for one president may not be a

priority for another. Congressional leaders change and with those changes, the

allocation of funding can change as well (GAO 2008, 18).

Given all of these issues with funding, there are some other basic underlying

causes which are major contributors to the scheduling problem. One of these issues

(which will be discussed later in this chapter) is that much of what NASA deals with is research and development. These types of projects are notoriously difficult to estimate because of all the unknowns. As the teams progress in the project, they gain more and more understanding and unknowns resolve themselves into increases in the requests for budgets and schedules (GAO 2014, 7, 2012, 12). The problem is that, whether legitimate or not, that initial budget declaration becomes an anchor point from which NASA and Congressional leaders base their perceptions (Kahneman 2011, 119). Projects that do not live within that perception can then run into funding issues when they request more money (GAO 2014, preface). This also gives the perception to Congress that the project is not under control, which makes Congress less likely to provide more money (GAO 1991b, 5, 1991a, 30–31, 2003, 8–9).

Another major contributor is what has been dubbed the "Hubble Psychology" (Martin 2012, 16). The Hubble telescope was a complete disaster from a project management perspective, exceeding cost and schedule estimates and initially being plagued by technical problems. Despite this, it continued to receive funding and schedule support and engineers were ultimately able to resolve issues. Now Hubble provides unprecedented views of our universe, making its project management failures pale in comparison to its technical success (Martin 2012, vi). This psychology has given rise to the belief that as long as a team can achieve technical success, sins against the more materialistic success criteria will be forgiven. This does not inspire project managers to be overly concerned with whether or not their projects come in on time and on budget as long as the project is a technological success (Martin 2012, 11–12). In general, NASA has a culture of optimism which

helps bring about these technological successes. Its "go forth and conquer" mentality

allows people to accomplish amazing things (Martin 2012, 37–38). An interesting

contrast to this culture of optimism, however, is NASA's culture of safety/mission

assurance-before-cost/schedule, but it results in the same prioritization of mission

success over project constraints. For all the wonderful things it has accomplished,

when NASA fails in its technological endeavors, it tends to fail spectacularly (or

worse). Missions often consist of one-of-a-kind payloads or, even more importantly,

human lives. In the event of a mishap, the former is difficult to recover from, the

latter, impossible. Because of these high stake missions, NASA must carefully

consider its management of project constraints (GAO 1988b, 18, 1977d, 60, 1977a, 9;

PMI 2013, para. 1.3; Martin 2012, 13,18). A quote from Walt W. Williams, the

Program Manager for X-15 and Mercury perfectly sums up the attitude of NASA

towards safety versus schedule: "You will never remember the many times the launch

slipped, but the on-time failures are with you always." (waynehale 2015) In an

environment such as this, it is highly unlikely that risk mitigation options will favor

relieving cost and schedule risks when those mitigations could potentially cause a

technological mission failure (GAO 2017, 15–17, 22–23; Mantel Jr. et al. 2004, 105).

Once a project is under way and effort has been expended on it, it becomes

much more difficult from a psychological perspective to give up on the project (Arkes

1985, 129). The longer the team works on the project and the more money invested,

the more attached team members and managers become, reflecting the concept of

"sunk cost" (Kahneman 2011, 345; Arkes 1985, 132). As resources are "sunk" into

the project, the attitude of, "we've already put so much into this, let's just finish it"

becomes harder to escape (Kahneman 2011, 354; Arkes 1985, 135). Some would argue to ignore what has been done and focus only on whether or not it makes sense to continue down the current path, although others would argue careful consideration of all factors is required (Kahneman 2011, 343; Mantel Jr. et al. 2004, 270; Farr 2012, 5–13; Arkes 1985, 124). A prevailing attitude at NASA, however, is that as long as the project continues to make technical progress, "someone" will find extra funding to keep the project alive (Martin 2012, vi). The problem with this, though, is that in some cases, the funding must come from other, lower priority projects (Martin 2012, viii). Part of the "sunk cost" struggle is that it means admitting defeat on a goal. NASA encourages a "can do" culture of optimism that translates over into its project management. When given a project, the tendency is to say "yes", despite possible funding and schedule challenges (GAO 1993a, 12; Martin 2012, iv). If the project managers cannot remain grounded in the initial stages of planning, then the project has little hope of meeting the already-unrealistic schedule once issues and challenges arise (Martin 2012, 12–13).

As previously mentioned, one of the major hurdles to successfully managing project constraints at NASA is the instability of available funding. The resulting uncertainty leads to issues not only with actual funding concerns, but also with another critical resource: people. The revolving funding door takes its toll on project members and their motivation to continue work knowing that at any moment, their project could be on the chopping block (GAO 1991a, 23, 29, 2008, 18, 1993a, 11, 1991a, 32). When people are worried about their jobs, they will be less likely to focus on solving the technical problems at hand. This in turn means that project

managers will need to spend more time focusing on managing personnel issues and less time managing project constraints (GAO 1991a, 32).

Even the fictional space research and development projects run into personnel problems. In the movie *Return of the Jedi*, the project manager in charge of Death Star construction insisted that the schedule could not be met because he needed more men (Marquand 1983; Ward 2015, 68). As previously mentioned, the aerospace career field is highly specialized and requires a very specific skill set, so even if there are enough people available, having the right skill set is equally important. A good project manager can help keep a project moving towards schedule completion, but, to quote David Mamet, "Old age [experience] and treachery [also experience] will always beat youth and exuberance" (Mamet 2015). Although research and development projects can be very different as far as requirements, experience can teach a project manager where to look for pitfalls and also how to "work" the system to get things done (e.g. where to get approvals, who to ask, good times to ask, bad times to ask, how to anticipate and mitigate personnel issues, etc.). NASA is facing a growing concern over its workforce development as its experienced project managers and engineers are beginning to reach retirement age. Those who know the ins-and-outs of the systems and who also know how to recognize a trend which can lead to a problem are starting to leave. Those who remain behind will become good project managers in their time, but they still need time to develop (GAO 2006a, 4, 2006e, 10).

The other problem facing NASA is the capability to backfill people once they retire or as new projects come online. Funding limitations make it difficult to hire on

new people, not just in management roles, but in technical roles as well (GAO 2006d, 6). Personnel are also challenged with performing work on multiple projects, forcing them to prioritize which projects receive attention. In these cases, trying to do more with few people usually results in work being put off until time is available. Personnel find it difficult to remain dedicated to side projects when their primary jobs are already consuming a significant amount of time (GAO 1991b, 27, 2006e, 10, 15, 1980b, 13). Further complicating this issue is the fact that NASA has outsourced much of its technical knowledge base which now rests more with contractors than it does with the government civilians (GAO 2006a, 3–4). Because of this shift, NASA must now provide technical and project oversight to new contractors who may or may not have experience with the types of projects NASA requires them to do. This inexperience can lead to costly delays as work must be re-done to meet the required standard (GAO 1991b, 27–28). In some cases, it is not only the contractor who lacks experience, but, as described above, the NASA project manager as well. This inexperience can affect how well the project is managed not only from a technical perspective, but from a project management perspective as well. Without the proper direction, contractors hired to do the job must fulfill requirements to the best of their understanding, but that understanding may be incorrect (GAO 2006f, 14, 1994a, 5, 1991b, 28). Schedule challenges are further complicated when funding is not available for outsourced work to be completed or when a contract cannot be definitized. When allocated funding is withheld, contractors cannot begin (or continue) to work. This can delay work to the point that it affects the overall completion of the entire project (GAO 2014, 17, 2009a, 14).

While funding is one of the major resources in short supply on a given project, other resources can also wreak havoc with planned schedules. In a specialized field such as aerospace, facilities can also be a cause for concern with respect to schedule. When several projects are vying for the same test facility, invariably someone must give way, which will usually result in a schedule delay (GAO 2008, preface, 2008, 13). Other times, facilities with the required capabilities are no longer in existence, having been shut down in previous rounds of budget cuts (GAO 2008, 14). In some cases, facilities are available, but there are no people to man the facilities (GAO 1976, i). Facilities are not the only material resources that can end up in short supply. Hardware and software can also delay schedules when it is either late in delivery or quality issues require re-work. This requires finding alternative ways to make the technology work which can, in turn, lead to more schedule delays. In some cases, equipment has become so obsolete that the technology required to put together the equipment no longer exists or is much harder to find which can also cause schedule delays (GAO 2004, 11, 1994b, 3–4; Martin 2012, 22–23). (GAO 2012, 31)

2.1.2 No overall plan (business case)

While some of the funding issues discussed in the previous section were out of the project manager's control, other issues may have been exacerbated by a failure to have a valid business case that adequately described the resource needs of the project (GAO 2014). One project management best practice states that prior to the start of any project, a business case should be developed to demonstrate the need for the project at hand (PMI 2013, para. 4.1.1.1). NASA goes on to define the business case as ensuring that project resources are matched to customer needs. Here, NASA

defines resources not only as time, money and people, but also knowledge (GAO 2006a, 10). As mentioned in the previous section, a major problem facing NASA right now is the retirement and outsourcing of its project management staff (GAO 2006a, 22). This exodus is a major concern for NASA because, as people leave, the knowledge leaves with them. Without this knowledge, it is much more difficult to accurately estimate how much something will cost or how long it will take to complete (GAO 2006a, 4). Both PMBOK and NASA state that cost and schedule estimates should be derived from past project's records and expert opinion, but when all the experts leave, the ability to make good estimates leaves with them (GAO 2007, 4, 2006a, 11,22-23, 2003, 7, 2004, 2, 2006b, 2–3, 2012, 4, 2011, 8, 2009c, 5–6; PMI 2013, para. 6.5.2.1, 7.2.2.1).

One GAO report recommends that NASA should implement policies which require better reviews before moving from one project development stage to the next. They refer to this approach as a "knowledge-based" approach to systems engineering. Basically, each project is required to prove that they have the "knowledge" needed to proceed to the next phase of development (GAO 2008, 16). This includes understanding the requirements and how the project will meet those requirements as well as (and this is stressed several times) whether or not the technology currently available to the project is capable of meeting those requirements. They further state that these projects should have good requirements and well defined cost and schedule estimates before progressing from "formulation" to "implementation" (GAO 2006a, 3, 2012, 5). This matches with PMBOK's recommendation of planning the project before moving on to the execution phase (PMI 2013, para. 3.4). Several GAO reports

mention that a failure to obtain the correct knowledge base prior to beginning a project or moving to the next phase significantly increases the probability of a "project management" failure of the project (GAO 2014, preface). One of the recurring themes in the later GAO reports is that the NASA teams seem to start projects without the knowledge required to truly evaluate the probability of success. It is almost a "figure it out along the way" mentality. Interestingly, this concept of knowledge-based engineering appears to be specifically called out more frequently only within the last ten to fifteen years. Prior to that, the general idea may have been mentioned, but problems were mostly blamed on the familiar culprits of inadequate funding, frozen budgets, and changing requirements.

NASA indicates that from their perspective, a business case must not only address the technical specifications of the program, but it must also show that the required technology is available and that the basis for the budget and schedule is reasonable (GAO 2009c, 6). PMBOK states that a business case is created to, "determine whether or not the project is worth the required investment" (PMI 2013, para. 4.1.1.2). Based on these GAO reports, NASA's business cases seem to have a slightly different purpose to them in that they seem to occur later in the project lifecycle than is discussed in PMBOK. According to PMBOK, development of the project charter occurs in the "Initiating" Process Group, which is the first process group in the lifecycle of a project. The project charter is the official approval to proceed for any project which means that no real work can begin on the project until it is approved (PMI 2013, para. 3.3). The business case is listed as one of the inputs to the project charter, meaning that the business case must be developed before any

work on the project officially begins. The business case itself is an analysis of a

statement of work which describes the high-level need and general scope of the

project. The business case then provides a high-level analysis of the statement of

work to determine if the benefit of undergoing the project has enough return to justify

the cost of the effort (PMI 2013, para. 4.1.1.2). At this early stage of the project, it

would be nearly impossible to have a good understanding of exactly what the project

would entail with respect to requirements, cost, and schedule. According to the

PMBOK model, only high level information would be available about the project at

this time. In fact, in some cases, a project manager has not even been assigned at this

stage (PMI 2013, para. 3.3, 4.1).

The next process group according to PMBOK is the "Planning" process

group. In this process group, the project manager and the team take the high-level

information of the project charter and begin to refine it into actionable parts. This

process should result in the Project Management Plan which should document every

aspect of what will be required to successfully meet the business need stated in the

project charter (PMI 2013, para. 4.2). The first step in creating the Project

Management Plan is to define the scope of the project (referred to as "Project Scope

Management") and one of the major steps of project scope management is to collect

the requirements (PMI 2013, para. 5.2). This step is crucial to the success of the

project as all project constraints will be tied to these requirements. The success or

failure of the project will also be judged in most cases by how thoroughly these

project requirements are met (PMI 2013, para. 3.4, 5.1.3.1-5.1.3.2). After

determining requirements, it is recommended that the project team define the scope

and create the Work Breakdown Structure (WBS) of the project. A basic scope has

previously been defined in the project charter, but now that the team has well-defined

requirements, the scope can be more accurately defined (PMI 2013, para. 5.3).

Defining the scope helps prevent "scope creep" where project stakeholders seek to

expand on the requirements. These expansions can wreak havoc with project costs

and schedules, but they can be difficult to challenge if they can be tied to something

already within the scope of the project (Mantel Jr. et al. 2004, 42). The final step in

the planning process group of the scope process is to establish a WBS. The WBS

translates the requirements into actions to be taken by the project team. These actions

can then be assigned a cost in terms of labor and materials and can also be assigned a

duration (how long it should take to complete the activity) and organized into a

schedule (PMI 2013, para. 5.4.2.2, 5.4.3.1). At this point, the project manager should

have what is needed to portray to "the powers that be" an accurate depiction of the

best estimate of what it will take to complete the project.

NASA's project development processes are defined in NASA Procedural

Requirement (NPR) 7120.5E. These processes are further described in a "best

practices" handbook called the *NASA Space Flight Program and Project

Management Handbook* (NASA/SP-2014-3705) which was released in September

2014. The document covers both program management and project management,

stating, much like PMBOK, that projects must fit into the overall strategic goals of

the organization (NASA 2014, 21; PMI 2013, para. 4.1.1.2). NASA's planning

processes break major projects into six phases (designated "A" through "F"), and in

some cases a "pre-phase A" for concept development.   Each phase concludes by

undergoing a boarded review, information from which is use in a "Key Decision

Point" (KDP). These KDPs provide senior leadership the chance to review the

project's current progress and determine whether or not to allow it to continue. Each

KDP is a gateway point that the project must pass before entering into that particular

phase, so "KDP A" will usher in Phase A (as opposed to concluding it) (NASA 2014,

114). Figure 2-1 (NASA 2014, 26) below shows the entire project process along with

the associated reviews and decisions points. Several of these will be described in the

following paragraphs.
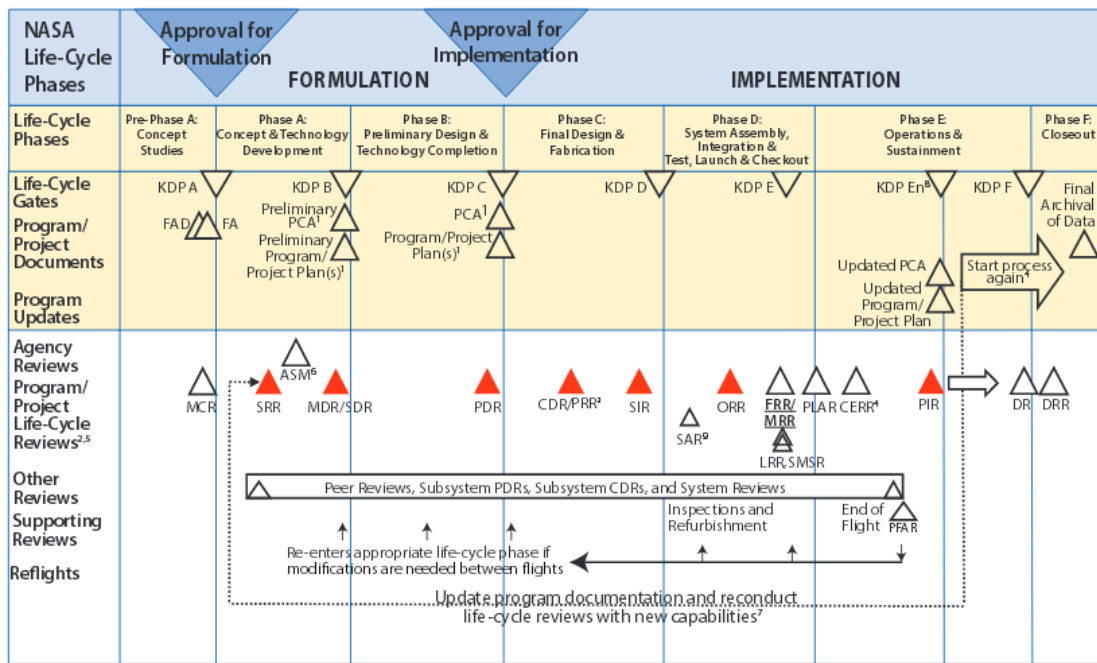
Figure 2-1: NASA Project Life Cycle

The six phases just discussed are further divided into two stages referred to as

"Formulation" and "Implementation". Prior to Formulation, the project engages in

"pre-phase A" activities where a need or concept is identified and analyzed to ensure

it aligns with the overall strategic goals of NASA. These projects then undergo a

high-level analysis to determine feasibility and potential challenges that could face

the program (NASA 2014, 138). These concept studies probably most closely match

the "business case" as described in PMBOK. They look at the different mission ideas

presented to upper management and determine which one is the most likely to

produce a good return on investment. Once a mission concept is selected, upper level

management at NASA develops the Formulation Authorization Document (FAD).

This document most closely matches a "Project Charter" as defined by PMBOK in

that it officially authorizes the project to begin and covers a wide variety of high-level

project characterizations such as scope, funding, authority, and constraints.

According to NPR 7120.5E, this document should contain, "requirements, schedules,

and project funding requirements." (NASA 2015a, 24, 2014, 141) The NASA Project

Handbook further clarifies that these should be project level requirements at this stage

and project-level cost and schedule, reflecting at least the completion date and

possibly broken down further into the cost and general schedule of each phase of the

project (NASA 2015a, 143, 146). The project team then responds with the

Formulation Agreement (FA) which is a preliminary plan to meet the requirements

described in the FAD (NASA 2015a, 25).

Once the project has been officially approved and passes KDP-A, it begins to

refine the mission concept. Throughout Phase A, a preliminary Project Plan should

be developed containing many of the same sections as a PMBOK recommended

Project Plan (NASA 2015b, 33, 2015a, 137–77). At the end of Phase A (at KDP-B),

the project requirements should be refined to at least the system level and the project

team should have an idea of what sub-system requirements will be (NASA 2014,

153). At KDP-B, the project team should be able to provide external stakeholders a general roadmap describing when and where the time and money will be spent (NASA 2014, 154). Within this phase, the team will conduct a Systems Requirements Review (SRR) which is meant to demonstrate that the project requirements as understood by the team will fill the need defined at the program level (NASA 2014, 32). Once the requirements are approved, the team will continue to develop its architecture and undergo a System Design Review (SDR)/Mission Design Review (MDR). These reviews communicate the team's plan of execution to the review board who will then provide an assessment as to whether or not the course of action will meet the approved requirements (NASA 2014, 153).   The cost estimates should be broken down into fiscal years expanding over the expected life of the project by this point (NASA 2014, 160).   At KDP-B, the team should have a good understanding of what the project should accomplish (requirements), how to accomplish that objective (technical plans), the resources needed to complete those plans (time, money, people, materials, etc.), and they should be reasonably certain that it can be accomplished within the provided estimates of those aforementioned resources (NASA 2014, 153–55, 165–66). All project planning to date should be consolidated into a preliminary Project Plan, which should be available for review by stakeholders by the SDR/MDR (NASA 2014, 173).

Once the team has successfully navigated through KDP-B, Phase B can begin. This phase is characterized by further refining the requirements and planned design. By the end of this phase, requirements should be baselined down to the sub-system level. Cost and schedule updates should be made based on the team's understanding

of the current risks facing the project and the Project Plan will be baselined prior to the Preliminary Design Review (PDR) (NASA 2014, 183,185). The team should also begin refining its time-phased cost estimates and comparing it to the project budget to be provided by Congress. The non-monetary resource requirements are also updated at this point to reflect the project team's better understanding of requirements and plans (NASA 2014, 182–183,185).

"Phase C" is characterized by further refinement of the plans in "Phase B". This is the last phase before full scale fabrication and testing of the system to be delivered by the project, so the team and review panel must ensure that details are understood (NASA 2014, 182–183,185). As stated before, the Project Plan has been baselined by this stage, so the team begins to implement the described execution plans. The team should also continue to provide updates on cost, schedule, risks, and resources throughout this phase. At this point, especially for large and expensive projects, the team must inform upper-level management of any milestone that is anticipated to be delayed over six months. They must also inform upper management of any cost growth in excess of 15%. For projects expected to have a life-cycle cost over $250 million, increases above 15% must be reported to Congress. Increases over 30% could be subject to re-authorization. In this phase, the team must undergo a Critical Design Review (CDR) to prove that the design is ready and also a Production Readiness Review (PRR) to prove that the team is ready to produce the systems required to successfully complete the project (NASA 2014, 189–98). "Phase D" is where the team actually implements all of the technical plans and begins to build and test the system. Drawings and technical documents reflect the "as-built"

configuration and are baselined. "Phase D" completes with the successful initial operational function of the project in question (NASA 2014, 196–205).

Ultimately, each lifecycle phase of the project is an expansion and refinement of the previous phase. As the team learns more about the project, requirements are better defined, which allows for more detailed designs, which allows for a better informed cost and schedule estimate. In the years prior to the NASA Program and Project Management Handbook previously described, the GAO criticized NASA for failing to follow good project management practices. Based on data from one report on the Constellation program, the major culprit seems to have been that the program/project manager did not fully develop the required information in the early phases of the project lifecycle. There was a lack of understanding by those involved, especially when it came to managing customer expectations such that they fit within the allowable resources of the project The report also stated that the project team lacked a good understanding of the requirements and exactly what resources would be required to meet those requirements (more will be discussed on requirements in the next section). It was also stated that the project team fell victim to its own optimism and failed to correctly estimate how much time and money it would take to successfully complete the project (GAO 2009c, 1,3,5-6).

It does appear that NASA has made great strides in its efforts to close the knowledge gaps called out in multiple GAO reports (GAO 2008, 8, 2006a, 13, 2006b, 2). In a report from 2006, GAO recommended implementing several different "Knowledge Points", where Knowledge Point 1 represented the point where the team could show that the requirements could be met with the available resources. It also

stated that it believed that NASA did not have a system in place which adequately

analyzed whether or not the current level of technology was adequate to meet the

requirements of the project (GAO 2006a, 3–4, 10, 13–15). NASA seems to have

taken this advice to heart and has updated its best practices as described in the

previous paragraphs. These updates include multiple reviews and plans that ensure

that the right people are looking at the project to ensure technology and other

resources are in place prior to making a major commitment to the project. In previous

versions of NPR 7120.5 (the version active when the above reports were written),

there were reviews required, but they were not nearly as extensive as the current

version. The NPR also did not have as many phases and "back down" points as the

current version (NASA 2015b).

While NASA's phases and definitions of project planning may vary from that

of PMBOK, the overall end-goal is the same. Both groups seek to clearly define how

a project will further the overall goals of the company/agency and both processes are

designed to help manage project constraints and ensure that stakeholders have a good

understanding of what is being asked of them. By following these best practices, the

project team is given its best opportunity to successfully complete a project (GAO

2006a, 11).

2.1.3 Changes, Uncertainty, and the "Experts"

The previous section described the best practice of developing a viable

business case. It also described the issues that were caused when a project failed to

develop this business case. There are many challenges NASA has faced in trying to

develop an overall viable business case, some having to do with failure to follow best-practices, some well out of control of the project manager.

One of the major struggles faced by many projects at NASA was the fact that requirements often were not well defined prior to the start of the development phase of the project (GAO 1993b, 4, 1993a, 11). It is nearly impossible to fully develop a complete requirements list early in the project and high level requirements rarely provide enough detail to develop a truly legitimate schedule (GAO 2014, 25). One report stated that a failure to adequately define requirements for both technical and management aspects of the program was the most significant cause of both cost and schedule growth (GAO 1993a, 11). When a system is not fully defined, the design team may need to spend a significant amount of both time and resources working re-designs (GAO 1991a, 4). NASA is aware of the struggles and consequences of a failure to develop detailed requirements and has even stated that it is expected that research and development projects are going to experience changes (GAO 1977b, ii). In some cases, this is simply a matter of requirements changing due to a better understanding of the system and how it will work, as opposed to an outright failure to adequately define requirements (GAO 1977c, 14). NASA's process even allows for this, as discussed in the previous section, where requirements are refined even after the official requirements review. Trying to develop a schedule in the midst of this uncertainty presents a challenge to project teams. Without fully knowing what changes will occur, it can be difficult to anticipate how long something will take GAO 2014m, 3). When project requirements are not well defined, the project team must make assumptions about the intent of the requirements as they are written.

34

When the team begins to design when requirements or other aspects of the project plan are still in flux, there is a very good chance changes will be required after the team has already invested significant time and effort into a plan. In some cases, the project will make it all the way to the Implementation Phase before design problems are discovered which can cause massive amounts of rework (GAO 2001, 4, 2001, 7). Another issue in developing requirements is ensuring that all stakeholders are able to review and discuss requirements before they are finalized. If key stakeholders are excluded from the reviews, costly re-work could become necessary when the project is more developed and less adaptable (GAO 1998, 16,19, 1992a, 3–4, 1982, 1,3).

In the case of the Ares I (rocket) and Orion crew transport (payload), although requirements actually were baselined at the project level, some uncertainty remained regarding the more specific requirements at the system level. These efforts were both separate projects, but were tied to one another and being developed together. When the team had uncertainty regarding specific technical requirements about the systems, it made it difficult to guess at the correct design that would be optimal for both projects, which led to re-baselining at least one of the projects (GAO 2008, 8). In another example of another major NASA project, the James Webb telescope ran into trouble because the launch vehicle was not selected until the telescope was already being designed. Once the vehicle was selected, it was discovered that the telescope would not fit. It can be inferred from the report that working this issue resulted in a one year delay of the mission GAO 2014p, 7–8). Time and again, it appears that this inadequate definition of requirements led to either a schedule delay or a cost increase. In some cases requirements were simply not well defined, while in other cases, the

requirements themselves actually changed (GAO 1991a, 14, 20). Either way, it

presented a challenge to the design team to ensure that the actual product produced

met the overall objective of the project (GAO 1992b, 2, 2003, 9, 2002a, 2, 2014, 21–

22, 1998, 16).

In other cases, the problem was not so much with the requirements, but with

the design itself. Beyond the struggle of contending with undefined requirements and

designs, some projects had to work around requirements/designs that changed mid-

stream (GAO 1991a, 4). Changes were sometimes caused by a better understanding

of how the technology would realize the end goal of the projects, but other times the

requirements were changed by direction from a higher power (for example a review

board or even Congress) due to budgetary and schedule concerns (GAO 1993a, 11,

1991a, 4). NASA has stated that one of its accepted best practices is to ensure that at

least 90% of the engineering drawings for a system are mature enough at the CDR

that they could, in theory, be released to the production team with minimal changes

required (GAO 2014, 7). Several GAO reports mentioned challenges with NASA

project personnel failing to stabilize the design of the system, which led to challenges

with both cost and schedule. Most reports mentioned a generic difficulty in

stabilizing designs, but in one report, the GAO stated that NASA had failed to follow

this best practice and that many projects had reached CDR without first stabilizing the

design. Another GAO report (written nearly two decades later) stated that the

majority of the projects that had conducted a CDR during the year assessed failed to

stabilize the system design prior to that review (GAO 1991a, 4, 2010, 5, 2003, 12,

1993a, 17, 2009b, 13).

Part of the problem with achieving a stable design was the complexity of many of the systems (GAO 1991b, 2–3). The teams would begin development based on what they thought they understood about the requirements, but as the design progressed, it became apparent that actually meeting the requirements would be a much more complicated endeavor than originally anticipated (GAO 1991c, 6, 1993b, 4, 1989, 21). Given that NASA is often pushing the boundary of what is defined as scientifically possible, project managers have stated that they struggle to discover how to achieve technical success, let alone project management success (Martin 2012, 17). This can affect schedule in a variety of different ways including re-design of implementation plans, delays in receiving parts, and problems selecting the correct contractor to implement the design (GAO 1991b, 23). One GAO report stated that in a study of 29 programs, "technical complexities" was one of the six major categories of reasons for cost and schedule changes (GAO 1993a, 11). One report completed by the IG nicely summed up the relationship between technical complexity and schedule delays. It stated that, based on past evidence, the more technically complex a given project is, the more likely it is that schedule-busting problems will plague the program (GAO 2013, 18).

This complexity contributed to another major struggle encountered by many projects which involved battling issues that arose during the design and testing of the system. (GAO 2010, 5) These technical challenges seem to occur over and over again, which is not surprising given the nature of the work performed by NASA. In several GAO reports, technical challenges are listed as the cause of cost increases and schedule slippages (GAO 2004, 10, 1991b, 15, 2006f, 18, 1991c, 4, 1993a, 17).

Several GAO reports simply refer to "technical problems" in an overarching term, but some reports specify things such as: failures during testing or testing restrictions/limitations (GAO 2008, 13, 2006c, 9, 1991c, 5, 1977c, preface), reductions in available tests that might have detected possible issues earlier (GAO 1977d, 6), problems with the actual technology itself (GAO 2006f, 18), and integration challenges (GAO 2013, 23, 2009a, 17).

One recurring major recurring theme that was specifically called out throughout these reports was the failure of the planned technology to meet an appropriate level of maturity. In one report in from the early 1990s, four of thirteen projects were cited for a failure to adequately mature the required technology prior to fabrication, implying that time and money were being spent to build something that had never been proven to work as expected. Any problems encountered would require re-work and a probable increase in schedule (GAO 1991a, 4, 2009a, 16). In the Ares/Orion example cited earlier, a report from 2008 predicted problems for the project because a design review for the entire rocket was conducted prior to the first stage, "demonstrat[ing] maturity" (GAO 2008, 12). The James Webb Space Telescope was also listed as being in danger of a schedule slip, with one of the primary causes listed as a failure to adequately mature technologies (GAO 2006c, 9). An IG report which covered several of these problems issued a recommendation as to when a project should be allowed to proceed. In this recommendation it listed "mature technologies" as a resource which was critical to success (Martin 2012, 20).

Another recurring theme was the difficulty in managing contractors hired to complete much of the technical work for NASA. While not a technical challenge per

se, it appears that much of the difficulty in managing the contractors arose from a

failure by the contractor to fully appreciate the difficulty of the work involved in the

project (GAO 2006c, 9). In some cases, contractors brought in to complete the work

underestimated the difficulty involved or did not have the skills or expertise to deal

with the technical challenges that arose (GAO 2009b, preface, 14, 2009a, 19, 2012,

12). Further compounding the issue is that NASA has struggled in the past with

providing the proper management and oversight of the contractors completing the

work (GAO 1993a, 16, 1991b, 2–3). When the contractors run into technical

problems, the overall project can suffer with delays in schedule and cost as more time

is required to resolve these issues (GAO 2006f, 11, 1991b, 27). Sometimes lack of

knowledge on both the contractor and NASA sides can result in issues. If NASA

does not provide good direction and the contractors do not have the required

knowledge, the likelihood of technical challenges which will cause schedule

problems will increase (GAO 1991c, 8, 1993a, 16). Some oversight challenges are

the results of a lack of personnel resources (i.e. personnel were busy trying to

complete other commitments and could not dedicate the time required to provide

adequate oversight to the contractor (GAO 1989, 4, 1980b, 13). It should also be

noted that when bidding a job, a contractor is going to be "in it to win it". One report

even suggests that the bids are deliberately understated in an effort to win the overall

contract (Martin 2012, 20). This will involve seeking ways to offer the lowest

possible bid, which may ultimately result in problems once the contract is awarded

because of overconfidence in capability or the assumption that past success in a

different field will translate into present success in the space field. (GAO 1991b, 19, 23).

Some challenges were not due to new technology but were caused by trying to retrofit heritage technology to make it useful to current projects (GAO 2010, 5, 1991a, 4). The theory is that heritage technology is already developed and tested. It is a "known quantity" that can help reduce uncertainty about technical performance as well as cost and schedule. Unfortunately, though, heritage technology is just that: heritage. Like trying to install new software on an older computer, sometimes there are compatibility issues that must be overcome.. In this case teams must integrate new technology with the old technology, which is bound to present some challenges. NASA must weigh the challenges of developing completely new technology against the challenges of developing integration solutions for a new/heritage mix. Take, for example, the SLS program. From the outside, the vehicle is very reminiscent of the Saturn V rocket used to launch the Apollo astronauts toward the moon. Despite the similarities, nearly fifty years separate the current vehicle from the first Apollo launch and many things have changed since then, including design standards. The design team must figure out how to integrate what NASA has already accomplished with what it still wants to accomplish (GAO 2009b, 11, 2014, 16–17).

In one GAO report, it was stated that problems with heritage technology were encountered in over half of the projects under review. In this case, the team underestimated the difficulty of using this technology, even though it had flown on previous missions. The result of that underestimation was a schedule slip of nine months (GAO 2009b, 14; Martin 2012, 23). As stated in the previous section, one of

the resource challenges faced by NASA is the inability to obtain required parts,

especially in situations where the use of heritage technology is required. Companies

that develop parts for spaceflight do not have the advantage of mass production to

increase profitability. If a certain technology is no longer needed, it can be difficult

for these companies to maintain enough profit margin to stay in business. If NASA

then decides to go back and use an older technology, there is a chance that the

original source no longer exists and that the knowledge base that developed that

original source disappeared with the company (Martin 2012, 22).

In some cases, this trade-off did not work. For example, the Ares and Orion

projects mentioned earlier originally tried to use heritage technology. Ultimately,

however, changes to the designs resulted in the team distancing itself from heritage

technology because newer development was deemed to be more cost effective. In

another case it was discovered during testing that heritage material that was originally

deemed acceptable for use did not fit the bill, forcing the team to look for other

options. This ultimately resulted in a schedule delay of nine months (Martin 2012,

23). Another challenge to using heritage technology was that the project team was

having trouble re-creating it. As discussed earlier in the previous section, this may

have been due to the retirement of knowledgeable personnel or the lack of facilities

still capable of manufacturing the required parts (GAO 2008, 6). In theory, it makes

sense to try and leverage past knowledge and previous designs to meet current goals,

but in practice, it tends to be more of a challenge than anticipated (Martin 2012, 22).

In some cases, design stability is further threatened by changes mandated by levels

above the project (Martin 2012, 27). In the reports reviewed, the primary driver for

these changes seemed to derive from one of two sources: either the project was seriously over cost/schedule estimates and the project was directed to re-design the system to reign it back in (GAO 1991a, 22, 1994a, 1–2) or there was a directive to remain within a predetermined budget profile which dictated that the system had to be re-designed to fit within the profile (GAO 1991a, 4). In the first case, projects bring the re-design on themselves. Significant technical problems call into question the feasibility of the program, causing Congress to question whether or not NASA has bit off more than they can chew (Martin 2012, 27). Projects also fall victim to the failure to define requirements. The project goes to Congress too early and too optimistically and once the project figures out what is really required, the increase in cost and schedule is no longer palatable to those who control the purse strings (GAO 1993a, 11; Martin 2012, 12). In an IG report, some interviewees even hinted that NASA's estimates to Congress were low-balled just to get the project out of the gate, meaning it was not just the contractors who were guilty of underestimating. The theory, as discussed before, was that if the project could just get started, it could probably get funding to continue as needed. If the cost was too high, it would not have a chance to start in the first place (GAO 2004, 11, 17; Martin 2012, 13, 20, 32).

In the second case, it was often Congress or even the President directing NASA to make changes to the design. The project's design would be reported to Congress who would then determine whether or not the proposed cost fit within the pre-determined funding profile. If it did not, NASA was directed to re-design the project to meet the funding limits (GAO 1991a, 4). In other cases, the prices quoted to Congress amounted to what was effectively sticker shock and NASA was sent back

to the drawing board to try again (GAO 1991a, 17). Design changes of this nature,

while helping to ensure fiscal responsibility with the limited resources available do

have a tradeoff. Redesigns lead to schedule increases, so there must be a careful

balance struck between cost savings gained from a new design versus the cost

increases derived from an increase in the project schedule (GAO 1991a, 25).

2.1.4 Concluding remarks

As can be seen throughout this section, scheduling challenges are nothing new

for NASA and its partners. While there are multiple causes for these schedule delays,

there also seem to be common themes weaving throughout the past four decades. In

order to have the best chance for finishing a project on time, one must first understand

what it is one is trying to do and how it fits into the overall grand scheme.

Requirements must be understood and clearly documented and funding and resources

must be available at the appropriate time. Once the team understands what is

required, they can begin to design and build the system. Herein is the difficult part.

Even if all requirements are fully understood and all resources are firmly in place,

problems will still occur as the team works through the design and fabrication phase.

How then, should a project team schedule these activities to allow for these problems,

but still keep within a reasonable constraint of how long a project should take? The

next section will discuss current recommended practices for creating a project

schedule and some of the challenges with implementing these practices.

This section describes the recommended best practices for developing a

project schedule, as well as some of the challenges with the currently proposed

methods. It also describes some alternative methods to the best practices designed to

help alleviate some of the noted challenges.

2.2.1 Developing the Schedule

Once the project is approved and requirements are defined, one of the first

steps of building a schedule is to take each element of the lowest level of the WBS

and break it down into its component activities (Mantel Jr. et al. 2004, 73; PMI 2013,

para. 6.2). When developing this activities list, it is recommended that the subject

matter experts and team members get involved early in the process. Personnel who

are familiar with the deliverable described by the WBS package will most likely be

the most knowledgeable about what activities will be required to produce said

deliverable (Mantel Jr. et al. 2004, 75; PMI 2013, para. 6.2.2). Given that each level

of the WBS further specifies the previous level, and that the activity list is the lowest

required specificity, if the project team successfully identifies each required activity,

then completion of those activities will roll up into its WBS package which will in

turn roll up into the next WBS level, ultimately resulting in the successful delivery of

the projects ultimate deliverable (Mantel Jr. et al. 2004, 73; PMI 2013, para. 5.4,

5.4.2.2).

Once project activities have been successfully identified, they must be placed

in the proper order. PMBOK refers to this as "sequencing" the activities (PMI 2013,

para. 6.3). Activities are arranged in a logical order and are connected to one another

in such a way that the team can tell which activities have predecessors (activities

which must be completed before the current activity can take place) and successors

(activities that must follow the current activities). Not all activities will be tied to one

another, but every activity will have at least one predecessor and one successor (PMI

2013, para. 6.3). Sequencing activities naturally lends itself to producing some type

of chart which can easily demonstrate the predecessor/successor relationships of each

of the activities. The current method for sequencing activities is referred to as

Activity on Node (AON). AON networks depict activities as "nodes" and

dependencies as arrows connecting the nodes (Mantel Jr. et al. 2004, 136).

After sequencing the network, resources are assigned to each activity which

then allows a project manager to begin working with the team to estimate how long

each activity will take. According to both PMBOK and the original developers of the

PERT system, these duration estimates should come from the people most familiar

with the work to be completed (the experts) (Malcolm et al. 1959, 650; PMI 2013,

para. 6.5.2). These estimates are typically informed by recorded durations of a

particular activity or project ("analogous estimating"), or, when that data has not been

recorded, it can be based on the previous experience of the project team member

(PMI 2013, para. 6.5.2.1-6.5.2.3). Durations estimates can be either deterministic or

stochastic, depending on what input a project manager is able to glean from the

project team (Mantel Jr. et al. 2004, 147; PMI 2013, para. 6.5.2.4). The latter will be

discussed in greater detail in the next section.

Now that the schedule has been sequenced, resources have been assigned, and

a duration has been determined, the project manager can determine the duration of the

entire project. A popular procedure to achieve this is referred to as the Critical Path

Method (CPM) and involves following the "path" of the activities based on their

sequencing from the beginning of the project to the end (Mantel Jr. et al. 2004, 138–

41). The completion time of each activity is basically the completion time of the

predecessor activity plus the current activity's duration. If an activity has two or

more predecessors, the largest predecessor completion time is carried forward as the

start time of the current activity. This procedure, called the "forward pass" is

completed for all activities and across all possible paths of the network schedule. The

result provides the earliest possible point at which the project could finish and also

provides the Early Start Time (EST) and Early Finish Time (EFT) of each activity.

Once completed, the same procedure is applied, but in reverse. Starting at the end of

the project with the previously calculated project duration from the forward pass,

each possible path is followed back to the start of the project, where the start time of

each successor activity becomes the completion time of the current activity. For

activities with two or more successors, the successor with the smallest start time

becomes completion time of the current activity. This result provides the Late Start

Time (LST) and Late Finish Time (LFT) of each activity and allows for the

calculation of "total float" for each path through the network as well as the

calculation of "free float" which shows how long an individual activity can be

delayed before it affects the EST of its successor. This calculation of float allows a

project manager to determine the "critical path" of activities . This critical path is the

longest possible path (also the shortest possible completion time) through the network

and has the smallest amount of total float (typically no float or negative float). If any

activity on this path is delayed, it will delay the overall completion date of the project (Mantel Jr. et al. 2004, 134–43; Malcolm et al. 1959, 654–57; PMI 2013, para. 6.6.2.2).

The preceding paragraphs provided a basic description of simple schedule development. In practice, project schedules will incorporate things such as lead/lag time (e.g. time for ordering materials early or required delays between the completion of one activity and the start of the next) and can have a variety of different predecessor/successor relationships such as finish-to-start, start-to-start, start-to-finish, and finish-to-finish. The basic method of calculating a schedule remains the same, but these nuances can complicate the development. For larger, more complex schedules, software is available that will allow the user to enter activities, durations, predecessor/successor relationships, lead/lag times, etc. and will calculate the critical path and project duration, as well as display the schedule in a Gantt chart for quick assessments of project progress. While all of these tools are extremely helpful, ultimately the accuracy of the schedule is going to be dependent on the accuracy of the duration estimates received from the "experts" and in an uncertain world, deterministic estimates probably will not fit the bill (Mantel Jr. et al. 2004, 141,162, 167; Regnier 2005b, 8; PMI 2013, para. 6.5.2.4, 6.7.3.2).

2.2.2 Dealing with uncertainty: Stochastic estimates

In the previous section, CPM was discussed as a way to organize the schedule and determine the estimated completion time of the project. It showed how much contingency time was available on each network path and within each activity. The Program Evaluation and Review Technique (PERT) created in the early 1960s by the

Navy to assist with the development of the Polaris system used a similar method for organizing its project activities (Regnier 2005a, 1). The creators of PERT went beyond the organizational tactics of scheduling, however, and introduced a method to try and account for the uncertainty in those deterministic methods. Their method used three estimates for each activity: most likely, best case (if everything went right), and worst case (if everything went wrong) (Malcolm et al. 1959, 650–51). These values were then combined in a weighted average using Equation 2-1 which provided the expected value of the duration of the activity (Mantel Jr. et al. 2004, 144; Malcolm et al. 1959, 651; PMI 2013, para. 6.5.2.4). This expected value could then be used within the network schedule to follow the procedure described above for determining project durations and float time (Mantel Jr. et al. 2004, 146).

$$T_e = \frac{BC + 4ML + WC}{6}$$ Eqn 2-1

where T_e is the expected duration of time , BC is the optimistic ("best case") duration, ML is the "most-likely" duration, and WC is the pessimistic ("worst case") duration.

The PERT formula can also be used to determine the standard deviation of the estimate distribution by using Equation 2-2. The variance can be found by squaring the value of σ found using Equation 2-2 (Mantel Jr. et al. 2004, 145; Malcolm et al. 1959, 652).

$$\sigma = \frac{WC - BC}{6}$$ Eqn 2-2:

where σ is the standard deviation , BC is the optimistic duration, and WC is the pessimistic duration.

The "6" in Equation 2-2 indicates the belief that the values between the two outside estimates (optimistic and pessimistic) cover over approximately 99% of all possible durations of an activity and that the duration of the activity will be outside of this range less than 1% of the time. For those who are less confident in their estimates, the dividend of Equation 2-2 can be altered to represent the different confidence levels (with 95% and 90% being other popular choices). When converted back to a variance by squaring the result of Equation 2-2, these individual variances can be useful in determining the overall variance of either the critical path or other paths of interest in the overall project network, assuming each activity can be treated as statistically independent. The variance can also help the project manager determine the level of uncertainty that went into the original estimates based on the size of said variance (Mantel Jr. et al. 2004, 145–46, 151). To determine this range, the creators of PERT looked to the Normal distribution as a guide. A Normal distribution is defined from (-∞, ∞), but truncating the distribution at a standard deviation of $\pm$ 2.66 encompasses 99.2% of the probability density and also results in the standard deviation equaling 1/6th of the range. Given the assumption that there was negligible density below the BC estimate or above the WC estimate, the creators of PERT decided that a good approximation for the variance of their beta distribution was to borrow from the Normal distribution and assume that the relationship between the standard deviation and the range was also 1/6th. (Clark 1962, 406; Regnier 2005b, 6; NIST 2017a).

2.2.3 Problems with PERT

The developers of the PERT method of schedule duration estimation were themselves working under a very tight deadline. They were tasked to provide a process to analyze a complex schedule and they were only provided one month to accomplish this task (Malcolm et al. 1959, 647). Given the short timeline, the team developed a basic methodology, but as the system became more widely used, some of the finer points of the methodology came into question.

One of the first questions involved the beta distribution itself. The creators of PERT did not have a particular distribution in mind, but in developing their risk concept, they felt that a unimodal distribution with low probabilities at the tails would adequately model the behavior of an activity duration (Malcolm et al. 1959, 650–51; Clark 1962, 406). These assumptions easily lent themselves to settling on a beta distribution as the chosen model (Malcolm et al. 1959, 651–52). Since that time, the beta distribution has become the generally accepted model used to account for uncertainty in duration estimates (Keefer and Verdini 1993, 1087; Pickard 2004, 1570–71; Bennett, Lu, and AbouRizk 2001, 513; D. Johnson 2002a, 457–58; David Johnson 1997, 387). Having said that, the fact remains that the beta distribution is an assumption and the true distribution of the activity durations is not known (D. Johnson 1998, 254–55; Grubbs 1962, 914–15; Bennett, Lu, and AbouRizk 2001, 513; D. Johnson 2002a, 463–64, 1998, 253; Pickard 2004, 1567).

A second concern involved the estimates obtained from the experts and their correspondence to true statistical values. The creators of PERT asked personnel to provide their estimates of the best case, worst case and most likely estimates. From

50

there, a bounded distribution was created with the most likely value representing the peak of the curve, while the best case and worst case values represented the bounds of the curve. Given these three numbers, the developers then calculated the mean and variance of the estimates. From a practical standpoint, this gives an idea of how long the person performing the activity thinks it should take, as well as a proxy measure of their uncertainty in their estimate (Malcolm et al. 1959, 650–51). From a statistical perspective, however, this method is problematic. Typically a distribution curve is based on multiple observed data points and the mean and variance are derived from this data. The PERT process creates a distribution using just three estimated numbers provided by personnel who may or may not have a background in statistics (Grubbs 1962, 914; Golenko-Ginzburg 1988, 770; Keefer and Verdini 1993, 1087; Pickard 2004, 1567). Without knowing the true underlying distribution, these estimates may or may not encompass the full range of possible duration values for an activity (Grubbs 1962, 914–15; Regnier 2005b, 8; Pickard 2004, 1569). A similar problem occurs with the estimate of the most likely (mode) value which is simple enough conceptually, but not easily estimable in the true statistical sense (D. Johnson 2002b, 457).

Because the ultimate goal is to determine the completion date of an entire project, the creators of the PERT methodology required the calculation of the expected time and variance for each activity. Assuming independence of each activity, at its most basic level this allowed the calculation of the total project duration by summing all of the activities along a given network path (Clark 1962, 406; Malcolm et al. 1959, 651–52). This assumption of independence allows a

decision maker to apply the Central Limit Theorem when summing the means and

variances of each activity along a given path which ultimately provides the mean and

variance of the total project duration (Keefer and Verdini 1993, 1086; Steyn 2001,

365).   Pickard stated that this is a prime example of an inverse statistics problem

where the decision maker wishes to know a certain parameter of a statistical

distribution, but he must derive that information given different parameters of the

distribution (Pickard 2004, 1567–68). In the PERT case, estimation of the beta mean

and variance is not intuitive, making it easier to derive these parameters from the

three estimates that are more easily understood (D. Johnson 2002b, 457). This case is

further complicated by the fact that the true distribution is unknown (Pickard 2004,

1567).   In this case, the desired statistical parameters are the mean and variance and

the available estimated parameters are the mode and extremes of the distribution as

provided by technical personnel working the activity (Malcolm et al. 1959, 648, 659;

Clark 1962, 406; Pickard 2004, 1569). Because the underlying distribution is

unknown and given the challenges with estimating the mode (most likely) and

extremes (best case/worst case), it is entirely possible that the values used to derive

the mean and variance do not accurately describe the true distribution of the variable

(Pickard 2004, 1573; Steyn 2001, 368). This in turn can lead to inaccurate estimates

of the true mean and variance of each activity which ultimately results in an

inaccurate project duration. Pickard has suggested that some of these challenges may

be overcome by supplementing the standard three estimates with information about

the supplier's previous experience with similar projects (i.e. the number of times the

estimator had worked on a similar project). Converting this into a likelihood, Pickard

was able to develop a method, using several assumptions, to fully characterize the beta distribution in a more statistically sound manner (Pickard 2004).

Further compounding the issues just discussed is the concern regarding the accuracy of Equations 2-1 and 2-2. To calculate the true mean and variance of a beta distribution, one must know the defining parameters of the curve, α and β which describe the distribution. Because this curve is developed based on estimates and not on observation of actual events, the defining parameters of the curve are not known (D. Johnson 2002b, 457). The mean and variance must therefore be derived based on available information, namely some combination of the mode/median and extreme estimates of duration (Pickard 2004, 1568). The creators of PERT made some assumptions regarding the values of α and β and developed Equations 2-1 and 2-2 based on those assumptions (Malcolm et al. 1959, 651–52; Mantel Jr. et al. 2004, 144; Grubbs 1962, 914; Golenko-Ginzburg 1988, 768; D. Johnson 2002b, 457). These equations provided a good approximation for specific values of α and β, but when later compared to a wide range of beta distributions with varying set values for α and β, it was discovered that Equations 2-1 and 2-2 did not provide good estimates for the true means and variances of the distributions as calculated using Equation 2-3 and 2-4 (Keefer and Verdini 1993, 1089; Regnier 2005b, 7; Grubbs 1962, 914).

$$\mu = \frac{\alpha}{\alpha+\beta}$$ Eqn 2-3

$$\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$ Eqn 2-4:

Keefer and Verdini consolidated several recommended modifications to the original PERT formula and compared their various estimating capabilities by

comparing the results of the approximating equations for the mean and variance to the

true mean and variance as derived by Equations 2-3 and 2-4. These values were

calculated using the inverse cumulative distribution function (CDF) of a normalized

beta distribution which falls between the interval 0<x<1. It was shown that the most

accurate approximation for the mean was a method developed by Pearson-Tukey,

which resulted in a maximum error of 0.07% for the mean and -1.6% for the variance

(as opposed to the original PERT formula which resulted in a maximum error of

451% for the mean and 5506% for the variance). A close second was the Swanson-

Megill approximation which resulted in errors of 0.33% and 11.1% for the mean and

variance respectively. These two approximations are provided below in Equations 2-

5 and 2-6 (Keefer and Verdini 1993; Pearson and Tukey 1965; Megill 1971).

$$\mu = (0.63x(0.5) + 0.185[x(0.05) + x(0.95)]$$ Eqn 2-5:

$$\mu = (0.4x(0.5) + 0.300[x(0.10) + x(0.90)]$$ Eqn 2-6

where x(a) represents the value of the inverse CDF calculated at the "a"th fractile

(Keefer and Verdini 1993, 1087–88). It should be noted that while the original PERT

formula uses the mode of the distribution, Equations 2-5 and 2-6 rely instead on the

median. It should also be noted that while the PERT formula requires personnel to

provide estimates encompassing nearly all possible durations of a given activity,

Equations 2-5 and 2-6 reign in the estimate to a smaller range. In this case, estimators

are required to provide estimates wherein the duration would fall 95% of the time (for

Equation 2-5) and 90% of the time (for Equation 2-6) (Keefer and Bodily 1983, 1090;

Mantel Jr. et al. 2004, 144; Regnier 2005b, 8).

Given that the underlying distribution for an activity time is unknown, it is possible that the random variable representing duration is not, in fact, beta distributed (D. Johnson 2002b, 464). Even if the underlying distribution is actually beta, it is highly unlikely that accurate parameters for the curve could be directly assessed (D. Johnson 2002b, 459). Without these parameters, it is impossible to know the exact location of the mean within the distribution, although Equations 2-5 and 2-6 have been shown to be good approximations. To help alleviate this issue, Johnson proposed the use of a triangular distribution whose descriptive parameters match the three values gathered during a PERT survey (i.e. two extremes and the mode/median). Johnson was able to show that for a beta distribution where p, q ≥ 2, parameters for the triangular distribution could be derived such that the error between the values provided by the beta distribution and the triangular distribution were minimized. He went on to show that by sacrificing some accuracy in matching the shape of the beta distribution, the descriptive parameters for the triangular distribution could be modified to better determine the mean and variance of the triangular distribution. These estimates of the mean and variance closely matched the estimates developed by Pearson-Tukey when using the 5% fractile , further confirming that this equation is best estimation of the mean/variance for an unknown, yet positively skewed beta function (D. Johnson 1998, 260). Johnson went on to prove that his method of estimating the triangular mean and variance works well for a variety of distributions making knowledge of the underlying distribution less critical since these equations could provide a good estimate of the mean and variance in the absence of absolute knowledge of the true distribution (D. Johnson 2002a, 464). Ultimately,

however, all of these distributions and estimates are still dependent on the raw data provided by the expert.

The final two challenges relating to the PERT estimation problem deal with the actual estimates provided. The first challenge involves estimating the extremes. Research has shown that it is extremely difficult to estimate the extreme range of a truly "best case" and "worst case" durations as described by the originators of PERT . These values are meant encompass over 99% of all possible durations , but it can be difficult to provide estimates which truly encompass this wide range. As that range is narrowed, it begins to fall more readily into the realm of experience of those providing the estimates who are then able to provide a more accurate assessment of what will occur "90% of the time" or "95% of the time" (D. Johnson 1998, 253; Mantel Jr. et al. 2004, 145; Moder and Rodgers 1968, B-76; Selvidge 1980, 502; Davidson and Cooper 1980, 67; Murphy and Winkler 1977, 790, 792). The second challenge involves estimating the mode. The original creators of PERT asked for the mode, but they did not specify where the mode should fall in relation to the extremes (Malcolm et al. 1959, 651; Pickard 2004, 1568). Research has shown that, when based on estimates provided by project personnel, the mode is often placed closer to the best case value than the worst case value (Moder and Rodgers 1968, B-82; Golenko-Ginzburg 1988, 767). It was further proven that not only does the estimated mode typically reside closer to the best case estimate, but that the relation between the mode and the extremes was (using the notation of this research): $ML = (2BC+WC)/3$ (Golenko-Ginzburg 1988, 770). Further complicating matters is the fact that the mode, while easily understood in theory, does not translate over as an

easily assessable probability, especially by those with no probability background (Golenko-Ginzburg 1988, 770). Without knowing the underlying distribution and being able to physically observe the peak of the curve, there is no way to easily place the mode in the appropriate fractile (D. Johnson 1998, 255, 2002b, 457). The median, on the other hand, translates easily into the language of probability, falling at the 0.5 fractile. It also translates easily into a survey question wherein the participant is asked for an estimate such that 50% of the time the activity duration will be smaller than the estimate, and 50% of the time it will be greater than the estimate. A another benefit is that it allows for a readily available measure of the calibration of the individual's estimating capabilities (D. Johnson 1998, 255, 2002b, 457). This makes Equations 2-5 and 2-6 even more attractive because not only are they better estimators of the true mean of a beta distribution across a wide range of α and β, but they also make sure of estimates which are more likely to be accurate (Keefer and Verdini 1993, 1087–88; Grubbs 1962, 914–15).

In summary, although the PERT equation remains is recommended practice for accounting for uncertainty (PMI 2013, para. 6.5.2.4), there are several problems with the formula. First, it cannot be proven that the beta distribution is, in fact, the underlying distribution of the activity. Assuming it is the correct underlying distribution, its defining shape parameters are not known. Because its shape parameters are not known, equations have been developed to estimate the distribution's mean and variance. These equations, however, have been proven as poor estimators of the true mean and variance of a beta distribution except in a very narrow margin. Beyond this, the three estimates typically required in a PERT

57

equation are very difficult to estimate in a true statistical sense. Despite all these

challenges, PERT remains a popular method for compensating for the unknown when

estimating schedule durations.

2.2.4 Other Alternatives

Other solutions to manage schedule risk have also been developed, focusing

less on the PERT equation itself and more on the path through the chains of activities.

One of the major issues with the PERT methodology is its deterministic assumption

of a single critical path (Keefer and Verdini 1993, 1086; Goldratt 1997, 157; Mantel

Jr. et al. 2004, 155). Any path could potentially become critical given a long enough

delay in one or more of the activities (Goldratt 1997, 157–59; Gould 2005, 262;

Mantel Jr. et al. 2004, 155–56; PMI 2013, para. 6.6.2.2, 6.7.2.1). With the rise of

computers and scheduling software, computer simulations provided a new way to

account for path duration uncertainty. Specifically, the Monte Carlo simulations, as

applied in a scheduling context, use the basic concepts of PERT, but take into account

the fact that some project activities will match their optimistic estimate, some their

"most likely" and some will even match the pessimistic estimate. An appropriate

range and distribution are chosen for each activity, and each activity is organized

within a network path. The simulation is then run multiple times (user's discretion)

using myriad values for each activity (within its previously stated parameters) along

the assigned network path, finally arriving the distribution for the overall project

completion time based on the longest path through the network. Because this

simulation accounts for multiple different scenarios involving activity completion

times, it provides a better picture of what could happen versus simply choosing one

value for each activity and calculating the project duration from that single data point (Mantel Jr. et al. 2004, 156–60). Simulation, especially in complex projects, requires the use of software which project managers may or may not have access to and requires an understanding of the background statistics to effectively use the software (Mantel Jr. et al. 2004, 161; Shih 2005, 744). Simulations are also completely dependent on the input ranges for each of the activities (Mantel Jr. et al. 2004, 157). Different input ranges could result in drastically different schedules, making it crucial to gather estimates from knowledgeable personnel. This, however, begs the question: what defines a knowledgeable person.

Another risk mitigation technique, suggested by Goldratt, is referred to as the Critical Chain. This method again uses the basic PERT methodology of developing a network of activities and eliciting estimates from project personnel. Goldratt maintains that project personnel typically pad their estimates to mitigate the risk of finishing late and that their "most likely" estimates are much closer to their "true" 80-90% completion times. These estimates build extra time, referred to as buffers by Goldratt", into each activity. Goldratt maintains that the risk mitigation happens in the wrong place. According to the Critical Chain methodology, estimates should be much closer to the 50% mark, where there is only a 50% chance of completing the activity on time. Risk mitigation is handled by a project buffer which consolidates each of the activity buffers as a dummy activity at the end of the project (Goldratt 1997, 45, 65–67, 154–55; Steyn 2001, 365–66). Personnel are no longer responsible for the due date of each activity, only for ensuring that the entire project is completed by the due date (Steyn 2001, 365). The basic concept is that each activity is a link in

the project chain and the project buffer is used to strengthen each link as it becomes

weak (i.e. gets behind schedule) (Goldratt 1997, 89–95). The method also adopts the

concept of feeding buffers. These buffers protect the critical path from delays in

network paths that merge into the critical path. Another key aspect of the Critical

Chain is that it factors resource constraints into the schedule. This allows project

managers to not only monitor the time of each activity, but also potential bottlenecks

in resources which are required by more than one activity (Goldratt 1997, 89–95,

157–58, 215–21; Steyn 2001, 366).

2.3 Decision Analysis and Expert Opinion

The previous two sections discussed why projects struggle at NASA and the

current best practices for scheduling a project. The scheduling best practices

described several methods for organizing inputs and developing statistical methods to

predict the expected value of an activity duration. The GAO reports describing

NASA's struggles to complete projects on time indicate that the best practices have

flaws. Where, then is the problem? Is the issue the method or is it the inputs into the

method (Regnier 2005b, 8; Shanteau 1992, 12; Meehl 1954, 136–38). From

discussions in the previous section, probably a little of both. The previous sections

described obstacles to project success and challenges with the method of PERT

scheduling. This section will provide information on the challenges of obtaining

good inputs for the PERT method.

2.3.1 Recognized Biases and Their Effects

Human beings must process vast amounts of information on a daily basis. Kahneman has proposed that in order to cope with this constant barrage of information, the mind has developed a two-tiered processing scheme. System 1 is the gate-keeper. It receives various stimuli and provides a quick judgment on the appropriate response to said stimuli based on its previous experiences (Kahneman 2011, 11,24,56-57,71). System 2 is the decision maker who does not like to be disturbed. For most decisions, System 1 sends along an information package with a sticky-tab saying "sign here" and System 2 is happy to oblige. When System 1 has trouble processing a new situation, it sends System 2 a package with a note saying, "Further information required." System 2 must then research, consider, deliberate, and ultimately make the final decision in how to act (Kahneman 2011, 24). If System 2 relies too readily on System 1, it may use past memories to inaccurately assess current situations. These faulty associations result in biases which can severely affect judgment (Kahneman 2011, 3, 6, 127; Winkler 1981, 482; Hogarth 1975, 284).

Working in concert, these two systems prefer to make sense of the world by using past experiences to fit a new experience into a recognizable picture. (Kahneman 2011, 24,66; Zajonc 1968, 23). Problems occur when System 2 blithely accepts inputs from System 1 without critical review of the input data. System 1 has a problem with overconfidence in its assessment abilities. In its efforts to make sense of a stimulus, it will explain away inconsistencies in the narrative it is trying to construct. It invests itself in the narrative and passes the information along to System

2 as the truth (Kahneman 2011, 114; Tversky and Kahneman 1981, 457). If System 2 accepts the narrative without applying critical thinking, the narrative becomes truth, and is established in memory as a good point of reference from which to make future decisions (Kahneman 2011, 45, 114; Tversky and Kahneman 1974, 1124). Because our minds are programmed to create a narrative which describes the stimuli we encounter, and because System 1 gets a first crack at organizing all of the information, it will pull information from memory that confirms its initial assessment (Tversky and Kahneman 1981, 453). Unless System 2 actively examines all of the data, including the memories which conflict with the current narrative, the story created by System 1 will get stronger and stronger, resulting in a bias from the truth (Kahneman 2011, 134, 247). This type of bias is referred to as "confirmation bias" because the brain will seek out stories which fit the narrative and reject those that do not (Kahneman 2011, 45, 61 80-81; Mumpower 1996, 196). This is tied closely to what researchers have dubbed the "priming effect", where one stimuli triggers the ability to remember other events that are deemed to be tied to the original stimuli. These memories can also go one to trigger other memories in what is known as the ideomotor effect (Kahneman 2011, 52–53). The more often one sees something, the more readily it will be brought to mind the next time (Kahneman 2011, 60; Benson, Curley, and Smith 1995, 1650; Whittlesea 1990, 716). Because System 2 can redefine System 1's definition of surprise, once System 2 has decided what is "normal", System 1 will work to retrieve memories to confirm that new opinion and discount conflicting evidence (Kahneman 2011, 122–24; Hubbard 2010, 222–23).

In some cases, however, recollection of memory is less a bias and better

defined as "learning". System 1 has learned to recognize that certain stimuli lead to

certain results. Kahneman states that, "Intuition is nothing more and nothing less

than recognition (Kahneman 2011, 235–36). In some cases, the experience gained by

an expert allows him to recognize stimuli and pull relevant information to predict a

result. Kahneman points out that in situations where feedback is frequent and

immediate, an expert's predictions can be better trusted because System 1 is

recognizing the stimuli and responding accordingly. Because the feedback is

immediate, the cues are less likely to be confounded by spurious data and the effect is

more closely tied to the actual cause (Kahneman 2011, 240, 242). In fact, one

proposed definition of an "expert" is that the person is both consistent with his

prediction when provided similar stimuli and that he is also in agreement with other

experts who have been provided the same stimuli (Einhorn 1974, 564–65; Winkler

1968, B-70). Others have pointed out, however, that one must be cautious when

considering group agreement since a group can be in total agreement and still be

wrong (Rowe and Wright 2001, 343). Different authors agreed with the premise that

the expert should be consistent in his predictions, but added the criteria that he should

also be able to distinguish among environmental cues that appeared similar. They

referred to these traits as discrimination and consistency and showed through

experiments that those deemed to be experts were more capable of demonstrating

discrimination among cues than lay-people. They also pointed out, however, that the

scale was only comparative among the experts within the group (Weiss and Shanteau

2014, 107, 108, 112; Dawes 1979, 573; Shanteau 1992, 16). They based their

definition of expertise on the behavior of the person in question as opposed to whether or not the person's prediction was correct or if the title was conferred by others (Weiss and Shanteau 2014, 104; Rowe and Wright 2001, 341–42; Önkal et al. 2003, 182; Shanteau 1992, 17).   They also maintained that expertise is non-transferable stating that a person who is an expert in one field may not necessarily be an expert in another field, which would seem to contradict the practice of requesting "almanac" type questions to determine the calibration of an expert (Weiss and Shanteau 2014, 105–6; D. V. Lindley 1982, 122; Shanteau 1992, 13).   Another study pointed out that when asking experts for their assessments, the value to be assessed should be within the domain of the tasks which the expert would typically perform (Rowe and Wright 2001, 343).   Lindley has also pointed out that an estimator's knowledge of an unknown variable and his skills as a probability assessor are two different things. (D. V. Lindley 1982, 121)

Another bias also caused by the priming affect is referred to as "anchoring". In this bias, a number was provided to a person, who was then asked to estimate an unrelated quantity. Because that initial number had triggered the brain to reference similar numbers, it was shown that the estimates provided were frequently in range of the initial number that was provided. Ariely goes on to point out that this initial anchor can continue to affect future decisions (Kahneman 2011, 119–20; Tversky and Kahneman 1974, 1128–29; Ariely 2009b, 26, 45; Tversky 1974, 154).

Closely related to the confirmation biases and the priming affect is the availability heuristic. In this case, when trying to estimate how often an event happens, the brain recalls information about your personal experiences with that

event. If System 2 does not catch this error, then personal experiences becomes

"truth" as opposed to verifying your experiences against the actual frequency of

occurrence (Kahneman 2011, 129–30, 132, 135, 159; Tversky and Kahneman 1974,

1127–28; Tversky 1974, 152).   This availability heuristic becomes even more

pronounced when combined with the affect heuristic. The affect heuristic occurs

when emotions are brought to bear on the situation and beliefs are determined not by

data, but by feelings (Kahneman 2011, 102, 237, 301). Ariely further demonstrated

the affect heuristic by proving that decisions made while a particular emotion was

triggered created, in Kahneman's parlance, a reference point for System 1. When met

with a similar situation in the future, participants in Ariely's study followed the same

decision patterns demonstrated during the initial scenario even though the strong

emotions were no longer present (Ariely 2009a, Loc. 3572, 3590, 3625, 3643).

　　　Kahneman also pointed out (and Meehl also suggested) that the brain is

performing a type of substitution. System 1 does not know the answer to the question

it has been asked, so it replaces the original question with a different question that it

is capable of answering (Meehl 1954, 111; Kahneman 2011, 97). It then passes this

answer on to System 2 as the correct answer, which, if not checked, will be accepted

and convert into a belief (Kahneman 2011, 100, 129–30, 138–39). The primary issue

with substitution is that the new question System 1 substitutes may not be an accurate

representation of the original question (Kahneman 2011, 97, 138, 149). Additionally,

heuristics will rarely withstand scrutiny under statistical analysis (Kahneman 2011,

149, 151). Kahneman warns that relying wholly on intuition will always make the

predictor overconfident because System 1 is working to form a complete picture that

will ignore any evidence to the contrary (Kahneman 2011, 194, 200, 249, 261; Benson, Curley, and Smith 1995, 1648; Hubbard 2010, 221–22).

2.3.2 "Your Overconfidence Is Your Weakness" (Marquand 1983)

Despite their drawbacks, Kahneman does point out that heuristics can be useful in guiding initial responses (Kahneman 2011, 151). When these heuristics dominate the discussion, however, it can result in stereotyping which Kahneman describes as, "…statements about the group that are (at least tentatively) accepted as facts about every member." (Kahneman 2011, 167) Beyond the obvious social implications, that basic description implies that in everyday decision making, people are depending on their past experiences in one case to describe what will occur every time a similar situation occurs (Kahneman 2011, 173; Ariely 2009b, 211, 2009a, Loc 3590, 3607).

In System 1's efforts to create a plausible story to explain the world around it, it will find an explanation that is most readily available. When things turn out well, the most ready explanation it that the expert was wise beyond his years. When things turn out poorly, it is due to incompetence. Kahneman warns, however, that we as humans tend to give ourselves too much credit. We ignore the elements that are outside our control and attribute everything to the skill sets (or lack thereof) of those involved in solving the problem (Kahneman 2011, 13, 199–204, 207). Narratives such as these become engrained in the psyche, as does the belief that the situation was under better control than it actually was. (Kahneman 2011, 209–11) When performing a project post-mortem, or even when the next similar project comes

around, System 1 searches its database to create a story that will quickly and easily explain the outcome. The simplest story is that the outcome was driven mostly by the decisions of the decision makers or experts in the groups. As the old saying goes, "Hindsight is 20/20" and because System 1 is searching for the simplest explanation, it begins to assign causes to the outcomes based on the benefit of hindsight (Kahneman 2011, 217; Mumpower 1996, 197). As Silver puts it, "…we are seeing signals in the noise" (Silver 2012, 7).

In our efforts to explain the world around us, research suggests that we have placed too much faith in the abilities of the experts.   Multiple studies, several of which are discussed in Meehl (Meehl, 1954, pg 83-128), have shown expert predictions do not offer significant advantages over outcomes predicted by simple formulas. (Kahneman 2011, 218–19, 222–23; Dawes 1979, 573; Ruland 1978, 441; Silver 2012, 51–52; Roebber and Bosart 2014, 554; Dawes and Corrigan 1974, 96–97; Meehl 1954, 119). A formula can typically outperform a human because it is not subject to the biases that plague humans. A formula is programmed to respond a certain way at a certain threshold, so it will process information in the same way every time (Kahneman 2011, 240). Humans, however, have a tendency to over-think the problem and try to accommodate for the nuances of each different situation (Kahneman 2011, 223).

Despite the fact that the formulas typically outperform the experts, the formulas still require inputs from the experts. One good measure of an expert is whether or not more information produces better predictions (Silver 2012, 99–100; Heath and Gonzalez 1995, 308). Unfortunately, studies showed that experts did not

increase in accuracy with more information, but their confidence in their estimates increased (Heath and Gonzalez 1995, 310; Tsai, Klayman, and Hastie 2008, 98, 102; Hubbard 2010, 225–26). Because System 1 continually insists that our intuitions are correct, they become entrenched to the point that to question the belief is to question competency (Kahneman 2011, 45; Silver 2012, 216). When confronted with dissenting opinions, one study found that participants actually increased their confidence in their estimates as opposed to decreasing it in the face of opposing information. They proposed that the reason for this was because participants had to prepare mental counterarguments for each conflicting viewpoint which convinced them even more of the rightness of their beliefs. The study also showed that while these interactions increased the confidence of the participants, they did not increase the accuracy of their decisions. This effect has been termed the "inference certification hypothesis", where new information is treated as a confirmation of previously evaluated information as opposed to actual new information   (Heath and Gonzalez 1995, 305–6, 310–11, 317, 321; Tversky and Kahneman 1974, 1125–26; Ariely 2009b, 53–54).

Another study went on to show that the more information an expert had available, the higher the confidence rating, even if the accuracy of the assessment did not improve. They asserted that humans are simply not capable of accurately processing large amounts of information, although they believe they are. Their research confirmed the opinion that an expert should be defined by his ability to successfully interpret relevant cues while discarding irrelevant information. They went a step further to show that experts are capable of recognizing relevant cues, but

they struggle with assimilating those cues to create a more accurate representation of the situation, similar to a computer reaching its processing capacity and being unable to accept new input (Tsai, Klayman, and Hastie 2008, 98,100-102) .

Experts also had a tendency to try and explain away their mistakes with a series of excuses. It was never the expert's fault, just the circumstances surrounding the prediction. (Kahneman 2011, 218–19; Tetlock 2005, 132). Tetlock went on to describe two different types of experts, based on the work of Berlin: the hedgehogs and the foxes. Hedgehogs are exceedingly confident in themselves and their ability. They make bold predictions and cannot fathom the thought that they could be wrong. The foxes are more careful. They include a wider margin in their predictions and their predictions are not as drastic (Tetlock 2005, 72, 80; Berlin, Hardy, and Ignatieff 2013, 15, 50–53). A fox's star may not shine as brightly because his predictions will be more tempered, but he also does not have as far to fall when his predictions are wrong (Kahneman 2011, 191; Tetlock 2005, 84–85).

Humans tend to think optimistically about things they are personally involved in (Ariely 2009b, 180; Hubbard 2009, 37). Kahneman and Tversky applied this tendency to the scheduling context by identifying another bias for the optimists of the project world. In what they termed the "planning fallacy", this bias causes project planners to provide a "most likely" estimate that is more descriptive of the "best case" scenario than the "most likely" scenario (Kahneman 2011, 246, 249; Buehler, Griffin, and Ross 1994, 366). Buehler et al. noted that in practice, the planning fallacy was not as apparent when estimators were providing durations for other people's work. It was hypothesized that these outsiders were not personally invested in the outcome of

the estimate and therefore had less need to prove their capability to complete a task quickly. It was also pointed out that outside-estimators could blame the people performing the work for failing to try hard enough while the people doing the work will blame circumstances beyond their control (Buehler, Griffin, and Ross 1994, 368, 378). Estimators exhibit overconfidence by not accounting for potential setbacks. (Kahneman 2011, 86, 258–60; Brenner et al. 1996, 212; Önkal et al. 2003, 177). That baseline may be a very detailed list of things which must happen, but it will most likely end by just creating the list without creating a probability tree of potential failures. They short themselves on time estimates and when the issues and setbacks invariably occur, there is no contingency to mitigate those delays (Kahneman 2011, 251–52, 255). In some cases, Kahneman maintains that this fallacy is driven by the same desire seen in the GAO reports: project teams know that optimistic schedules are more likely to get approved, and once the project is approved, it is much harder to shut the project down despite any issues. (Kahneman 2011, 250). This behavior is reflective of another fallacy referred to as the "sunk cost" fallacy where project team members focus more on what has already been invested than the potential benefits of finishing the project (Kahneman 2011, 344–45). The defense against the planning fallacy is to pay attention to past projects of a similar nature. Actual durations provide a much better baseline than biased personal assessments. Actual durations will include all the challenges, setbacks, and excluded activities that have been encountered in the past. The fact that many, even experts, do not pay attention to these historical data points is known as base-rate neglect and will be discussed further in Section 2.4 (Kahneman 2011, 251–52; Tversky 1975, 164). It has been suggested

70

that one reason for the neglect of base rates is that estimators struggle to match the current situation with past experience. There is always just enough variance in the present case to make it "special" and therefore incapable of being compared to the past (Buehler, Griffin, and Ross 1994, 367).

There is also some research which shows that overconfidence is a problem no matter the level of expertise such that and that experts could be even more guilty of it than less experienced assessors (Mosleh, Bier, and Apostolakis 1988, 66; Baecher 1999, 5; Murphy and Winkler 1977, 42; Önkal et al. 2003, 176–77, 181). In a study closely applicable to the present research, Mosleh et al. showed that, when asked to estimate the median duration of a task, subjects provided estimates that smaller than experience would dictate by a factor of three. A study by the same group went on to show that when asked for duration estimates, subjects demonstrated a moderate degree of overconfidence in their estimations (Mosleh, Bier, and Apostolakis 1988, 70, 79). Another study showed that, once again, when asked how long an activity would take, those polled would fall victim to the planning fallacy and under-estimate the time required. This study had the added advantage of polling for continuous time estimations of familiar events as opposed to point-estimate almanac-type questions (Buehler, Griffin, and Ross 1994, 367). In another study conducted by Buehler et al. an optimistic bias in providing estimates was again discovered. Most respondents to this study failed to complete their tasks by their "most likely" estimates and fewer than half finished before their "worst case estimate". In variations of this study, Buehler et al. discovered that these optimistic biases did not seem to be solely driven by a desire to appear able to complete the task quickly. Even in cases where the

participants were unaware of the intent of the experiment, the bias still manifested

itself. One suggested reason for this bias is that people are typically in a success-

oriented mindset when planning. Obstacles to success are much more difficult to

imagine than the simple and clear path. They maintained that estimators must be

made to associate past experiences with present prediction problems, or they will not

make the connection (Buehler, Griffin, and Ross 1994, 369, 371–72, 376). Hubbard

went on to point out that experts can be guilty of the confirmation bias when it comes

to assessing their confidence levels. Without maintaining accurate records of the

results of their assessments, experts will rely on their memory to determine whether

or not their decisions have been good ones, and their memories confirm that their

decisions have usually been good (Hubbard 2010, 225). Hubbard's assertion falls in

line with other research on human behavior since, as Ariely points out, people tend to

look favorably on their own performance and since, as Kahneman points out, there is

a tendency to only remember things that confirm a person's already-held opinion.

    2.3.3 "Your Faith in Your Friends Is Yours" (Marquand 1983)

    Because System 1 wants all data to confirm its initial impression, it is no

surprise that people seek out others who will confirm their beliefs. When enough

people band together these beliefs get reinforced and an us-vs.-them mentality is

established (Kahneman 2011, 217; Ariely 2009b, 216; Silver 2012, 3; Heath and

Gonzalez 1995, 323). When one group has had a bad experience with another group,

the affect heuristic comes into play, and the divide becomes even stronger. As

members of the same group provide back each other up with more evidence of the

duplicity of the other group, the distrust basically enters a feedback loop which is very difficult to escape (Ariely 2009b, 258, 265, 268). In some of his experiments, Ariely showed that the distrust had become so engrained that participants in the study were incapable of recognizing true statements from a distrusted source (Ariely 2009b, 261).

When members of each group consider themselves experts, it stands to reason that members of the other group are considered "lay-people" who do not have the knowledge necessary to make an accurate decision. Unfortunately for the experts, however, research has shown that experts are not immune to the biases previously described (Kahneman 2011, 140). Beyond the biases already discussed is another bias unofficially termed the "Not Invented Here" bias (Ariely 2009a, Loc. 1443). In this case, ideas and concepts developed outside the group are considered less valid than the solutions proposed by the group (Ariely 2009a, Loc. 1467, 1571, 1608; Ruland 1978, 441).

If there are multiple groups who consider themselves experts, what then, are the possible sources of difference driving the experts? Hammond attributes these disagreements to one of three causes: incompetence, venality, and ideology. Incompetence can be described by a baseline lack of knowledge or a judgment based on misinformation. Venality describes decision based not on data, but on acting in one's best interest. Ideology could be described by the confirmation bias developed by Kahneman and Tversky and combines with base-rate neglect to cause experts to ignore their statistical accuracy rate and base their confidence on cases where they remember being correct (Mumpower 1996, 193–94; Buehler, Griffin, and Ross 1994,

368; Hammond 1996, 272, 290; Tetlock 2005, 40; Tversky and Kahneman 1974, 1124; Kahneman 2011, 80–81; Einhorn and Hogarth 1978, 399–401, 413–14). Mumpower, however, maintains, that in some cases, differences are due less to incompetence or malice and more because people simply perceive information differently from one another. Differing assessments could be the result of different ways of organizing the available information, priming, differences in assessments of the relevance of the cues, and differences in personal biases. The priming effect describes the case where information received first drives the perception of future information (Kahneman 2011, 52). Mumpower goes on to say that a further source of expert disagreement is because experts are not assessing the situation that "is", but are assessing what they believe the result of the situation ought to be. When these beliefs differ, the resulting assessments will differ (Mumpower 1996, 195–96, 201). Bram goes a step further and points out that in the case of management, personnel are making the best decisions they can based on the information provided, but in some cases, that information may be skewed to a positive light. No one wants to give the boss bad news, so it is feasible that reports provided to management may not accurately represent problems or challenges for fear of being perceived as ineffective (Bram 2011, 95, 112).

2.3.4 Options for Overcoming Bias

Given all of the biases, differing opinions, and general pitfalls of assessing the probability of an uncertain event, it seems unlikely that any one person will have all of the necessary tools to produce a valid judgment (D. V. Lindley, Tversky, and Brown 1979, 147). Differences in the order in which the data was received can affect

74

which memories are brought to mind first and cascade information can relegate important information to background clutter. Mumpower states that poor feedback or poor/missing information can affect an expert's ability to provide a valid assessment. Differing personalities, analytical styles, and social norms can also affect how difference experts process data (Mumpower 1996, 196, 203, 205–6; Surowiecki 2005, 37) . Each individual stakeholder, though, may possess new information or a different perspective about the problem. These perceptions can also be influenced by previously held beliefs. As Ariely points out, a previous belief can strongly influence the perception of a present stimulus (Ariely 2009b, 204, 206). The literature of expert aggregation maintains that a group assessment will provide a more accurate assessment than any one person's estimate under specific circumstances (Kane 1995, 63; Surowiecki 2005, 4–10; D. V. Lindley, Tversky, and Brown 1979, 149; Budescu and Rantilla 2000, 373–74; Sniezek and Henry 1990, 77–78). There is some research to suggest, however, that expert collaboration does have it limits and will eventually start decreasing the accuracy of the decisions being made (Hubbard 2010, 225). One suggested aggregation option is to aggregate the opinions of teams as opposed to the opinions of individuals. This allows different groups to interact with one another to incorporate different knowledge areas. Each team input is treated as if they were developed by a single entity which is then aggregated as it would be if the estimate had truly come from a single person. This method increases the chances of cancelling out biases while still allowing team members to feed off each other's knowledge base. They point out, however, that informal, non-structured elicitation sessions may not result in the mathematically rigorous results that are desired in such assessments,

typically because experts are relying on heuristics. (Mosleh, Bier, and Apostolakis 1988, 72, 78–79; Surowiecki 2005, 10) Obtaining the opinions of multiple stakeholders with different viewpoints can help ensure that all perspectives are considered without looking the group into one particular point of view. (Hubbard 2009, 47–48; Surowiecki 2005, 29; Winkler 1968, B-71)

If humans are incapable of fully processing all available data, then having several experts examine the problem would seem to alleviate that issue. Experts should be able to pick out relevant cues, but people pick out cues differently based on how they are hardwired and in what order they receive the information. Bringing experts together results in a better overall judgment. (Tsai, Klayman, and Hastie 2008, 100–103; Surowiecki 2005, 29; Sniezek and Henry 1990, 77–78).

Before aggregating expert judgment, one must first gather the judgments of several experts. There is a wide field of study on the elicitation of expert judgment with many different methods proposed. Several of these methods propose a regimented process where an analyst works with the subject matter expert to help him refine his assessments and point out inconsistencies. Benson et al. also point out that even when a decision analyst works with an expert, it can be very difficult to overcome the biases exhibited by the expert (Benson, Curley, and Smith 1995, 1640, 1642).   There are some elicitation methods, however, that provide better results than others, including asking the expert to actively assess cases that contradict his point of view or to break down the problem into more smaller, more easily assessable estimates (Mosleh, Bier, and Apostolakis 1988, 64, 67). As in the single expert case, there is evidence to support that aggregation formulas will outperform consensus

gathering methods where the possibility of one stakeholder dominating the assessment is much greater (Mosleh, Bier, and Apostolakis 1988, 67–68). Mosleh et al. maintain that proper elicitation methods in a group setting can protect the stakeholders agreeing with each other simply for the sake of agreeing (Mosleh, Bier, and Apostolakis 1988, 81).

2.3.5 Loss and Risk Aversion

While not a bias per se, humans do tend to exhibit aversions to certain things which could negatively impact their lives. Two aversions which can have an effect on scheduling estimations are loss aversion and risk aversion. Loss aversion refers to the desire to avoid the perception that one has taken a step backward from the status quo, no matter what that status quo may be (Ariely 2009b, 172, 177). Risk aversion is a tendency to invest extra resources to avoid falling victim to certain risks. In utility theory, this behavior is represented by a utility curve that is concave at all points (Kahneman and Tversky 1979, 264; Raiffa 1968, 68). Risk-seeking behavior, on the other hand is represented by a convex region of the utility curve (Raiffa 1968, 94–97). Tversky discovered that for many people, behavior near in the tails of the probability spectrum takes on what he termed a "certainty effect." This effect describes the human tendency to take disproportionately large measures to ensure that a small risk probability will drop to zero and, conversely, take these same large measures to make a highly probable opportunity a certainty (Tversky 1975, 167; Kahneman and Tversky 1979, 265).

Ariely points out that this concept of loss aversion applies not only to things, but also to beliefs and opinions (Ariely 2009b, 60–61, 177). Earlier in this

discussion, it was pointed out that the harder one works for an idea, the more

ownership one takes. The more one feels one owns an idea, the more personally one

will take any criticism of the idea due to the perception of losing face (Ariely 2009a,

Loc 1328, 1518). In a scheduling context, merging the concepts of loss aversion and

risk aversion could be thought of as actions taken which help the participant save

face. For example, if a manager is responsible for ensuring a project finishes on time

and on budget. In order to save face, the manager may take actions to ensure the

project is completed within the allotted time and budget. According to the certainty

effect, this indicates that actions taken will be for the purpose of reducing a risk with

negligible probability to zero probability or an opportunity with high probability to a

certainty. Kahneman and Tversky pointed out that the standard utility curve does not

account for this certainty effect and that the effect must be considered from a

reference point. The function below this reference point will be concave, indicating

risk aversion and the desire to minimize the risk of losing the status quo, while the

function above this point will be convex, indicating risk-seeking behavior that will

increase the probability of gaining an advantage above the status quo. They also

pointed out that this curve tends to be, "steeper for losses than for gains", indicating

once again that humans will work harder to prevent negative changes to the status quo

than they will to produce positive changes (Kahneman and Tversky 1979, 274, 277–

79; Tversky and Shafir 1992, 307; Ariely 2009b, 174, 2009a, Loc 3770; Kahneman

2011, 348). Ultimately, this results in an S-shaped utility curve where the status-quo

is the point of reference about which the curve revolves (Tversky and Wakker 1995,

1255). Once a decision maker's utility curve is established, it can be used by a

different person as a point of reference for how a decision maker would respond in different circumstances (Raiffa 1968, 69–70). Dawes and Corrigan maintain that linear models can provide a similar capability and provide a general guideline for how a manager can be expected to respond in certain situations. These rules help mitigate the effects of primacy and other mental cues, since an average behavior of the manager will be comprised of multiple situations that were managed under different sets of information (Dawes and Corrigan 1974, 100–102).

2.4 Experts as Data in a Bayesian Model

From the previous sections, while experts do provide a wealth of knowledge, they are subject to both personal and environmental biases. Past experiences, political beliefs, and even the order in which information was received can affect the estimating process. The project manager is then faced with two choices: pick the estimate of one stakeholder and base the schedule on that stakeholder's opinion or aggregate the inputs of all available stakeholders into one final number that can be used in a network schedule. If the project manager chooses the second option, an aggregation method must be selected.

From the literature, there appear to be two popular methods for expert aggregation. The first is a weighting scheme where opinions of different experts are weighted according to how much faith the decision maker places in the expert (West and Crosse 1992, 286; Winkler 1986, 300). Winkler provides several suggestions for weighting methodologies which can be applied to the distributions provided to the Decision Maker (DM) by the experts. These methods include both subjective

assessments regarding the abilities of the expert and assessments based on performance data (Winkler 1968, B63–64).

The second method involves the application of Bayes Theorem. In this model, a decision maker forms a prior probability based on the information she already possesses. According to Bayes' Theorem, a DM makes an initial probability assessment about an unknown variable based on the information she has available to her. This assessment is referred to as the "prior" and essentially describes the point at which a person is willing to take action based on their assessment (Silver 2012, 255; Chaloner and Duncan 1983, 174; Simon French 1980, 45; Mumpower 1996, 198–99; Baecher 1999, 3). Experts are then treated as if they were data points by which a DM (DM) can update her beliefs (Morris 1977, 680; Clemen 1987, 373–74; S. French 1986, 315; Morris 1974, 1235; Winkler 1986, 299; Morris 1983, 24; D. V. Lindley, Tversky, and Brown 1979, 146; Savage 1971, 797; Simon French 1980, 44; D. V. Lindley 1982, 117–18; Roberts 1965, 52–58; Simon French 1985, 188–91). The concept is that there is a "truth" to each unknown variable and that a DM can approach the truth by gathering more and more data and updating her beliefs based on that data (Silver 2012, 232, 241; Benson, Curley, and Smith 1995, 1644–45, 1647). Another advantage to the Bayesian method is that it allows for the messiness of the real world and the potential for human biases through the use of a prior distribution (Silver 2012, 252–55). It also accounts for how difficult it can be to gather real-world data by allowing for a constant updating process when new information is obtained (Silver 2012, 409).

As new data is received, the DM updates her beliefs resulting in a new probability assessment referred to as the "posterior." Formally, this can be described mathematically using Equation 2-7 ("Bayes' Theorem" 2017; Gelman et al. 2013, 7; Dennis V. Lindley 1983, 1; West and Crosse 1992, 285–86; Winkler 1981, 479; S. French 1986, 315).

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{p(y)}$$
<div align="right">Eqn 2-7</div>

In this method, each expert provides an assessment on an unknown variable, θ, based on his previously held knowledge, preferably before talking to any of the other experts (Silver 2012, 245). According to Morris and Gelman, for a continuous variable, the variance of the expert's assessment can be used as a gauge for the certainty the expert has in his estimate where a smaller variance suggests higher certainty (Gelman et al. 2013, 32; Morris 1977, 688).

Earlier in this chapter, it was mentioned that humans are prone to fall victim to a fallacy described as base-rate neglect (Kahneman 2011, 88; Brenner et al. 1996, 217) The base-rate is a measure of the number of entities in a particular class in relation to the entire population. In a Bayesian context, this is the p(y) term from Equation 2-7 (Kahneman 2011, 146; "Bayes' Theorem" 2017). According to Kahneman, there are two types of base rates: statistical (which describes overall populations) and causal (which describe information about a specific case). He maintains that people will typically ignore the statistical base rate when given information that is specific to the situation (Kahneman 2011, 166–67). He also maintains that in the midst of uncertainty and the absence of any other data, the base

rate should rule the day (Kahneman 2011, 152). In a scheduling context, a study by Buehler et al. showed that observers (those outside the project) were much more likely to use the statistical base rate of the actor's past performance than the actors (those involved directly in the project). The actors in this study seemed to ignore their past performance in favor of anticipated future performance. As part of this study, Buehler et al created a condition where they attempted to get their subjects to associate past performance with future results. Although participants still resisted the association, bringing up past struggles did seem to cause the participants to at least consider some of the challenges which they could face (Buehler, Griffin, and Ross 1994, 367, 376, 379)

Given the human tendency of ignoring base rates and basing probability assessments on memory, it becomes necessary to calibrate the human assessment machine to account for biases. Just as with any measuring instrument, when compared to a "truth" source, if the measurement is off, the results must be calibrated (e.g. a thermometer that always reads ten degrees higher than the actual temperature). Calibration refers to the ability of a forecaster to make accurate forecasts (Weiss and Shanteau 2014, 105, 109; Morris 1983, 24). In a forecasting context, when a forecast predicts that there is an "X" probability of an event happening, over an extended period of time given the same circumstances, that even should happen X% of the time (DeGroot and Fienberg 1983, 13; D. V. Lindley, Tversky, and Brown 1979, 147; Morris 1983, 24). For a continuous variable, a similar concept applies, except that instead of a point estimate, the estimate provided should correspond with the appropriate point in the cumulative distribution function (CDF) (Morris 1983, 28–29).

Based on this assessment method of calibration, it is possible for forecasters to "game the system" by providing estimates which, over time, allow them to appear well calibrated. (DeGroot and Fienberg 1983, 14) Weiss and Shanteau maintain that in order to be properly calibrated, a probability assessor must demonstrate both discrimination (recognizing changes to the situation) and consistency (providing the same assessment in the same situation) (Weiss and Shanteau 2014, 109). Onkäl et al. agree with the assessment of discrimination and add the requirement to apply the cues provided in the situation. (Önkal et al. 2003, 179).

From Section 2.3.2 and a study conducted by Alpert and Raiffa (Alpert and Raiffa 1982, 294–305), research has shown that humans tend to be overconfident in their probability assessments. It has been suggested that this trait can be used as a marker for calibration of experts when requesting probability assessments. (Mosleh, Bier, and Apostolakis 1988, 71) It has also been shown that training can help mitigate some of this overconfidence and that training can help people become overall better assessors of probability (Mosleh, Bier, and Apostolakis 1988, 80; Dennis V. Lindley 1983, 9; D. V. Lindley 1982, 125; Selvidge 1980, 502; Lichtenstein, Fischhoff, and Phillips 1977, 294,316-317).

Knowing that humans are subject to some form of bias, according to Morris' method, the DM will first establish her prior, based on her past experiences and the data available to her at the time (D. V. Lindley, Tversky, and Brown 1979, 146). Each expert will also individually establish his prior based on this same information and provide that information to the DM (Morris 1977, 680). If empirical calibration data is available, the DM can use that information to modify the expert's prior to

more accurately reflect his ability to perform a probability assessment. If empirical data is not available subjective calibration is an alternative option (S. French 1986, 316; Morris 1983, 24, 1986, 327, 1977, 688–89). Using subjective calibration, the DM encodes her beliefs about the expert's ability as a probability assessor by determining the likelihood that the actual value will fall within the tails of the expert's prior distribution (Morris 1974, 1235, 1977, 691). For independent experts, the posterior distribution is then described by the prior of the DM multiplied by the calibrated priors of the experts. This calibrated prior is effectively a likelihood assessment describing the likelihood that the expert will provide that particular probability assessment, given the revealed value of the unknown variable. If the DM believes that the expert overstates his level of knowledge, she can modify the prior such that the variance is wider (Winkler 1981, 482). If she believes he understates his knowledge, she can modify the prior such that the variance is smaller. The degree of modification reflects her belief in the expert's ability as a probability assessor. The final result is a single posterior assessment that provides a decision maker with a recommended course of action based on Bayesian principles of probability (Morris 1977, 679; Silver 2012, 327).

One of the major concerns with any expert aggregation problem is the dependence of the experts. Statistical independence is difficult to achieve in the expert aggregation case due to a variety of factors including similar expert backgrounds and review of similar data (Clemen 1987, 373; Winkler 1968, B-65; D. V. Lindley 1982, 120; Winkler 1981, 480; Clemen 1986, 313; Winkler 1986, 302; Morris 1974, 1236, 1983, 24; Clemen 1987, 374-375; 378-379; Harrison 1977, 320).

The concern derives from the fact that the posterior distribution will change based on the level of dependence of the experts (Clemen 1987, 374; Winkler 1981, 487). Morris also points out that the simple multiplicative rule described in his 1977 paper (Morris, 1977) is only applicable under certain conditions, one of which is independence. In cases of dependence, a more complex method is required (Morris 1986, 322). He goes on to say, however, that in a case where the expert is treated as "…a measurement instrument measuring data from a physical experiment", the multiplicative rule is appropriate (although still requiring modification when modeling dependence) (Morris 1986, 325; D. V. Lindley, Tversky, and Brown 1979, 149).

One specific concern with respect to dependence is that subjective calibration of the experts can never be truly independent of the decision maker performing the calibration (S. French 1986, 320; Clemen 1986, 313–14; Winkler 1986, 299, 302; West and Crosse 1992, 287, 291; Genest and Schervish 1985, 1198; Simon French 1980, 43–46; Schervish 1984). Morris points out, however, that the expert-decision maker dependence problem may not be a major issue. He maintains that in most cases, the expert will have significantly more information than the decision maker or that the decision maker will refrain from expressing an opinion to prevent swaying the judgment. In both cases, Morris maintains that the degree of dependence then becomes negligible (Morris 1986, 326).

A second concern is dependence not on the decision maker, but on the shared backgrounds of the experts. It is highly unlikely that personnel from the same work area involved in the same project will not be privy to the same information, training,

or experiences. From earlier in this chapter, it was seen that perceptions can differ

based on various factors (e.g. time of receipt of data, order of receipt of data, mood

when data was received, etc.). The actual data, however, will probably be the same,

even if it is processed differently by each expert. These shared backgrounds and data

sets complicate the Bayesian updating process due to, as Clemen and Morris phrased

it, "double counting" the data (Clemen 1986, 314; Morris 1986, 324; Schervish 1986,

309; Simon French 1980, 46). The DM cannot update her prior with the expert's

information if they both are using the same information. This does not provide the

DM with any new data, only a rehash of the data she already has (T. R. Johnson,

Budescu, and Wallsten 2001, 137–39; Budescu and Rantilla 2000, 374). While this is

certainly an issue for further exploration, for the purposes of this research, it is

assumed that the experts are all statistically independent of one another.

Chapter 3: Methods and Materials, Data Collection

Determining differences in the priorities and estimating practices of project

stakeholders required a diverse set of both projects and stakeholders. As an

operational facility, NASA's WFF provided both of these conditions. Once the data

were collected, it was organized and masked to the best extent possible. Masking was

required to protect the anonymity of the research subjects to the best extent possible.

The chapter begins by describing the surveys used to collect data from the subjects

who consented to participate in the research, followed by a description of how these

inputs were processed and consolidated. Following this is a description of the

analysis performed on the processed data, as well as a high-level description of the

Design of Experiments (DOE) factorial analysis method used to analyze the data.

The chapter concludes with a description of how the DOE process can be used to

predict the response of an estimator within a particular demographic and a description

of a proposed estimate aggregation method.

3.1 Data Collection

Data were collected from on-going projects planned, managed, and executed

by personnel working at NASA's WFF. Projects were selected based on whether or

not they were active during the data collection period and whether or not they had

timelines which seemed to be a source of conflict among the stakeholders. For

example, given the nature of the work at WFF, certain projects, although still projects

by definition, have become routine in their execution and the durations for each activity seem to be known and accepted. In these cases, the schedule is mostly driven by the project documentation process, so project team members are given the due date and are left to take the actions required to meet that due date. Other projects, however, involve more planning and are subject to more scrutiny from project stakeholders. These projects often leave stakeholders asking, "why does it take this long" or "why is the schedule so compressed", depending on the job and perspective of the stakeholder. The data collection period for this research project was approximately 26 March 2014 – 20 March 2015 and the projects selected typically fell into one of three categories: operations, maintenance, and engineering. Operations is defined as a project which is directly tied to launch campaign support, including set up and testing of all equipment. Engineering is defined as an effort to implement new or upgrade existing technical systems. Maintenance is defined as actions taken to ensure the operational systems remain in good working order. While maintenance is technically not a project since it has no defined beginning or ending, certain aspects of the maintenance performed at WFF lend themselves to the project definition (PMI 2013, para. 1.2).

Initial project selection was made by determining which projects were either currently in work or were just entering (or just about to enter) the planning phase. Personnel assigned to the project were contacted individually using "Recruitment E-mail" (Appendix A.1). If the person responded that he/she was willing to participate, a meeting was set up to complete the consent form required by the Institutional Review Board (IRB) and to provide instruction on filling out the various surveys.

Four surveys were created:

1) Traits/Opinions Survey,

2) Scheduling Survey,

3) Follow-on Survey,

4) Course of Action Survey.

3.1.1 Traits/Opinions Survey

The first survey was called the "Traits/Opinions" survey (Appendix A.2) and was used to gather basic demographic information about the participant as well as get his/ her opinion on managing project constraints and attitudes toward risk. This survey was created using GoogleDrive™ and could be accessed using a website provided to the participant. Once the participant completed the survey, the results were consolidated into a separate web page. During the consent process, each participant chose a random three-digit number that would be used for identification when filling out this survey. This number was stored on an iPad ™ using a program called SafeNote™. When filling out the Traits/Opinion survey, each participant would begin by filling in this random 3-digit number. All responses of the participant would then be associated with that number during data compilation. This number was then transferred to a spreadsheet which associated the name of the participant with the number selected. Given that the scheduling surveys were submitted through e-mail, this "linking" spreadsheet was critical to associating the results supplied by e-mail (identified by the sender) to the results supplied in the GoogleDrive™ survey (identified by the random 3-digit number).

3.1.2 Scheduling and Follow-on Surveys

The Scheduling Survey and Follow-on Survey (Appendix A.3 and A.4, respectively) were administered over e-mail. These two surveys were project specific and allowed subjects to: state how much time they thought should be allotted for each activity of the project, if they believed adequate resources were assigned, if they believed the activity list accurately described the scope of the project, and provide an estimate of the overall project duration. The activity list was developed with the assistance of a knowledgeable technician or manager directly involved with the project. Once the activity list was populated, the survey was e-mailed out to appropriate members of that project team who had consented to participate in this study. For the technicians, activity lists were sometimes broken down by different task sections so that subjects only had to provide estimates on applicable project tasks (as opposed to all tasks in the project). Managers were asked to complete estimates for a full project task list, since typically a manager is expected to oversee all activities within a project (PMI 2013, para. 1.7). All subjects were asked to complete and return the surveys prior to starting any of the activities on the list so that the values provided would be estimates and not a recording of what had already happened. A "Follow-on" survey was also e-mailed along with the "Scheduling Survey" to each subjects directly involved with project execution. This Follow-on Survey allowed subjects to record the actual durations of each of the listed activities on the Scheduling Survey, as well as describe major changes in team size, work hours, or availability that occurred while executing the project. This survey also allowed the participant to comment on challenges encountered during the project.

Section 1.5 mentioned that data from three types of projects were collected for this study (operations, engineering, and maintenance). Due to the close-knit nature of the community at Wallops Flight Facility, the projects discussed in these results are not divided by type to help maintain the anonymity of the subjects who provided data for this research.

When collecting estimates from various subjects, organizational strategies and project risk tolerances were not explicitly accounted for in their affects on the estimation process. Subjects were asked to provide scheduling estimates based on their understanding of project requirements, while also factoring in their understanding of possible constraints both within the project and other organizational commitments.

3.1.3 "Course of Action" (COA) Survey

The COA survey (Appendix A.5) delved into the question of different perceptions of "risk-mitigation" versus "gold plating", where "gold plating" is defined as going over and above a requirement to provide unnecessary capability or performance. This survey was developed and uploaded to GoogleDrive ™, and the address of the survey was e-mailed to each of the subjects. This survey was intended to be anonymous, so no survey identifiers were collected. After initially clicking on the link, subjects were asked to choose "management" or "technician" based on the choices they picked in the Traits/Opinions survey. This selection directed the participant to a new page where the primary question was phrased in such a way to make it applicable to the participant (whether management or technician). The primary question involved a generalized scenario stating that a piece of equipment

91

was within specifications but just barely. It went on to say that fixing the equipment would cause a schedule delay. Two questions were then posed to each participant: should they fix the equipment or leave it alone and: would they consider taking the extra time to fix the equipment risk mitigation or gold plating (see Appendix A.5 for the exact verbiage). A final question on this survey asked subjects to describe why they believed projects in general fell behind schedule.

3.2 Data Processing

This section describes the methods used to organize and aggregate the raw data collected from the surveys. It also describes the concepts behind the survey questions described in Section 3.1 as well as some challenges encountered during the data collection process.

3.2.1 Categorizing the Subjects

Results from the "Demographics" portion of the Traits/Opinions survey were categorized by three demographics which were then broken down into different levels. The demographics chosen were: position (technical or management), years of experience (0-7, 8-15, 16-23, 24+), and completed level of formal education (high school, associate's/technical, bachelor's, master's).

The Position demographic was labeled as an "M" or a "T". Subjects were categorized as "management" if they were primarily responsible for managing personnel or project constraints. Subjects were categorized as "technical" if they were primarily responsible for completing the technical work required to achieve the technical project objective.

The Years of Experience (YoE) demographic was labeled as a number between one and four (with "1" being 0-7 years of experience and "4" being 24+ years of experience). The Level of Formal Education (LoE) as either an "H", "T", "B", or "M" corresponding to the levels of education described above. For example, a manager with 0-7 years of experience and a master's degree would appear as "M1M". Given that some subjects shared the same demographic traits (i.e. there was more than one manager who had 0-7 years of experience and a master's degree), it was necessary to assign another designator to distinguish all of the subjects. Table 3-1 summarizes the demographics and their representations.

Demographic	Levels
Position	M: Management
	T: Technical
Years of Experience	1: 0-7 years
	2: 8-14 years
	3: 15-23 years
	4: 24+ years
Level of Formal Education	M: Master's
	B: Bachelor's
	T: Associates/Technical
	H: High School

Table 3-1: Demographic Identifiers

As results from the Traits/Opinions surveys came in, each respondent's selected 3-digit number was converted to a second, randomly selected number (to help protect anonymity). This number was then linked to the appropriate demographics designator which was linked to the responses from the other surveys in this study. This process ensured each participant was assigned a demographic category while still ensuring that results of the other surveys were associated with the correct participant.

When collecting data for this research, training in the field of project management was not directly taken into consideration. Subjects were asked to provide their level of completed formal education, but were not asked specifics about the field of study. Participants were also asked to provide the number of years of experience in their field, but were not asked if any of this experience was related to technical training, management in general, or specifically project management.

3.2.2 Risk Tolerance

As a measure of risk behavior, participant responses to various gambles were used to create a utility curve (Dennis V. Lindley 1983, 2). Based on a method developed by Raiffa, in a situation with two outcomes, one considered a "win" and one considered a "loss", a decision maker would be handed a "basic reference lottery ticket" (brlt). This ticket contained a value between zero and one, and represented the probability of winning. This probability was represented by the variable "π". The probability of losing was "1- π". This "π" designator is the reason the general name of the "ticket" was referred to as a "π –brlt" (Raiffa 1968, 57–60). Although there are many uses for the π-brlt, especially in decision trees involving non-monetary values, this research focused primarily on their use in developing utility curves for those who did not behave according to the principles of Expected Monetary Value (EMV) (Raiffa 1968, 61–65). EMV is calculated by multiplying a monetary value (whether positive or negative) by the probability of "winning" said monetary value. This will result in a straight line passing through the points [x_{min},0] and [x_{max}, 1] (Raiffa 1968, 8-9,66-67). When a decision maker feels that the EMV line is either too aggressive or

not aggressive enough, a similar procedure can be used to determine a new curve that more accurately represents the decision maker's risk thresholds (Raiffa 1968, 51–53). This curve is referred to as the π-indifference curve and is created by plotting the π-brlt value on the y-axis and the monetary value at which a decision maker will trade the chance of winning for actual money on the x-axis. Starting with $[x_{min}, 0]$ and finishing with $[x_{max}, 1]$, points in between these two values describe the shape of the curve (Raiffa 1968, 66–67).

The general case of the π-indifference curve is referred to as the utility curve. Since these values are used for decision makers who do not follow an EMV line, the π-brlt value provides a conversion point which can be used in a decision tree to determine the best course of action. In this context, the π-brlt value can be referred to as the utility of the monetary value in question (i.e. the point at which a decision maker is indifferent between on-the-table money and the probability of winning the full prize). The curve is described by $\pi_i = u(x_i)$, where "u" is the shape of the utility curve, and π_i is the value of the curve evaluated at a given monetary value, x_i (Raiffa 1968, 86–89). The shape of this curve can provide an indication of the risk tolerance of a participant. A concave curve indicates risk aversion while a convex curve indicates risk-prone behavior. The more pronounced the curve, the more risk-averse/risk-prone the participant (Raiffa 1968, 68–70, 94). Using the example survey provided to subjects in this survey, an example of the three basic risk attitudes (EMV, risk averse, and risk prone) can be seen in Figure 3-1. Figure 3-2 shows two subjects who demonstrate different levels of extremity for both risk averse and risk prone behavior.

95
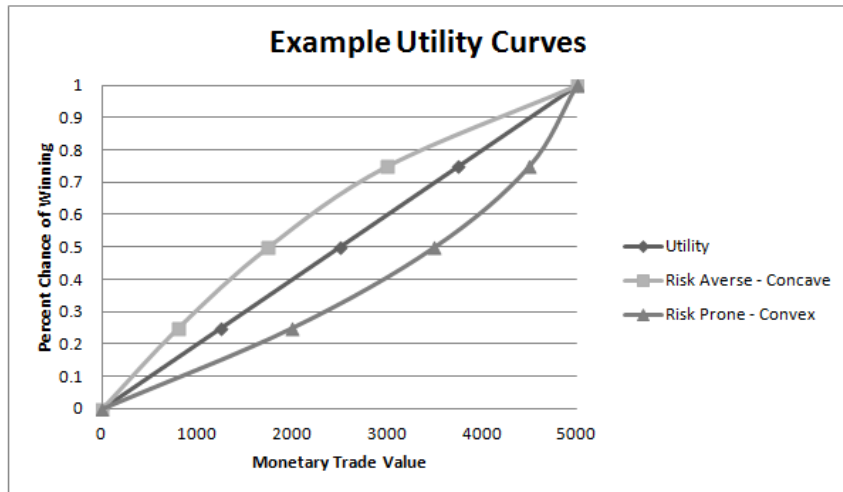
Figure 3-1: Example Basic Utility Curves
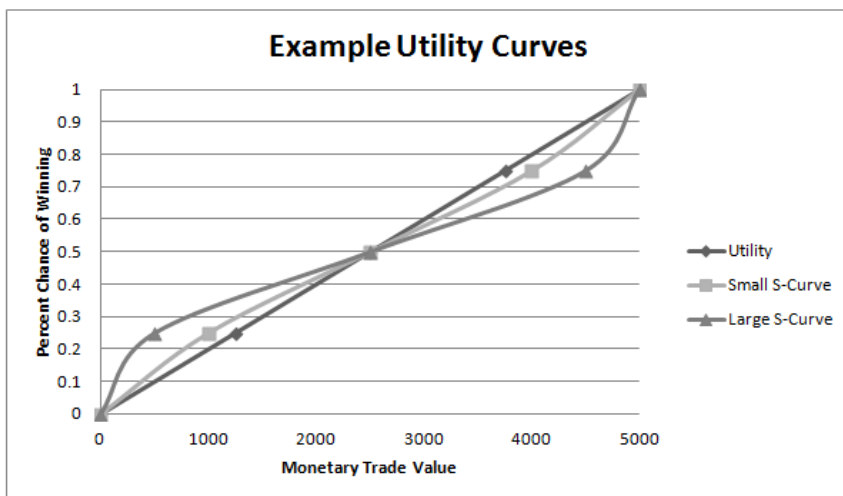
Figure 3-2: Example Risk Averse and Risk Prone Behavior

In order to model the risk tolerance of the subjects, a scenario was presented wherein the participant was handed an hypothetical piece of paper that gave him/her a certain probability of winning $5000. The participant was then asked how much "cash on the table" would be required to trade the chance at winning $5000 for the immediate cash-out (see "Risk Tolerance" in Appendix A.2). The percentages provided were: 10%, 35%, 50%, 68%, and 87%. These percentages were chosen such that they would cover a wide spectrum of probabilities of winning, but not at

such easily calculated intervals that they would lend themselves to choosing utility. Given the extreme curvature of some of the responses (discussed further in Chapter 5), only the responses at the 50% probability were analyzed. Risk aversion was measured by distance from the EMV value of $2500 (i.e. $5000*(0.5)). To the left of this point, risk aversion increased as the trade-in value decreased. To the right of the EMV point, risk aversion decreased as the trade-in value increased (Raiffa 1968, 68–69).

Once the subjects submitted their answers, an Excel ™ table was developed where the first column listed out the probabilities of winning and the second column listed the EMV monetary value (Raiffa 1968, 8–9). The following columns listed responses from each participant where each response was listed in the row corresponding to the appropriate probability of winning (see Appendix A.7). Once this table was established, the Excel ™ graph function was used to determine the curve of each of the participant's response in relation to an EMV utility curve (a straight line passing through [0,0; 0.5, 2500; 1, 5000]) (see Figure 5-1 through Figure 5-5). The monetary values represented by the participant's responses served as the x-axis, while the probabilities of winning served as the y-axis. The format of the lines themselves (dots, dashes, etc.) were based on the 3-digit number assigned to the participant and were used to help distinguish one line from another. These lines were then plotted using the "connect the dots" feature in Excel ™.

3.2.3 Constraint Preference

One of the suppositions in this research is that stakeholders, based on their demographics, have different priorities when determining which project constraints

97

(PMI 2013, para. 1.3) are more important than the others. The hypothesis was that

certain demographics will tend to sacrifice performance on one particular constraint

for better performance in another preferred constraint.

To test this theory, a set of questions was developed which set four of the

major project constraints against one another (Appendix A.2, "Preference Analysis),

- Cost

- Schedule

- Quality

- Risk

Based on concepts derived from the Analytic Hierarchy Process (AHP), the

survey pitted two constraints directly against one another and then asked the

participant to choose which one should be sacrificed for the sake of the other (e.g.

subjects were asked to choose between having an increased cost or an increased

schedule on any given project). (Winston 2003, 785–91; Mantel Jr. et al. 2004, 5–6;

PMI 2013, para. 1.3)

Responses to each question in this survey resulted in each of the four chosen

constraints being compared to one another. In their responses, subjects were asked to

write down which option they chose, as well as a preference between one and nine,

indicating how strongly they preferred their choice (Winston 2003, 787). In this

scale, a "1" response meant that the participant was indifferent between the two

options and a "9" meant the participant strongly preferred one option over the other.

The other numbers represented various levels of preference between those two

extremes (Zio 1996, 129). A scale was provided on the survey itself that explained

the meaning behind each number on the scale.

Once the subjects' responses were received, a grid was developed in Excel ™

to compile the results. For each participant, a 4x4 matrix was created which listed the

four project constraints in the same order down the rows and across the columns (see

Table 3-2 for a partially populated example matrix). The cells along the diagonal

were populated with a "1" since a constraint cannot be compared to itself. The

participant's preference ranking was then placed in the appropriate cell where the two

constraints under consideration intersected. The reciprocal of this value would be

placed in the reciprocal row and column. For example, the first question in the

survey asked if the participant preferred an increased cost or an increased schedule.

If the participant responded, "Increased Cost, 7" indicating he had a moderate

preference for increased cost over increased schedule, then a "7" would be placed at

the intersection of the "Cost" row and the "Schedule" column. Conversely, a "1/7"

would be placed in the "Schedule" row and the "Cost" column (see Table 3-2). This

process was repeated for each constraint comparison and ultimately resulted in each

constraint being compared to all the others.

XXX	Cost	Schedule	Quality	Risk
Cost	1	7		
Schedule	1/7	1		
Quality			1	
Risk				1

Table 3-2: Example Preference Matrix

Once the entire grid was filled out, a weight for each constraint was calculated

using Equation 3-1 (Winston 2003, 788). This weight represented a subject's

willingness to sacrifice that constraint for the sake of achieving success in meeting

the other constraints. For example, based on the responses of the subject, if the

schedule constraint had a calculated weight of 0.42 and the quality constraint had a

calculated weight of 0.12, that indicates the subject is more willing to sacrifice the

schedule for improved quality than he is to sacrifice quality for a faster completion

time.

$$W_i = \frac{\sum_{j=1}^{n} \frac{aij}{\sum_{i=1}^{n}(aij)}}{n}$$ Eqn 3-1

where "i" is the row number, "j" is the column number, and "n" is the total number of

constraints under consideration. See Table 3-3 for a fully populated example matrix

and its resulting calculated weights. The generalized form of this table can be seen in

Table 3-4.

	Cost	Schedule	Quality	Risk	W_i
Cost	1	4	4	6	0.55
Schedule	1/4	1	4	4	0.27
Quality	1/4	1/4	1	2	0.11
Risk	1/6	1/4	1/2	1	0.07

Table 3-3: Example Matrix

	C constraint 1	Constraint 2	...	Constraint n	W_i
Cost	a_{11}	a_{12}	...	a_{1n}	W_1
Schedule	a_{21}	a_{22}	...	a_{2n}	W_2
Quality	...	...	...	...	...
Risk	a_{n1}	a_{n2}	...	a_{nn}	W_n

Table 3-4: Generalized AHP Matrix

Equation 3-2 calculates the consistency of the participant preferences. The

result of this equation is referred to as the Consistency Index (CI).

$$CI = \frac{\left(\frac{\sum_{i=1}^{n}\frac{\sum_{j=1}^{n}a_{ij}*W_{i=j}}{W_i}}{n}\right) - n}{n-1}$$
Eqn 3-2

Based on a table provided in Winston, because there were four preferences

under consideration, a Random Index (RI) value of 0.90 was selected (Winston 2003,

788–89). The CI value was then divided by the RI value to determine the ratio of

CI/RI.   According to Winston, if this ratio is less than 0.10, then the participant can

be considered "consistent" and their preferences considered valid (Winston 2003,

789). For this research, the consistency for each participant was only calculated as a

point of interest to determine whether or not the subjects were behaving in a

consistent manner. Consistency is a measure of comparison among each of the

constraints. If a participant weights Constraint B twice as much as Constraint A, and

Constraint C four times as much as Constraint A, then, by definition Constraint C

should be weighted twice as much as Constraint B. If not, then the participant is not

consistent in his preferences. The ratio of CI/RI provides a measurement of the

consistency of the participant. A perfectly consistent matrix will result in a CI of zero

(Winston 2003, 786). Appendix A.8 provides the compiled list of project weights and

consistency ratings for each participant.

3.2.4 Schedule Survey Data

Responses to the scheduling surveys were compiled using Excel™

spreadsheets where participant numerical/demographic designators were listed in the

same row as their responses (Appendix A.9). Projects that had multiple independent parts were split up and treated as separate projects, except in one case where the subjects for the two separate parts were all the same; those responses were merged and treated as one "project". Data provided by each participant consisted of a "most likely" (ML) estimate describing how long the participant believed the activity should normally take, a "best case" (BC) estimate describing how long it should take if everything went well, and a "worst case" (WC) estimate describing how long it should take if everything went poorly. Subjects were also asked to provide an estimate of how confident they were in their ML estimate (further discussion on this to be provided in Chapter 5).

Question 5 on the Scheduling Survey (Appendix A.3) asked subjects to provide either a completion date or a recommended start date for the project, depending on whether or not the completion date for a given project had already been set. To further clarify, at WFF, some projects are given a completion date and the team must work to be ready by said completion date, so the schedule is effectively built by first conducting a "backward pass". Other projects are more traditional in that they would be built using the "forward pass" first in order to determine the completion date. This question was an effort to compensate for the fact that an actual network schedule with activity relationships was not developed. It was hoped that a final completion date would give an idea of the total expected duration of the project regardless of how the activities were related to one another within the project itself. Unfortunately, for various reasons, surveys for each project were not necessarily sent or completed at the same time (e.g. time of participant consent, operational

102

commitments, etc.) and the format of responses varied among subjects, requiring several assumptions regarding the intent of the response. For these reasons, these results were not included in data analysis or Appendix A.9.

3.2.5 Follow-on Survey

The Follow-on Survey was intended to compare the estimates provided by each participant to the actual duration of each activity. Unfortunately, only a few follow-on surveys were returned. Of the responses received, it would have been challenging to confirm the reported values. In some cases, the reported values differ among project subjects, indicating that the perception of what constituted task completion may not have been standard across all project team members. Because of these challenges, results from the Follow-On surveys were not included in data analysis or in Appendix A.9.

3.3 Data Analysis – Characterization

The following section describes the methods used to characterize the subjects in this study. It describes the method used to compare project constraint preferences and differences in total project duration estimates, without regard to demographics. It goes on to describe several questions that were developed to compare differences in the responses of stakeholders of varying demographics. It then describes the method used to analyze the results of the project constraint questions, as well as the method used to analyze the confidence and risk aversion levels of the subjects with respect to their stated demographics. It provides a high-level description of the Design of Experiments (DoE) process used to analyze the data. Table A-1 through Table A-5

are provided to allow for the recreation of the experiments conducted here using DesignExpert™ and the data found in Appendix A.7 through Appendix A.9. Those tables show the inputs that were used in that program to set up the experiment prior to analysis. Following this section is a description of the method used to determine the correlation between personality traits and project estimating practices.

3.3.1 Constraints Analysis – by Constraint

After developing the normalized matrix for the project constraints as described in Section 3.2.3, results for each participant were organized by constraint such that all "cost" weights from all subjects were grouped together, all schedule weights, etc. The average weight for each constraint was then calculated. To determine whether or not the differences seen in the average weights among each constraint were statistically significant, a simple t-test was performed using the Data Analysis package on Excel™. The "t-statistic with unequal variances" was selected with H_0: $\mu_1 = \mu_2$, and alpha = 0.05. If the one-tailed p-value was below 0.05, the difference in the averages was considered statistically significant.

3.3.2 Network Path Standard Deviation

Within each project, the total duration estimates for each participant were compared to determine whether or not there were differences in the way different stakeholders estimated duration times. It was hypothesized that if, given the same information, all stakeholders in a project were in agreement about how long the project should take, the standard deviation among the project duration estimates would be zero. Because network schedules were not built for these projects, each

104

project is assumed to have a single path where each activity begins upon the completion of the previous activity. Given that assumption, the total duration of the project is the sum of each activity's PERT average and is represented by the variable T_e. Once T_e was calculated for each participant on each project, the standard deviation of T_e among each of the subjects per project was calculated. Demographics of the subjects were not considered in this part of the analysis. A histogram plot was created using Excel™ to display the results.

3.3.3 Comparison Questions

The three baseline estimates provided by each participant were further analyzed to determine whether or not certain demographics exhibited similar estimating trends. For each of the three demographics considered, eleven questions were developed. Questions 1-7 and 10 were derived from the Scheduling survey. Questions 8, 9, and 11 were derived from the Traits/Opinions survey.

Table 3-5 lists out the questions for the Position demographic. For the YOE demographic, the questions remain the same, but the word "management" is replaced with "fewer years of experience" and the word "technician" with "more years of experience". The same is true for the LOE demographic, where "management" is replaced with "more formal education" and technician with "less formal education". The questions were answered by comparing the results of each of the surveys. The YOE and LOE demographics each had four levels of comparison as opposed to the Position demographic which only had two. If a certain project had members of only one level of the demographic in question, that survey was excluded from that group (e.g. if a project had two respondents, both of which fell under the "management"

105

category, that project was excluded from the Position group because there would be

no point of comparison on that demographic).

Question 1: Is management's total project duration estimate T_e (based on PERT) lower than the technician's?
Question 2: For all demographics, is the separation between the ML and BC estimates smaller than the separation between the ML and WC estimates?
Question 3: Does management have a smaller separation between the ML and BC estimates than technicians?
Question 4: Does management have a smaller separation between the ML and WC estimates than technicians?
Question 5: Is management's ML estimate higher than the technicians'?
Question 6: Is management's BC estimate higher than the technicians'?
Question7: Is management's WC estimate higher than the technicians'?
Question 8: Are management personnel less risk averse than technicians'?
Question 9: Is management's variance smaller than technicians'?
Question 10: Is management's confidence greater than technicians'?
Question 11: Is management's willingness to sacrifice schedule less than technicians'?

Table 3-5: Comparison Questions

For each of the questions in Table 3-5, the Excel ™ "IF" command was used

in one of the following ways, depending on how the question was asked:

=IF(X<Y, "Yes", "No")   or   =IF(X>Y, "Yes", "No").

In the command above, "X" and "Y" are represented by the responses

provided by each participant in each project. For projects with only two subjects, a

simple comparison was performed for each project activity using the formula above.

When a project had more than two subjects each participant was compared to the

others in its group, as long as they did not share the same demographic. For example,

for a project where the respondents consisted of two managers and two technicians,

responses from Manager 1 would be compared to Technician 1 and then the

procedure would be repeated with Technician 2. Manager 2 would then be compared

to Technician 1 and then to Technician 2, with each comparison providing a different data point. This procedure ultimately resulted in comparing each participant with every other participant in that project, as long as they did not share the same demographic.

Each "Yes" answer was tallied along with the total number of responses to determine the total percentage of times the data supported the premise of the questions seen in Table 3-5. To determine the statistical significance of the results, RStudio™ was used to conduct a binomial test by using one of two commands listed below (R Core Team 2014):

binom.test(#of successes, # of tests, 0.5, alternative = "greater")
binom.test(#of successes, # of tests, 0.5, alternative= "less")

The null hypothesis was that if all stakeholders behaved the same, each group should agree with the premise of the questions in Table 3-5 about 50% of the time. For example, all things being equal, 50% of the time management should have a higher total duration estimate and 50% of the time technicians should have a higher total duration estimate (i.e. 0.5 in the RStudio™ command).   For those cases where the number of "Yes" answers divided by the total number of answers was greater than 0.50, the binomial test determined whether or not the true probability of success (a "Yes" answer) was greater than 50%.. For cases where it was less than 0.50, the test determined whether or not the true probability of success was less than 50% ("Binomial Distribution" 2016; Gelman et al. 2013, 29–32). Statistically significance was defined as p<0.5 at a confidence level of 95%.

3.3.4 Design of Experiments

In several of the sections below, results are analyzed using Design of

Experiments (DoE) using the DesignExpert™ software. DOE allows an experimenter

to monitor how certain factors affect the response of a system and which of those

factors has a significant part in driving the response (Montgomery 2008, 162–64).

For example, if a machine had three levers, each with a high and low setting, an

experimenter could run eight experiments, testing the response of the system at each

combination of each lever. These results could then be analyzed to determine which

of the three levers, if any, had the most significant effect on the performance of the

machine. The experimenter could then run each of these experiments again to

confirm the results. These runs would be referred to as replicates and they help

ensure that the results are not due to random error or factors outside of the study.

Repeat measurements are measurements that are taken within the same experimental

run (Montgomery 2008, 12–13). The advantage of the DOE process is that it

minimizes the number of runs necessary to determine which, if any, of the factors are

significant. In the above example, this would mean that the experimenter would not

necessarily need to complete all eight experiments to determine the significant factor.

Each participant was treated as an experimental run and his or her

demographics represented the different settings of the human machine. Responses to

several surveys were then gathered and analyzed to determine how the different

demographic settings changed the response of the human machine. Different subjects

within the same demographic category were treated as replicate measurements.

When a participant provided responses for multiple surveys of the same type, the

responses were treated as repeat measurements and were averaged to create one overall response for that participant (Montgomery 2008, 607).

Although humanity is a machine with myriad factors, this research concerned itself with only three: position (management or technical), years of experience (YoE), and level of formal education (LoE). The position factor had two levels, while YoE and LoE had four settings each. The breakdown of these factors is described in Section 3.2.1. The goal of this part of the research was to poll several subjects, each representing different levels of the three factors and determine which, if any of those factors was driving the differences in response. This process also predicted the expected response of someone outside the study who shared the same demographic factor.

Analysis on the data for this research was conducted using a mixed factorial design, specifically a 2^k and 4^k factorial design. A 2^k factorial design studies the effects and interactions of "k" number of factors, each with two settings: high and low. A 4^k factorial design is a modification of the 2^k design (Montgomery 2008, 162, 207, 382–85). According to the process, the experimenter would determine the factors he wished to study and develop a run-sheet where he systematically altered each factor between its high and low settings. Once completed, the run sheet would represent all possible combinations of all possible settings for all factors. The sheet would then be randomized to reduce the chance of a hidden variable affecting the results (Montgomery 2008, 12). The experiments would then be completed with the run-sheet telling the experimenter how to configure the system for each experimental

run. Results from each run would then be recorded in the results column of the

appropriate row and the entire sheet would be analyzed when completed.

For this research, the collected results were effectively randomized by the fact

that subjects were independent of one another and were polled after signing the

consent form. Constructing experiments using this method allows researchers to

determine not only the effects of each chosen factor upon the results using the

minimum number of experiments, but also if interactions among the factors have an

effect (Montgomery 2008, 4, 162–64).

Two of the factors under study had four levels instead of two levels (YoE and

LoE), requiring a slightly modified version of the 2^k design referred to as a mixed

level design. To accomplish this, a method called replacement was used where each

level of the four-level factor was described using two replacement variables. This

effectively converts the 4-level factor into two 2-level factors. Each level of the 4-

level factor is represented by two replacement variables. The replacement variables

are set at their low value for the first level of the 4-level factor, at their high level for

the fourth level of the 4-level factor, and at alternating high/low settings for the two

levels in between. This effectively turns the mixed design into a higher-order 2^k

design. For example, in this case, if each of the three factors had been at two levels,

this research would have used a 2^3 design which would require eight runs for a full

factorial analysis. Because two of the factors were at four levels for this research, the

total number of runs increased. Both of the 4-level factors (YoE and LoE) were

treated as if it were two 2-level factors, plus the actual two level factor (the position

demographic), effectively resulting in five 2-level factors. To test each combination

of each factor, a 2^5 design would be needed which would required 32 experimental runs (Montgomery 2008, 382–85).

For a full factorial design, upon completion, the run-sheet would show the response of the system at every possible factor combination. In order to reduce the number of runs required, the experimenter could also use a fractional factorial design. In this case, the number of required runs can be reduced based on the principle of sparsity of effects. This principle states that responses are most likely to be driven by the main effects and/or the lower-order interactions. When reducing the design, terms become aliased with on another where the result of that factor level combination is described by the main effect plus its aliased factors (interaction factors which share the same sign as the main effect). In these cases, the sparsity of effects principle states that the result is most likely being driven by the main effect and the higher-order interactions can be ignored (Montgomery 2008, 290–93). There are several methods available to reduce the number of runs required to successfully analyze the data (Montgomery 2008, 380, 450–53).

Because this research used a mixed design of both 2- and 4- level factors and did not have enough data points for a full factorial design, the DesignExpert™ software was used to create a D-optimal design using a coordinate exchange (Montgomery 2008, 380). A D-optimal design creates a run-sheet where the selected experiments to be completed will minimize the variance of the regression coefficients of the model (Montgomery 2008, 254, 336, 452). This is accomplished by performing matrix multiplication on a matrix created from the factor level settings of each selected combination on the run-sheet by its transpose and then maximizing the

determinant of that matrix (Montgomery 2008, 253–54). The coordinate exchange is an algorithm that searches the design space based on the initial parameters selected by the experimenter (i.e. the factors, their levels, and the estimated standard deviation of the response) for the best combination of experimental runs that meet the D-optimality criteria (Montgomery 2008, 453, 456–57).

The Design Expert™ software sets the minimum number of runs which must be completed in order to successfully analyze the data, but options for extra experimental runs are left to the experimenter's discretion based on the availability of data. Additional runs beyond those required by for basic analysis help improve the accuracy of the predicted results and increase the power of the analysis (Montgomery 2008, 12; *DesignExpert* (version 9.0.6.2) 2015, n. Help File). Design Expert™ then creates a run-sheet with the desired number of experimental runs based on the input from the experimenter and adhering to the requirement to minimize the variance of the regression coefficients. If the design is orthogonal, the coefficients are independent of one another and all have equal variances. If the design is not orthogonal, the coefficients could change based on the selection of the model terms (*DesignExpert* (version 9.0.6.2) 2015, n. Help File). The run-sheets created by DesignExpert™ using this method created combination of factors that were not available in the data set (e.g. a manager whose highest level of formal education was high school). Because of this, modifications to the run-sheet were required which meant that the design no longer met the optimality criteria, but could still be used for analysis.

112

Once the experiment was designed and the results from the subjects were matched to the appropriate lines in the run-sheet, a predicted response could be developed using the regression model in Equation 3-3 (Montgomery 2008, 164).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \ldots + \beta_i x_i + \beta_{12} x_1 x_2 + \cdots \beta_{1*\ldots*i} x_1 * \ldots * x_i + \varepsilon \qquad \text{Eqn 3-3}$$

where y is the expected response, β_0 is average of all responses collected (also known as the intercept), β_i is the coefficient of the factor, and x_i is a coded variable representing one of the factors in the experiment. The first part of the equation represents the main effects of the actual factors, the second part of the equation represent interactions between the factors (e.g. $\beta_{12} x_1 x_2$ is a term describing the results of an interaction between Factor 1 and Factor 2), and ε is a term to cover the random error. The "x" terms in Equation 3-3 are derived from the desired level of each factor. For this research, each factor was qualitative, so the coded variables are set to either -1 or 1 to represent the "high" and "low" levels of each factor. DesignExpert™ represented the 4-level factors by using replacement variables whose different settings represented the required four levels.

In order to calculate the β coefficients for the main effects, the sum of the results when the main effect is at the "low" level is subtracted from the sum of the results when the main effect is at the "high" level and then divided by 2 (Montgomery 2008, 163, 208–9). The interaction coefficient is calculated by summing the results obtained for one factor when another factor is at its low level and subtracting that value from the summation of the first factor when the second factor is at its high level, and then taking the average. The magnitude and direction of the resulting coefficients will provide a relatively good indication of whether or not the

effect or interaction is driving the result. If the term does not have a significant effect on the final result, it can be removed from the equation, leaving only those terms that do significantly affect result (Montgomery 2008, 164, 210; *DesignExpert* (version 9.0.6.2) 2015, n. Help File).

The error term from Equation 3-3 can be used to determine whether or not the model is a good fit and conforms to the basic assumption that the errors in the model are, "…normally and independently distributed with mean zero and constant but unknown variance". These errors are analyzed by determining the residuals of the model (Montgomery 2008, 75). The residuals are calculated by solving for Equation 3-3 at each factor setting from the run-sheet and subtracting that value from the actual response recorded for that setting. For experiments with replicates, this process would be repeated with each replicate (Montgomery 2008, 213–14). These residuals are then plotted versus several comparison criteria to check the assumption of normalcy and to ensure that there are no significant factors hidden within the error term. A correct model with no hidden factors in the error term will show a "fat-pencil" straight line in the normalcy test and a random scatter on the remaining tests. In the event of non-constant variance, several transformations are available to mitigate this effect. The DesignExpert™ program automatically provides a recommended transform if it detects an issue (Montgomery 2008, 75–83).

In order to determine whether or not a factor is statistically significant, analysis of variance (ANOVA) using the sum of squares and mean squares is used. The sum of squares for each factor is calculated by squaring the contrast (i.e. the sum of the results of all observations at the high setting of that factor subtracted from the

sum of all the results of all observations at the low setting of that factor) divided by total number of observations multiplied by the square of the contrast coefficients (i.e. the value of the x_i variable for that particular factor setting, either -1 or 1). For orthogonal designs, the sum of squares of the contrast coefficients will be equal across all main effects and interactions. This will not be the case for non-orthogonality.

The total sum of squares is calculated by first summing together the results of each factor combination treatment, squaring that sum, and dividing it by the square of the contrast coefficients multiplied by the total number of experiments. This value is then subtracted from the result obtained by squaring each result in the experiment summing them together. The sum of squares of the error is calculated by subtracting the sum of squares of each factor and interaction from the sum of squares of the total. The remainder is the sum of squares of the error (Montgomery 2008, 70, 87–90, 208–12). The mean square error is calculated by dividing the sum of squares for the model/factor/interaction in question by the total number of degrees of freedom used by that model/factor/interaction (Montgomery 2008, 67, 223). The F-statistic is then calculated by dividing the mean square of the model/factor/interaction in question by the mean square of the error of the model. The p-value can then be calculated from this F-statistic to determine the statistical significance of the model/factor/interaction in question (Montgomery 2008, 69).

The DOE method in general, and the factorial design in particular allows an experimenter to not only determine whether an effect is statistically significant, but also allows him to make statements about the response of a population based on

115

sample data collected (Montgomery 2008, 213). By determining the response of a system based on the levels of certain factors, an experimenter can produce the desired results by only setting the system at the levels which produce the desired results. In a human context, these experiments provide a reference point for what to expect from personnel at various levels of the different demographics.

3.3.5 Constraints Analysis/Risk Aversion – by Demographic

In Section 3.2.3, a procedure was described to compare project constraints to one another with no respect to the demographics. This section looks at each project constraint individually and analyzes the weights assigned by each participant with respect to the three demographics described in Section 3.2.1. This is done to determine if certain demographic traits affect how subjects view that particular constraint. It also describes the process used to determine whether or not any demographic factors had a significant effect on a participant's risk tolerance (i.e. utility).

For the project constraints (cost, schedule, risk, and quality), a total of 36 data points were available from the 36 weights obtained for each constraint calculated using Equation 3-1. Thrity-eight data points were available for the risk-aversion analysis. Table A-1 in the Appendix describes the parameters used to configure DesignExpert ™ for the project constraints model. Table A-2 in the Appendix describes the parameters used for the development of the risk aversion model.

DesignExpert™ then produced a matrix of 36 model points comprised of various levels of the three demographic settings. Normally this program is used to help inform the collection of data points, where an experimenter would set each factor

to the level recommended by DesignExpert™ , and then observe and record the response. In this case, the data had already been collected, so the design matrix was modified to accurately reflect the demographic levels of the pool of subjects. For each constraint, the weights calculated using Equation 3-1 (see Appendix A.8) were then entered into the "results" column of the design matrix to allow the program to perform its analysis.

3.3.6 Confidence Analysis

As part of the Scheduling survey, each participant was asked to provide a "most likely" estimate for each activity, as well as their confidence in that estimate (see Appendix A.3 for the verbiage used in the survey). For each project, the average of each participant's confidence levels across all activities in that project was calculated. This average was interpreted as the confidence in the total project duration for that participant. Because some subjects provided estimates for more than one project, a second average consisting of all responses from that participant was calculated. This was necessary to avoid "repeat measurements" which would affect the DOE process (Montgomery 2008, 13). This resulted in each participant having one confidence value that represented their confidence in their estimating ability.

DesignExpert™ was again used in the same manner as described in Section 3.3.4. Design parameters used for this analysis are listed in Table A-3. A total of 26 data points were used in this model. The total number of responses is different from the previous analysis because not all subjects who provided responses in the Traits/Opinions survey provided responses to the Schedule survey.

117

3.3.7 Correlating the Results

Five questions were developed to study whether or not personality traits (e.g. confidence and risk tolerance) and project constraint preferences had any bearing on the final duration estimates provided by the participant. The questions and rationale behind the questions are listed in Table 3-6. The "C" after each question is to identify it as a separate list from the eleven questions in Table 3-5.

Question #	Question
1C	Are confidence levels and standard deviation negatively correlated (i.e. a lower confidence results in a larger standard deviation)? Rationale: Each participant was asked to give their confidence level on their "most likely" estimate. It was effectively asking "how confident are you that the activity will take exactly the time you listed as your most likely estimate". Based on this, it would seem the more confident one was in the "most likely" estimate, the smaller the standard deviation would be since there would be less need to compensate for uncertainty. Method: Each activity had its own standard deviation and confidence, providing a 1-to-1 comparison. A correlation coefficient was calculated for each participant on each project. The average of all of these coefficients was then calculated to provide one final value across all projects.
2C	Does higher confidence positively correlate to higher utility values (i.e. are people who are confident in their estimates less worried about risks? Rationale: Someone who has high confidence in their estimates is optimistic they will succeed. A higher utility value is someone who believes they can beat the odds and will succeed. Method: Calculated the average confidence value for each participant for each project such that each participant had one value for confidence and one value for utility (per project). The average of all of these coefficients was then calculated to provide one final value across all projects.

Question #	Question
3C	Are utility values and standard deviation negatively correlated (i.e. a smaller utility value means a wider standard deviation)? Rationale: The standard deviation for a PERT average is described by Equation 2-2. Using utility as an indicator of optimism, a person with a higher utility value is going to be more likely to believe things will go well. Once they have established a "most likely" estimate, the "best case" and "worst case" estimates will be closer to the "most likely" estimate because they will not feel as much need to "hedge their bet", where the "bet" is whether or not the advertised completion time is correct. Method: For each participant, calculated the variance for each activity and summed them together. Took the square root of the sum to determine the project standard deviation (Keefer and Verdini 1993, 1086). This resulted in one standard deviation and one utility value for each participant for each project, providing a 1-to-1 comparison. The average of all of these coefficients was then calculated to provide one final value across all projects.
4C	Are utility values and T_e negatively correlated (i.e. does a smaller utility value mean a higher T_e value? Rationale: T_e is the sum of the PERT weighted average of the activity durations. A smaller utility value corresponds to "risk aversion". If risk is defined as "failing to complete the project within the projected time", then someone who is averse to risk will have a higher T_e estimate to allow for things to go wrong, but still have a chance of finishing the project by the advertised completion time. Method: Straight comparison. When provided by the participant, each participant had only one utility value and one T_e per project, providing the 1-to-1 comparison for each participant on each project. The average of all of these coefficients was then calculated to provide one final value across all projects.
5C	Does a higher Schedule AHP weight positively correlate to a higher T_e value (i.e. does willingness to sacrifice schedule mean a higher duration estimate)? Rationale: The way the survey asked the question, a larger AHP value indicated more willingness to sacrifice the schedule in favor of preserving other project constraints (cost, risk, and quality). Working under the assumption that higher quality requires longer durations (as indicated by subjects stating they are always too rushed to do their jobs), someone who is ready to sacrifice the schedule for

Question #	Question
	the sake of quality will provide higher duration estimates because they want to make sure they have enough time to do the work to their satisfaction. Someone less willing to sacrifice the schedule will have lower PERT estimates because a project completed quickly keeps end users happy

Method: Straight comparison. When provided by the participant each participant had only one Schedule AHP value and one T_e per project, providing the 1-to-1 comparison for each participant for each project. The average of all of these coefficients was then calculated to provide one final value across all projects. |

Table 3-6: Correlation Questions

In order to calculate a correlation coefficient for each of the comparisons listed in Table 3-6, for each project, two arrays were created, one with the results of the first value in question, the other with the second value. The Excel™ function "=correl(array1, array2)" was then used to determine the correlation coefficient within each project of the two values in question. An average correlation coefficient was then calculated across all projects. A second correlation coefficient was calculated by removing all results where only two subjects provided values and re-calculating the average correlation coefficient. In those cases where only two subjects provided estimates, the results were either fully positively correlated or fully negatively correlated (i.e. "1" or "-1"). When these results were factored into the average, it was believed that they were unduly influencing the results. Removal of these binary data points was an effort to remove this influence. When constructing the arrays used in the "correl" function, if a value was missing, it was replaced with the letters "BLNK" to indicate that the value was blank.

The previous section characterized the subjects based on the responses provided to the various surveys and matching those responses to behaviors in relation to scheduling. This section delves into the estimates themselves and compares the duration values provided to the demographics of the subjects. This was done to determine the expected behavior of subjects within a particular demographic group. This section also describes a Bayesian aggregation method that uses the prior distributions of a Decision Maker (DM) and several experts to develop a single posterior distribution which can be used to in a network schedule.

3.4.1 Participant Behavior in Estimating Durations

The methods described in Section 3.3.3 looked at "fact of" differences of the responses provided by the subjects. They analyzed whether or not estimates for one particular demographic were higher than another, but they did not analyze the magnitude of the difference, nor were they able to determine if one particular demographic had the largest effect on the responses provided. This section uses the three point estimates provided by the subjects to determine whether or not there are any trends in the estimation practices of stakeholders based on their demographics.

For each participant on each project, activity estimates were added together to get a total BC estimate, a total ML estimate, and a total WC estimate. It can be shown that the total network path duration (T_e) calculated by summing the PERT average across all activities in a network path results in the same duration as summing each of the three point estimates across all activities in the path and then calculating a PERT average using these totals (Keefer and Verdini 1993, 1086):

121

$$\sum_{i=1}^{n} \frac{BC_i + (4*ML_i) + WC_i}{6} \quad = \quad \frac{(\sum_{i=1}^{n} BC_i + \sum_{i=1}^{n} 4(ML_i) + WC_i)}{6} \qquad \text{Eqn 3-4}$$

where "n" is the total number of activities in the path. Note that this same principle

does not hold when calculating the variance.

For this section, the ML, BC, and WC values refer to the sum of those values

across all activities, as opposed to the value for each individual activity. To test the

skew of the estimates, the sum of the ML estimates, the BC estimates, and WC

estimates were calculated for each participant for each project. A comparison was

then made of the ratio of the separation between the ML and BC and the sum of the

separation between the ML and BC estimate and the WC and ML estimate (see

Equation 3-5).

$$\frac{(ML-BC)}{(ML-BC) + (WC-ML)} \qquad \text{Eqn 3-5}$$

If the values of the two terms in the denominator of Equation 3-5 are equal, then the

ratio created by the numerator and the denominator would be 0.5, since each term in

the denominator is contributing half of the total and the numerator consists of the first

term in the denominator. If the (ML-BC) value is slightly smaller than the (WC-ML)

value, it will contribute slightly less to the overall total, thus changing the ratio (e.g.

45% as opposed to 50%). Smaller values of (ML-BC) result in smaller ratios for

Equation 3-5.

After computing this ratio for each participant for each project, the results

were organized as described in Section 3.3.5. Multiple results from the same

participant were averaged into one value such that each participant was represented

by a single ratio, resulting in 29 data points. These results were then entered into

DesignExpert™ to determine if there were any significant factors driving the results. Design parameters are listed in Table A-4.

These results, however, only told half the story. Given the ML value, any number of BC and WC values could result in the expected values predicted by DesignExpert™. To further narrow down the possible results, a second set of ratios was developed. This time, for each participant for each project, the BC estimates were summed and ML estimates were summed to provide a single value for each of the two estimates in a project. Two ratios were then calculated using Equation 3-6 and Equation 3-7 to determine the ratio of the outlying estimate as compared to the sum of the outlying estimate and the ML estimate. For Equation 3-6, a result close to 0.5 meant that the BC estimate was relatively close to the ML estimate. As the separation between the BC and ML value increased, the result of Equation 3-6 would get smaller and smaller. For Equation 3-7, a result of 0.5 meant that the WC estimate was relatively close to the ML estimate. Larger values indicated a wider separation between the WC estimate and the ML estimate.

$$\frac{BC}{(ML+BC)} \qquad \text{Eqn 3-6}$$

$$\frac{WC}{(ML+WC)} \qquad \text{Eqn 3-7}$$

The results from Equation 3-6 and Equation 3-7 were organized as described in the previous section. Multiple results from the same participant were averaged such that each participant was associated with a single value, resulting in a total of 29 values.

These values were then entered into DesignExpert™ with the design parameters

shown in Table A-5.

3.4.2 Calculation of PERT Beta parameters

Chapter 2 describes how the beta distribution is widely accepted as descriptive

of the probability distribution of activity duration times (Malcolm et al. 1959, 651).

The probability density function (pdf) of the beta distribution is described by

Equation 3-8:.

$$f(x; \alpha, \beta) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1 - x)^{\beta-1}; \ 0 \leq x \leq 1 \qquad \text{Eqn 3-8:}$$

where α and β describe the parameters of the function and "B" is the Beta function
("NIST/SEMATECH E-Handbook of Statistical Methods" 2016, 1.3.6.6.17, "Beta
Distribution" 2016).

Because the pdf of the beta function is described only on the interval of zero to one, t

Equation 3-9 converts "x" to "t", where "t" is the actual time estimate provided by the

participant (Grubbs 1962).

$$t = BC + (WC - BC)x \qquad \text{Eqn 3-9}$$

Using the value of "t", the mean (T_e) and the mode (ML) can be calculated using

Equation 3-10 and Equation 3-11 (Grubbs 1962).

$$T_e = BC + (WC\text{-}BC)\left[\frac{\alpha}{\alpha + \beta}\right] \qquad \text{Eqn 3-10}$$

$$ML = BC + (WC\text{-}BC)\left[\frac{\alpha-1}{\alpha + \beta-2}\right] \qquad \text{Eqn 3-11}$$

Using a system of simultaneous equations, and assuming Equation 2-1 is a valid

approximation of T$_e$, these two equations were manipulated to solve for the two beta

parameters, α and β using Equation 3-12 and Equation 3-13.

$$\alpha = \left(\frac{-1+2(\frac{ML-BC}{WC-BC})}{\left(\frac{ML-BC}{WC-BC}\right)+\left(\frac{WC-Te}{Te-BC}\right)\left(\frac{ML-BC}{WC-BC}\right)-1} \right) \qquad \text{Eqn 3-12}$$

$$\beta = \left(\frac{(WC-Te)}{(Te-BC)}\alpha \right) \qquad \text{Eqn 3-13}$$

3.5 Duration Estimate Modeling and Expert Aggregation

The following sections describe the proposed new method for modeling

duration estimates and aggregating those estimates based on the work Roberts

(Roberts 1965) as refined by Morris (Morris 1977). In light of some of the challenges

described in Section 2.2.3 with the PERT beta model, Section 3.5.1 provides a

recommendation for a new model. It also describes the method for converting the

three baseline estimates obtained in the PERT methodology into the shape parameters

required to describe the new model. Section 3.5.2 describes a method for calibrating

estimates provided by expert stakeholders. Section 3.5.3 describes Morris'

aggregation method as applies specifically to the problem of aggregating the duration

estimates of multiple project stakeholders.

3.5.1 Determining the Prior

When using the opinion of one person to develop a schedule, the beta

distribution serves as a good model for activity durations and the approximations of

its mean and variance provide a quick way to determine the values required for

network scheduling techniques (Malcolm et al. 1959, 659). When multiple opinions are available, however, the beta distribution becomes more problematic because the distribution is zero outside the defined range. The method to combine these different estimates will be described in more detail in Section 3.5.3, but to summarize, the distribution functions of each estimator (the "priors") are evaluated across a large number of durations ranging from the smallest estimate to the largest estimate among all estimators. The results are then multiplied together to produce a new distribution (the "posterior") which is a single distribution with a new mean and variance to be used in the network schedule. In the case of a beta distribution, the posterior will have density only between duration values shared among the estimates; all values outside this range will drop to zero. To resolve this issue, a distribution was needed that, like the beta distribution, had a single mode and was capable of handling left- and right-skewed data, as well as data without any skew. Unlike the beta distribution, this alternate distribution had to be defined on the entire positive real number line to ensure it could accommodate all possible estimates. A Type I Generalized Extreme Value (GEV) distribution met nearly all of these criteria; its "maximum" form handled the right-skewed estimates and its "minimum" form handled the left-skewed estimates. A Normal distribution could model a case with no skew ("Statistical Distributions" 2016).

Once a new distribution was selected, the standard ML, BC, and WC estimates were modeled using this new distribution. The intent was to match the shape of the beta distribution as closely as possible given that it has been used as a model for over fifty years and, as the creators pointed out, does seem to provide a

126

reasonable approximation of the behavior of activity durations (Malcolm et al. 1959, 651). A GEV distribution is defined by three parameters: ξ, k, and, μ where ξ defines the GEV Type, k is the shape parameter, and μ is the location parameter. For this research ξ = 0 since a Type I distribution was used to model the data. The μ parameter was defined as the participant's ML estimate. This left only the shape parameter, k, undefined. The desire was to match the form of the beta distribution as closely as possible, so with the mode of the new distribution already defined, the next step was to define the "end points" of the distribution (NIST 2016b, 1.3.6.6.16; "Generalized Extreme Value Distribution - Wikipedia" 2016).

For most subjects' estimates, the separation between the BC and ML estimates was less than the separation between the ML and WC estimates, indicating that the subjects were compensating more for things going wrong than assuming things would go right. This group was referred to as the "pessimists" because they were planning for the worst. With these characteristics, the GEV distribution was able to model the estimates due to the right-skew of the estimates. In some cases, however, the separation between the BC and ML estimates was larger than the separation between the ML and WC estimates. This group was referred to as the "optimists" because they appeared to be less inclined to believe things would go badly and more inclined to believe things would go well as indicated by their left-skewed distribution. The form of the GEV distribution used for the pessimists (referred to hereafter as GEV Max) did not accurately model this left-skewed distribution, but by negating the "x" value and subtracting from one, the GEV Max right-skewed distribution could be converted into a left-skewed distribution (referred to hereafter as GEV Min)

(("Generalized Extreme Value Distribution - Wikipedia" 2016, *MATLAB* (version 9.1.0.441655) 2016, n. Help File).

In cases where the separation between the ML and BC estimates and the ML and WC estimates was equal, the Normal distribution served as the model. In this case the two parameters, μ and σ, are pulled directly from the estimates of the subjects. The first parameter, μ, is set to the ML value provided by the participant. The second parameter, σ, is the separation between the ML value and either endpoint (e.g. WC – ML or ML-BC; either will work since the separations are equal) divided by 3. This ensures that 99.7% of the density falls between the BC and WC estimates (Farr 2012, 29). Between the GEV Max, GEV Min, and Normal distributions, all subjects' estimates were modeled. One particular case that occurred in a few of the estimates cannot be modeled by any standard distribution curve. This was a case where the participant provided the same estimate for both the BC and the ML values. In these cases, it is recommended that the decision maker request a new estimate where the BC value must be less than the ML value.

Because the GEV distribution is defined from (-∞,∞), in order to mimic the PERT beta model, it was necessary to solve for "k" such that most of the density fell between the BC and WC estimates. For the GEV Max case, when graphing various estimates it was noted that the graph of the PDF began to rise above the x-axis when "k" was set such that the cumulative distribution function (CDF), when evaluated at the BC estimate was equal to 0.0001. For the GEV Min case, the same behavior was noted at the WC estimate when it was evaluated at 0.99995. Equation 3-14 through Equation 3-19 provide the PDF and CDF of each distribution model ("Generalized

Extreme Value Distribution - Wikipedia" 2016, "Normal Distribution" 2016,

MATLAB (version 9.1.0.441655) 2016, n. Help File)

$$PDF\ GEV\ Max = \frac{1}{k}e^{\frac{-(x-\mu)}{k}}e^{-e^{\frac{-(x-\mu)}{k}}}$$ Eqn 3-14

$$CDF\ GEV\ Max = F(x) = e^{-e^{-(x-\mu)/k}}$$ Eqn 3-15

$$PDF\ GEV\ Min = \frac{1}{k}e^{\frac{(x-\mu)}{k}}e^{-e^{\frac{(x-\mu)}{k}}}$$ Eqn 3-16

$$CDF\ GEV\ Min = F(x) = 1 - (e^{-e^{\frac{x-\mu}{k}}})$$ Eqn 3-17

$$PDF\ Normal = \frac{1}{\sqrt{2\sigma^2\pi}}e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$ Eqn 3-18

.

$$CDF\ Normal = F(x) = \frac{1}{2}[1 + \text{erf}(\frac{x-\mu}{\sigma\sqrt{2}})]$$ Eqn 3-19

Setting the probabilities to these extreme values in both cases also helped
ensure that most of the probability density would fall within the outlying estimates as
the creators of PERT intended (Malcolm et al. 1959, 651). It also meant, however,
that unlike PERT, there was a non-zero probability that the duration could be outside
the extreme estimates, which helps account for the cases when the estimator is wrong
about the location of the extremes of the distribution.

To solve for "k" in the GEV Max model, Equation 3-15 was set to equal

0.0001 and evaluated at the BC estimate, where x = BC and μ = ML. Rearranging the

variables to solve for k resulted in Equation 3-20 for the GEV Max case.

$$k = \frac{(BC-ML)}{(-(\ln(-(\ln(0.0001)))))} \qquad \text{Eqn 3-20}$$

It was noted that there was a relationship between the k parameter and the

separation between the ML and BC estimates. When plotted using Excel™ and using

the "trendline" option, it was discovered that this relationship was linear and, for a

GEV Max model, solving for k simplifies to:

$$k = 0.45038 * (ML - BC) \qquad \text{Eqn 3-21}$$

To solve for "k" in the GEV Min model, Equation 3-17 was set to 0.99995

and evaluated at the WC estimate, where x = WC and and μ = ML. Rearranging to

solve for "k" resulted in Equation 3-22.

$$k = \frac{(WC-ML)}{(\ln(-\ln(-(0.99995-1))))} \qquad \text{Eqn 3-22}$$

There was also a linear relationship between the "k" parameter and the

separation between the WC and ML values for the GEV Min model. For this model,

solving for k simplifies to:

$$k = 0.43613 * (WC - ML) \qquad \text{Eqn 3-23}$$

Solving for "k" using the endpoint estimates meant that all three distribution

parameters were defined. The subjects' beliefs regarding the probability of activity

duration could then be described by the PDFs of the GEV Max distribution , the GEV

Min distribution, or the Normal distribution as seen in Equation 3-14, Equation 3-16,

and Equation 3-18, respectively.

Unfortunately, this meant that for the GEV Max model, while the "ML and

the BC estimates matched a beta distribution reasonably closely, the WC value did

not necessarily match up as well (see Figure 6-1). The opposite was true of the GEV

Min model, where the ML and WC estimates matched, but the BC estimate did not.

This is because the GEV distribution is determined by two parameters, so only two of

the three values can be set. The Normal distribution matched closely on both ends

due to its symmetrical nature. A discussion on this limitation of the model will be

provided in Chapter 5.

3.5.2 Calibrating the Experts

Using the method described in Section 3.5.1, a project manager can take the

three estimates provided by any given project team member (including herself) and

model his belief about the probability of the task duration. It has been shown,

however, that in general, people are not good estimators of probability for many of

the reasons discussed in Chapter 2 (Morris 1977, 682; Hubbard 2010, 57). To

compensate for that limitation, the prior distribution must be calibrated, just as a

weight scale must be calibrated to compensate for any off-set in the measurement.

Since empirical calibration data was not available for this study, a subjective

calibration scheme was used instead. In Morris' work, the CDF of the expert's prior

distribution is run through a calibration function which is defined by the decision

maker's beliefs about the expert (Morris 1977, 682, 684–88). If the decision maker

believes the expert has understated his knowledge, the calibration function will take

on a different form than if the decision maker believes the expert has overstated his

knowledge. When a decision maker believes an expert has understated his

knowledge, it is as if the decision maker looks at the estimate and says, "I believe you know more about this than you think you do. We don't need to account for such a wide range and can tighten up this variance." Conversely, if the decision maker believes an expert has overstated his knowledge, the decision maker would say, "I don't believe you know as much about this as you think you do. We need to account for a wider range of values and widen this variance". Based on this belief, the calibration function will modify the expert's prior distribution to alter the variance and make it either smaller or larger, depending on how the decision maker feels about the expert (Morris 1977, 688; Savage 1971, 796).

For this research, the beta distribution was chosen to calibrate the experts. The beta distribution was able to accommodate both the overstated and understated experts in the left-skewed, right-skewed, and symmetrical cases. It can also handle the case of a fully calibrated expert. In this case, the beta distribution is treated less as a distribution function and more as a filter through which the signal of the expert's prior distribution is processed. To avoid confusion, it will be referred to as the beta filter from this point forward.

Morris recommends a calibration scheme based on whether or not the expert will be surprised by the revealed value of the variable, in this case, the actual duration of the activity. He defines surprise as the revealed value occurring below the 0.1 fractile or above the 0.9 fractile (Morris 1977, 691). The probability the decision maker assigns to the likelihood of this event (i.e. the actual value falls in the tails of the expert's distribution) is the basis for the shape of the calibration curve.

The decision maker's belief about the expert can be modeled by altering the parameters of the beta filter. The table below provides a general guideline for the relationship between the two beta filter parameters so that the model will reflect the decision maker's belief about the expert.

Expert's Prior Skew	**DM's Belief about the Expert**	**Beta Parameter Setting**
Left, Right, or Symmetrical	Fully Calibrated	$\alpha = \beta = 1$
Left	Understated	$\alpha < \beta$; $\alpha > 1$; $\beta > 1$
Right	Understated	$\alpha > \beta$; $\alpha > 1$; $\beta > 1$
Left	Overstated	$\alpha < \beta$; $\alpha < 1$; $\beta < 1$
Right	Overstated	$\alpha > \beta$; $\alpha < 1$; $\beta < 1$
Symmetrical	Understated	$\alpha = \beta > 1$
Symmetrical	Overstated	$\alpha = \beta < 1$

Table 3-7: α and β Beta Filter Parameters

By providing estimates, experts are indirectly providing information on their uncertainty about the duration (Gelman et al. 2013, 32; Morris 1977, 688). When those beliefs are modeled as the "prior", the expert indirectly informs the decision maker of the revealed value that would surprise him. The lower bound of "surprise" is determined by setting the CDF to 0.1 and solving for "x", where ML and "k" have already been set. The upper bound of "surprise" is determined by setting the CDF to 0.9 and again solving for "x" (Morris 1977, 683,691; Keefer and Verdini 1993, 1087). If the decision maker believes the expert is fully calibrated, the sum of the area under the two tails of the beta filter curve will be 0.2, where the area between zero and 0.1 is 0.1 and the area between 0.9 and 1.0 is also 0.1 (Önkal et al. 2003, 181; Yates

133

1990, 21–23, 69–71). This is effectively saying that the decision maker believes that the likelihood of the revealed value of the duration falling in one of the tails of the expert's prior is 20%.   If the decision maker believes that the expert is understating his knowledge, she believes the sum of the areas under the tails will be less than 0.2. If she feels the expert is overstating his knowledge, she believes the sum will be greater than 0.2 (Morris 1977, 688; Winkler 1981, 482).

Once the decision maker's beliefs about the expert are determined, the values of the parameters can be set. The mode of a beta distribution (and therefore the beta filter) is described by Equation 3-24 ("Beta Distribution" 2016).

$$Beta\ Filter\ Mode = \frac{\alpha - 1}{\alpha + \beta - 2}$$ Eqn 3-24

For the three distribution models used (GEV Max, GEV Min, and Normal), when the CDFs of these distributions are evaluated at the ML value (i.e. the mode), the result is always the value shown in Table 3-8. Since the intent was to only alter the variance of the expert's prior distribution and not its location on the number line, the values of α and β were set such that solving Equation 3-24 using those values of α and β would result in the values below.

Expert's Prior Distribution Shape	Value of the Beta Filter Mode
GEV Max	0.3679
GEV Min	0.6321
Normal	0.5

Table 3-8: Beta Filter Modes

With that constraint in place, the values of α and β were then adjusted so that, while still meeting the constraint of Equation 3-24, the values of α and β also set the resulted in the desired likelihood of surprise (see Appendix A.10 through A.12).

The CDF of the beta filter is described by the incomplete beta function which is numerically challenging to evaluate ("Beta Distribution" 2016). As a practical implementation, the Excel™ function "=betadist(x,a,b) was used, where "x" is the value at which the CDF of the beta distribution (in this case the beta filter) is evaluated, and "a" and "b" are the shape parameters of the distribution. To calculate the Likelihood of Surprise (LoS), Equation 3-25 was entered into an Excel™ spreadsheet.

$$LoS = (\text{betadist}(0.1, \alpha, \beta)) + (1 - (\text{betadist}(0.9, \alpha, \beta))) \qquad \text{Eqn 3-25}$$

where α and β are the shape parameters of the filter. The first term calculates the likelihood that the revealed value will fall below the expert's prior's 0.1 fractile and the second term calculates the likelihood that the revealed value will fall above the expert's prior's 0.9 fractile. Together, they represent the likelihood that the revealed value will surprise the expert. Using the "Solver" application on Excel™, Equation 3-25 was systematically set to all values from 0.1 to 0.99 in increments of 0.1 (i.e. likelihoods from 1% to 99%), subject to the constraint that the values of α and β had to satisfy Equation 3-24 when set to the values shown in Table 3-8. Appendix A.10 through A.12 shows the different combinations of α and β that result in the desired likelihood. This appendix consists of tables for all three prior distribution models, as well as the values of 1/(B(α,β) which will be required for Eqn 3-25.

135

With the expert's prior and the beta filter both fully determined, the process of calibrating the expert is relatively straight forward. First, the CDF of the expert's prior is calculated based on their original estimates by using Equation 3-15, Equation 3-17, or Equation 3-19 as appropriate. Once these values are determined, they can be processed through the beta filter described by Equation 3-26 (the beta filter) ("NIST/SEMATECH E-Handbook of Statistical Methods" 2016, 1.3.6.6.17, "Beta Distribution" 2016)).

$$\Phi(F(x)) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} \qquad \text{Eqn 3-26}$$

where "x" is the value of F(x) as calculated by Equation 3-17, or Equation 3-19, α and β are the filter parameters (previously determined), and B(α,β) is the beta function (used to normalize the equation) evaluated at α and β ("Beta Function" 2016). This result is then multiplied by the original PDF of the expert to produce the calibrated prior function.

$$f_{cE_i(x)} = \left(\frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}\right) * f_{E_i}(x) \qquad \text{Eqn 3-27}$$

where $f_{E_i}(x)$ is the prior distribution of the "i"th Expert, as described by either Equation 3-14, Equation 3-16, or Equation 3-18, depending on the original estimates provided by the Expert and "x" is the duration estimate at which the function is evaluated. The result is a calibrated expert prior that can then be combined with the decision maker's prior to calculate the posterior probability of the activity duration (Morris 1977, 683, 685–86).

3.5.3 Calculating the Posterior Probability

With the prior probability of the decision maker defined and the expert

defined and calibrated, a posterior probability distribution for the activity duration

could be calculated. In a Bayesian belief-updating model, a decision maker forms a

prior probability distribution based on the available data (Simon French 1985, 189;

Jeffreys 1983, 33–34). As new data are received, this distribution is updated to

reflect the new data (Silver 2012, 241, 247–48). Using a Bayesian belief-updating

model and treating each expert's estimation as new data, the decision maker's belief

about the probability of the activity duration is updated based on the information

provided by the experts (Morris 1977, 680). For the purposes of this research, the

assumption is made that the decision maker and all experts are independent of one

another in the statistical sense.

According to Morris' method, assuming independence, the posterior

probability curve is described by Equation 3-28

$$f_p(x) = c * f_{DM}(x) * f_{cE_i}(x) * \ldots * f_{cEe_n}(x); \quad \text{for i} = 1\ldots\text{n} \qquad \text{Eqn 3-28}$$

where $f_p(x)$ is the posterior distribution, "c" is a normalizing constant, $f_{DM}(x)$ is the

decision maker's prior probability distribution, and $f_{cE_i}(x)$ is the ith expert's

calibrated prior probability distribution for "n" total experts (Morris 1977, 687).

Evaluating each term on the right side of the equation over a range from zero to some

value larger than the largest WC value (to compensate for the tail) will produce a

fully defined posterior curve. To normalize the curve, each value is then multiplied

by the reciprocal of the area under the curve.

The full process for calculating the posterior is described below using an example activity, a DM, and one Expert. The prior distributions of both the DM and Expert are modeled using a GEV Max distribution. Table 3-9 explains the methods used to arrive at the example values shown in Table 3-10.

	Value in the column represents:	Calculated using:
Column A	Duration in increments of 0.1	N/A
Column B	DM's prior distribution PDF	Equation 3-14, where x = value in Column A
Column C	DM's prior distribution CDF	Equation 3-15, where x = value in Column A
Column D	Expert's prior distribution PDF	Equation 3-14, where x = value in Column A
Column E	Expert's prior distribution CDF	Equation 3-15, where x = value in Column A
Column F	Calibration of the expert	Equation 3-26, where $\alpha = 1.44$ and $\beta = 1.76$, and x = value in Column E
Column G	Expert's Calibrated Prior	Column D multiplied by Column F
Column H	Normalized aggregated posterior distribution PDF	Column B multiplied by Column G and the normalizing constant, c
Column I	Aggregated posterior distribution CDF	Column H multiplied by 0.1
c	Normalizing constant	The reciprocal of the sum of all values in Column I

Table 3-9: Calculating the Aggregated Posterior Distribution

From Table 3-10, it can be seen that the maximum value of the posterior distribution (Column H) occurs at a duration of 21.8 (Column A).

A	B	C	D	E	F	G	H	I
	Decision Maker Prior		Expert #1 Prior		$\Phi(x) = beta(F1(x),1.44,1.76)$			
	f(x) -> PDF	F(x) -> CDF	f1(x) -> PDF	F1(x) -> CDF	Beta Filter	Calibrated Prior	Posterior PDF	Posterior CDF
Duration	GEV Max		GEV Max		$\Phi(F(x))$	$\Phi(F(x))*f(x)$	DM * E1	c = 38.120
21.2	0.0643	0.7681	0.0496	0.1327	1.0959	0.0543	0.13316	0.0133
21.3	0.0628	0.7744	0.0505	0.1377	1.1090	0.0560	0.13409	0.0134
21.4	0.0613	0.7806	0.0514	0.1428	1.1219	0.0577	0.13484	0.0135
21.5	0.0599	0.7867	0.0523	0.1479	1.1344	0.0593	0.13542	0.0135
21.6	0.0584	0.7926	0.0532	0.1532	1.1466	0.0610	0.13583	0.0136
21.7	0.0570	0.7984	0.0540	0.1586	1.1584	0.0626	0.13607	0.0136
21.8	0.0556	0.8040	0.0549	0.1640	1.1700	0.0642	0.13613	0.0136
21.9	0.0543	0.8095	0.0557	0.1695	1.1812	0.0658	0.13603	0.0136
22	0.0529	0.8148	0.0565	0.1752	1.1921	0.0673	0.13577	0.0136
22.1	0.0516	0.8201	0.0572	0.1808	1.2026	0.0688	0.13535	0.0135
22.2	0.0503	0.8252	0.0580	0.1866	1.2128	0.0703	0.13478	0.0135
22.3	0.0490	0.8301	0.0587	0.1924	1.2226	0.0717	0.13406	0.0134
22.4	0.0478	0.8350	0.0594	0.1983	1.2321	0.0731	0.13319	0.0133
22.5	0.0465	0.8397	0.0600	0.2043	1.2412	0.0745	0.13219	0.0132
22.6	0.0453	0.8443	0.0607	0.2103	1.2500	0.0758	0.13105	0.0131
22.7	0.0441	0.8488	0.0613	0.2164	1.2583	0.0771	0.12979	0.0130
22.8	0.0430	0.8531	0.0619	0.2226	1.2664	0.0784	0.12841	0.0128

Table 3-10: Example Full Process Calculations

To avoid the algebraic quagmire of finding a general equation to describe the curve in Column H, it seemed best to use an approximation with known equations for the probability density, cumulative probability, mean and variance. It was noted that plotting the results of Equation 3-28 resulted in a curve that still closely resembled either a GEV Max, GEV Min, or Normal distribution. The end-points and mode were in new locations, but the general shape remained similar. To determine the required shape, Equation 3-29 subtracted the sum of the area above the mode from the sum of the area below the mode.

$$S = (\sum_{d=0}^{d=m-1} CDF_d) - (\sum_{d=m+1}^{d=n} CDF_d) \qquad \text{Eqn 3-29}$$

where "d" is a duration estimate, "m" is the mode, "n" is the total number of activities and CDF_d is the is the value of Equation 3-28 multiplied by the value by which the duration is incremented (e.g. 0.1 in this example).

Once this value for "S" is calculated, the following Excel™ command was used to quickly determine the best approximation for the posterior curve.

139

=IF(AND(S < 0.13,S > -0.13),"Normal",IF(S > 0.13, "GEV Min", IF(S < -0.13,"GEV Max")))

After determining the general shape of the resulting posterior distribution curve using the Excel™ commands just described, it was necessary to once again determine the parameters that defined the mean and the variance needed for development of a network schedule. The new mode (i.e. ML estimate) was determined by finding the largest value of $f_p(x)$ from Equation 3-28. This was accomplished by visually scanning the results of the spreadsheet (Column H from Table 3-10 in this example) with the assistance of Excel's™ color-scaling feature. This value represented the peak of the posterior distribution curve. The "x" value from Column A (i.e. the duration) associated with that peak was set as the mode of the posterior distribution, which in turn defined the "μ" parameter of the GEV and Normal approximations. This left only "k" or "σ", depending on the form of the approximating curve, to be defined. After some experimentation, it was determined that the best method for matching the GEV approximation to the posterior distribution as calculated using Equation 3-28 was to set the Equation 3-14 (for GEV Max) or Equation 3-16 (for GEV Min) to $f_p(mode)$, which is the value of Equation 3-28 evaluated at the mode, and solve for "k". When evaluated at the mode, Equation 3-14 and Equation 3-16 reduce to Equation 3-30.

$$k = 0.367879/f_p(mode)$$ Eqn 3-30

With both "μ" and "k" defined, the GEV approximation for the decision maker's posterior probability is fully characterized. This, in turn, allows for the calculation of the mean and variance of the posterior probability by using Equation 3-31 through Equation 3-33 (NIST 2016b, 1.3.6.6.16; "Generalized Extreme Value Distribution -

Wikipedia" 2016, "MinStableDistribution—Wolfram Language Documentation" 2017).

$$Mean_{GEV\ Max} = \mu + k\gamma \qquad \text{Eqn 3-31}$$

$$Mean_{GEV\ Min} = \mu - k\gamma \qquad \text{Eqn 3-32}$$

$$Variance = k^2 \frac{\pi^2}{6} \qquad \text{Eqn 3-33}$$

where "μ" is the mode of the GEV approximation, "k" is the shape parameter, and "γ" is Euler's constant (approximated at 0.57721) ("Euler–Mascheroni Constant" 2016).

In the event the closest model of Equation 3-28 is a Normal distribution, the same procedure is followed as for the GEV approximations, but Equation 3-30 is replaced by Equation 3-34 which defines the variance, and mean is equal to the mode ("Normal Distribution" 2016).

$$\sigma = \frac{\sqrt{(\frac{1}{f_p(mode)})^2}}{2\pi} \qquad \text{Eqn 3-34}$$

For a typical network schedule, the mean and variance of each activity distribution are sufficient to calculate the total project duration and project variance. In some cases, however, it may be desirable to determine a BC and WC value for the posterior distribution, perhaps for use in a Monte Carlo simulation (Mantel Jr. et al. 2004, 156–60). Because the GEV Max and Min approximations only have two parameters, only one of the extremes can be set, but the other can be approximated.

For a GEV Max prior distribution, the BC estimate was used to solve for the shape parameter "k" by setting the CDF of the distribution to 0.0001 and evaluating at the BC estimate. For the posterior distribution, the shape parameter has already been set as described above. With the shape parameter set, Equation 3-20 and

Equation 3-21 can be rearranged to solve for the BC value as seen in Equation 3-35 or, for a simplified version, Equation 3-36.

$$BC = -k(\ln(-(\ln(0.0001)))) + ML \qquad \text{Eqn 3-35}$$

$$BC = ML - \left(\frac{k}{0.45038}\right) \qquad \text{Eqn 3-36}$$

For the GEV Min case, the WC parameter was used to solve for k by setting the CDF of the distribution to 0.99995 when evaluated at the WC value. For this distribution model, rearranging Equation 3-22 and Equation 3-23 solves for the WC value as seen in Equation 3-37 and, for the simplified version, Equation 3-38.

$$WC = (\ln(-\ln(0.00005)))k + ML \qquad \text{Eqn 3-37}$$

$$WC = ML + \left(\frac{k}{0.43613}\right) \qquad \text{Eqn 3-38}$$

The remaining extreme estimate (WC for the GEV Max distribution and BC for the GEV Min) cannot be calculated since there are no remaining parameters in the distribution equation, but the values can be set such that the density between the BC and WC values is at the desired level. The original creators of PERT intended that most of the density should fall between the BC and WC estimates (Malcolm et al. 1959, 651). Given that 3σ of the Normal distribution comprises 99.7% of the density between the BC and WC estimates, it is recommended to set the remaining parameter for both models of the GEV distribution such that 99.7% of the density will also fall between the BC and WC value (to standardize across all models) (Farr 2012, 29).

Equation 3-39 and Equation 3-40 calculate the probability density between the BC and WC values for the GEV Max and GEV Min distributions, respectively.

$$\Delta_{GEV\ Max} = \left(e^{-e^{-\frac{WC-\mu}{k}}}\right) - \left(e^{-e^{-\frac{BC-\mu}{k}}}\right) \qquad \text{Eqn 3-39}$$

$$\Delta_{GEV\ Min} = (1 - \left(e^{-e^{\frac{WC-\mu}{k}}}\right)) - (1 - \left(e^{-e^{\frac{BC-\mu}{k}}}\right)) \qquad \text{Eqn 3-40}$$

For Equation 3-39, $\Delta_{GEV\ Max}$ is set to 0.997 to force 99.7% of the density to fall between the BC and WC estimates. The final term of that equation is equal to 0.0001 based on the discussion in Section 3.5.1. Rearranging Equation 3-39 to solve for the WC value results in Equation 3-41

$$WC = ML - k(\ln(-(\ln(0.9969)))) \qquad \text{Eqn 3-41}$$

For the GEV Min case, $\Delta_{GEV\ Min}$ is once again set to 0.997 and the first term on the right of the equation reduces to 0.99995 based on the discussion in Section 3.5.1. Rearranging Equation 3-40 and solving for the BC value results in Equation 3-42.

$$BC = ML + k(\ln(-(\ln(0.99705)))) \qquad \text{Eqn 3-42}$$

In the event of a Normal approximation, the ML value will equal the mean, and the BC and WC values can be found by solving Equation 3-43 and Equation 3-44.

$$BC = ML - (3\sigma) \qquad \text{Eqn 3-43}$$

$$WC = ML + (3\sigma) \qquad \text{Eqn 3-44}$$

Ultimately, the method just described provides a means for incorporating the beliefs of multiple team members while still using the network scheduling methods that have been developed and refined over the last fifty years. It also provides a way for project managers to show a basis for estimate when presenting the schedule to

143

senior leadership. Examples of this process and its results will be demonstrated in

Chapter 7.

Chapter 4 Results – Opinions on Scheduling Issues

The previous chapter described the methods used to categorize subjects, gather their inputs on activity durations, and explore some of the thinking behind those estimates. This chapter presents the results of the "essay" questions from the "Course of Action" (COA) survey and the Scheduling surveys. Section 4.1 consolidates the response from the COA survey which asked why subjects believed projects fall behind schedule. Section 4.2 provides the results of the second part of the Scheduling survey. These questions were related to whether or not subjects believed the provided task list covered all required activities, if any activities were missing, and if the resources provided to the project were adequate.

4.1 COA Survey – The Results

The following section describes participant responses to the question "why do projects fall behind schedule. Subjects were identified only by their Position demographic (management or technical). Responses are organized by the ways in which the two groups agreed, the ways in which they disagreed (along with some "editorial" comments which provide further insights into the perceptions held by members of each group), and finally, a summary of the results.

4.1.1 Why do projects struggle? – Agreements

From Part 2 of the survey found in Appendix A.5, subjects were asked why in general, given their professional experiences, they believed projects fell behind

schedule. Although the answers varied, several themes emerged among the subjects and even across the management/technical boundary. In some cases, the same theme appeared across this boundary, but the sides took opposing views. Below is a summary of the responses organized by general theme. Note that throughout this section, the thoughts and opinions expressed are those of the research subjects. Some phrases are taken directly from the responses of the subjects while others are paraphrased, but all of the ideas and concepts are derived from the anonymous surveys submitted by the subjects.

One of the themes mentioned by both management and technical subjects was a perceived inability to properly plan out a project. Several subjects simply listed out "poor planning" or "inadequate planning" as an explanation for why projects fail, which in this context, is interpreted to be primarily focused on the failure to truly capture all activities required to complete a project. Other subjects expanded this definition of "poor planning" by explaining that resources (which seemed to refer mostly to people) were not properly managed because adequate time was not spent on resource allocation. Still others went on to clarify that this failure to plan resulted in schedule slips because oversights in the planning phase resulted in problems during the execution phase which led to re-work and re-design. Interestingly, while technicians listed funding as a cause of schedule slips, subjects in the management category did not. The specific funding issues mentioned by the subjects were: lack of funding, timeliness of funding, and end-of-fiscal-year spending driving purchases before requirements and design were completed (i.e. a system is purchased and the

project is planned around the already-purchased system as opposed to planning a project and purchasing a system to meet the needs of that project).

Another cause of schedule slips which fell under the "poor planning" category was a project's inability to deal with unforeseen circumstances. Some subjects listed "unforeseen issues" as a cause while others more specifically called out weather or equipment failures, the latter notably mentioned only by the technicians and not by management. Still others listed logistics problems or delays by organizations outside of the project team's control. Subjects in both the management and technical categories mentioned the need to build contingency into the schedule to handle these unforeseen issues. One participant contended that most projects do not include contingency because those with approval authority are more likely to approve projects that initially show a quick completion date.

Another theme, closely tied to "poor planning" and mentioned mostly by management subjects, but also by at least two technicians, is a failure to adequately define requirements. The concerns mostly centered on the fact that a project rarely has a complete understanding of everything that is required at the outset. Teams develop requirements as they understand them, but things are complicated by unknowns (especially in research and development projects), changing requirements, and poor communication with stakeholders at the beginning of the project. As one participant mentioned, the failure to fully define requirements at the beginning of the project hurts the project team in later project phases, as they must incorporate updated requirements during development and execution. Another participant pointed out that project teams are sometimes hamstringed by processes that do not necessarily match

147

the project and that it is sometimes impossible to fully define requirements at the outset of the project unless the team were to be eternally stuck in requirements development. The participant pointed out that the danger of remaining in requirements definition too long was that technology would speed past the development team, who would then be stuck in an endless loop of updating requirements to match technological capabilities.

One cause mentioned multiple times by subjects in both categories was the belief that most project schedules are too aggressive from the start. Both management and technical subjects felt there was a problem with unrealistic expectations/goals being placed on the project team to complete the project by a given date. One management participant even went on to say that the schedules were aggressive because of the need to meet an already-unachievable target date and a technical participant stated that adequate time was not allocated from the beginning of the project. One of the reasons listed by both management and technical subjects was that schedules were created and assumptions were made at the milestone level and that these milestones did not adequately describe the level of work that needed to be completed. One participant went on to say that it was not possible to properly allocate resources using only a milestone schedule. The participant stated that this allocation could only be completed with a fully realized schedule, but creation of detailed schedules was rarely accomplished. Further complicating matters was the belief by both managers and technicians that schedules were developed and presented in a way that would ensure customer approval to proceed (or continue) as opposed to a schedule that accurately reflected how long it would take to complete the project.

148

Technical subjects also believed that there was a problem with non-technical personnel dictating the schedule. The belief seemed to be that these non-technical personnel did not have enough understanding of the details of the work and were therefore not able to develop accurate schedules. These responses also suggested that the technical team had not had the opportunity to review these schedules before they were presented outside the team.

Probably the most frequently mentioned cause of schedule delays across both groups of subjects was the fact that personnel assigned to a project were pulled off the project prior to completion or they were not allowed to focus solely on the project at hand due to having several other projects that also needed attention. Management subjects pointed out that when schedules are created, they are based on the known available resources. When those resources are decreased, the schedule will also slip. Another subject pointed out that even if the resources are re-assigned to the project, there is a learning curve associated with getting them re-acquainted with the project and caught up on current progress. Technical subjects seemed to focus more on resources being over-tasked. Their contention was that there were too many activities required to be completed by too few people and that schedules do not account for the realities of a matrixed organization.

An interesting theme that manifested itself mostly in the technical group was too much interference by personnel who were not directly involved in the project. These subjects believed that too many people who were not directly involved in the project and did not have intimate knowledge of the details were able to affect the schedule. One believed that the constant review process hampered progress and one

149

subject went so far as to say these reviewers were putting up roadblocks that the team must then overcome. Another subject pointed out that project personnel may be following certain project methodologies because they have been directed to by a higher authority, and not because it is a good fit for the project. As mentioned before, one complaint was that inexperienced personnel were shrinking the schedule because they did not understand the full scope of the project. This theme was also reflected in two of the management responses. One believed there was a problem with the schedule planner being too inexperienced to know to build contingency into the schedule and the other believed that the weak-matrixed nature of the organization did not allow for a project manager to understand what other projects his/her resources were also assigned to. This second issue is reflective of another issue mentioned by one subject: poor communication. On a more positive note, one technical response did state the belief that as project managers were gaining experience, they were learning to manage contingency instead of simply using it for the sake of using it.

Several other causes were mentioned which fell outside of the major themes just described. Poor estimating was mentioned by both management and technical subjects, with one subject stating that it is difficult to learn from mistakes because there were no records of how long previous projects actually took. One management subject also mentioned poor execution and poor teamwork as causes for schedule delays. Technical subjects brought up inadequate training on equipment and no concrete delivery date (which meant no accountability for the completion). This second cause was further clarified to say that those who will accept the project may

150

not necessarily feel any pressure to approve the work if there is no official due date set.

4.1.2 Why do projects struggle? – Disagreements and Editorials

The previous section described several similarities between the perspectives of management and technical subjects regarding why schedules fall behind. Despite these similarities, there were a few marked differences. Two management subjects pointed out that they believed schedules were being lengthened because team members kept trying to improve the system in question "just a little bit more". One subject quoted that the "enemy of good is better" and believed that it was okay to try and optimize the system as long as the project could absorb the effort. Otherwise, the subject feared that the project teams would get caught in an improvement loop. Another management subject stated the belief that there was too much "gold plating" in projects overall. This subject specifically mentioned the customer and technical leads pushing to continue testing even once the system had proved operational capability.

On the other side of the fence, the technicians believed that the problem was on the other end. One subject stated that [technical people] prefer to have all systems as close to perfection as possible and will therefore usually push for as much testing as the project is willing to give them. This statement was offered without hinting whether or not the subject thought this was good or bad. Other subjects, however, believed that the attitude of "close enough" was too prevalent and that this ultimately caused problems later on in the project. One subject stated that the survey question itself was contradictory because if the system had "issues" then, by definition it

151

needed to be repaired. The subject went on to say that with a management hat on and without all the facts (implying that managers typically do not have all of the technical details) that system acceptance would be acceptable given that the system was meeting requirements. Another subject echoed these sentiments by saying that in the given scenario, one should make the system better in order to allow for system adjustments later on. One management subject did state that it was important to note the trending behavior of the system. This subject stated that if one could predict the system would be out of specification by the time it was needed, then one should take the extra week at the onset as opposed to accomplishing several rounds of testing and then having to re-do the work when the system failed. Two technical subjects brought up a conflict between funding to replace/upgrade/repair the system versus the funding it would take extend the project to fix the system. Both subjects advocated increased funding for the systems to mitigate potential future schedule delays driven by equipment failures.

Although not related directly to why schedules fail, subjects did provide some insight into what seem to be prevalent attitudes between management and technical personnel. A management participant stated that there can be a wide variation in estimates among management, functional supervisors, and project personnel. The participant also stated that, based on experience, the longest estimates came from the project personnel (i.e. the technical people). Speaking somewhat to that point, one of the technical subjects hinted that, in the past, estimates had been exaggerated with respect to how long it should take to complete a given task. The participant stated that in an effort to curb this trend, activity durations were cut significantly, with the

suggestion that the cuts may have been too severe. The participant also stated that the

durations are starting to lengthen again as project managers gain experience.

4.1.3 Summing Up

To summarize the responses to the question "why do projects fall behind

schedule", it would appear that multi-tasking of project resources is a major offender,

followed by overly aggressive initial schedules which are then subject to revision by

those who are not directly involved in the day-to-day details. There was a belief that

technical personnel were not given adequate input into the schedule, but also that,

when given input, their estimates were typically the longest. Lack of funding and the

allocation of that funding were also mentioned, as well as a debate regarding the

criticality of certain project activities.

4.2 Scheduling Surveys – Beyond the Duration Estimates

The following section provides the results of the second part of the Scheduling

survey. In the first part of that survey, subjects were asked to provide duration

estimates for a list of activities that were deemed necessary to finish the project

described in the survey. The second part of the survey asked three questions:

- Are the resources assigned adequate to successfully complete the project?

- Are all activities on this list required for successful project completion?

- Are there any activities missing from the list that are required for successful project completion?

These questions gave each participant the opportunity to provide comments on

human resource levels and legitimacy of the activity list provided. This section

analyzes those results to determine if there are any patterns that could help explain

why estimating continues to be an ongoing challenge.

4.2.1 Adequacy of Resources Assigned

The first question asked if more personnel needed to be assigned to the

project. Of the 70 responses received, 48 responded that no additional personnel

were required and that the suggested number of project personnel was adequate to

meet the requirements. Eight other subjects answered with a qualified no. The

qualifiers generally fell into one of two categories. One qualification was that the

number of people assigned was adequate as long as those assigned were able to

dedicate their time primarily to the task at hand. The other category focused more on

project unknowns. If training was involved or if someone was unable to work, then

the participant would have preferred to have an extra person available.

Five subjects believed that either one or two more project team members were

needed to successfully execute the project. Four subjects provided a qualified "yes",

with three stating that more people would be needed if the desire was to decrease the

schedule. One of the subjects who responded with a "yes" may have misunderstood

the intent of the question. When reading the clarification statements, it appeared that

the "yes" response focused on activities that were outside the scope of the project

phase in question. Two subjects were undecided on whether or not more people were

needed. Both of these subjects mentioned overtasking personnel with too many

competing priorities. One participant stated that either more people needed to be

added or those already working on the project needed more time to focus solely on

the project. The other stated that the roles needed to be clarified to ensure the correct

additions or it would not matter who was added. Three subjects did not provide a

response.

4.2.2 Activity Necessity

In the next question, subjects were asked if they believed each of the listed

activities needed to be completed in order to successfully complete the project. Of

the 70 responses, 45 stated that all activities listed in the survey were required for

successful completion. Seven responses could be categorized as a "qualified yes"

with reasons including some activities were only needed due to a special case on that

project, a statement that the participant was "still learning", and a statement of the

belief that one of the activities had already been completed when the survey was

filled out. Twelve responses indicated that that not all items listed in the survey

needed to be completed or that activities listed had already been accounted for in

other activities. Two subjects provided a "qualified no" with one indicating that if no

problems were found some activities would not be necessary and the other stating that

some activities were not technically required, but should be given a "best effort".

One response was categorized as "undecided" because the participant was unsure if

another activity would be needed for testing. Three subjects did not provide

responses.

4.2.3 Activity List Completeness

The next Scheduling Survey question asked subjects if they believed any

activities were missing from the activity list provided. Of the 70 responses, 28

responded that they believed the activity list provided needed to be expanded. Most

subjects who responded "yes" stated that only 1 to 3 activities needed to be added,

although some requested 4 or more. Additional close-out activities and reviews,

along with activities required to close out those reviews were mentioned several times

as extra activities that needed to be accounted for in the activity list.   One participant

stated that they were anticipating additional requirements while another stated that

there was work to be done that, "was not explicitly called out in the schedule". This

participant stated that the work was probably covered under activities that were listed,

but that extra time was added to those activities to account for those not specifically

called out. Another participant stated that there were activities that were not

essential, but that would assist the project team.

Three subjects were listed in the "qualified yes" category, with one participant

listing roughly 15 activities to be added, but upon closer inspection, a case could be

made that these activities were expansions of overarching activities already listed in

the survey. Another participant factored in extra time for unaccounted for activities

and also seemed to suggest that additional activities were recommended, but were not

explicitly added. The third "qualified yes" participant replaced one activity with

another, removing an erroneously duplicated activity in the original survey and

replacing it with a new activity which had been left out.

Twenty-five subjects responded that they believed the list provided was

adequate. Nine subjects provided answers that were categorized as a "qualified no".

Some subjects listed activities that needed to be completed, but they were either

associated with a different project or out of the scope of the project phase covered by

the survey. One participant added activities that had previously been completed.

Another participant listed management activities, but stated these would not cause

additions to the schedule. One participant in this category listed unknowns that could

potentially increase the level of activities and another commented that there were

complications to listed activities, but that no new activities needed to be added. Five

subjects did not provide a response. Of those five, one participant's response is

unknown. In the raw data consolidation spreadsheet, the response says "comment" to

signify further clarification written elsewhere beyond a "yes"/"no" answer, but the

comment could not be located.

To clarify the number of responses, in some cases, projects could be broken

up into several different independent sections that were all needed to successfully

achieve project success, but were not necessarily dependent on one another. In these

cases, subjects responsible for managing multiple sections were given one survey

with all of the different sections, while those responsible for execution of the project

were given only sections applicable to their assignments. In the results, each of these

different sections were broken up and treated as separate projects. If a manager

responded "no" to the overall survey on any of the questions, it was assumed that the

answer applied to each of the different sections of that survey. The manager's

response, therefore, is counted independently for each section. For example, if a

project had activities for Team A, Team B, and Team C, each member within that

team would receive a survey specific to that team, but the manager would receive a

larger survey with activities for all three teams. If the manager then responded "no"

157

to any of the questions just described, that "no" response would be tallied in each of the individual surveys of Team A, Team B, and Team C.

4.2.4 Summarizing the Results

The results from the first question, regarding whether or not more team members were needed, indicate that the number of assigned project personnel is not the major concern of project teams. Factoring in both the "no" and "qualified no" responses, 80% of respondents believed the number of personnel was adequate. Of the approximately 13% of those who said that more people needed to be added, all factor levels of all demographics were represented with the exception of the 24+ YoE factor and the High School LoE factor. What does seem to be a major concern, however, is allowing personnel who are assigned to a project to focus primarily on that project. The implication here is that if personnel who are assigned to multiple projects are allowed to focus entirely on one project, then extra personnel will be needed to backfill the other projects, or those other projects must resign themselves to a delayed schedule until personnel are again available.

For the next question which asked whether or not all activities on the list needed to be completed, accounting for both the "yes" and "qualified yes" answers, 74% of the subjects agreed that all activities on the list were required for successful project completion. Of the 20% ("no" and "qualified "no") that believed activities could be removed, all factor levels of the three demographics were represented. Based on the data collected, it would appear that most stakeholders, regardless of demographic, are in reasonable agreement when presented with a list of activities for a project. These results show slightly less agreement among stakeholders than was

seen on the first question, but it does appear that disagreement on proposed activities is not driving the disagreements regarding schedule duration. This, however, is only one half of the coin.

The second question discussed whether or not stakeholders believed listed activities needed to be completed. The final question asked whether or not any activities were left off of the task list. This question seemed to be the point of most disagreement among stakeholders. Factoring in both the "no" and "qualified no" answers, approximately 48.5% of respondents believed the provided list was adequate and additional activities were not needed. On the other hand, factoring in both the "yes"/"qualified yes" responses, approximately 44% of respondents believed additional activities were needed to successfully complete the project. All factor levels of all demographics were covered in both categories. Of the three questions just discussed, this question represents the most likely driver behind differing duration estimates. While it may not be a significant contributor, if one stakeholder's assumptions regarding required activities differ from another's within the same project, it could cause disagreements on how long the project should take, especially if those assumptions are never discussed and one set of assumptions is driving the schedule.

Chapter 5 Results – Priorities, Personalities, and Predictions

The previous chapter described the opinions of several subjects regarding why they believe projects struggle to finish on time. It also described opinions regarding the activity lists from the Scheduling survey and whether or not both the activity list and assigned resources were adequate. This chapter covers the remaining survey questions and investigates how the three demographics chosen for study, Position, Years of Experience (YoE), and Level of Formal Education (LoE), relate to personality traits such as confidence and risk aversion. It also investigates how these demographics relate to schedule duration estimating practices.

The DesignExpert™ software was used to set up factorial experiments which used ANOVA to determine which, if any, of the demographic factors were driving the results. The software was also used to determine the expected response of a stakeholder with a given set of demographics based on the results seen in this study. Correlations between personality traits and estimating practices were also examined. Seventy subjects were contacted regarding participation in this study. Of those 70, 45 signed the consent form and agreed to participate. Throughout the different surveys, the total number of respondents differed because not all subjects responded to the surveys and of those who did respond, not all subjects answered all questions. The total number of responses for each part of each survey is listed in the sections below.

5.1 "Course of Action" Survey: Is it really necessary?

Subjects were given a "Course of Action" COA survey asking what they would do given a situation where equipment was barely within specifications, but repairing it would cause a schedule delay (see Appendix A.5). The purpose of this question was to determine whether or not managers and technicians perceived the criticality of an activity differently from one another. "Gold Plating" is defined as unnecessarily going above and beyond stated requirements.

It was hypothesized that one possible cause of scheduling disagreements was due to differing perceptions of what constituted "necessary" work. This survey received a total of 27 responses, 11 from those identifying as "management" and 16 from those identifying as "technical". The breakdown of responses from management subjects and technical subjects can be seen below in Table 5-1 and Table 5-2, respectively. The rows describe the action recommended by the participant and the columns describe whether or not the participant believed the extra effort was truly necessary. For example, in Table 5-1 five subjects recommended taking an extra week to bring the equipment up to full operating specification, as opposed to leaving it in its "barely operational" status. They considered this work a necessary action to mitigate the risk of system failure. In contrast, four subjects believed the equipment should be left alone, believing that any troubleshooting efforts constituted unnecessary work (i.e. "gold plating").

Management	Risk Mitigation	Gold Plating
Take extra week to fix	6	0
Leave "As Is"	1	4

Table 5-1: Management COA Response

Technician	Risk Mitigation	Gold Plating
Take extra week to fix	8	0
Leave "As Is"	6	2

Table 5-2: Technician COA Response

When the survey was initially provided, it was believed that the subjects would respond in one of two ways: take the extra week to fix the problem as a risk mitigation strategy or leave the system "as is" because further work would be unnecessarily going beyond the required work. As can be seen from Table 5-1 and Table 5-2, a third option was also selected: the subjects stated the belief that extra time spent working on the system would constitute risk mitigation, but chose to leave the system "as is". This particular selection was favored more by the technicians than the managers.

Knowing that technicians are primarily responsible for ensuring the equipment is functioning, it is interesting that some technicians, believing that an extra week would help mitigate a potential risk, would still forgo that extra week, thereby allowing the project to meet its schedule. This would seem to contradict results of an experiment described later in this chapter regarding whether or not

schedule should be sacrificed for the sake of quality. It should also be noted that the

split between risk mitigation and gold plating for the managers was 64/36 while the

split on the technician side was 87.5/12.5. This indicates that while managers

somewhat disagree about what constitutes a necessary fix, technicians are more

united in their opinions. Although the technicians are in better agreement than the

managers regarding opinions about the necessity of the work, both groups are nearly

evenly divided as to whether or not to take the extra week to repair the system or

leave it as is, with the managers showing a very slight preference to take the extra

time to fix the system.

5.2 Traits/Opinions Results

From Section 5.1, it was seen that there are differences in the way managers

and technicians perceive what constitutes "necessary" work. This section expands on

that line of inquiry. The Traits/Opinions survey organized subjects by the

demographics of Position, YoE, and LoE. Beyond this basic categorization, this

survey also gathered information about each participant's level of risk aversion and

also their preferences for what project constraint to sacrifice first when things go

wrong. The Scheduling survey collected estimates from subjects on activity durations

across several different projects. The results of both surveys were then compared to

the demographic results to determine whether or not stakeholders in different

demographics respond differently from one another. Correlations between the

personality traits of risk aversion/confidence and schedule estimates were then

calculated. For example, did a lower risk tolerance correlate with a wider standard

deviation in the schedule estimate? Or did rating schedule slips as a low priority

correlate to higher estimates in the scheduling surveys?

5.2.1 Constraints Analysis – by Constraint – The Results

After gathering basic demographic information, the Traits/Opinions survey

asked subjects to rate which project constraint they would sacrifice first should a

project start to falter. Because it can be difficult to assign a quantitative number to a

preference, this method gauged whether or not subjects treated each project constraint

equally. If they did, for each subject, the calculated weights for each constraint

would be equal. As can be seen in Table 5-3, however, this is not the case. Thirty-six

subjects responded to this survey. After gathering the initial constraint rankings as

described in Section 3.2.3 from the data collected from the survey in Appendix A.2, a

weight was calculated for each constraint for each subject using Equation 3-1 (see

Appendix A.8 for each subject's individual constraint weights). A higher weight

indicates more willingness to fail at meeting that constraint for the sake of

successfully meeting the others. For each constraint, the weights provided by the 36

subjects and the results are provided in Table 5-3.

Constraint	**μ**
Schedule	0.40
Cost	0.35
Risk	0.15
Quality	0.10

Table 5-3: Average weight per constraint

From the table, it can be seen that the average weights for each constraint are

not equal and that the average weight for the Schedule and Cost constraints are higher

than the average weight for the Risk and Quality constraints. This indicates more

164

willingness to sacrifice cost and schedule for the sake of minimizing project risk or

decreasing project quality that there is a clear preference for increasing cost and

schedule before decreasing quality or increasing risk.

As described in Section 3.3.1, the 28 responses per constraint and Excel™ "t-

test with unequal variances" function were used to determine whether or not the

differences seen in the averages shown in Table 5-3 were statistically significant,

where significance is defined as $p<0.05$, H_0: $\mu_i=\mu_j$, and "i" and "j" are the two

constraints being compared as seen in Table 5-4.

Constraints Compared	Difference in Average Significant?	P-value
i = Schedule; j = Cost	No	0.087
i = Schedule; j = Quality	Yes	2.63E-13
i = Schedule; j = Risk	Yes	4.63E-10
i = Cost; j = Quality	Yes	9.82E-12
i = Cost; j = Risk	Yes	3.86E-08
i = Quality; j = Risk	No	0.062

Table 5-4: Statistical Significance of Weight Differences

Based on the p-values in the last column of Table 5-4, the null hypothesis

(equal constraint averages) can be rejected when the either the Cost or Schedule

constraint is compared to either the Risk or Quality constraint. It cannot be rejected,

however, when the Cost and Schedule constraints are compared or when the Quality

and Risk are compared.

From these results, it can be inferred that quality is the most important

constraint, followed by risk (defined as minimizing the risk of project failure), then

cost, then schedule. This matches what was found in the GAO reports: technical

success is the key indicator of project success (Martin 2012, vi). Project concerns

such as cost and schedule increases will be forgotten as long as there is technical

success and no one got hurt. Based on these responses, if problems arise, the schedule will take a hit to ensure the overall technical quality is maximized.

There are some issues concerning the returned data that should be considered when interpreting the results. First, as described in Section 3.2.3, the weights among each preference should have a consistency rating of less than 0.1 in order to be considered valid. Out of all of the subjects who responded to the survey, only 32% (12/36) were below the 0.1 threshold for consistency. One participant ranked all preferences equally, thus resulting in perfect consistency, but providing little insight to the actual preference. This response was removed from the data set and was not included in the analysis. Another 32% (12/36) of the subjects exhibited slightly inconsistent behavior, with their consistencies falling above the 0.1 threshold, but below 0.2. The remaining 32% of subjects (12/36) were very inconsistent among their preferences. Having said that, this exercise was only a gauge meant to provide insight into how employees at WFF view the importance of meeting different project constraints.

Another more serious issue resulted from a potential misunderstanding of what was asked in the survey. Some subjects regarded the rankings as a sliding scale where a "1" meant Constraint A was preferred over Constraint B, a "9" Constraint B was preferred over Constraint A, and a "5" meant there was no strong preference either way. When it was obvious this mistake was made (usually because the participant only provide a numerical ranking with no constraint associated with it), the participant was asked to resupply answers using the correct ranking system. The

166

error may not have been as obvious, however, if some subjects listed both a constraint and a numerical ranking.

Another potential misunderstanding revolved around what exactly was meant by the project constraint "risk". When the survey was originally developed, risk was listed as one of the project constraints, but the meaning of the term was not clearly defined. The intent was a somewhat vague concept intended to depict the risk of project failure or risk to personal safety. Some subjects were unsure what was meant by risk in that it is usually tied to one of the other project constraints (e.g. risk of schedule increase, risk of cost increase, etc.). Given this ambiguity, some subjects may have understood the meaning behind "risk" differently from one another, which could have affected their rankings.

Finally, some subjects struggled with the fact that the question asked what the preferred constraint was "in general". The participant who rated everything equally said that it was impossible to pick unless project specifics were known (e.g. for some projects, schedule is very important, for others cost is very important), so it is impossible to know what to sacrifice without knowing the nature of the project.

5.2.2 Constraints Analysis – by Demographic – The Results

Section 5.2.1 provided the results of the constraint preferences across all participant responses. This section looks at each constraint individually to determine whether or not one of the three demographics (Position, YoE, or LoE) is a significant factor driving the differences seen in the responses. If subjects in a particular demographic regarded all project constraints as equally important, there would be no statistically significant difference among the average weight for different levels of the

167

demographic for that particular constraint. For these results, statistical significance

was defined as $p<0.1$ where $H_0: \mu_1 = \mu_{2 = \ldots =} \mu_n$, where "n" is the number of factors

under consideration (Montgomery 2008, 70–71). Table 5-5 consolidates the results.

Constraint	**Significant Factors?**	**Factor P-value**
Cost	None	N/A
Schedule	Position	0.0706
Quality*	Level of Formal Education	0.0621
Risk*	None	N/A

*Inverse Square Root Transformation

Table 5-5: Significant Factors per Constraint

Based on the data collected, Table 5-6 shows the expected weight for the

Schedule constraint for stakeholders in the two Position categories. It also shows the

expected weight for the Quality constraint for stakeholders in the four LoE categories.

	Management		Technical	
Schedule	0.33		0.43	
	Masters	Bachelors	Tech/Associates	High School
Quality	0.147	0.069	0.07	0.098

Table 5-6: Expected weights per factor level

From these results, it can be expected that a technician will be more willing to

sacrifice schedule than a manager, as indicated by the larger expected value of the

weight.   For the Quality constraint, it can be expected that those with a Master's

degree will be more ready to sacrifice quality, with the remaining three categories

significantly less willing.

5.2.3 Utility/Risk Tolerance – The Results

Sections 3.2.2 and 3.3.5 described the process for obtaining the risk tolerance

of each participant and determining which (if any) of the three demographics under

study drove the response. To summarize that procedure, in the survey, subjects were asked to provide the monetary value required to trade in their chance of winning $5000 for cash-in-hand. Each participant was asked for that value against 5 different probabilities of winning. There were 38 total responses for this question.

When plotting the results of this survey, the plots showed that most subjects did not exhibit risk-averse behavior across the entire spectrum of probabilities as would be indicated by a curve that was concave at all points (Raiffa 1968, 68). This is not entirely unexpected, however, as previous research has demonstrated, utility curves that are both concave and convex are actually very common and reflect changing preferences as the risk/rewards ratio varies (Raiffa 1968, 8–9, 94–95).

Some subjects provided responses that resulted in an extreme curvature which may be tied to the way the question was asked. The questions in the survey mentioned only the possibility of winning $5000 and did not conclude with the statement, "…or of walking away with nothing." It is believed some subjects may have anchored on the possibility of winning the full $5000 total, meaning that anything less than that total would be considered a loss. From Chapter 2, it was shown that most people will focus more on a possible loss than a potential gain (Kahneman 2011, 119, 281–84). In this case, turning in the ticket represented a loss of the difference between the initial $5000 and the trade in value. In some cases the prospect of that loss appeared to cause the subjects to demand a trade value greater than utility. Given the questionable nature of the resulting curves and the issue described with the question itself, the results were simplified as described in the next

paragraph. The complete responses for each participant can be found in Appendix A.7.

The results below reflect only three points: (0,0), (X, 0.5), and (5000, 1). These points corresponded to the minimum monetary trade value and the minimum probability of winning, the participant's monetary trade in value at the 50% chance point, where X is the participant's monetary trade value, and the maximum possible monetary trade value and the maximum probability of winning. These simplified curves are shown in Figure 5-1 through Figure 5-5. In these figures, the top two graphs show the utility curves described above. Each point represents a response from the participants in the demographic category as described in the chart title. The bottom two graphs are a histogram of the frequency of a particular response.

Figure 5-1 describes the behavior along the management/technical divide, Figure 5-2 and Figure 5-3 describe the responses among the different ranges of the YoE demographic, and Figure 5-4 and Figure 5-5 describe the responses among the LoE demographic. Utility curves were created using the MatLab™ "fit" command with the 'power1' fit option with one exception. One participant responded that the trade-in value at 0.5 percent chance of winning was the full $5000 offered. An acceptable fit curve was not found to match this data, so that plot simply connects the three points [0,0], [5000,0.5], [5000,1]. These results can be seen in the top-left of Figure 5-1, the top-right of Figure 5-2, and the top-right of Figure 5-4.

170
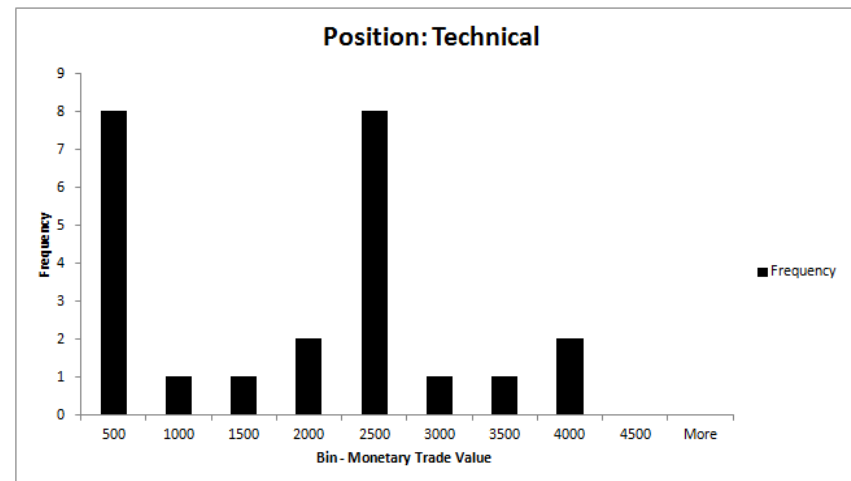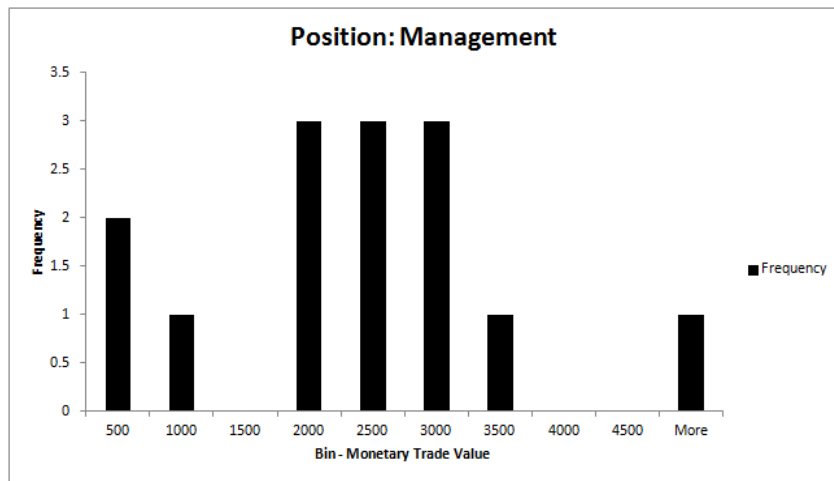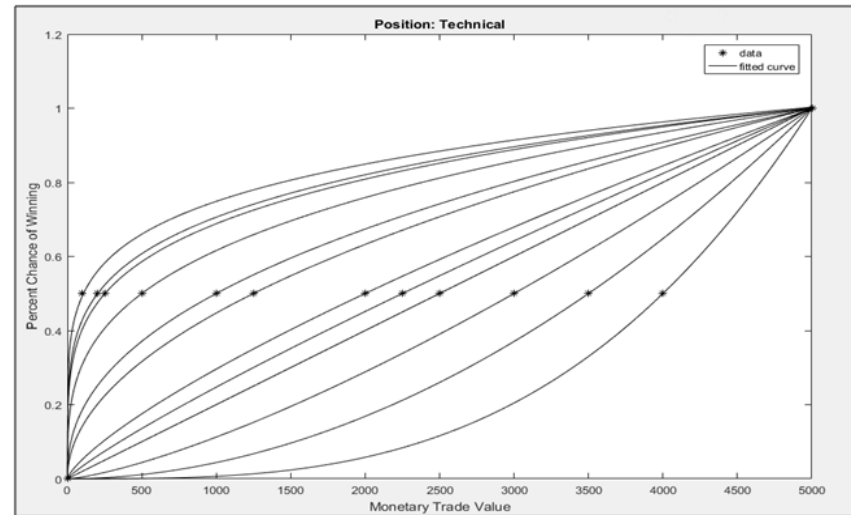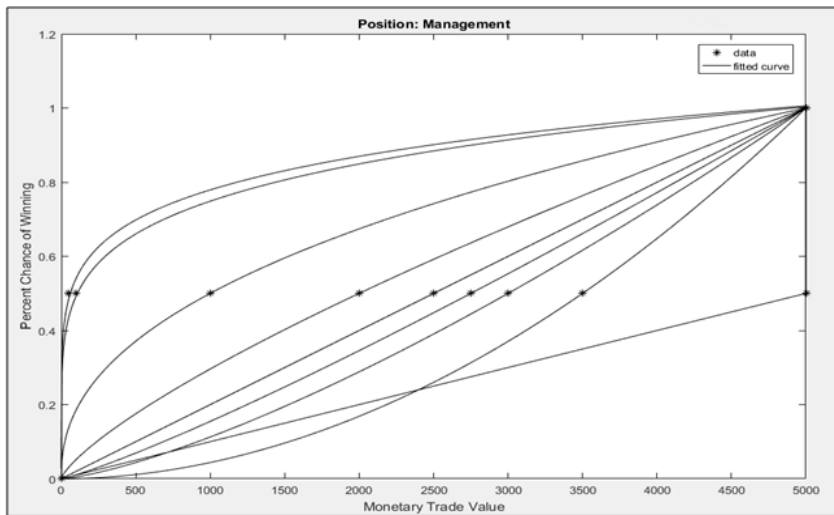
Figure 5-1: Utility Curve – "Position" Demographic
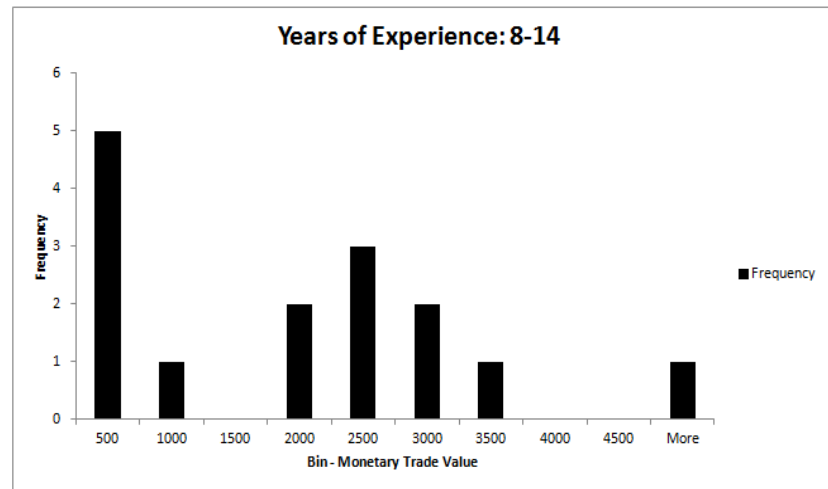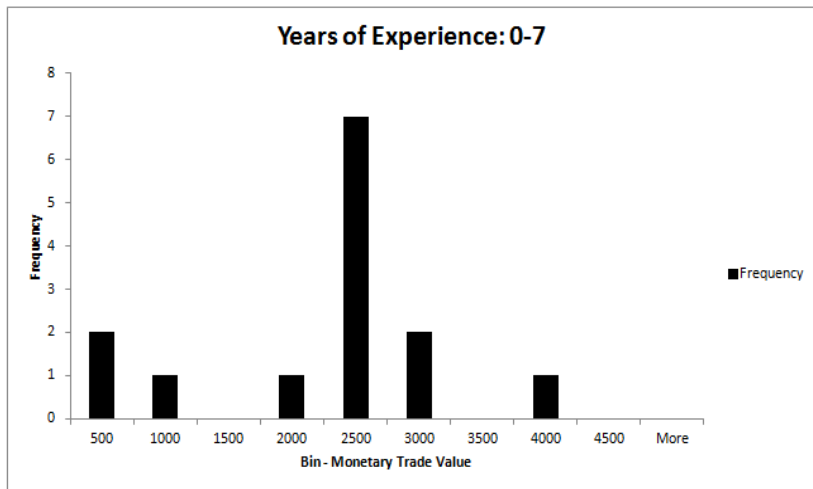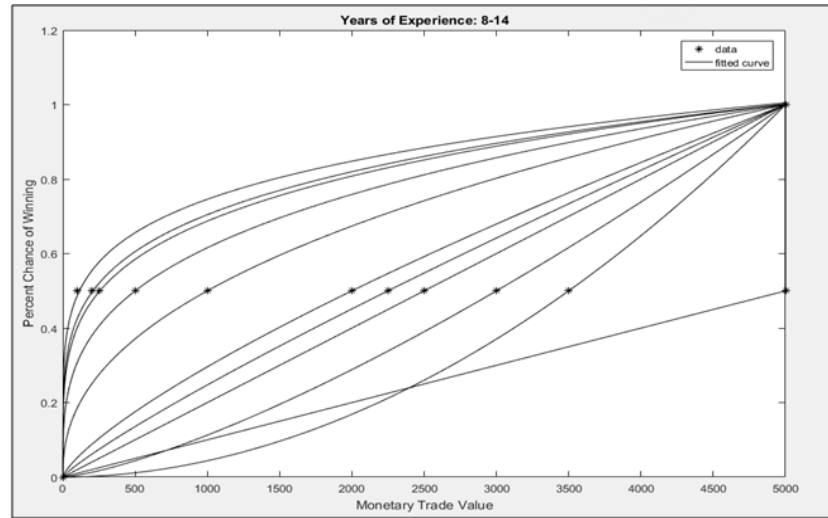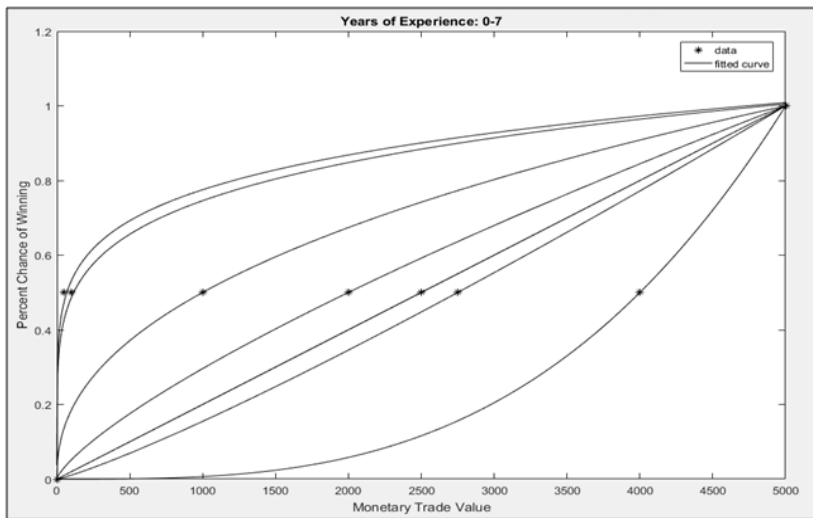
Figure 5-2: Utility Curve – "Years of Experience" Demographic
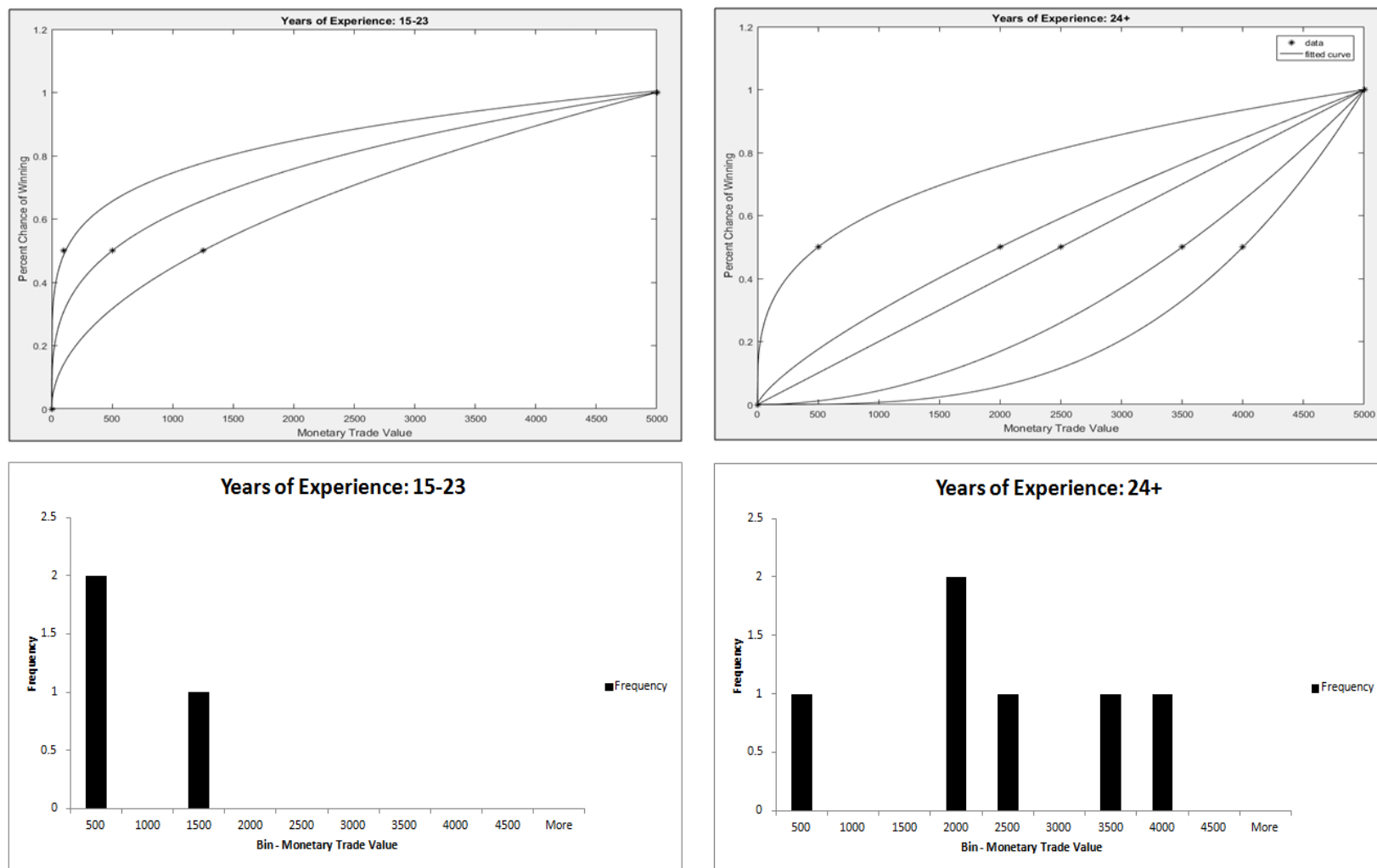
Figure 5-3: Utility Curve – "Years of Experience" Demographic (continued)
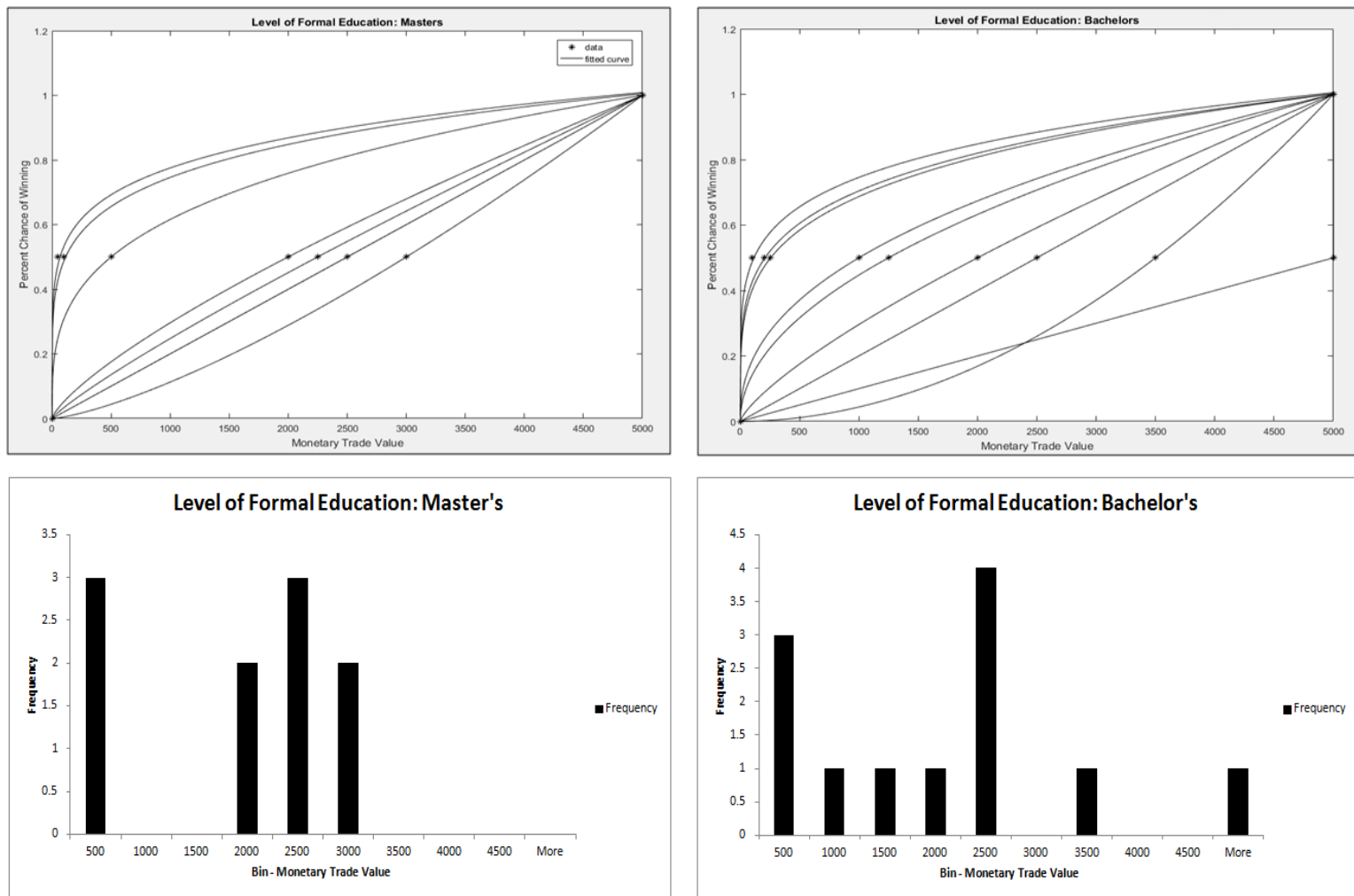
Figure 5-4: Utility Curve – "Level of Formal Education" Demographic

174
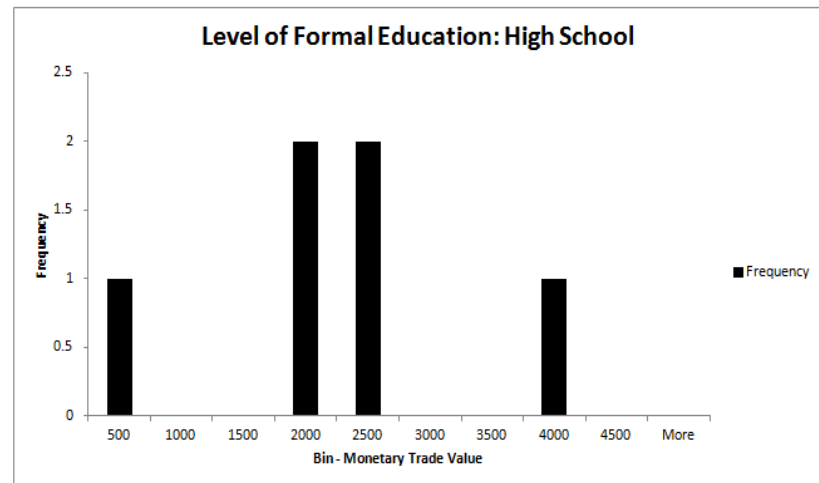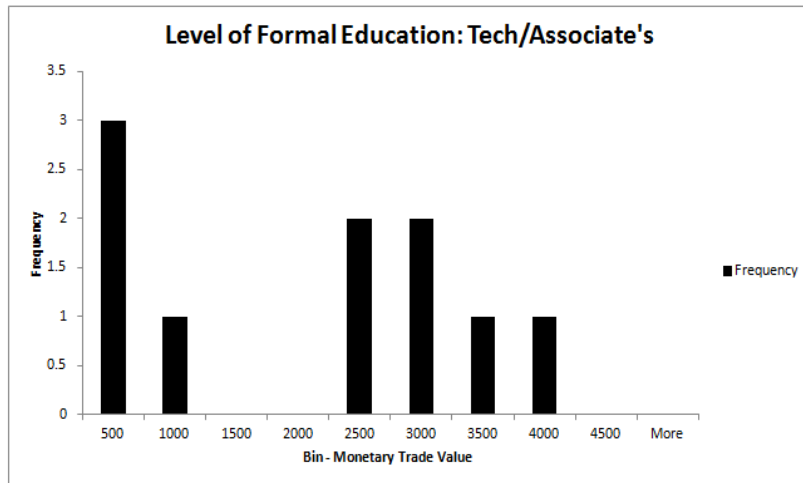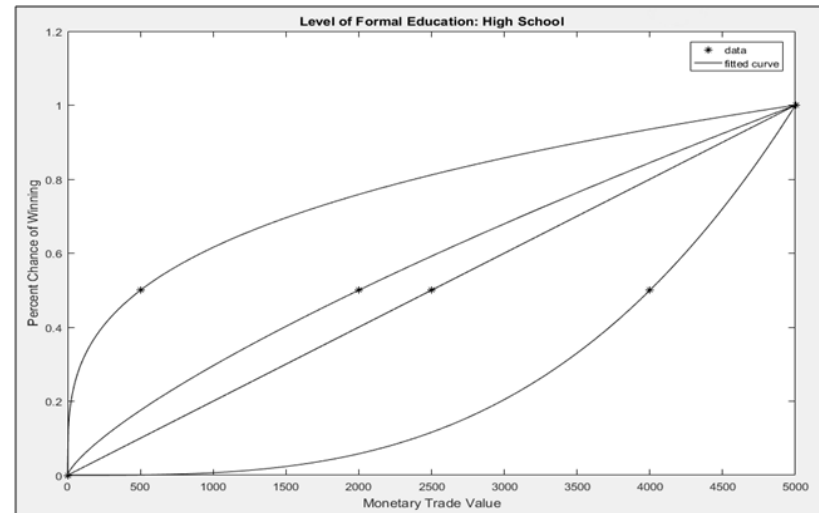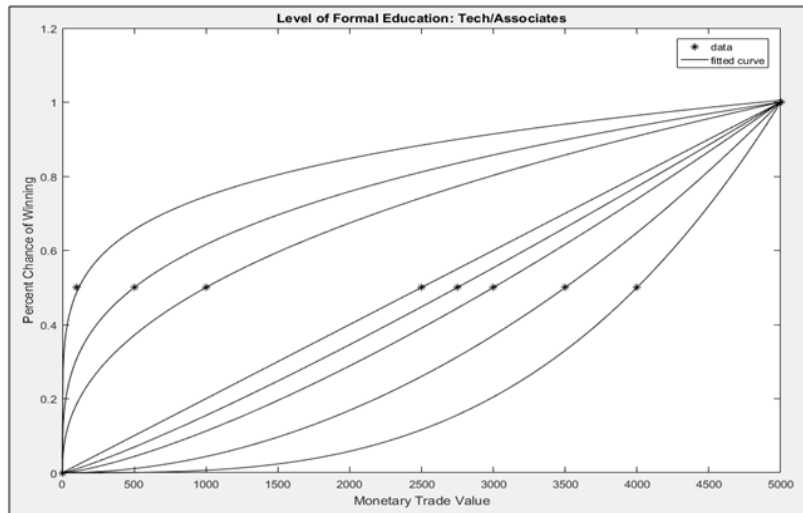
Figure 5-5: Utility Curve – "Level of Formal Education" Demographic (continued)

While differences can be seen in the responses among the different demographics, the model was not statistically significant and of the three demographics, there were no significant factors driving the results. The effect of risk aversion on estimating practices is provided in Section 5.3.3and discussed further in Section 7.2.4.

5.2.4 Confidence Analysis – The Results

Based on the methods described in Section 3.3.6, the model describing the confidence estimates was significant at a p-value of 0.02 and the YoE factor was the significant factor driving the participant responses (p<0.02). There were 26 data points used in the analysis. Table 5-7 provides the expected confidence level for subjects at each of the YoE factor levels:

Factor Level	**Expected Response**
0-7	0.725
8-14	0.767
15-23	0.858
24+	0.891

Table 5-7: Expected Confidence Values

These results indicate that stakeholders gain confidence in their estimates as they progress through their careers. Based on these results, it would appear that experience is driving confidence when it comes to schedule estimation. A discussion on the meaning of "confidence" as applies to a single value will be provided in Chapter 7.

5.3 Scheduling Results

The previous section provided results on stakeholder project constraint
priorities and stakeholder personality traits. The following section provides the
results of the Scheduling surveys and analyzes those results in the light of the
demographics of the respondents.

5.3.1 Network Path Standard Deviation Results

For each participant response within a particular project, a total project
duration (T_e) was calculated by summing the PERT average duration for each activity
in the project. Within each project, the standard deviation among the T_e value for
each participant was calculated. The resulting standard deviations are provided below
in Figure 5-6. It was hypothesized that if stakeholders agreed about the total duration
of a project, within that project, the standard deviation of the total time to completion
should be zero.
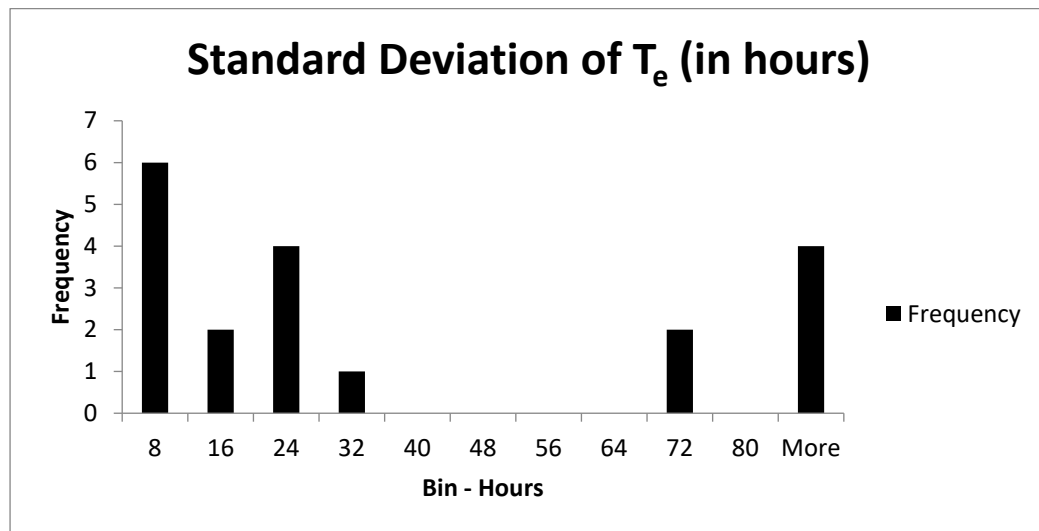
Figure 5-6: Standard Deviation of T_e

177

From this chart, it can be seen that for the 19 projects used in the analysis, the standard deviation in total duration for six projects was less than eight hours (i.e. a standard work day). On the other end of the chart, the standard deviation was several days, with the most extreme deviation being over a month. For projects below the "80 hours" bin, total project duration did not seem to be a driving factor of the standard deviation (i.e. longer projects did not necessarily always have a larger standard deviation). For projects above the "80 hours" bin, there did appear to be a correlation between project length and standard deviation, but the correlation was not linear and did not hold for all projects. The gap in the middle is reflective of the estimated durations of the individual activities. For projects to the left of the gap, most activities were estimated to take ten hours or less. On the opposite end, activity estimates are much higher, especially the "worst case" estimates. With more room to maneuver, subjects had a wider variety of opinions regarding how long things should take, resulting in a wider standard deviation.

These results show that there are differences in the estimates provided by stakeholders; otherwise the standard deviation within each project would be zero. While some of these differences are nearly insignificant (less than half of a standard work day), other differences are quite extreme. The question then becomes what is driving these differences in estimation.

5.3.2 Comparison Results

Surveys for several different types of projects were created and provided to subjects assigned to those projects who had agreed to participate in the study. Thirty-nine individual surveys were created, along with eight "collective" surveys consisting

of compilations previously created surveys, where independent parts of the same project were combined (i.e. all parts had to be completed for the overall project to be successful, but the individual parts did not interact with one another). These "collective" surveys were provided to management subjects who were responsible for more than one aspect of a project, but contained the same activity lists as the individual surveys. When analyzing the data, responses from the "collective" surveys were broken out and associated with their original individual survey.

One survey was left out of the final analysis because the responses of the subjects were so disparate in their format, it would have been extremely challenging to accurately compare them without making several assumptions. Taking into account the information just provided, of the 39 surveys created, usable responses were received from at least one participant on 30 of the surveys. In the raw data provided in Appendix A.9, it should be noted that there are thirty-five projects listed with their associated estimates. Five of these projects are "dummy projects" and are being used to help further mask the subjects. Data from these projects was not used during the analysis process.

Out of the 45 subjects who agreed to participate in the study, 31 provided responses to the scheduling surveys. Several subjects provided responses to more than one survey. Appendix A.9 provides a summary of each of the responses of each participant on each survey and provides a description of how to read the consolidated data.

The data from Appendix A.9 (and some from Appendices A.7 and A.8) were organized using the questions listed in Table 3-5. The results of the analysis for each

179

demographic are listed below in Table 5-8. The first column indicates the question

number from Table 3-5. The second column indicates the number of successes (i.e.

the total number of "Yes" answers for that particular question). The third column

indicates the total number of trials. The fourth column indicates the sample

probability of success and is calculated by dividing the first column by the second

column (R Core Team 2014).   For example, in the management demographic, for

question #1, the 0.88 value indicates that 88% of the time a management participant

provided a T_e value higher than a technician. The fifth column indicates the

alternative hypothesis for each question, where the alternative hypothesis is either that

the true population success rate is greater than 50% or less than 50%, depending on

the results of the sample success rate. The sixth column indicates the p-value for each

binomial test and the seventh column indicates whether or not the results are

statistically significant (p<0.05).

Q#	# of successes	Total Tests	Sample success rate	Alternative Hypothesis	p-value	Statistically Significant?
Position Demographic						
1	21	26	0.81	$\mu1 > 0.5$	0.001247	Yes
2	447	602	0.74	$\mu1 > 0.5$	$<2.2 \times 10^{-16}$	Yes
3	106	217	0.49	$\mu1 < 0.5$	0.393	No
4	136	217	0.63	$\mu1 > 0.5$	0.0001	Yes
5	55	217	0.25	$\mu1 < 0.5$	9.92×10^{-14}	Yes
6	73	305	0.24	$\mu1 < 0.5$	$<2.2 \times 10^{-16}$	Yes
7	65	305	0.21	$\mu1 < 0.5$	$<2.2 \times 10^{-16}$	Yes
8	17	33	0.52	$\mu1 > 0.5$	0.5	No
9	200	305	0.66	$\mu1 > 0.5$	2.897×10^{-8}	Yes
10	62	152	0.41	$\mu1 < 0.5$	0.014	Yes
11	23	28	0.82	$\mu1 > 0.5$	0.0004	Yes
Years of Experience Demographic						
1	20	40	0.50	$\mu1 > 0.5$	0.5627	No
2	447	602	0.74	$\mu1 > 0.5$	$<2.2 \times 10^{-16}$	Yes
3	179	367	0.49	$\mu1 < 0.5$	0.3381	No

Q#	# of successes	Total Tests	Sample success rate	Alternative Hypothesis	p-value	Statistically Significant?
Years of Experience Demographic (cont.)						
4	165	367	0.45	$\mu1 < 0.5$	0.0300	Yes
5	147	367	0.40	$\mu1 < 0.5$	8.171×10^{-5}	Yes
6	183	482	0.38	$\mu1 < 0.5$	7.103×10^{-8}	Yes
7	191	482	0.40	$\mu1 < 0.5$	3.025×10^{-6}	Yes
8	25	49	0.51	$\mu1 > 0.5$	0.5	No
9	244	482	0.51	$\mu1 > 0.5$	0.4099	No
10	71	241	0.29	$\mu1 < 0.5$	7.61×10^{-11}	Yes
11	35	40	0.88	$\mu1 > 0.5$	6.913×10^{-7}	Yes
Level of Formal Education Demographic						
1	25	41	0.61	$\mu1 > 0.5$	0.1055	No
2	447	602	0.74	$\mu1 > 0.5$	$<2.2 \times 10^{-16}$	Yes
3	161	361	0.45	$\mu1 < 0.5$	0.02268	Yes
4	185	361	0.51	$\mu1 > 0.5$	0.3969	No
5	153	361	0.42	$\mu1 < 0.5$	0.0022	Yes
6	158	416	0.38	$\mu1 < 0.5$	5.405×10^{-7}	Yes
7	158	416	0.38	$\mu1 < 0.5$	5.405×10^{-7}	Yes
8	18	45	0.40	$\mu1 < 0.5$	0.1163	No
9	220	418	0.53	$\mu1 > 0.5$	0.1522	No
10	106	264	0.40	$\mu1 < 0.5$	0.0008	Yes
11	28	37	0.76	$\mu1 > 0.5$	0.001282	Yes

Table 5-8: Binomial Analysis by Demographic

For the results focused solely on the schedule estimates (Questions 1-7,9), the Position demographic has the smallest probability of occurring by chance (i.e. smallest p-value) with one exception (Question #2 is the same for each demographic and is therefore discounted): for Question #3, which looks at the separation between the ML and BC values, the number of successes in the LoE demographic is statistically significant while the others are not.

The results for the remaining questions (Question 8,10-11) were not as clear-cut. These results were based on answers from the Traits/Opinions survey and the confidence estimates from the Scheduling Survey. These results show that there are no significant factors driving risk-aversion in stakeholders among the different

demographics. This general result matches with the results achieved using DOE, but the p-values seen using this binary yes/no method, do not correspond to the significance order seen in DOE. DOE showed that the LoE demographic was the factor with the least effect on the results provided by the subjects. These binary comparisons would indicate it is the largest (although still not statistically significant). With respect to confidence, the Years of Experience demographic produced the most significant results, which correlates with the results found using DOE with the actual estimates. When determining readiness to sacrifice the schedule for other project constraints, the Years of Experience demographic once again exhibited the most significant results. These results differ from those found using DOE where the Position demographic was determined to be the most significant factor driving the results.

Across all demographics, for Question #2, when examining the separation between the ML and BC estimates versus the separation between the ML and WC estimates, nearly 75% of subjects are allowing more time for things to go wrong than they hope that things will go right. This result correlates with the literature that states most people fear loss more than they appreciate gain (Kahneman 2011, 281–84). These results indicate that subjects were compensating more for unknowns by providing a larger WC estimate than they were assuming things would go well which would be indicated by a small BC estimate.

5.3.3 Correlation Results

One objective was to determine whether or not certain demographics exhibited traits and, if so, did those traits have an effect on project duration estimates.

182

The results described in previous sections of this chapter have discussed some

different characteristics such as risk aversion and project constraint preferences. This

section provides the results to the correlation questions described in Section 3.3.7. In

Table 5-9, the third column displays the data using all available data. The fourth

column displays the results if only projects with three or more subjects were included.

Correlation Question	Correlation Factors	Correlation Coefficient (all projects)	Correlation Coefficient (projects with 3 or more subjects)
QC1	Confidence and standard deviation negatively correlated	-0.30	N/A*
QC2	Confidence and Utility positively correlated	0.1	0.14
QC3	Utility values and standard deviation negatively correlated	0.15	-0.21
QC4	Utility and T_e negatively correlated	-0.04	-0.29
QC5	AHP and T_e positively correlated	0.02	-0.10

* In this case, a correlation coefficient was calculated for each participant within a project. These correlation coefficients were then averaged across all projects to provide the value shown. Because the coefficient was calculated per participant and not per project, there was no case where only two values were used in the correlation.

Table 5-9: Correlation Results

Based on the results of the chart above, using the results from the fourth

column where only projects with three or more subjects were considered, the

following conclusions were drawn as shown in Table 5-10:

Question #	Conclusion
QC1	There is a weak negative correlation between subjects having a larger standard deviation and a lower confidence. This indicates that project stakeholders who are less confident in their ML estimate will probably provide a wider range between their BC and WC values to compensate for that uncertainty.
QC2	There is a very weak positive correlation between confidence in the ML estimate and Utility. This would indicate that level of risk aversion does not significantly affect confidence levels regarding the

Question #	Conclusion
	ML estimate. This confidence level may be more driven by familiarity with the project as opposed to an overarching personality trait.
QC3	There is a weak negative correlation between Utility values and standard deviation. This would indicate that personnel who exhibit risk-averse behavior will manifest that behavior in a scheduling context by compensating for the unknown with a wider range of possible activity completion times. This wider range increases the probability that the actual time will fall somewhere within the provided estimate, thus mitigating the risk of failing to provide a good estimate.
QC4	There is a weak correlation between Utility and T_e. This would indicate that personnel who exhibit risk-averse behavior in general will manifest that behavior in a scheduling context by compensating for the unknown with a higher T_e. This behavior increases the chance that the final completion time will fall below the expected value as calculated by summing activity times using the PERT average.
QC5	There is a very weak correlation between willingness to sacrifice schedule (as measured by the AHP weight) and T_e. It also indicates that what little correlation exists is negative. This indicates that willingness to sacrifice schedule in the event of problems on the project does not significantly affect the initial T_e estimate. It also hints that those willing to sacrifice schedule first are providing smaller T_e estimates.

Table 5-10: Correlation Conclusions

5.3.4 Data Collection Challenges

When collecting the data, some challenges arose which may have affected the

estimates provided. Every attempt was made to gather inputs prior to the beginning

of project execution such that the estimates provided were true estimations and not

after-the-fact reconstructions of the actual events. Some subjects provided estimates

a day or two after the project started (cells highlighted in gray in Appendix A.9), but

the estimates are included because these were all management subjects and it is

believed that very little information about activity completion had been reported by

184

the time the estimate was provided. In many cases, the projects under consideration had already been planned out by the time the estimates were gathered. Subjects were asked to provide estimates based on what they would recommend if they were completely in charge, but the already-planned timelines may have affected the estimates. Another factor which may have affected the outcome was the availability of resources on the project (i.e. percentage of time project personnel were allocated to work on the given project). Some project surveys stated that assets should be considered to be allocated at 100%, so these should not be an issue. For other projects, the assumption of allocated time was less than 100%. Some respondents may have provided an estimate based on the provided availability, but others for that same survey may have assumed 100% availability of all resources. These differences could have affected the total durations provided for each activity. In rare instances, subjects provided BC estimates that were larger than the ML estimates. These activity estimates were removed from the data set. If a PERT duration could not be calculated for any activity in a project, that activity was removed from the T_e summation for all subjects in an effort to standardize the number of activities used in the summation.

5.4 Predicting T_e

The previous section discussed how demographics can affect personality traits and even how one estimates activity durations. This section compares the results of the duration estimates based on the demographics of the subjects who provided the estimates. This was done to not only determine which demographic drives the

185

response, but to also predict the future estimates of stakeholders belonging to that particular demographic. If a project manager can only get one estimate from a stakeholder, the results of this section will allow her to calculate the remaining two estimates needed to determine a PERT average, as long as she knows which demographic category the stakeholder belongs to.

5.4.1 Worst-Case Estimate as Related to Most Likely

The first part of Section 3.4.1 described the method for studying the skew of a participant's prior distribution, assuming a PERT beta model. If subjects accounted equally for things going well and things going poorly, the distribution would have no skew, indicating that the separation between the ML and WC estimates should have been equal to the separation between the ML and BC estimates. With the BC and WC estimates being equidistant from the ML estimate, performing Equation 3-5 on the estimates should result in a value of 0.5 because the numerator will always be half of the denominator. If the result of Equation 3-5 was less than 0.5, it would indicate a positive skew, where the smaller the value, the larger the skew. A value greater than 0.5 indicates a negative skew, where a larger indicates a larger skew. After performing Equation 3-5 on the estimates from each participant, consolidating, and analyzing the data, it was determined that the significant factor affecting these results was the Position demographic ($p < 0.015$). It was also seen that, based on the calculated expected value, both managers and technicians exhibited some level of positive skew in their estimates. The predicted results for the Position demographic are listed below in Table 5-11.

Factor Level	$\dfrac{(ML-BC)}{(ML-BC)+(WC-ML)}$	$1-\left(\dfrac{(ML-BC)}{((ML-BC)+(WC-ML)}\right)$
Management	0.3882	0.6118
Technician	0.293	0.707

Table 5-11: Separation Weight Ratio

These results show that management subjects exhibit a roughly 40/60 split in the separation between the ML and BC estimates and the ML and WC estimates, respectively. Technicians, on the other hand, exhibit a roughly 30/70 split. This would indicate that, in general, technicians have a higher positive skew than their manager counterparts and are compensating for what could go wrong more than the managers.

This result tells only part of the story, however, since many different estimates could combine to produce these ratios. To narrow down the possible values, Equation 3-6 and Equation 3-7 were applied to each participant's estimates, consolidated, and analyzed. Once again, it was shown that the "Position" demographic was the significant factor driving the results which are summarized below in Table 5-12.

Result	Model Significant?	P-value	Significant Factor	P-value
BC/(ML+BC)	No	0.194	N/A	N/A
WC/(ML+WC)	Yes	0.048	Position	0.048

Table 5-12: Outlier Weight Significant Factors

Given that the "WC" ratio was significant, the expected response for the two Position demographic levels are listed below in Table 5-13.

Demographic	WC/(ML+WC) =
Management	0.5577
Technician	0.6315

Table 5-13: Outlier Weight Ratio

5.4.2 Expanding the Results – T_e Assessment

The results from Table 5-12 and Table 5-13 and Equation 3-5, Equation 3-6 and Equation 3-7 resulted in the system of simultaneous equations listed below for the management demographic:

$$\frac{WC}{(ML+WC)} = 0.5577 \qquad\qquad \text{Eqn 5-1}$$

$$\frac{(ML-BC)}{(ML-BC)+(WC-ML)} = 0.3882 \qquad\qquad \text{Eqn 5-2}$$

Solving the first equation for the WC value in terms of the ML value, and then substituting that term for the WC value in the second equation, provided the following results for the management case:

$$WC = 1.2609*(ML) \qquad\qquad \text{Eqn 5-3}$$

$$BC = 0.8345*(ML) \qquad\qquad \text{Eqn 5-4}$$

For the technicians, the following equations were used:

$$\frac{WC}{(ML+WC)} = 0.6315 \qquad\qquad \text{Eqn 5-5}$$

$$\frac{(ML-BC)}{(ML-BC)+(WC-ML)} = 0.293 \qquad\qquad \text{Eqn 5-6}$$

This resulted in the following values for the technician demographic:

$$WC = 1.7137*(ML) \qquad\qquad \text{Eqn 5-7}$$

$$BC = 0.7042*(ML) \qquad\qquad \text{Eqn 5-8}$$

From these equations, without knowing anything except the demographic of the stakeholder and the ML estimate of an activity from that stakeholder, it is possible to make a reasonable assumption on the value of the BC and WC values. These values also provide some insight as to what in the estimate is driving the higher duration times. Based on these ratios, assuming both a technician and a manager provide the same ML value, the expected value of the activity duration as calculated by Equation 2-1 will always result in the technician having a higher estimate than the manager. These results also indicate that the expected standard deviation as calculated using Equation 2-2 will be larger for a technician than for a manager.

5.4.3 Duration Estimate Skew

The results from Section 5.4.2 were calculated based on the summation of each of the estimates for each project. From these results, at the project level, the relation of the BC, ML, and WC estimates typically resulted in a positive skew. When looking at each individual activity estimate and performing a simple comparison of the separation between the ML and BC values and the ML and WC values, the results tell a slightly different story. Using the equation (WC-ML) – (ML-BC) on each individual activity estimate, the following results were obtained:

The result of (WC-ML) – (ML-BC) is:	< 0 Negative Skew	= 0 No Skew	> 0 Positive Skew	**Total**
Total	50	105	447	602
Management	32	71	201	304
Technical	18	34	246	298

Table 5-14: Skew Results

From Table 5-14, it can be seen that across the board for all subjects, most subjects provided an estimate with a positive skew. It can also be seen, however, that several estimates had no skew at all. In some rare cases, some subjects even had a negative skew, indicating there was more uncertainty about the BC estimate than the WC estimate. When these numbers were broken down into the two Position demographic factors, it can be seen that the technicians heavily favored the positively skewed distributions. Managers also favor this distribution, but they are more likely to provide estimates resulting in either no skew or negative skew.

Chapter 6 Results – Aggregating the Estimates

The previous chapter described how stakeholders from different demographic categories respond differently from one another with respect to project constraint preferences and schedule estimating practices. It also showed how stakeholders differ in personality traits, even if there was not one specific demographic driving those differences. The correlation analysis showed that these different personality traits did appear to have some bearing, however slight, in how stakeholders estimated activity durations. Given these results and the decision analysis literature which shows how biases and perceptions can affect assessments of the unknown, this chapter describes a method which allows a project manager to aggregate all of the duration estimates provided by the team into one final estimate that can be used in a network schedule.

6.1 Determining the Prior

Section 3.5.1 described the method for converting the estimates provided by project team members into a probability distribution. In Bayesian statistics, this distribution is referred to as the "prior" because it is based solely on the individual's prior state of knowledge. This section will step through the method described in Section 3.5.1 using example estimates and compare those results to what would have been derived using the PERT beta distribution.

For this process example, it is assumed that a project manager, also known as the Decision Maker (DM) is developing a schedule and has asked for inputs from three other people: a fellow manager who has worked several similar projects in the past and two technicians who are currently assigned to the project. These three people will be referred to as Expert #1 (manager), Expert #2 (technician), and Expert #3 (technician). Table 6-1 shows the estimates (in hours) for one particular activity in this project and the resulting mean and standard deviation. In Table 6-1, k is the shape parameter of the GEV distribution (Max or Min) as calculated using Equation 3-20 and Equation 3-22, the location parameter is determined by the ML estimate, and Delta is the difference between the GEV CDF evaluated at the WC value and the GEV CDF evaluated at the BC value. For Expert #3, k is replaced by σ, the standard deviation of a Normal distribution, and Delta is the standard value of 3σ for a Normal distribution. The mean for the GEV Max and GEV Min case were calculated using Equation 3-31 and Equation 3-32, respectively, and the standard deviation was calculated by taking the square root of Equation 3-33.

	ML	BC	WC	Type	k or σ	Delta	Mean	Std Dev
DM	17	10	31	GEV Max	3.15269	0.988	18.82	4.04
Expert #1	25	13	51	GEV Max	5.40461	0.992	28.12	6.93
Expert #2	15	8	19.5	GEV Min	1.96259	0.972	13.87	2.52
Expert #3	20	11	29	Normal	3	0.997	20	3

Table 6-1: Prior Distributions

As a point of reference, the PERT means and standard deviations are provided in Table 6-2. In this table, k is the shape parameter and μ is the location parameter for the GEV distributions, with the means and standard deviations calculated as described above. The Normal distribution is described by the location parameter μ and the shape parameter σ. The PERT beta distribution is described by the α and β shape parameters as calculated using Equation 3-12 and Equation 3-13. The means and standard deviations of the PERT examples were calculated using Equation 2-1 and Equation 2-2. In order to solve for the two beta parameters, α and β using Equation 3-12 and Equation 3-13, an assumption was made that the PERT mean (as calculated using Equation 2-1) was reasonably close to the true beta mean as calculated by Equation 2-3.

	GEV Approximation				PERT Beta Approximation			
	k or σ	μ	Mean	Std Dev	α	β	Mean	Std Dev
DM	3.15269	17	18.82	4.04	2.33	3.67	17.33	2.67
Expert #1	5.40461	25	28.12	6.93	2.26	3.74	27.33	6.33
Expert #2	1.96259	15	13.87	2.52	3.43	2.57	14.58	1.92
Expert #3	3	20	20	3	3	3	20	3

Table 6-2: Mean/Std Dev Comparisons

By definition, the entire density of a beta distribution must fall between the BC and WC estimates since the distribution is zero outside that stated range ("Beta Distribution" 2016; Grubbs 1962, 913). For the two GEV distributions and the Normal distribution, there is some density outside the chosen range ("Generalized Extreme Value Distribution - Wikipedia" 2016).

Figure 6-2 through Figure 6-4 show plots of the distributions described in

Table 6-2. In the legend, "c" is the normalizing constant used on the beta distribution

so it could be easily compared to the GEV and Normal distribution models.
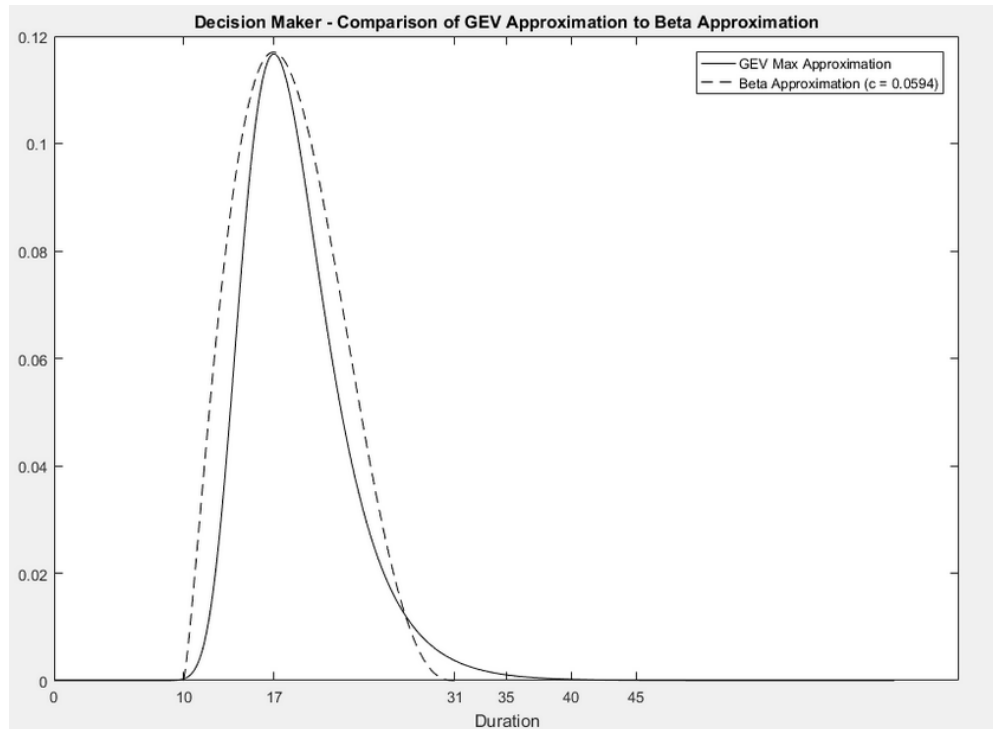
Figure 6-1: Decision Maker – GEV and Beta Distribution Models
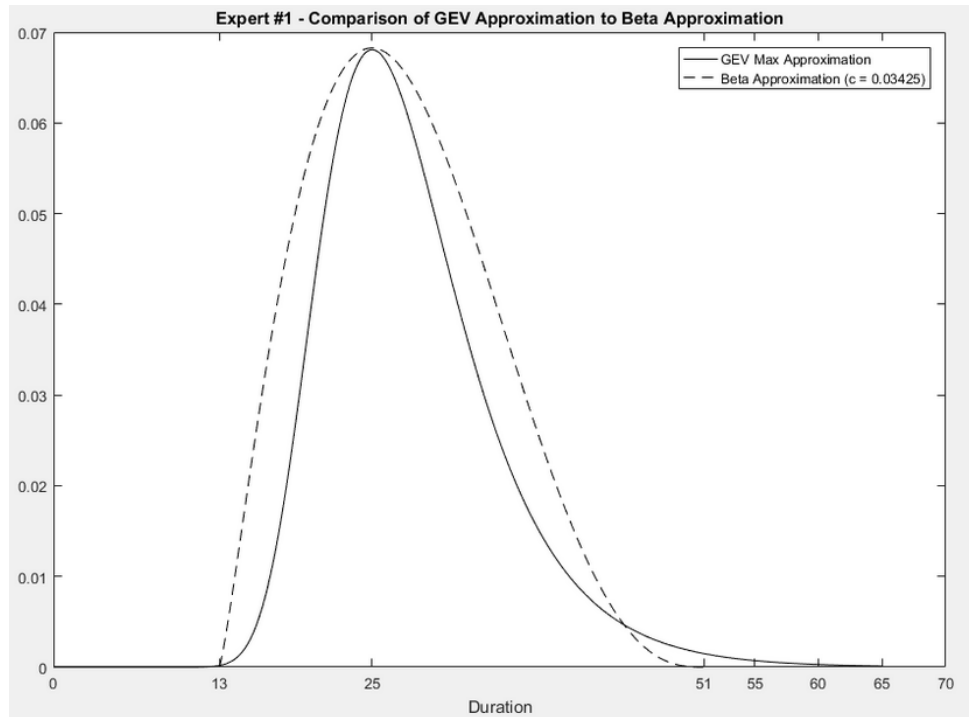
Figure 6-2: Expert #1 – GEV and Beta Distribution Models
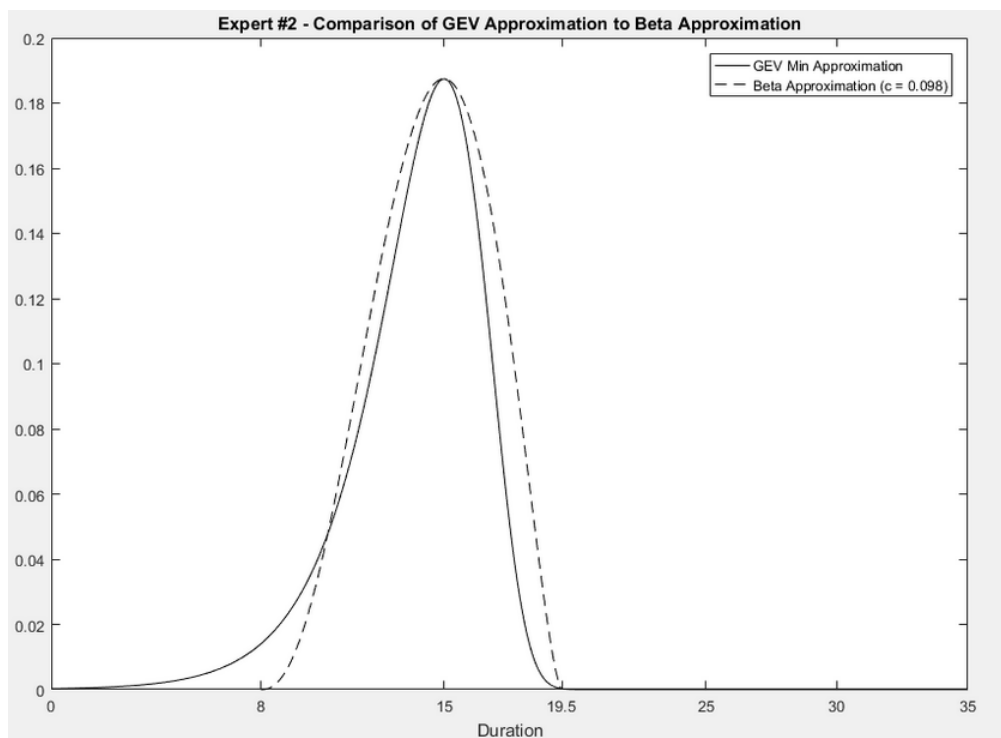
Figure 6-3: Expert #2 – GEV and Beta Distribution Models
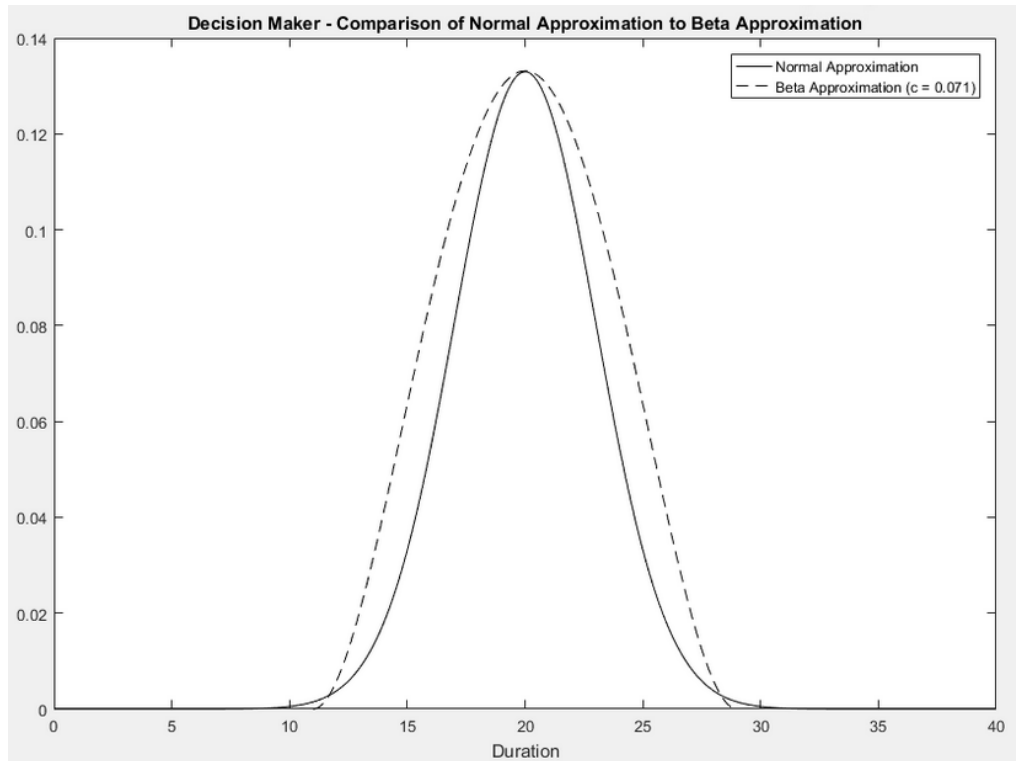
Figure 6-4: Expert #3 – GEV and Beta Distribution Models

As can be seen from Figure 6-2 through Figure 6-4, the graph of the GEV

Max approximation begins to move away from the x-axis (i.e. gain appreciable

density) at the location of the BC estimate and the GEV Min begins to move away

from the x-axis at the WC estimate. It can also be seen that the mode of both the

Beta/GEV and Beta/Normal approximations occur at the ML value. The major

differences between the two approximations occur at the WC value (for the GEV

Max distribution) and the BC value (for the GEV Min distribution). With the GEV

distribution defined on the entire real number axis, only two of the three parameters

could be set. The choice was made to set the parameter which seemed to exhibit less

uncertainty (the BC estimate for the GEV Max model and the WC estimate for the

GEV Min model). This resulted in the remaining parameter not matching as closely to the PERT Beta approximation.

6.2 Calibrating the Expert

Once the prior distribution had been established, it could be subjectively calibrated based on the DM's belief about the Expert by using the charts in Appendix A.10 – Appendix A.12). Multiplying together Equation 3-26 (using the appropriate values of α and β) and the appropriate prior shape as described either Equation 3-14, Equation 3-16, or Equation 3-18, causes the variance of the original prior to shrink or grow. The examples below show the affects of different calibrations schemes on the three prior distribution types. The parameters for the beta filter are provided in Table 6-3 as a reference (the entire list can be found in Appendix A.10 – Appendix A.12).

Figure #	Distribution Type	Calibration Percentage		α	β
Figure 6-5	GEV Max	5%	- Understated	2.03	2.77
	GEV Max	10%	- Understated	1.44	1.76
	GEV Max	15%	- Understated	1.17	1.29
	GEV Max	30%	- Overstated	0.80	0.65
	GEV Max	35%	- Overstated	0.73	0.54
	GEV Max	40%	- Overstated	0.68	0.45
Figure 6-6	GEV Min	5%	- Understated	2.77	2.03
	GEV Min	10%	- Understated	1.76	1.44
	GEV Min	15%	- Understated	1.29	1.17
	GEV Min	30%	- Overstated	0.65	0.80
	GEV Min	35%	- Overstated	0.54	0.73
	GEV Min	40%	- Overstated	0.45	0.68
Figure 6-7	Normal	5%	- Understated	2.09	2.09
	Normal	10%	- Understated	1.53	1.53
	Normal	15%	- Understated	1.22	1.22

Figure #	Distribution Type	Calibration Percentage	α	β
Figure 6-7 (cont)	Normal	30% - Overstated	0.71	0.71
	Normal	35% - Overstated	0.60	0.60
	Normal	40% - Overstated	0.52	0.52

Table 6-3: Calibration Examples

In Figure 6-5 through Figure 6-7, the figure on the left represents an

understated expert calibrated at the 15%, 10%, and 5% levels. The graph on the right

shows an overstated expert calibrated at the 30%, 35% and 40% levels. As can be

seen in these figures, the variance of the calibrated expert will change while the mode

remains in the same location along the x-axis.
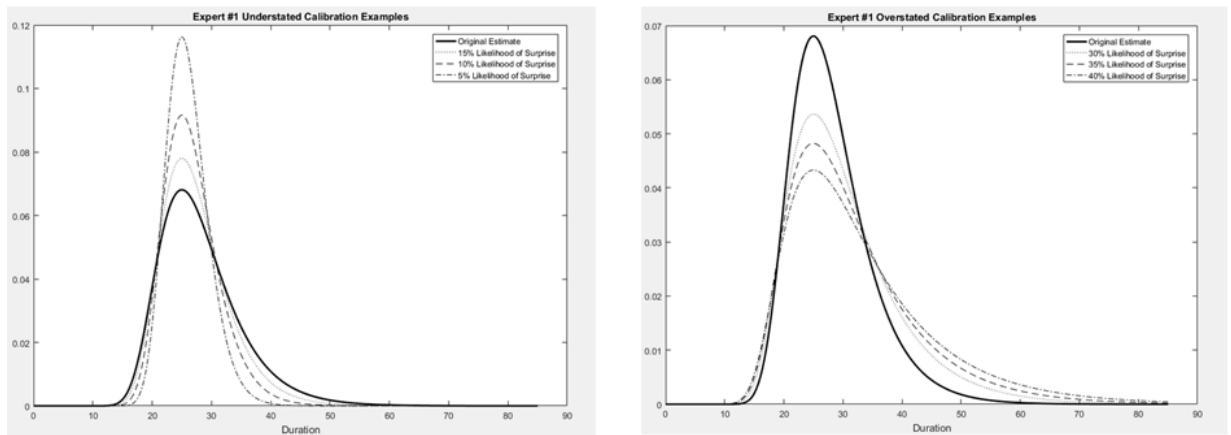
Figure 6-5: Expert #1 - GEV Max Calibration Results
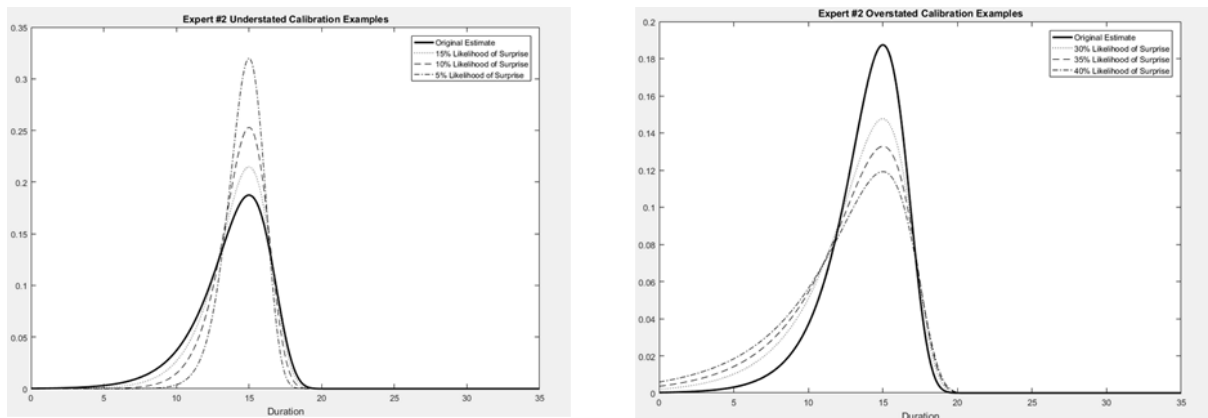
Figure 6-6: Expert #2 - GEV Min Calibration Results
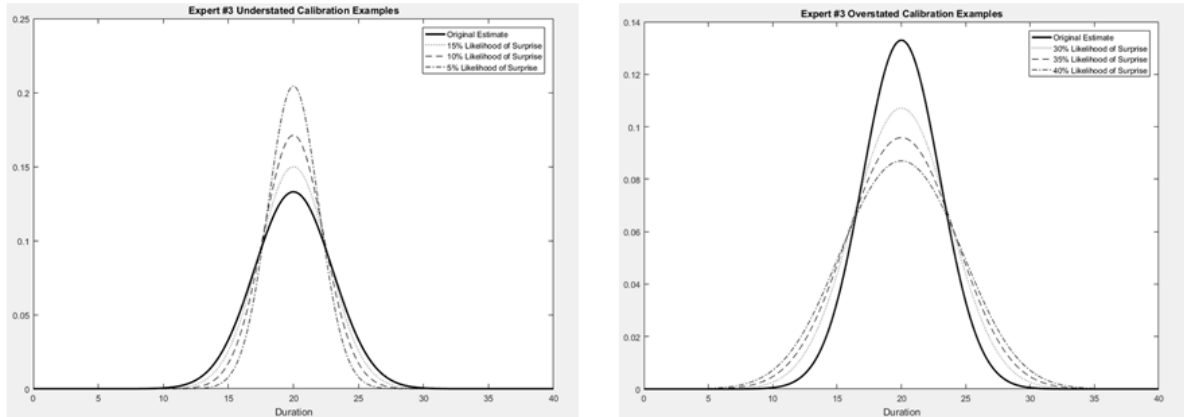
Figure 6-7: Expert #3 - Normal Calibration Results

6.3 Calculating the Posterior

Once the Expert's prior has been calibrated, the final posterior distribution can be calculated. The posterior is calculated by multiplying together the DM's prior with all of the calibrated Expert estimates that have been provided. Because the final equation describing the resulting curve is unknown, an approximation is used to calculate the mean and variance of the posterior. Using Excel™ and evaluating the posterior curve from zero to some value larger than the largest WC estimate (among all estimates provided) allows the Decision Maker to determine the maximum value (i.e the mode) of the resulting posterior distribution. Once normalized, Equation 3-30 can be used to solve for k, the shape parameter of the distribution. With the mode and shape parameter determined, the mean and variance of the posterior distribution can be calculated.

Using the example estimates shown in Table 6-1 (reproduced in part below in Table 6-4, it can be seen how the various parameters change with different combinations of expert opinion and different calibration levels. Table 6-5 provides a

summary of several example combinations. In Table 6-5, the first column describes

the combination of prior estimates and the resulting posterior distribution model (i.e.

GEV Max, GEV Min, or Normal). In the subsequent columns, ML is the resulting

posterior mode, c is the normalizing constant used when multiplying together the

priors, k is shape the parameter for the GEV posterior approximation, and σ is the

standard deviation of the Normal model. Note that the mean and standard deviation

for the GEV model are calculated using Equation 3-32 and Equation 3-33, while the

mean and standard deviation for the Normal distribution match the ML estimate and

the σ parameter.

	ML	Type	k	σ	Mean	Std Dev
DM	17	GEV Max	3.15269	N/A	18.82	4.04
Expert #1	25	GEV Max	5.40461	N/A	28.12	6.93
Expert #2	15	GEV Min	1.96259	N/A	13.87	2.52
Expert #3	20	Normal	N/A	3	20	3

Table 6-4: Summary Example Estimates

Posterior Combination/Approximation	ML	c	k or σ	Mean	Std Dev
Expert Calibration = No Calibration					
DM * E1 – GEV Max	20.8	38.905	2.93712	22.50	3.77
DM * E2 – Normal	15.6	15.748	1.36851	15.6	1.37
DM * E3 – Normal	18.8	13.227	2.41226	18.8	2.41
DM * E1 * E2 – Normal	16.7	3520.98	1.05089	16.7	1.05
DM * E1 * E3 – Normal	20.4	451.428	2.0966	20.4	2.10
DM *E2 * E3 – Normal	16.3	345.924	1.14922	16.3	1.15
DM * E1 * E2 * E3 – Normal	17.1	50249.4	0.94969	17.1	0.95
Expert Calibration = 10%					
DM * E1 – GEV Max	21.8	38.120	2.70253	23.36	3.47
DM * E2 – Normal	15.4	17.426	0.93837	15.4	0.94
DM * E3 –Normal	19.2	11.855	2.16911	19.2	2.17
DM * E1 * E2 – Normal	16.7	11402	0.87696	16.7	0.88
DM * E1 * E3 – Normal	20.8	426.25	1.72798	20.8	1.73
DM * E2 * E3 – Normal	16.3	475.547	0.95227	16.3	0.95
DM * E1 * E2 * E3 – Normal	17	185451	0.78244	17	0.78

Expert Calibration = 30%					
DM * E1 – GEV Max	20.1	41.533	3.03118	21.85	3.89
DM * E2 – Normal	15.7	17.989	1.49371	15.7	1.49
DM * E3 – Normal	18.4	14.154	2.68723	18.4	2.69
DM * E1 * E2 – Normal	16.7	2468.3	1.1748	16.7	1.17
DM * E1 * E3 – Normal	20	531.53	2.39821	20	2.40
DM * E2 * E3 – Normal	16.3	336.75	1.30241	16.3	1.30
DM * E1 * E2 * E3 – Normal	17	35256	1.0787	17	1.08

Table 6-5: Posterior Duration Results

Figure 6-8 through Figure 6-14 provide a graphical representation of the data shown in Table 6-5. The graph in the top left shows the prior distributions of each participant. The remaining graphs show the posterior distribution when calculated using Equation 3-28 and also the resulting GEV Max, GEV Min, or Normal approximation, as appropriate. The top right graph shows the resulting posterior distribution if all experts are fully calibrated, the bottom left graph shows a 10% calibration level, and the bottom right graph shows a 30% calibration level.
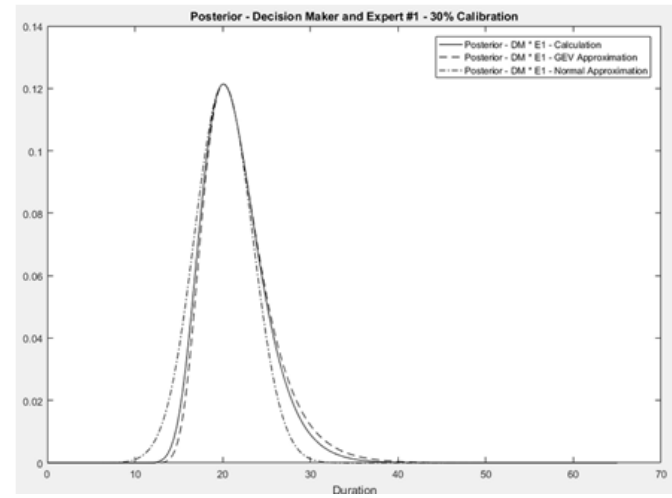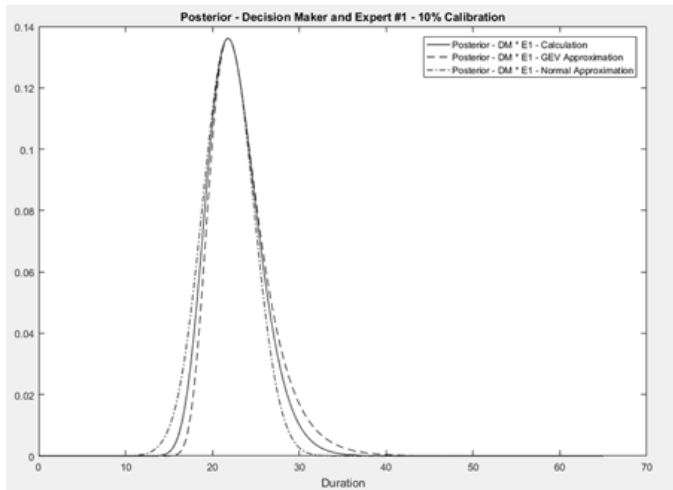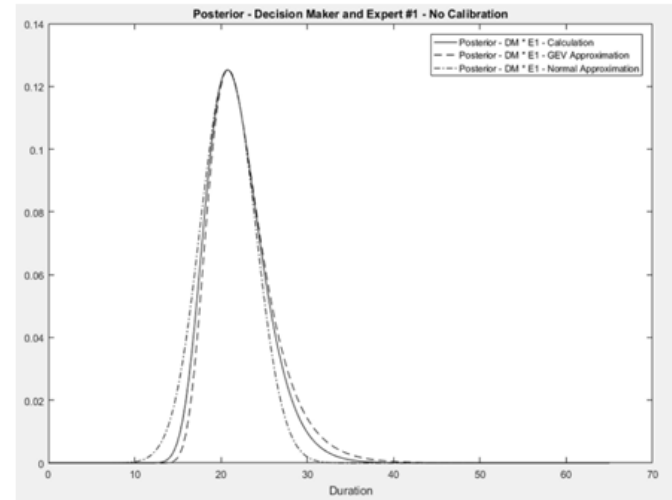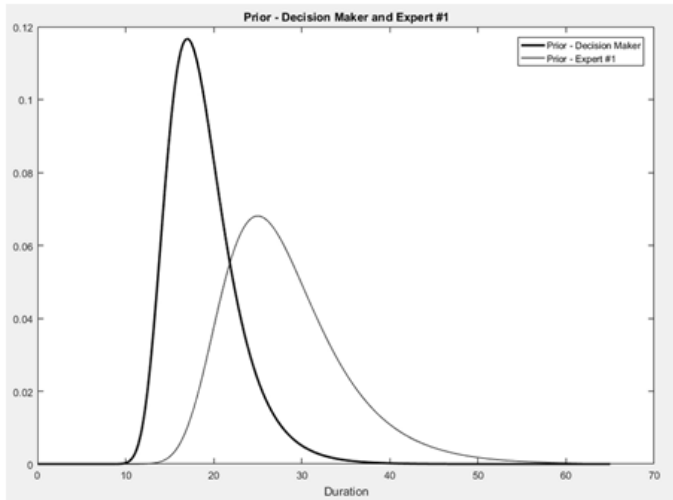
Figure 6-8: Decision Maker and Expert #1
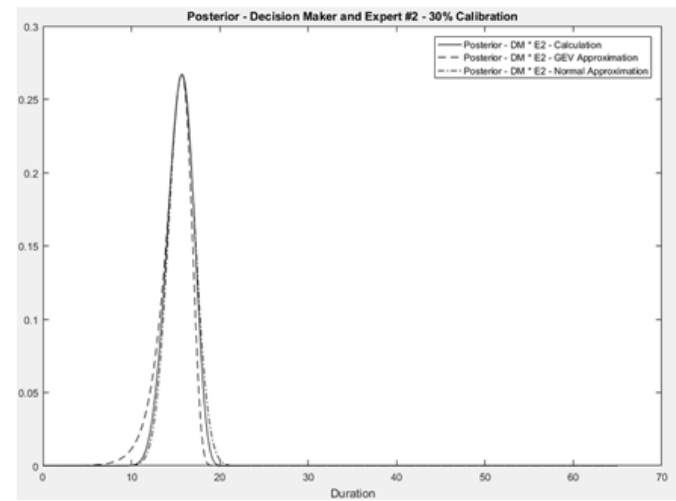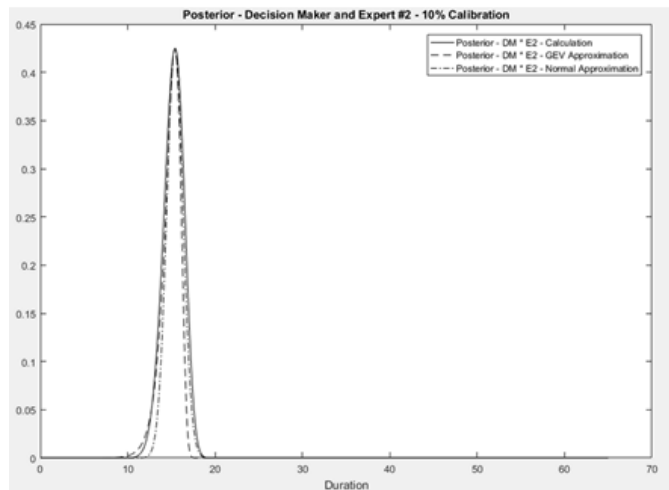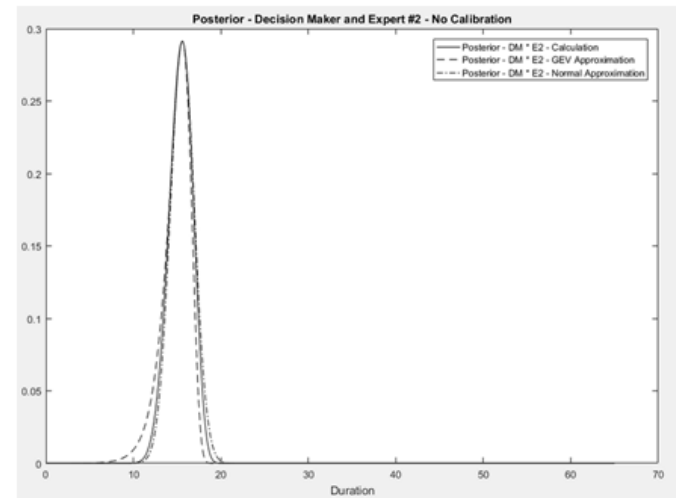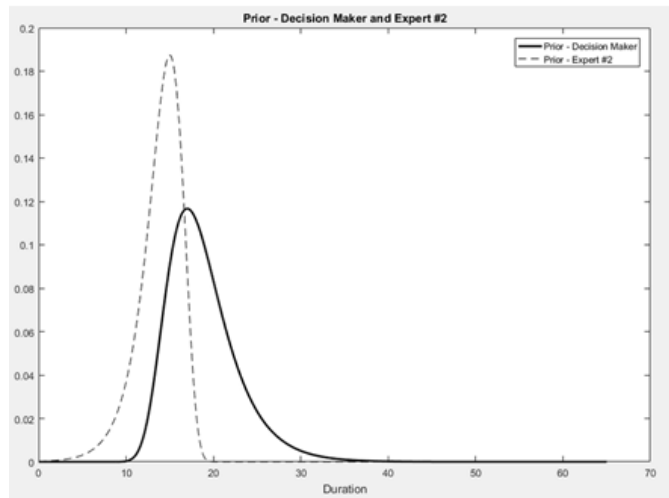
Figure 6-9: Decision Maker and Expert #2
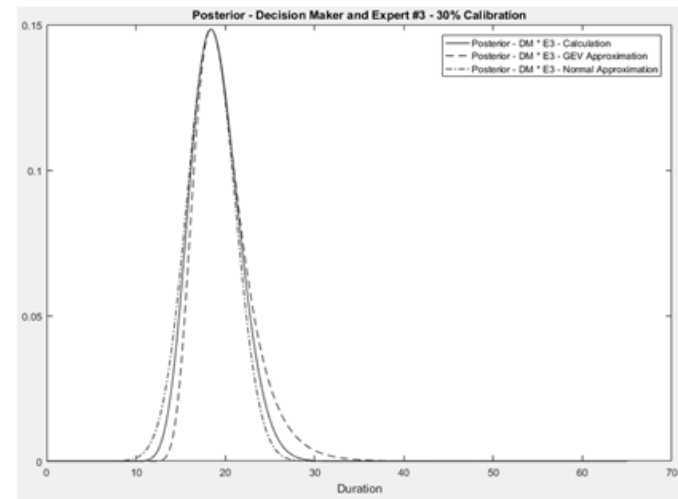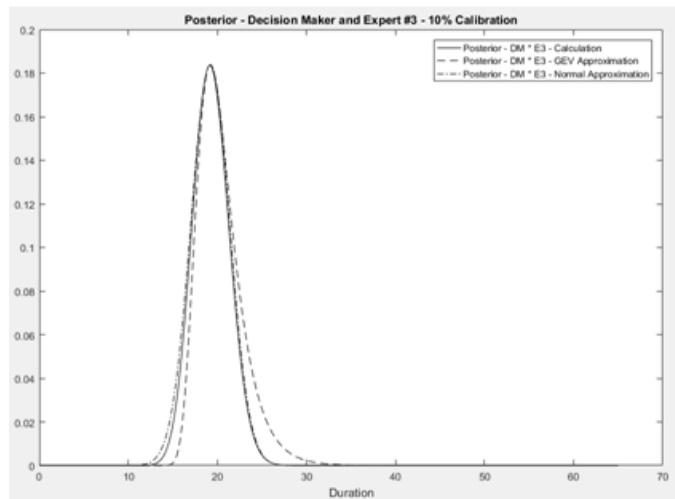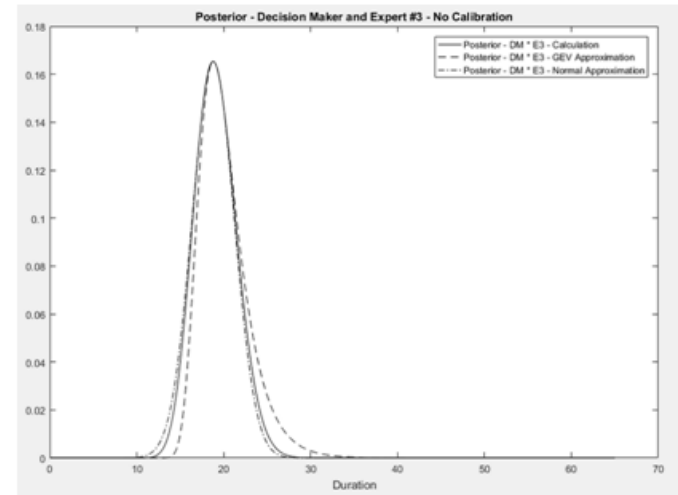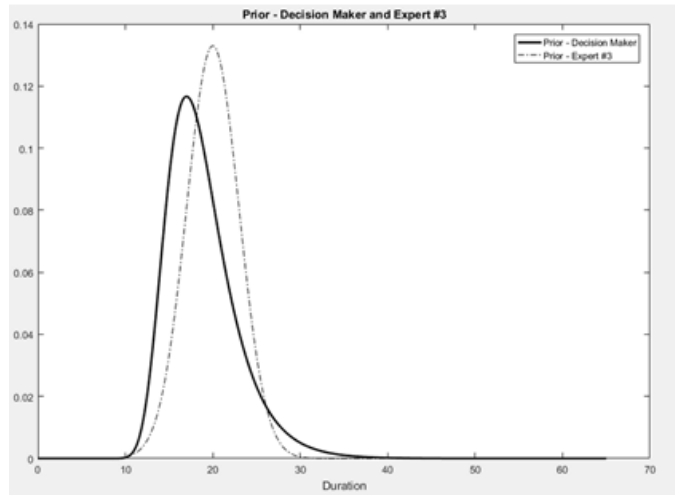
Figure 6-10: Decision Maker and Expert #3
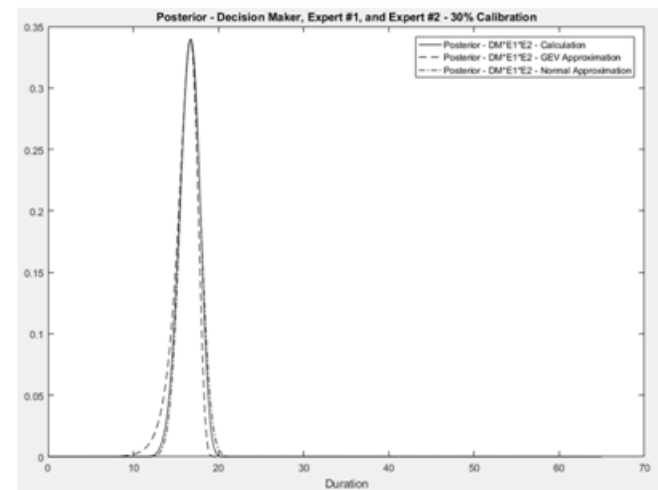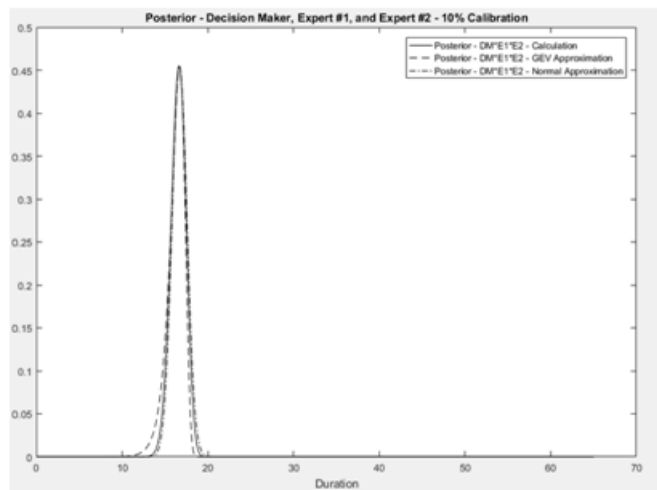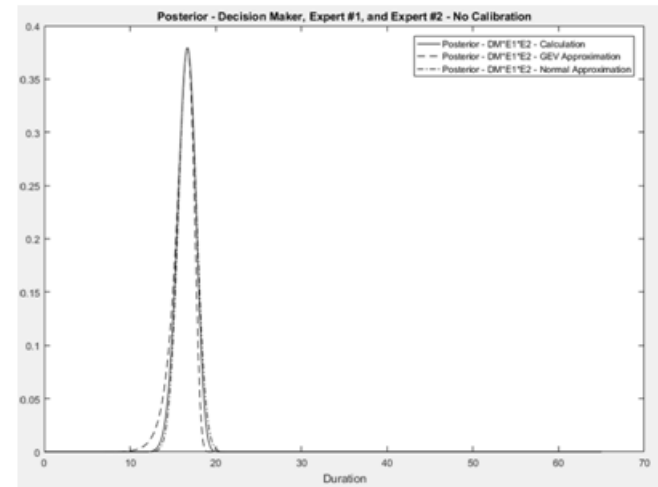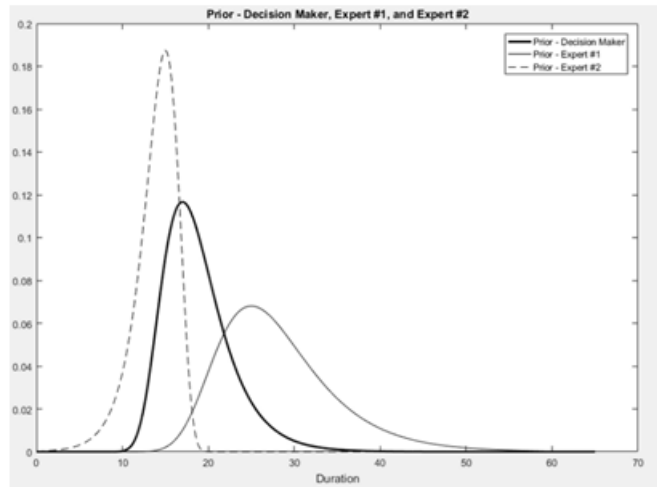
Figure 6-11: Decision Maker, Expert #1, and Expert #2
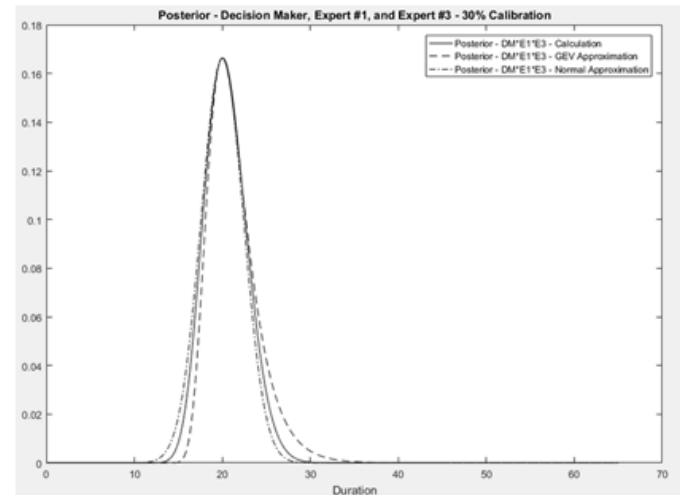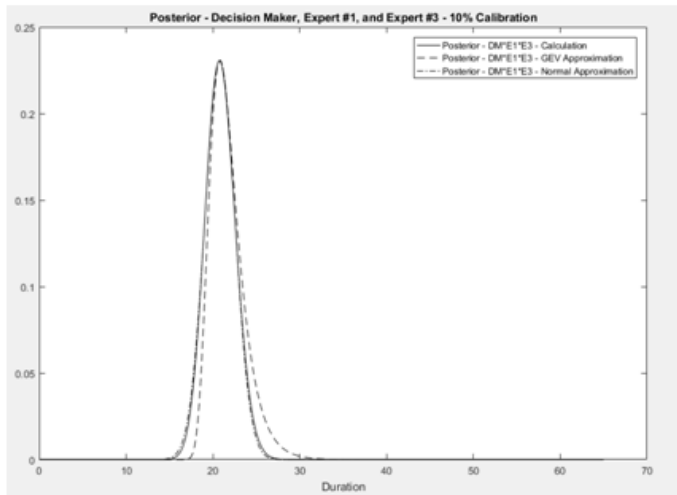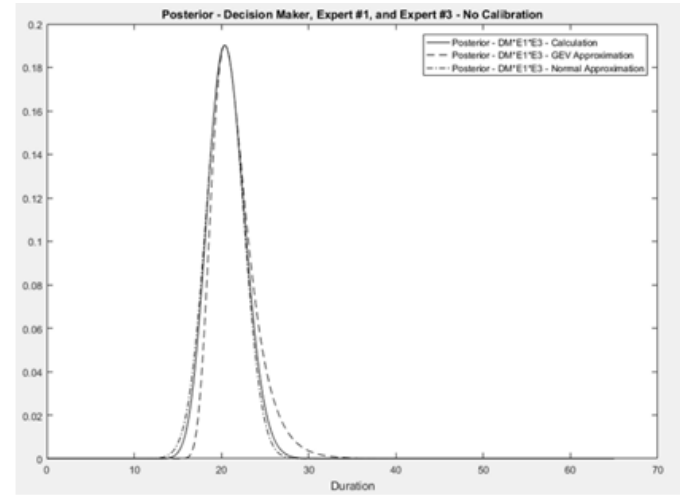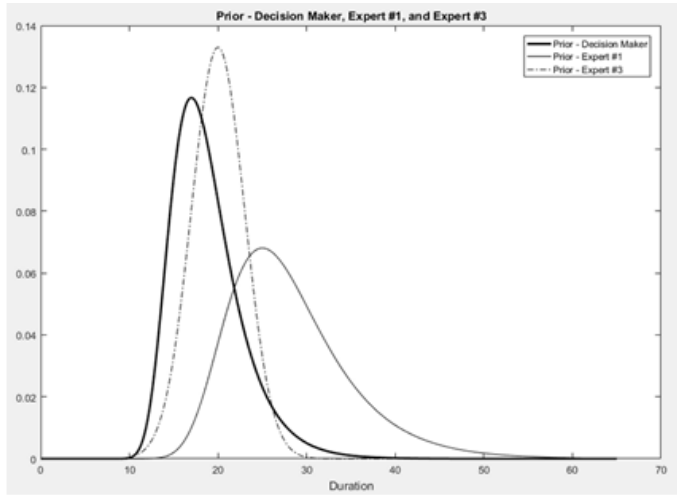
Figure 6-12: Decision Maker, Expert #1, and Expert #3
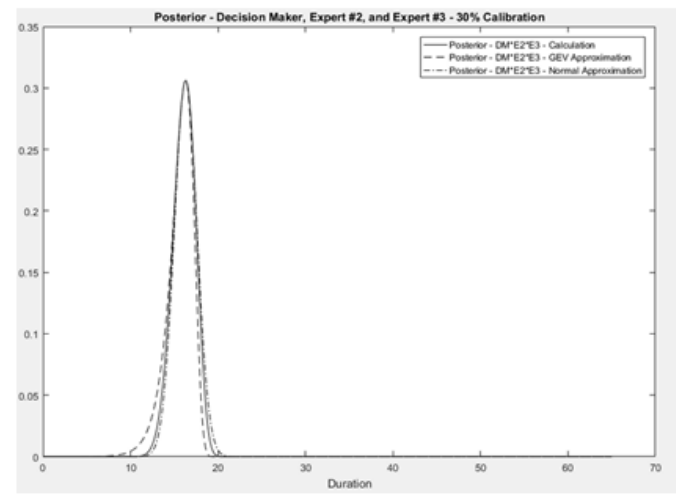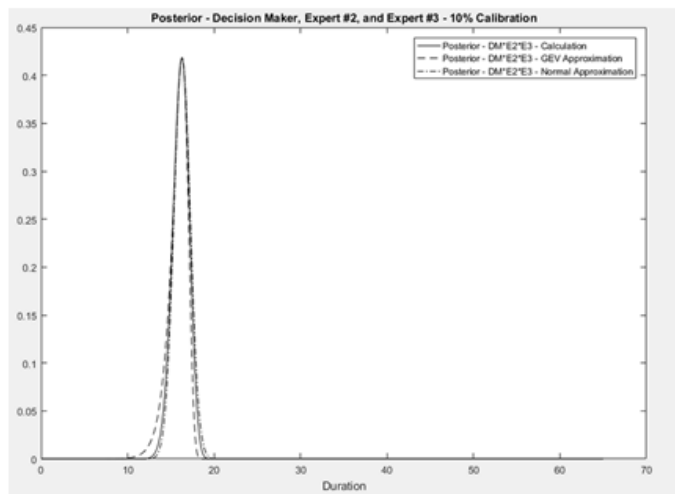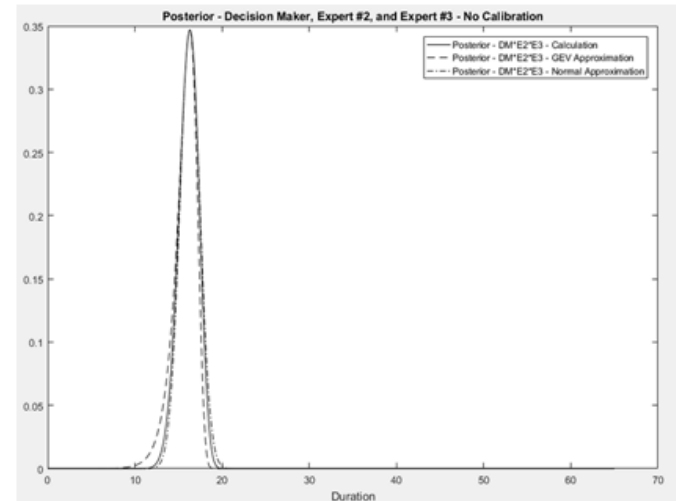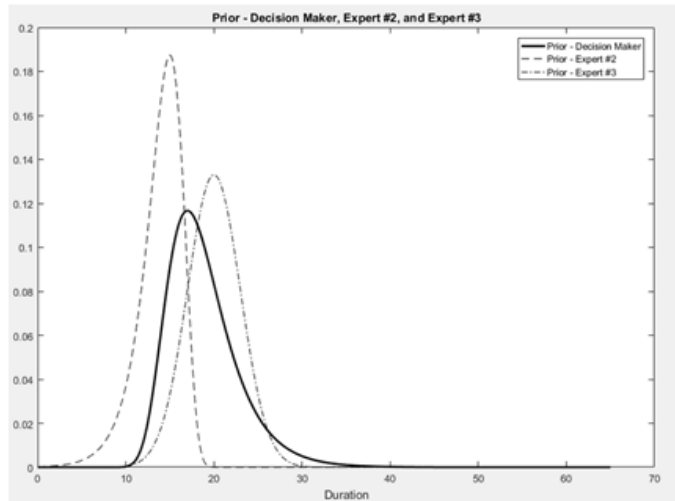
Figure 6-13: Decision Maker, Expert #2, and Expert #3
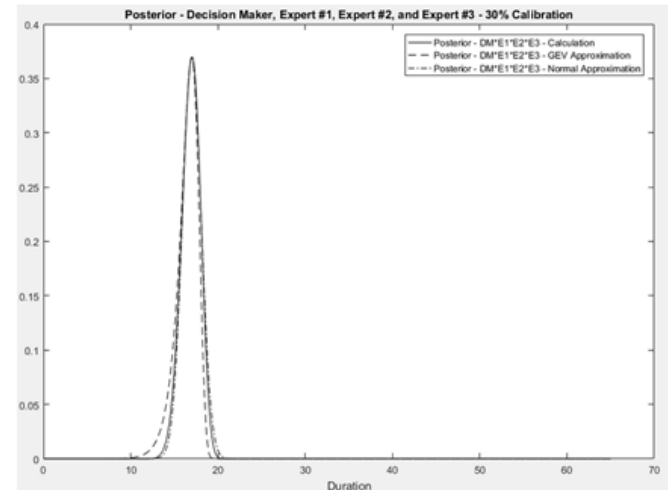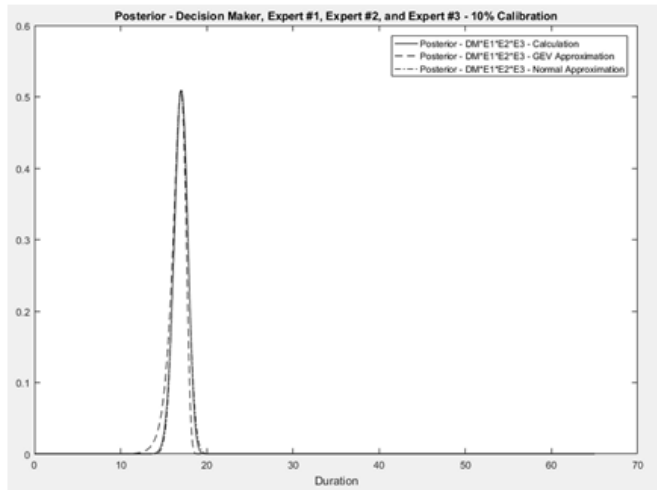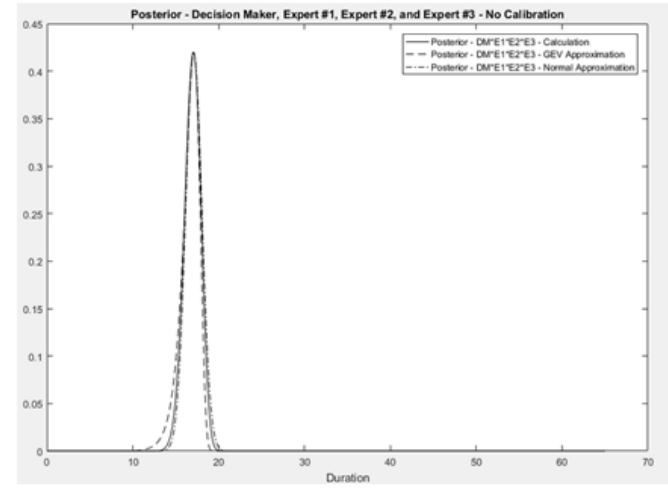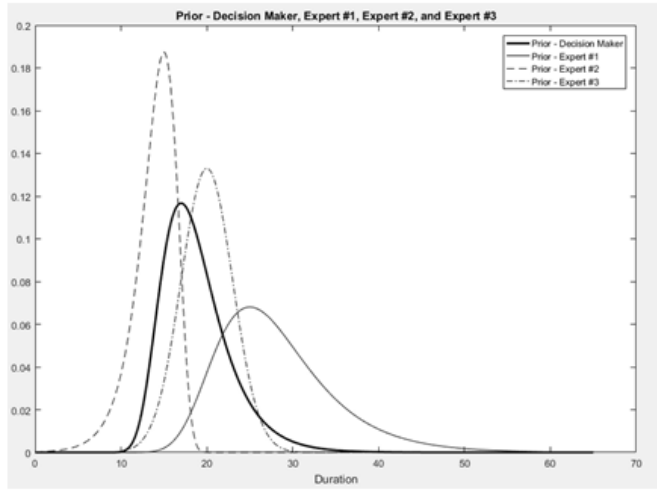
Figure 6-14: Decision Maker, Expert #1, Expert #2, and Expert #3

6.4 Further Examples

Section 6.3 provided several examples of how the posterior distribution changes based on various combinations of calibration schemes and expert inputs. Because the example estimates in Table 6-1 represented all three distribution models, the resulting posterior distribution was often most closely modeled by a Normal distribution. This section will provide several examples of the posterior distribution when the prior distributions all share the same from and one case where the DM and the Expert provide wildly disparate estimates. To simplify these examples, it is assumed that all experts are fully calibrated.

In Section 6.3, when different types of distributions were multiplied together, the tails effectively cancelled each other out, resulting in a posterior that most closely resembled a Normal distribution. As seen in Table 5-14, however, most people favored a positively skewed distribution which would be most closely modeled by a GEV Max distribution. When all of the prior distributions are of the same type, the resulting posterior distribution will maintain its GEV shape unless there are a large number of experts providing estimates Equation 3-29 and the Excel™ command described just below Equation 3-29 will provide an indicator of when one should switch from the GEV Max approximation to the Normal approximation. Table 6-6 shows the initial estimates and prior distribution parameters for a DM and two Experts. It also shows the resulting parameters of the posterior distribution, calculated using Equation 3-28. The priors for all three stakeholders, as well as the resulting posterior can be modeled by a GEV Max distribution. Figure 6-15 provides a graph of the priors, the result of Equation 3-28, and the GEV Max approximation

used to describe the curve that results from Equation 3-28. It also shows the Normal

approximation for reference. In Table 6-6, the normalizing constant, c, is not

required for the prior distributions. For a GEV Max approximation, the BC estimate

can be calculated using Equation 3-35 and the WC estimate can be approximated

using Equation 3-41.

	ML	BC	WC	c	k	Mean	Std Dev
DM Prior	17	10	26		3.15269	18.82	4.04
Expert #1 Prior	25	13	49		5.40461	28.12	6.93
Expert #2 Prior	15	9	35		2.70230	16.56	3.47
Posterior	18.7	14.1	30.8	972.32	2.09243	19.91	2.68

Table 6-6: GEV Max Example Prior Distribution
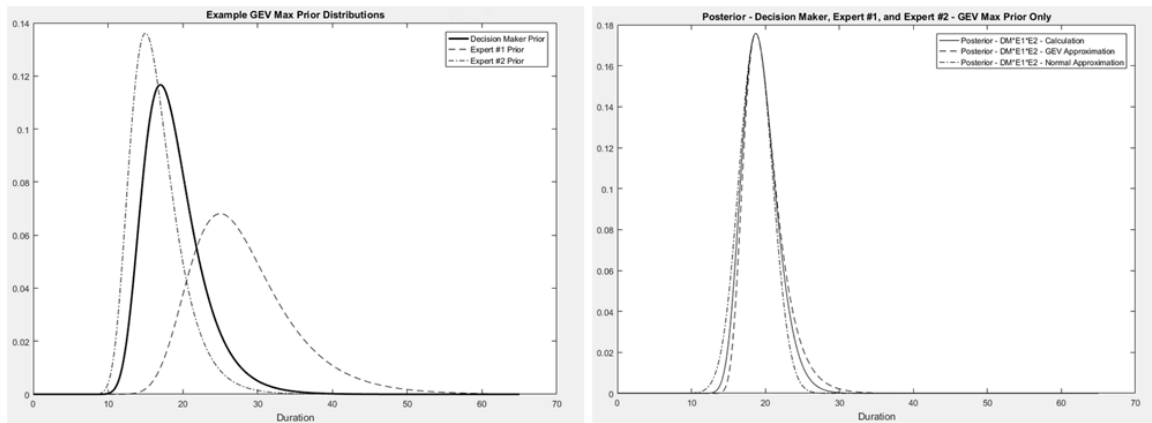
Figure 6-15: GEV Max Example Priors and Posterior

The same results can be demonstrated in the GEV Min case. Table 6-7

provides similar information to Table 6-6, except the prior estimates and posterior are

all described by the GEV Min distribution. Figure 6-16 provides a graph of the priors

on the left and the posterior on the right. The three posterior graphs represent the

result of Equation 3-28 as performed on the priors, the resulting GEV Min

approximation, and the Normal approximation, provided for reference. For the GEV

Min case, the BC value can be solved using Equation 3-42, and WC value can be

approximated using Equation 3-37.

	ML	BC	WC	c	k	Mean	Std Dev
DM Prior	35	19	40		2.18066	33.74	2.80
Expert #1 Prior	25	15.3	30		2.18066	23.74	2.80
Expert #2 Prior	40	32	46		2.61679	38.49	3.36
Posterior	27.2	20	30	103860	1.23115	26.49	1.58

Table 6-7: GEV Min Example Prior Distribution
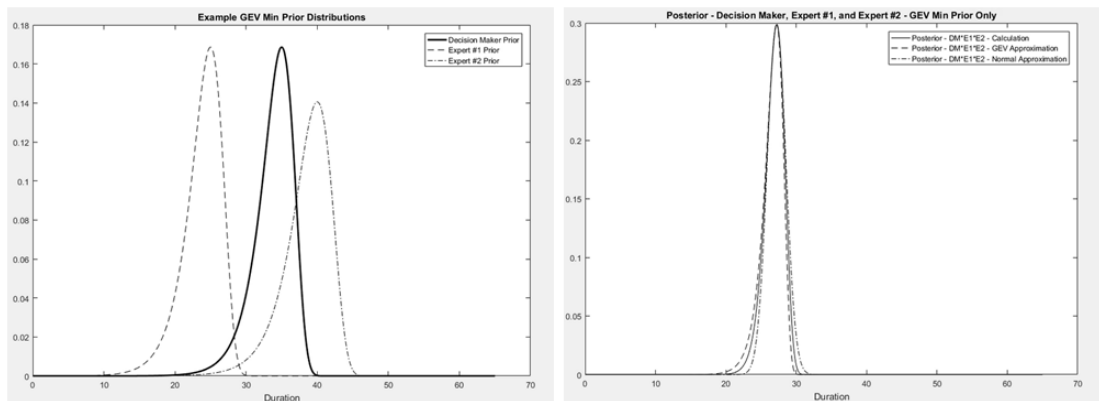
Figure 6-16: GEV Min Example Priors and Posterior

The final examples in this chapter deal with two extreme cases. The first case

shows the results of multiple experts in complete agreement with one another and the

second case shows the results of a DM and Expert in complete disagreement with one

another. For each of the prior distribution models, Table 6-8 shows the parameters

for the DM and nine experts, all of whom are in complete agreement with one

another, so their prior distributions are all exactly the same. Table 6-8 also shows the

parameters of the resulting posterior distributions (with rounded ML,BC, and WC

values) in these cases of extreme agreement. In each of the three posterior cases, the

resulting posterior curve is most closely modeled by a Normal distribution. The

shape parameter, mean, and standard deviation are all reflective of this Normal

model. Figure 6-17 through Figure 6-19 show the resulting posterior distributions in

each of the three cases.

	ML	BC	WC	c	k or σ	Mean	Std Dev
DM & 9 Experts: GEV Max	17	10	26		3.15269	18.82	4.04
DM & 9 Experts: GEV Min	25	15.3	30		2.18066	23.74	2.80
DM & 9 Experts: Normal	20	11	29		3	20	3
Posterior – GEV Max Model - Normal	17	14	20	847946379	1.00531	17	1.01
Posterior – GEV Min Model - Normal	25	23	27	30727596	0.69535	25	0.70
Posterior – Normal Model - Normal	20	17	23	243164795	0.94868	20	0.95

Table 6-8: DM and Expert Complete Agreement
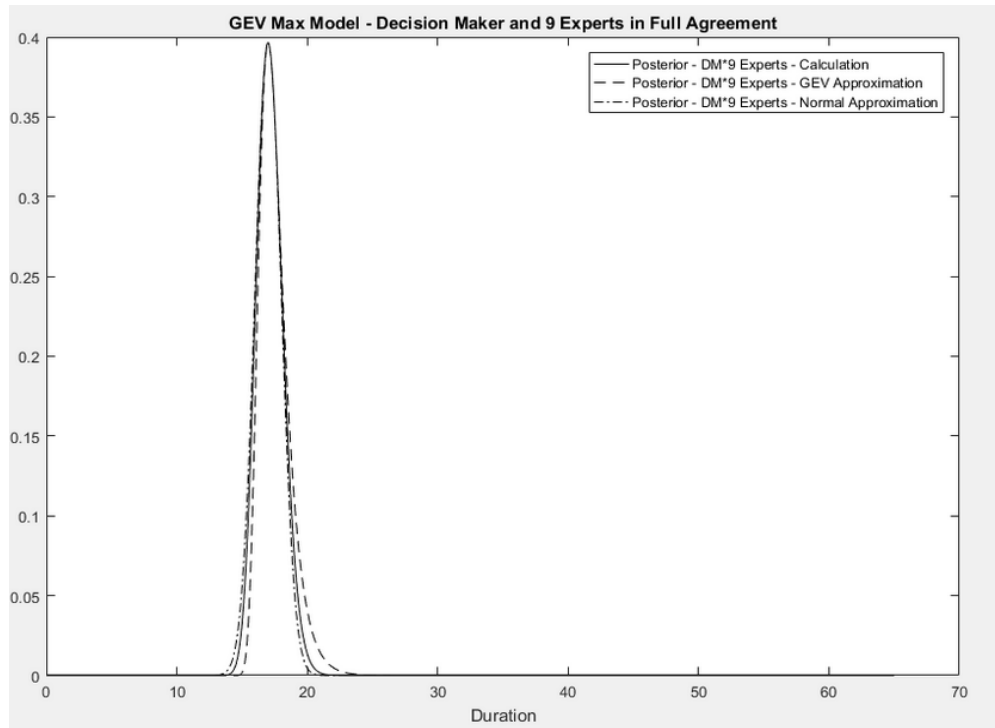
Figure 6-17: Posterior: Decision Maker and 9 Experts; Full Agreement – GEV Max Model
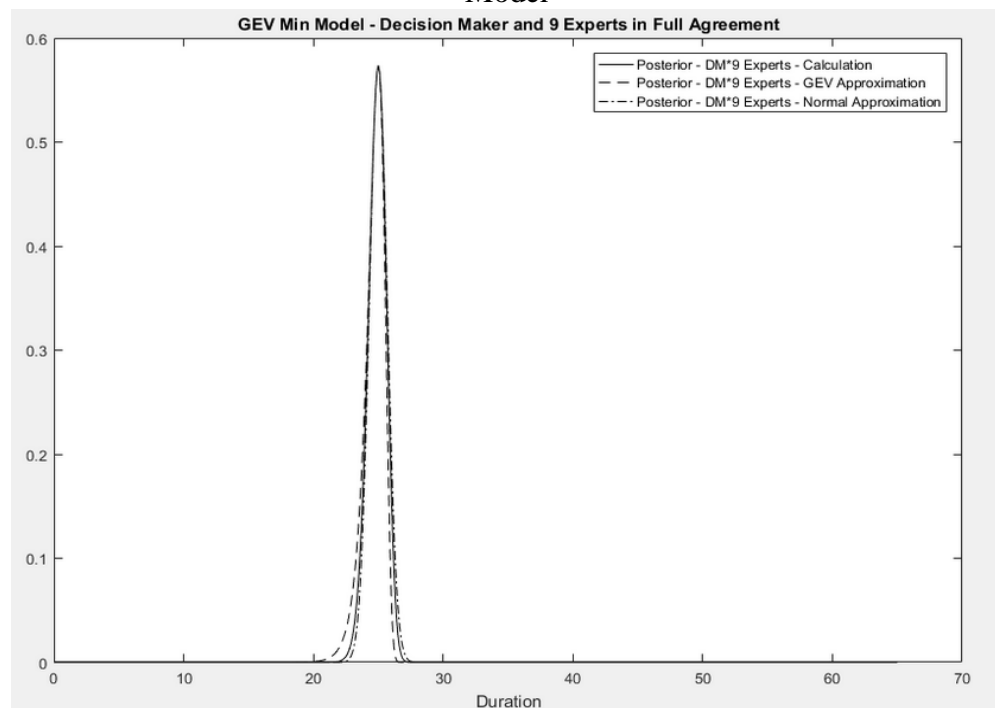
Figure 6-18: Posterior: Decision Maker and 9 Experts; Full Agreement – GEV Min Model
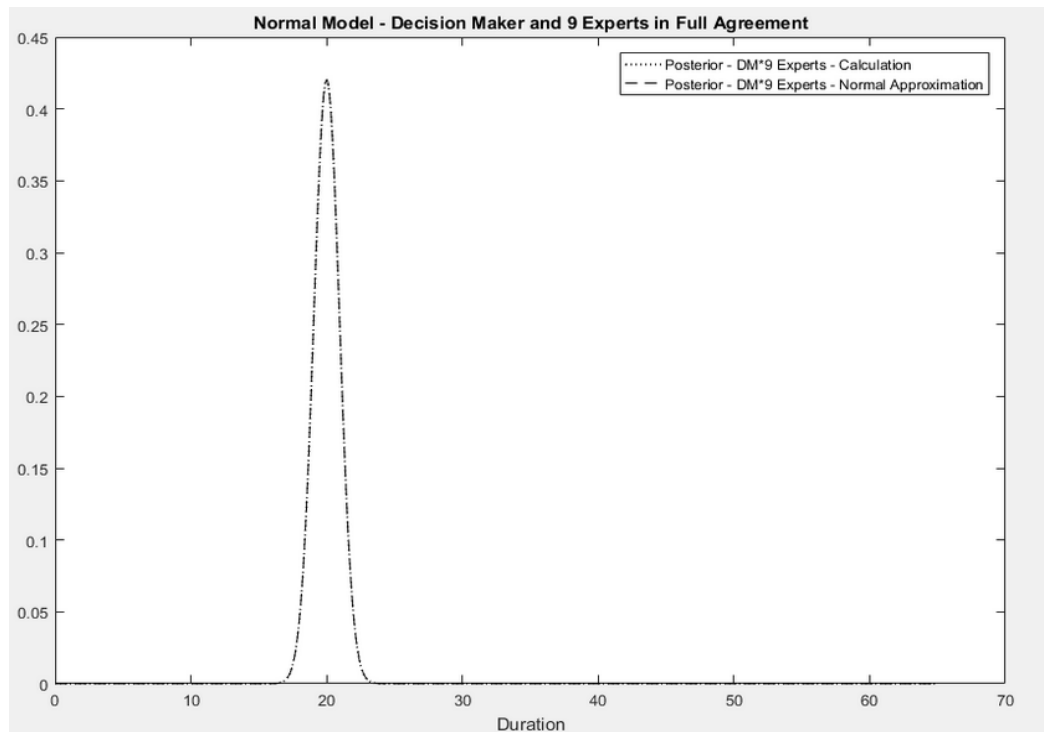
Figure 6-19: Posterior: Decision Maker and 9 Experts; Full Agreement – Normal Model

The final example shows a case when the DM and Expert are in complete disagreement with one another. Table 6-9 provides the prior distributions for a DM and Expert, as well as the resulting posterior distribution as calculated by Equation 3-28 and the resulting GEV Min approximation. The BC value and WC value are once again solved/approximated using Equation 3-42 and Equation 3-37, respectively.

	ML	BC	WC	c	k or σ	Mean	Std Dev
DM - GEV Max	17	10	26		3.15269	18.82	4.04
Expert #1 – GEV Min	40	30	46		2.61679	38.49	3.36
Posterior – GEV Max	35.4	0.18	49	1188.1	6.04750	31.91	7.76

Table 6-9: DM and Expert Severe Disagreement

214
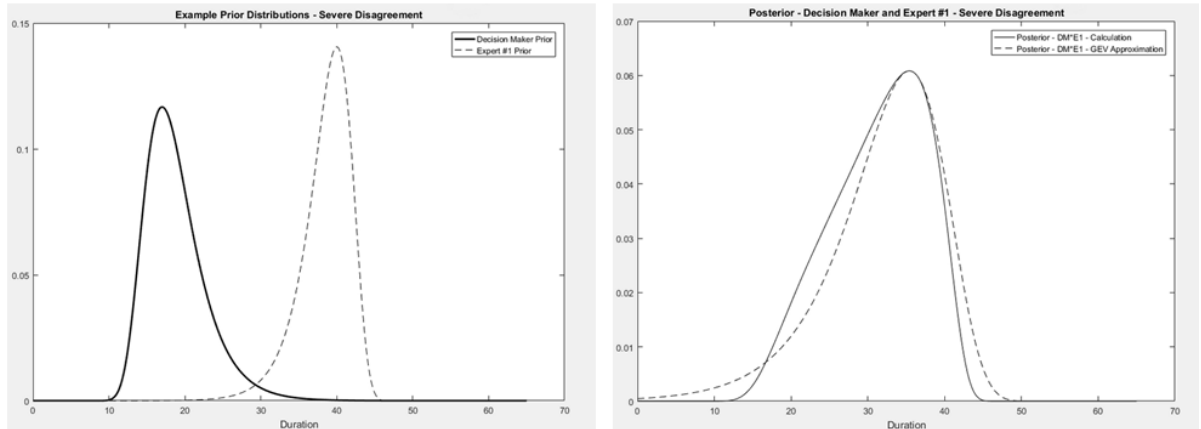
Figure 6-20: Decision Maker and Expert #1 – Severe Disagreement

Figure 6-20 it shows that in cases of extreme disagreement, the GEV

approximation of the curve calculated by Equation 3-28 begins to break down. This

graph also illustrates another concern with the GEV Min model. Looking closely at

the dashed line in the left graph, it can be seen that the graph has not quite collapsed

to the x-axis at X=0. This indicates that, using the GEV Min approximation, there is

a small, but non-zero probability that the activity duration will be negative. Further

discussion on this issue will be provided in Chapter 7.

Chapter 7: Discussion

Scheduling challenges are neither unique to Wallops Flight Facility (WFF), nor are they localized to the recent past. Many practices have been developed to describe how one should schedule a project, but despite these recommendations, many projects fail to meet deadlines. Based on this research, it appears there are trends in estimation practices among stakeholders in the three demographics studied. How can this information be used to develop better schedules? What can be learned from the GAO reports that can be augmented by scheduling best practices? What is the best way to incorporate the estimates of a diverse group of stakeholders? This chapter seeks to tie together previous chapters to answer these questions.

7.1 Past is Present: GAO Reports vs. Current Results

The GAO reports from Chapter 2 provided a quick summary of project history at NASA. As a government agency, projects at NASA are routinely analyzed to determine what went right and what went wrong ("About GAO" 2015). This analysis coupled with long-standing research in decision making biases helps shed light on the challenges faced by project managers in developing an accurate schedule. This section brings together the results of the GAO reports and the responses of subjects in this research and discusses those results through the lens of biases identified in the decision-making literature to determine reasons why projects struggle to finish on time.

216

7.1.1 External Influences

In an uncertain world, there are many roadblocks to successfully completing a project on time. Some of these can be managed within the project, but some are outside the control of anyone involved in planning or execution. The effects of these external constraints could be seen throughout the GAO reports and also in the responses of the subjects in this study.

One external influence that can affect the accuracy of the estimating process is the retirement of the stakeholders who posses background history and knowledge (GAO 2006a, 4, 22, 2006e, 10, 2008, 6). Chapter 2 provided several methods for defining an expert, but typically, these are people who have, "…special knowledge about an uncertain quantity or event" (Morris 1977, 679). Without the knowledge base of these experts, there will be gaps in the estimating processes. This is why it is critical to account for the opinions of multiple stakeholders to ensure missing pieces to the knowledge base are minimized and also to ensure the correct mix of project stakeholders within the project team. Each piece of information is a data point by which the decision maker can update her beliefs.

Review boards served as another external influence on a project's schedule. The purpose of these boards is to review the plan and question the project team to determine whether or not the plan is mature enough to continue. A GAO report from 2006 recommended more reviews and project-stop points throughout the cycle. In theory, this would lead the project teams to develop more thorough plans to ensure approval from the review board. It seems, however, based on the responses of some

subjects in this research that input from these review boards is not always appreciated. In some cases, they were referred to as stumbling blocks to the project.

Project teams may perceive that these reviews are required as proof to outsiders that the project team knew what it was doing. These beliefs and perceptions may be indicative of the "not invented here" bias identified by Ariely (Ariely 2009a, Loc. 1443). When the project plan is presented to a board, questions can be perceived as personal attacks, and since board members are an entity outside the project team, the project team is less likely to want to incorporate the suggestions since they were not invented within the project team itself. If those suggestions cause extra work that was not originally accounted for, the perception of the project team could be that the delay was caused by a force beyond their control and that they are no longer responsible for the resulting schedule delays.

The GAO reports studied pointed out that the culture of NASA is generally two fold. Both the "can do" attitude and the culture of safety could almost be merged into one general goal: "Make it happen, but make it happen safely" (Martin 2012, 11–17; GAO 2017, 15–17, 21–22). While not a constraint per se, this culture does have an effect on project schedules. From the Demographics survey, it could be seen that across all subjects, schedule and cost took a back seat to reducing risk and ensuring technical success. The desire to preserve technical success over other constraints can also have a significant effect on the schedule. From the project preferences survey, when comparing the different project constraints to one another there is a clear preference for sacrificing the material constraints of time and money in order to reduce risk of project failure/personal injury or decrease the quality of

support. The GAO reports point out that there is a prevailing attitude at NASA that states that as long as technical success is achieved, no one will care whether or not the project came in on time or under budget, or at least that memories of those project management failures will fade in the light of the technical success (Martin 2012, 11– 12). Several subjects in this study mentioned the same mentality, confirming that this same attitude is also prevalent at WFF. (Kremer 2017c). The implication for scheduling practices is that if problems arise in a project, the schedule will be the first thing to go, and as schedule increases, the cost will most likely increase as well.

7.1.2 Internal Influences

Some constraints are beyond the project manager's control, but others are within her sphere of influence. One of the major complaints in both the GAO reports and across both the management and technical subjects in this research was a failure to adequately define requirements (GAO 2009c, 1,3,5-6, 1993b, 4, 1993a, 11). According to PMBOK, in order to manage a project successfully, all activities must be tied back to a requirement (PMI 2013, para. 5.4, 6.2). If requirements are only developed at a high level, there may be confusion about what is actually required in order to meet a requirement (GAO 2014, 25; Mantel Jr. et al. 2004, 82). As was seen in the COA Survey, perceptions differ between stakeholders in the management/technical divide. An activity that one stakeholder believes is necessary to meet a high-level requirement may be perceived as "gold-plating" to another.

Additionally, stakeholders may be planning to different definitions of loss aversion where loss aversion is defined as saving face for the project constraint they

are most responsible for (Ariely 2009a, Loc 1328, 1518). For the project manager, saving face means finishing the project on-time and on-budget, so she may not be as concerned with repeated testing as long as it works. For the technician, saving face means ensuring the system performs as advertised. Each group, however, still shoulders some responsibility for those other constraints. The project manager must still deliver a working system or risk losing face and the technician can still be held accountable if the system is not ready when advertised. In the latter case especially, this could be a major reason for the oft cited complaint that the schedules were too compressed to begin with (GAO 1991c, 6, 1993b, 4, 1989, 21). When the technician provides an estimate, he may be factoring in the additional testing that will prove he did everything in his power to deliver a functioning system that meets requirements.

At a project level, when looking at individual estimates in the surveys provided, it was noted that several of the estimates followed a pattern where the same or similar best case (BC), worst case (WC), and most likely (ML) estimates were provided for several activities in a row. This could be a form of anchoring where a subject is trying to formulate an estimate quickly, and once an initial number is settled on, that value will is used over and over for similar activities and may, in fact, be affecting the estimates of all subsequent activities. This could unknowingly influence all of the values in the estimate, depending on how carefully the estimate is made. Although it would be only ineffective to allow an estimator to provide one estimate at a time, the anchoring effect is something all estimators should be mindful of as they provide the estimates.

As stated in the GAO reports and in the responses from the subjects in this study, there is a tendency to plan to high-level schedules (GAO 2014, 25). If these schedules are too high level, then the activity anchoring problem can be compounded by estimators making different assumptions about what is meant by the activity title. From the scheduling survey responses, it was shown that the biggest disagreement among the survey subjects came from determining whether or not all required activities were included in the activity list. Many subjects felt that activities were left out, while still others believed that the activities were included, but that they may have been rolled up into one overarching activity. By not explicitly calling out specific activities, assumptions from different stakeholders can result in significantly different estimates (Mumpower 1996, 194). This would be akin to estimating the drive time from College Park to Dulles International on Friday afternoon at 4:00pm by only taking until account the miles to travel. Without having "technical" knowledge of the area and the associated traffic patterns, this would seem like a good measure. Experts, however, know differently.

It is this expert knowledge of the nuances of the activities that is critical to accurately assessing how long an activity should take (Moder and Rodgers 1968, B-79). And when these experts are also responsible for completing work on multiple different projects, it is less likely that they will be able to spend their time carefully planning out those nuances. As a remedy to this, even if the project team cannot be involved in the initial planning of the project, it may be helpful for the project manager to develop the initial schedule and then provide it individually for each project stakeholder to review as they are available.

221

From the Scheduling surveys, it was seen that subjects were in general agreement that the provided lists were good starting points for the schedule (i.e. very few felt there were extraneous activities on the lists provided). One option is for project managers to provide the high-level schedules individually to the project team members and ask them to list out the activities they believed were required to meet the high-level milestone. Private initial assessment ensures that no one person can dominate the discussion and squelch the quieter members of the group Baecher 1999, 20; Surowiecki 2005, 29). Before obtaining the estimates, however, it is helpful to understand some of the expected estimating behaviors of those who provide the estimates.

7.2 Stakeholder Responses: What to Expect

This section considers a discussion on how demographic traits influence personality traits and scheduling duration estimations. This section combines the results of the Demographics and Scheduling surveys and interprets results of Chapter 5.

7.2.1: The Influence of Demographics

The present results suggest that a stakeholder's Position Demographic exerts a heavy influence. This was shown in multiple analyses of the estimation data collected subjects. Analysis showed that managers traditionally provide lower duration estimates than do technicians. For each project, when activity durations were summed to find the total project duration (T_e), the sample population showed

that 81% of the time, managers estimate a lower T_e than technical subjects did.

Performing a binomial test on the results indicated that the probability of observing

this behavior in the general population was statistically significant ($p<0.001$). The

other two demographics did not show any statistically significant differences among

demographic levels.

The pattern repeated itself when looking individually at each of the three

estimates which compose a PERT mean. In this case, when the estimates were

compared within each demographic, each case showed statistically significant

differences among the levels of each demographic. For example:

- Managers provided smaller estimates than technical subjects

- Those with fewer years of experience provided smaller estimates than
  those with more years of experience

- Those with a more formal education provided smaller estimates than
  those with a less formal education.

Because these comparisons were made at the activity level of each project, there

should not be any concern that the results are skewed by comparison of dissimilar

activities, especially in the case of the ML estimate. The BC and WC estimates may

have bias, since they are based on the ML estimate, but they demonstrate the same

pattern. When using Equation 2-1 to solve for a project duration mean, even if the

ML values are identical between two estimators, a smaller BC and smaller WC

estimate will result in a smaller overall mean. It also stands to reason that if the ML

value of an estimator is smaller, in most cases the two outlying estimates will also be

smaller, again resulting in a smaller activity duration.

223

Although each of the demographics showed statistically significant differences, given that the smallest p-value belonged to the Position demographic by several orders of magnitude, it would appear that demographic is the driving force behind the differences in estimates. DOE could not be used to confirm these results because the sampled projects were very different from one another in both size and complexity. For example, one project may have a larger total duration simply because it is a larger project and not because the subjects are responding in a certain manner.

When the magnitude of the difference between the ML and BC estimates were compared, only the LoE demographic produced statistically significant results. In this case, the results showed that the magnitude of the difference between the ML and BC estimates of a subject with more formal education is likely to be a smaller than a subject with less formal education. Given an equal ML estimate, this means that estimators with less formal education cluster their BC estimates more closely to their ML estimates than the estimators with more formal education.

When the difference between the ML and WC estimates were compared, the Position demographic again was the clearest effect. In this case, the results showed that technicians are less optimistic than the managers as indicated by the wider separation between the ML estimate and the WC estimate. This spread is representative of the contingency the estimator believes is required to account for project unknowns. These results indicate that technicians are accounting for more things to go wrong than the managers. The other two demographics were also statistically significant, but with larger p-values. These results show that the largest

estimate will most likely be provided by a technician with many years of experience and modest formal education.

After the model was selected (GEV Max, GEV Min, or Normal) and after the parameters were set, the probability density between the BC and WC values was evaluated. When the differences were evaluated using DOE, the Position demographic once again showed statistical significance ($p<0.0490$). In this case, the expected value for the managers was 0.97 and 0.98 for technicians. While these values are separated by only a very small margin, they are another indicator that technicians are accounting for a wider range of possible durations than the managers. The technicians are leaving less density in the tails of their estimates, so they are less likely to be surprised when the project is actually completed.

When comparing the variance as calculated by the squared value of Equation 2-2, the results showed once again that the Position demographic was driving the differences in estimates. A smaller variance indicates less uncertainty in an estimate (Morris 1977, 688). The estimator is not compensating for things going wrong or for things going right. The larger variances in the technical group correlate with the results seen earlier when looking at each individual estimate. With a larger ML estimate than the managers, and a larger separation between the ML and WC estimates, these results again demonstrate that technicians are providing larger estimates and also accounting for more uncertainty in their estimates.

When the contribution ratios of the outlying estimates were calculated using Equation 3-5 through Equation 3-7, the Position demographic was once again the driving factor, except in the case of Equation 3-6; the contribution ratio using the BC

estimate showed no significant factors. Equation 4-3, Equation 4-4, and Equation 4-7, Equation 4-8 went on to show that to calculate the BC and WC values, given only an estimator's ML value, the multiplying constants for technicians for the WC estimate was larger than that of the managers. Conversely, the managers had a larger BC multiplying constant than the technicians. While the multiplying constant to solve of the BC estimate was relatively close between the managers and technicians, the constant to solve for the WC estimate had a much larger separation. These results contradict Question 4 in Table 5-8, which showed that technicians were more likely to have a smaller BC estimate than technicians, but they confirm Question 1 (overall PERT estimates) and Question 9 (variance analysis). With a smaller BC multiplying constant and a larger WC multiplying constant, assuming equal ML values, the variance of a technician will be wider than that of a manager. Again assuming equal ML values, these multiplying constants will result in a larger PERT activity average when calculating the average using Equation 2-1. Results from Question 1 showed that typically the manager ML estimates are smaller than the technician's ML estimates, so the difference in the two estimates will be even larger.

The calculations described above also provided an indication as to how skewed an estimation is. If the separation between the ML and BC estimates is equal to the separation between the ML and WC estimates, the distribution will not have any skew. If the separation between the ML and BC estimates is smaller than the separation between the ML and WC estimates, the distribution model will be skewed to the right. Chapter 4 described that the expected value of the ratio calculated in Equation 3-5 was approximately 39% for managers and 30% for technicians,

indicating that technicians have a heavier positive skew than the managers. In this context, the heavier positive skew indicates that the technical estimators are compensating for more adverse uncertainty than the manager estimators. Adverse uncertainty is meant to describe project challenges as opposed to things that will clear the way for project success. The managers seem to have more hope that the good things could happen as well as the bad. The technicians seem less hopeful in their estimates, with their uncertainty manifesting itself as a larger WC estimate. Another interpretation of the data is that the technicians are less sure of their estimates. It may not be that they believe everything will go wrong, but uncertainty breeds caution and caution includes providing a larger WC estimate that is more likely to incorporate unknown issues (Goldratt 1997, 152; Golenko-Ginzburg 1988, 767). This ties into the discussion regarding the confidence estimates provided by the subjects which will be presented later in this chapter.

The expected value of Equation 3-5 was the result of performing DOE on the aggregated data of each of the subjects. The final expected value provided by DOE showed that, in general, estimates resulted in a positively skewed distribution. Breaking down this analysis, however, to look at individual activities tells a slightly different story. Looking at estimates for all of the activities provided, the subjects heavily favored the positively skewed model. This seems to indicate a general belief that things are more likely to go wrong on a project than they are to go right. When these results are further decomposed into the two factors of the Position demographic, the results are more interesting. These results show that managers are twice as likely (34% vs 17%) to provide an estimate that has either no skew or negative skew as the

227

technicians. This could be an indicator as to why managers and technicians disagree

on project durations. In this context, a distribution with no skew or negative skew

indicates a higher level of optimism than a distribution with a positive skew.

Managers are effectively saying that they believe things are either equally likely to go

right or wrong on the project, or that they even believe that there is a better chance

that things will go right instead of going wrong. This may also be an indicator that

managers feel they have more control of the schedule and can therefore drive the

schedule to ensure it meets the originally advertised completion time. (Kremer

2017a). Technicians on the other hand seem to believe things are much more likely to

go poorly and that more time should be allotted for impending doom. From the

results seen in this study, however, it is important to at least be aware of the

possibility of a negatively skewed distribution, especially when working with

managers.

7.2.2 Discrete vs. Continuous Confidence Assessments

When given the Scheduling Survey, subjects were asked to provide a value for

how confident they were in their assessment of the ML estimate, described as a

probability. The survey treated the ML estimate as a discrete value with an

associated probability that the activity would finish within an operating window about

that time (Tetlock 2005, 40; Önkal et al. 2003, 182–83). In hindsight, the survey

should have been more specific about the definition of "confidence," but none of the

subjects questioned the legitimacy of the way the confidence estimate was asked.

An indicator that subjects were treating the ML estimate as a discrete value could be seen in a few of the confidence levels provided. Most subjects provided confidence levels in the 70%-90% range, but a few provided confidence rates much lower, some even below 50%. It would seem that if a person provided a probability of less than 50%, then there should be some other value on the positive number line that they believed would more accurately represent the actual duration of the activity. By assessing a probability of less than 50%, these subjects seemed to indicate that they were treating the ML estimate as an event. An event either happens or does not, as opposed to a continuous variable that can move within some range (Murphy and Winkler 1977, 45). These low probabilities show the subject treated the ML duration as an event finishing within the rounded-off time. If the subjects were not treating the ML estimate this way, they would have adjusted their ML estimate to a different value that they believed was closer to how long the activity would take. Another possible interpretation of the confidence estimates is that they were tied less to the actual estimate and were actually reflective of the estimator's confidence in his estimating ability. In cases where the estimated probability was very low, it could be that subjects were so uncertain about the activity that they were effectively providing a uniform probability for all possible durations. If this is a more accurate interpretation of the confidence estimate collected in this study, Hubbard suggests using Fermi decomposition to approach the desired confidence range (Hubbard 2010, 10–12). It should also be noted that Kahnemann and Tversky's anchoring concept seemed to apply here as well. In most cases, the same confidence level was used for

all activities estimated by the subject. The first estimate provided may have heavily influenced subsequent estimates as the subject anchored on that first estimate.

The variance, as represented by the BC and WC estimates is indicative of an estimator's level of uncertainty (Morris 1977, 688). The larger the separation between the BC and WC estimates, the larger the uncertainty. Assuming the first interpretation of the confidence estimate is correct, if the subjects were treating the ML estimate as an event with a probability of occurring, then the BC and WC estimates are compensating for the uncertainty in that estimate. For example, if a subject estimates that an activity will most likely take 10 hours and she is 90% certain that it will take 10 hours, then she believes that there is a 10% chance the event will take on some other discrete value. The BC and WC estimates are an attempt to account for that other 10%. The less certain an estimator is about the ML value, the wider the spread of the BC and WC estimate should be to account for that uncertainty.

In testing the correlation between the confidence level (interpreted as above) and the standard deviation (calculated using the PERT approximation), there was a -0.3 correlation between the confidence estimate and standard deviation.

When the confidence estimates were analyzed using DOE it was discovered that this was one of the few cases where the Position demographic was not the significant factor driving the results. YoE was the driving factor. Assuming the interpretation of confidence as a probability of a discrete event is correct, averaging across projects and subjects revealed that confidence hovered around 75% for subjects in the first half of their career, but rose to 85-90% in the second half of their

careers. When broken down further, confidence started out at an average of 73% in the first eight years, increased to 77% in the second eight years, increased again to 86% in the third eight years, and finally topped out at 89% in the 24+ years category. This pattern suggests that as subjects progress in their careers, they gain confidence in their estimating abilities (which may not be warranted) (Shanteau 1992, 12; Trumbo et al. 1962, 68–71). This could be for several reasons, but the most likely explanation is that as subjects gain experience, they learn what to expect on certain types of projects. As subjects learn what to expect, their uncertainty decreases because they began to get a feel for problems that were originally unknowns. As the uncertainty decreases, they become more confident that the estimate provided is sufficient to account for uncertainties and also matches historical completion times as experienced by the subjects. Because the significant factor driving confidence was YoE, this would indicate that this increase in confidence levels applies to both managers and technicians (Kremer 2017a).

7.2.3 Risk Aversion

If confidence is a measure of uncertainty, is high confidence indicative of knowledge or does it reflect a subject's belief about his or her ability as an estimator? The original hypothesis was that management subjects would be less risk averse than technicians, by dint of the presumed personality types of managers and technicians.

Risk aversion manifests in a concave utility function (Raiffa 1968, 68). Both management and technicians were presumed to exhibit risk aversion, but the supposition was that management would exhibit less than technicians, and thus have less concave utility functions.

231

The results did not confirm these hypotheses. From the full set of results in Appendix A.7, subjects, regardless of their demographic, tended to be to the left of the CME line, indicating risk averse behavior. Most subjects, however, demonstrated some convexity in their curve, indicating risk prone behavior.

As Raiffa points out, this behavior has been documented in several studies and is not unexpected. (Raiffa 1968, 95) When the results were reduced to reflect only the mid-point estimates, the results still did not show a clear delineation among the different demographics. As can be seen in Figure 5-1 and the results in Table 5-8, there was no statistically significant difference between the responses of the management group or the technical group, nor did the other two demographics show any statistically significant differences (see Figure 5-2 through Figure 5-5. It appears that subjects with differing levels of risk aversion, as measured through a monetary bet, can be found at different leadership levels, experience levels, and education levels.

While Utility did not have any significant factors driving the results, and confidence appeared to be driven by the YoE demographic, there was a weak correlation between risk aversion and confidence. In this case, it was hypothesized that risk-seeking subjects are, in general, confident by nature because in order to seek out risk, one must believe one will succeed (Raiffa 1968, 94–95). A person with a high level of confidence in their general abilities will probably also have high confidence in their estimating abilities. With no regard to demographics, the correlation between the two traits was 0.14 for cases with three or more projects, indicating that these two personality types may be weakly linked.

Although this Utility assessment was developed assuming the use of personal money, and the projects studies were funded through government money, it is believed that a Utility curve created by using personal money is still a good assessment, because as subjects provided estimates for the projects, they became personally invested in those estimates. Now, instead of money, reputation is attached to the project, so it translates the government project to the personal realm. Ariely conducted a study where subjects were asked to perform a menial task. Upon completion, for some subjects the results of the work were destroyed as soon as they were completed while others' work was left alone. Those who saw their work destroyed were less motivated to continue working. This simple experiment is an indicator of how quickly people can become invested in their work. (Ariely 2009a, Loc 883-1015)

7.2.4 Risk Aversion as Applies to Scheduling

If schedule risk is defined as the project finishing either above or below the extreme estimates provided by an estimator, and if risk averse subjects typically take action to mitigate a risk, then it was hypothesized that subjects who exhibit risk averse behavior would have a larger separation between their BC and WC estimates in an effort to compensate for as much uncertainty as possible. If variance is regarded as an, "I told you so" buffer where an estimator can claim success as long as the actual duration falls between the BC and WC values, then it stands to reason that a risk averse person will make that range wider to increase the chance of the actual duration falling within that range.

While Table 5-8 shows that there are no significant factors driving the levels of risk aversion, it also shows that for standard deviation, the Position demographic was the only one of the three where one group was consistently behaving differently from the others within the demographic. In this case, when looking at each individual activity, the managers were providing smaller variances than their technician counterparts. When the results of these two surveys were tested for correlation, the resulting correlation coefficient was -0.21 when an average correlation coefficient was calculated using only projects with three or more subjects. Although the correlation is weak, it indicates that at some level, personnel described as risk averse based on the Utility model are providing wider estimate ranges. Risk-averse estimators may believe this helps to ensure that any unknown contingencies are accounted for and decreases the probability the estimator will be blamed for a bad estimate.

Closely related to the correlation between utility and standard deviation is the correlation between Utility and T_e. The T_e value is heavily influenced by the ML estimate since, as can be seen in Equation 2-1, the ML value is given four times the weight of the other two estimates. As was seen in Table 5-8, the Position demographic is the driving force behind the differences in responses for both the variance and the T_e estimates, while Utility was not being driven by any particular demographic. From those results, it was shown that typically managers were providing smaller T_e values and also providing smaller variances than the technicians. As was shown previously, there was a negative correlation between utility and

standard deviation. A slightly stronger negative correlation (-0.29) was also observed when comparing Utility and T_e for projects with three or more subjects.

While T_e is partly composed of the BC and WC values that are also used in the calculation of standard deviation (see Equation 2-2), the value is dominated by the ML value since its weight is four times more than the other two estimates. While the correlation between Utility and standard deviation reflects how subjects compensate for uncertainty by widening their estimated range, this value also factors in the magnitude of the estimate itself. Given that the ML estimate is the driving force behind the T_e value, it can be inferred that not only are risk averse personnel providing a wider range in their estimates, but they are also providing larger ML estimates. A larger ML estimate negatively correlated with risk aversion indicates that personnel who are risk averse are more concerned with a project finishing late than they are with it finishing early. For these people, success is defined as finishing before the deadline. The best way to mitigate the risk of finishing after the deadline is to provide a larger ML estimate that will be less likely to be exceeded.

7.2.5 Summary

The past sections described how personality traits and estimating practices could be observed at varying levels of the three demographics. It also showed how personality traits and estimating practices relate to one another. In a scheduling context, the data indicate that subjects in the technical career fields provide higher estimates and wider variances than those in management.

7.3 Aggregating Estimates

There is considerable literature on both expert aggregation and improvement of schedule estimations, but the two fields do not seem to intersect. The literature on scheduling provided many ways to improve on the PERT equation, but seemed to focus on estimates derived from a single person. Although there will be overlap, each stakeholder in a project brings a slightly different view with slightly different information (Surowiecki 2005, 9–10; Budescu and Rantilla 2000, 373–74). The "unknown unknown" of one stakeholder may be a known issue to another (Silver 2012, 420). Using all available information gives a decision maker the best possible chance of accounting for uncertainties. Aggregating these estimates, however, presents challenges.

7.3.1 The PERT (Beta) Prior

The creators of PERT needed a way to account for uncertainty in schedules and they only had a month to develop that method. They believed that the best person to provide duration estimates was the "technical man" who was actually doing the work. They also knew, however, the "technical man" was busy doing technical work and did not have much time to devote to the development of a schedule (Malcolm et al. 1959, 648, 650, 659). They were also aware that most people do not have training in probability assessment and would need an easier way to express their uncertainty (Chaloner and Duncan 1983, 174; Clark 1962, 406). To that end, the creators of PERT settled on a method that required an estimator to simply provide BC, WC and ML estimates (Malcolm et al. 1959, 651). These were used to develop a distribution curve for the activity using the beta distribution.

236

The recommended three-point estimates in PERT for calculating the expected value and variance of an activity duration are given in Equation 2-1 and Equation 2-2 (Malcolm et al. 1959, 651–52; Mantel Jr. et al. 2004, 144; PMI 2013, para. 6.5.2.4). These are simple equations, easily remembered and applied. The creators of PERT did not start out with a particular distribution, but, as Clark pointed out, they needed something to model the probability distribution of the activity duration and a beta distribution fit the need (Clark 1962, 406). It is flexibly, can be uni-modal, modeling the tendency of an activity to have only one ML value, and it tapered off at the tales, which modeled how the probability decreased as it moved away from the mode (Malcolm et al. 1959, 651).

To fully define a beta PDF the two hyper-parameters, α and β must be known. Having these parameters allows the calculation of the mean and variance of the distribution. Without proper training, it can be very difficult to estimate these parameters, so Equation 2-1 and Equation 2-2 were developed in an attempt to create an approximation to the actual mean and variance of the beta distribution (Regnier 2005b, 6).

While the PERT approximations served their creators relatively well, once statisticians began to take a hard look, the model began to break down (Grubbs 1962, 913). The PERT distribution is based on an estimate provided by an imperfect human. As the PERT methodology was analyzed more closely, it was pointed out that the estimates provided by the experts were just that: estimates. There was no way to know if the BC estimate used in Equation 2-1 was actually the true absolute shortest duration of an activity or if it was just the belief of the expert that it was the

shortest (Grubbs 1962, 914–15). The same logic applied to both the ML and the WC estimates. Pickard provided a solution to that issue, but the data collection and approximating algebra are much more complex than the simple PERT formula (Pickard 2004, 1571–74).

These approximations are only exact when the sum of the hyper-parameters equals six (Grubbs 1962, 914). Keefer and Verdini did a demonstration comparing how well Equation 2-1 and Equation 2-2 approximate the true mean and variance of the beta distribution. They confirmed that the PERT approximation worked when the hyper-parameters summed to six, but as the summation value increased, the approximation is less close. The variance approximation was even worse (Keefer and Verdini 1993, 1087–88). Grubbs pointed out that this constraint put a limitation on the beta distribution which had been chosen for its versatility (Grubbs 1962, 914).

Several researchers have proposed new approximating equations for the beta distribution which allow for a wider range of hyper-parameter values (Pearson and Tukey 1965; Golenko-Ginzburg 1988; Megill 1971). Keefer and Verdini provided a summary of several of these approximations and their capability to accurately approximate the true beta mean and variance (Keefer and Verdini 1993, 1087–88). The two most accurate approximations (Equation 2-5 and Equation 2-6) were no more mathematically complicated than the PERT estimation, but use the median and varying endpoint fractiles. Unfortunately, they still ran into the same problem as all the other approximations: complete dependence on the estimates of flawed experts coupled with an unknown underlying distribution (Regnier 2005b, 8).

The creators of PERT chose the beta distribution because they believed its general shape matched what could be expected for an activity duration (Malcolm et al. 1959, 651; Clark 1962, 406). Its versatility also had the added benefit of enabling it to model what in this research is referred to as the optimists, the pessimists, and the neutrals simply by adjusting the hyper-parameters. While the creators of PERT appeared to be leaning towards a Bayesian construct by modeling the belief of the estimator (Malcolm et al. 1959, 651), others approached the model from a frequentist perspective and struggled with the fact that the true underlying distribution of the activity was unknown (Pickard 2004, 1568–73; Grubbs 1962, 914–15; PMI 2013, para. 6.5.2.4). Given that activity durations must have a lower limit, but could technically remain uncompleted for an extended period of time, it would seem that a frequentist distribution of an activity would be skewed to the right. From this research, however, it was seen that estimators would sometimes provide BC, WC, and ML estimates that were more accurately modeled by a left-skewed or no-skew distribution. In the frequentist view, the estimates provided may not accurately reflect the true mode or extreme estimates of the distribution of historical durations. In a subjective context, the model reflects the beliefs of the expert.

7.3.2 Bayesian Prior

Estimation of the mode has been criticized because it can be difficult to accurately estimate the mode of an unknown distribution from data (Golenko-Ginzburg 1988, 770). In the Bayesian (degree-of-belief) sense, the mode is easy to determine because it reflects the estimator's belief in the ML duration (Chaloner and Duncan 1983, 175). If the estimator believed that a different value had a higher

239

probability of occurring, then the mode would move to that value. The end-points also become reflective of a belief about the activity instead of the actual smallest and largest durations of an activity.

Interpreting the three baseline estimates as a function of Kahneman's System 1 and System 2, may explain why estimating the median may not necessarily be easier than estimating the mode. The three PERT baseline estimates could be interpreted as a function of System 1, whereas the other approximations are more a function of System 2. For example, when someone asks an expert how long something should take, a number will immediately pop into his head, based on past experiences or analogous estimating (Kahneman 2011, 24; Clark 1962, 406; PMI 2013, para. 6.5.2.1-6.5.2.2). The same holds true for asking for a BC or WC estimate. Asking an expert to provide a number for which half the time the duration is below that number and half the time the duration is above that number, however, involves more careful thought (Chaloner and Duncan 1983, 175). A number will probably present itself in the expert's mind, but he will need to stop and consider it in relation to all other possible durations to determine whether or not it is the median. Even if the median is considered a better statistical assessment (Keefer and Verdini 1993, 1088), it is still an assessment.

Golenko-Ginzburg has argued that estimating the mode maybe be unnecessary if the two endpoint estimates are provided. He showed that when asked to provide duration estimates, the mode typically fell in a location calculated by (2BC + WC)/3, that is, at the one-third point (Golenko-Ginzburg 1988, 770). This experiment was repeated in the present project, and similar results were found: there was no

statistically significant difference between the calculated mode and the actual mode, even when comparing the management subjects who had a more varied skew type. These results were calculated using Excel's ™ "t-test: Two-Sample Assuming Unequal Variance" in the Data Analysis Add-On which compared the mean of the ML estimates provided by the subjects as compared to the results that would be obtained by using the calculation described above (Golenko-Ginzburg 1988, 769).

Golenko-Ginzberg asserted this meant that the ML estimate was unimportant in calculating the expected value of activity duration (Golenko-Ginzburg 1988, 769). In the present research, when providing estimates, several estimators omitted the ML estimates on their survey sheets. The reason is unknown, but when an estimate was left out, it was always the ML estimate, perhaps further indicating that the location of the mode is not considered a useful measure. When using the distributions proposed in this research (GEV or Normal), however, the mode serves as the location parameter of the curve and becomes more important in modeling the estimator's beliefs.

Despite the criticisms of the mode, it is useful to remember the reason PERT was originally developed. The technique developed by the creators of PERT allowed the "technical man" a way to provide a quick estimate without having to take the time for a thorough statistical analysis (Malcolm et al. 1959, 659; Clark 1962, 406). Certain activities must be scheduled based on a set number (e.g. delivery dates, work crew arrival dates, etc.). Because most people do not spend their days calculating the expected value of an unknown quantity based on their perceived range of values, they have only their "ML" estimate by which to live.

7.3.3 A New Prior Model

For the all of the beta distribution's versatility, it has two major drawbacks. The first is the requirement to estimate the hyper-parameters which describe the shape. The second is its limited domain. Because the beta distribution is only defined between its two endpoints and is zero elsewhere, using it in an expert aggregation context is a challenge unless the experts have the same BC and WC estimates. In its standard form, the beta distribution is only defined on the interval from [0,1] ("Beta Distribution" 2016). This interval can be converted using the change of variables described by Grubbs such that the points [0,1] are redefined as the BC and WC estimates, but this involves extra conversion steps when aggregating expert opinion (Grubbs 1962, 913). For example, one expert could have a BC estimate of 10 hours and a WC estimate of 25 while another could have values of 15 and 40. Plotting both of these distributions on the interval [0,1] would yield an inaccurate representation of the distributions. To accurately display the distribution, the estimates would need to be converted to their actual locations on the number line. In addition, given the definition of the beta domain, it can be seen that, using Morris' aggregation method, there would be no probability density for anything below 15 or above 25 because one or the other of the two distributions is zero beyond those points.

The fact that PERT remains a popular method of accounting for project schedule uncertainty seems to confirm the belief that the overall shape of the model is good. These considerations were the driving force behind the selection of the GEV Max, GEV Min, and Normal distributions as the new models for the subjective

duration assessment. These three distributions maintain the general unimodal/tapering shape of the beta distribution, but have the added advantage of being defined along the entire real number line. Between the three models, it was also possible to account for all three skew types seen in the expert estimates. The necessity of using three different distributions complicates matters slightly, but if a decision maker knows what she is looking for, these complications can be overcome. These distributions present advantages over the beta distribution in the aggregation. First, all three distributions are defined on $(-\infty, \infty)$ so, even though the density may be negligible, there is no point at which the density is completely zero. In the aggregation, this is critical to ensuring that the posterior density exists everywhere.

Another advantage of using the GEV Max-GEV Min-Normal model is that its defining parameters are directly related to the estimates provided by experts. For the GEV Max and GEV Min models, the distribution is defined by three parameters. The first parameter, ξ, was set to zero which allowed the distribution to be defined along the entire number line. This meant that at least some density could occur below the BC estimate and above the WC estimate, acknowledging that the person providing the estimate may not have fully accounted for extreme cases. The second parameter, μ, is simply equal to the mode of the distribution which is the ML estimate provided by the expert. No special considerations are needed for this parameter as it translates directly from the estimate to the distribution and locks the distribution into the appropriate location on the number line. The shape parameter, referred to as "k", is directly dependent on the initial estimates provided by the expert as can be seen in Equation 3-20 and Equation 3-22.

With these parameters defined, the mean and variance for the GEV Max/GEV

Min distributions can be directly calculated without the need for an approximating

equation. For the Normal distribution, defining the distribution is even simpler. The

location on the number line is once again defined by the ML estimate and the

variance is defined by solving for the separation between the ML estimate and either

endpoint and dividing by three.

While the defining parameters for the GEV Max/GEV Min distributions are

tied directly to the baseline estimates, it should be noted one major assumption was

required in order to determine the shape parameter used to define the distribution.

The intent of the PERT creators was that there should be negligible density below the

BC estimate and above the WC estimate (Malcolm et al. 1959, 651). Based on the

acceptance of the general shape of beta distribution, the goal was to make the new

distribution models match the shape of the beta distribution as closely as possible.

From Equation 3-15 and Equation 3-17 the value of F(x) was set at 0.0001 for the

GEV Max distribution and 0.99995 for the GEV Min distribution.

For the GEV Max distribution, setting the CDF to 0.0001 ensured that the

density below the BC estimate was negligible, but not zero. The value of 0.0001 was

selected by plotting the GEV model in MatLab™ and selecting the value at which the

curve begins to rise above the x-axis. As can be seen in Figure 6-1, this is roughly

equivalent to the beta distribution starting to have a density at its BC value. The same

logic was applied to the GEV Min case, where the value of 0.99995 was roughly the

point where the curve began to collapse back down to the x-axis (see Figure 6-3).

The advantage of this method is that if there is a belief that more density should be

below the BC or above the WC estimates, the calculation of "k" is still possible. The decision maker would simply need to change the value in the denominator of Equation 3-20 or Equation 3-22, depending on the model in question (and noting that Equation 3-21 and Equation 3-23 would need to be reworked to reflect the new density value selected). If the intent is to match the general shape of the PERT distribution, however, these values should be reasonable estimates.

One disadvantage of the GEV Max/GEV Min distributions is that there are three baseline estimates to model, but only two distribution parameters to describe the location and shape. Because one of those parameters is taken up with the mode, that leaves only the shape parameter to match either the BC estimate or the WC estimate. For the GEV Max case, the choice was made to match the BC estimate. Given the results seen in the GAO reports, and the opinions provided by subjects in this research, it was deemed unlikely that the project would finish earlier than the BC estimate. Using these assumptions and looking again at Figure 6-1, it can be seen that the GEV Max model matches the BC estimate relatively closely and matches the mode exactly, but that the curve does not collapse back down to the x-axis until considerably after the provided WC estimate.

For a typical network schedule, unless the decision maker is using a Monte Carlo simulation, the only required values from the distribution are the mean and the variance. For the GEV Max model, the mean and variance are calculated using μ and "k" and "k" is only dependent on the BC and ML values, which are provided by the expert. Because of this, the fact that the WC value is less accurately modeled is not as critical.

The skew of the expert estimates was taken as indication of the certainty in an expert's outlying estimates. For example, a right-skewed estimate (as modeled by the GEV Max distribution) indicated that an expert was relatively confident that the duration would not be less than the BC estimate, but that there was more uncertainty about the WC estimate. The longer tail to the right of the mean accounts for that uncertainty. A left-skewed estimate (as modeled by the GEV Min distribution) was an indication that there was relatively high confidence that the duration would not exceed the WC estimate, but that there was less certainty about the BC estimate, leading to the longer tail to the left of the mean.

For the GEV Min distribution the shape parameter was calculated slightly differently, based on the discussion in the preceding paragraph, the desire to continue to match the beta distribution, and the fact that trying to match the BC estimate using the 0.0001 value resulted in a distribution that did not appear to accurately model either end-point estimate. For the GEV Min distribution, the decision was made to match the ML and WC estimates and leave the BC estimate as the "floating" estimate. In this case, the shape parameter is defined by the ML and WC estimates, so the mean and variance are defined by the two parameters in which the expert seems to hold the most confidence. In these cases, it is believed that the left-skewed nature of the distribution indicates the estimator feels less need to account for the duration exceeding the WC estimate. Although this is not typically the case, experts who provide estimates of this type are either extremely optimistic or they have some information that gives them more confidence that they will not exceed their WC estimate.

The GEV Min model has a drawback in Monte Carlo simulation. Because the domain of this distribution is the entire real number line, if the activity duration is relatively short (roughly less than ten time units), the distribution could result in appreciable probability density assigned to negative duration values. A solution for this issue is suggested in Chapter 8.

7.3.4 Calibrating the Experts

One problem with eliciting expert probabilities is that experts are rarely calibrated (Morris 1977, 682; Baecher 1999, 4). Part of the process in Morris' method requires the decision maker to calibrate the expert either empirically or subjectively. For this research, data was not available to empirically calibrate the experts, so a subjective calibration method was developed. While the beta distribution's limited domain was problematic for aggregating the experts, it was perfect as a calibration filter. Because the CDF of any probability is only defined on the interval [0,1], the beta distribution was an ideal match for processing the input data ("Beta Distribution" 2016). To calibrate the expert, Morris' method passes the CDF of the prior through the calibration filter (Morris 1977, 682–86, 689).

In an example provided in Morris' 1977 article, he used curves that were described by polynomial equations to describe experts who were either over-stated or under-stated in their knowledge, but the general shape resembled a symmetrical beta distribution (Morris 1977, 691). Using that example as a baseline, it was determined that the beta function's versatility allowed for a single equation to describe the entire calibration range, including the case where no calibration was required, simply by adjusting the hyper-parameters of the defining equation of the beta curve.

Empirically, for a continuous variable, a person is fully calibrated if actual values fall within the correct fractile. In a scheduling context, the expert is well calibrated if across several assessments actual values occur at the frequency predicted by the expert (Tversky and Kahneman 1974, 1128–29). For example, across several activities, actual durations estimated to be below the 0.1 fractile should only occur less than 10% of the time. If they occur more often than 10% of the time, then the expert requires calibration to make reality match the prediction (Baecher 1999, 5–6).

Morris contends that subjective calibration is also possible in the absence of empirical data (Morris 1977, 689). In this case, the decision maker assesses her opinion of the expert to determine whether or not she believes the expert will be surprised by the actual results, where surprise is defined as the actual value being anything below the 0.1 fractile or above the 0.9 fractile (Morris 1977, 691). She can calculate the duration value that falls within a particular fractile using the CDF of the prior distribution and determine from that value whether or not she believes the expert will be surprised by the results. To that end, the intent of the calibration function is effectively to adjust the variance by altering the fractiles of the extreme estimates.

Passing the CDF of the expert's prior through the beta filter alters the duration values that fall at the 0.1 or 0.9 fractile. It widens the values for experts the decision maker believes to be overstating their knowledge and shrinks the values of those she believes to be understating their knowledge (Morris 1977, 691). Another advantage of using the beta curve as a calibration function was that it allowed the expert's mode to remain unaltered as it passed through the filter. Because the intent was to only

affect the extreme values where the expert could be "surprised", the values for the

two beta hyper-parameters provided in Appendices 8-10 were calculated with the

constraint that the filter could not affect the mode of the prior distribution. Based on

the above discussions regarding the importance of the mode in subjective probability

assessments, it was believed that the mode should not be altered by filtering. If both

the mode and the endpoints were altered drastically, then the purpose of eliciting an

expert's opinion would be rendered moot.

Although subjective calibration was used for this research, Equation 3-39 and

Equation 3-40 could possibly be used as a point of reference for how well calibrated

the expert is. These equations are used to determine the density between the BC and

WC estimates for the GEV Max and GEV Min distributions, respectively (a Normal

distribution will always have a density of 0.997 based on the 3-sigma assumption

used in this research (NIST 2017a). For the GEV Max model, due to the assumption

that the density below the BC estimate was 0.0001, the value of Δ from Equation 3-39

can only change based on the location of the WC estimate. The further away the WC

estimate is from the ML estimate, the more density is accounted for between the two

extreme estimates. If the decision maker has a point of reference for what is

considered an appropriate value of Δ, the location of the WC estimate could possibly

be used to calibrate the expert. For the GEV Min case, the same holds true, except

that due to the assumption that the density in the range (-∞, WC] is 0.99995, Equation

3-40 can only change based on the location of the BC estimate. As in the GEV Max

case, for a GEV Min prior, if the decision maker has a point of reference for her

beliefs regarding how much density falls between the BC an WC estimates for a well-calibrated estimator, she can use these as a point of reference for calibration.

Current project management software allows for the adjustment of the weights typically used to calculate a PERT Mean (Equation 2-1) (Microsoft 2017). This adjustment allows a project manager to account for uncertainty by changin the weights of the three estimates (e.g. if the pproject manager anticipates problems, the weight of the WC estimate can be increased and the weight of the ML estimated decreased). Changing the weights, however, may violate the assumption of the mean of a beta distribution as approximated by Equation 2-1, so the method should be used with caution.

The method proposed in this research is intended to model the subjective probability of the assessor. Because these are subjective assessments, a DM can best account for uncertainty through the placement of the ML estimate. Instead of influencing the mean by weighting one estimate more heavily than another, the DM should consider her beliefs and determine a ML estimate accordingly. The mean for the two GEV models is then calculated using this ML value and the shape parameter (as calculated using the extreme estimates; BC for GEV Max and WC for GEV Min). For the Normal model, the mean will equal the ML value. With respect to the Expert's assessments, the DM can account for uncertainty through the calibration process. The calibration function will alter the mean of the posterior distribution depending on the DM's assessment of the Expert's estimating abilities. This allows the DM to effectively weight her faith in the expert because, due to the assumption of

independence, the estimate with the tigher variance will have the heavier the influence on the mean of the poseterior.

7.3.5 Posterior Distribution

Although the aggregation method was able to combine the estimates of multiple experts, there are things to consider when using the method. When aggregating expert inputs, it was noted that with the addition of each input, the variance of the posterior distribution gets smaller (see Figure 6-8 vs. Figure 6-17). From a sampling theory view, this is correct: based on the assumption of independence, as more experts are added, the variance shrinks because their "errors" cancel (Gelman et al. 2013, 32). If multiple experts are in relative agreement with one another, then the expert's uncertainty regarding the actual duration can be drastically reduced. Instead of accounting for a wide range of possible outcomes, the agreement among the experts means that the decision maker no longer needs to account for huge amounts of uncertainty.

In gathering multiple opinions, the decision maker has followed the advice to, "Trust, but verify". As the number of experts increases, the variance of the posterior distribution may become unrealistically narrow. This occurs as a function of the multiplication in Equation 3-28, but it may not accurately reflect the uncertainty of the decision maker, especially if the experts were not in reasonable agreement with one another and a wider variance is warranted to account for the disagreement.

Another issue noted in the calculation of the posterior distribution was the tendency of an outlying prior distribution to dominate the shape and location of the posterior. As with the shrinking variance, this phenomenon is a result of the

multiplication performed in Equation 3-28. In order to match the intent of PERT, an effort was made to concentrate most of the probability density between the BC and WC estimates of each stakeholder (both decision maker and experts). Although the density outside these estimates is non-zero, it is still very small. Because the shape of the curve of the posterior distribution is created by multiplying together the various PDFs of each stakeholder, for any given duration value, one small result of Equation 3-14, Equation 3-16, Equation 3-18 can dominate the result of Equation 3-28.

The dominating outlier will depend greatly on the prior probabilities of each stakeholder. For example, assume each stakeholder's prior is modeled by a GEV Max distribution. For each stakeholder, the density below the BC value is negligible, but there may be considerable density above the WC estimate. If there is a cluster of prior probabilities with their modes at the lower end of the number line with one outlier with its mode at the higher end, the negligible density of the outlier will collapse the larger densities of the clustered probabilities, even though they are in agreement with one another. On the other end of the spectrum, the three clustered priors may still have considerable density as the density of the outlier begins to substantially increase. Because each stakeholder has non-negligible density at that point, the posterior curve will also have non-negligible density which will cause the posterior distribution to more closely resemble the outlying prior distribution.

The exact opposite case will happen in the GEV Min distribution, where the prior with the smallest mode will dominate the posterior. For an outlier modeled by a Normal distribution, the variance of the outlier will play a role in how heavily it dominates the posterior. The wider the variance, the smaller the affect the outlier will

have on the posterior because the density of the outlier will be spread across a wide region and will be more likely to still have significant density when it encounters the cluster of expert priors probabilities exists elsewhere on the number line.

As was seen in Figure 6-20, if stakeholders are in severe disagreement with one another where one expert is modeled with a GEV Max distribution and a mode on the lower end of the number line and another is modeled with a GEV Min distribution and a mode on the upper end of the number line, the outcome of Equation 3-28 does not cleanly result in any of the three recommended approximation models. This is due to the multiplication of the priors from Equation 3-28. Because the two priors are "facing" one another, their peaks are on opposite ends of the number line, but each still has considerable density when they intersect in the middle. This results in an oddly shaped posterior that cannot be closely approximated by any of the recommended models (Winkler 1986, 302). In cases such as these the recommendation is to either base the approximation on the general skew of the calculated posterior curve or to approach the stakeholders with the disagreement and request a reassessment of the estimate (Winkler 1968, 70).

Chapter 8: Conclusions and Future Work

Projects are different, but the problems are the same. External influences such as resource constraints and unforeseen issues can cause a project to fall behind schedule, but the biggest issues come from within. Differences in opinion regarding how long an activity should take plague project managers as stakeholders vie to alter the planned duration based on past experiences and current constraints. This research sought to understand the differences among stakeholders regarding activity and project durations and to develop methods to bridge those differences in opinion.

8.1 Conclusions

Project stakeholders come from a wide variety of backgrounds, experience levels, and agendas. A major part of this research was to determine how these differences manifested in scheduling practices. To that end, project stakeholders from Wallops Flight Facility (WFF) were surveyed to gather information on demographics, risk aversion, project constraint preferences, and schedule estimation practices. A method was then developed which allowed the aggregation of scheduling estimates from multiple stakeholders. This aggregate estimate represented the sum knowledge of all participating stakeholders and allowed the project manager to create a plan based on input from her entire team.

8.1.1 Influence of Demographics

Several constraints were analyzed by comparing responses of stakeholders in different demographics to determine if different demographics responded in

predictable ways. When it came to project constraint preferences, Section 5.2.1 showed there was a clear preference among those polled to sacrifice schedule and cost for the sake of preserving quality and reducing risk. This correlated well with the literature from GAO reports seen in Section 2.1.1 which consistently showed the same results given the technical focus and culture of safety seen at NASA. This prevailing attitude contributes to scheduling difficulties in the face of challenges because any challenge that threatens technical quality or is perceived as a potential threat to personnel safety could cause a schedule slip (Martin 2012, 11). When constraints were looked at individually in Section 5.2.2, the results showed that only Schedule and Quality had significant factors driving the results. The Position demographic drove the AHP weights calculated for the Schedule constraint, where Table 5-6 showed that technicians were more willing to sacrifice Schedule for the sake of Cost/Risk/Quality than the managers. The LoE demographic drove the Quality constraint, where Table 5-6 showed that the those with more formal education were more ready to sacrifice Quality for the sake of Cost/Schedule/Risk than those with less formal education.

With respect to risk aversion, the results seen in Section 5.2.3 did not appear to be driven by any particular factor. Subjects across all demographic divides demonstrated a wide variety of risk behaviors, although several demonstrated the S-curve as described by Kahneman and Tversky (Kahneman 2011, 282; Raiffa 1968, 94–95).

Confidence was treated not as confidence intervals, but as probability assessments of a binary yes/no event. Section 5.2.4 showed that, when averaging

255

each participant's confidence estimates, confidence increased as experience increased. From these results, it can be concluded that personnel become more confident in their estimates and that they put more trust in their own judgment as they gain experience.

When it came to the actual schedule estimates, Table 5-8 showed that, with only one exception, the Position demographic was the major demographic driving the different responses. For both the PERT average and each element making up the PERT average, managers could typically be depended upon to provide smaller duration estimates of work activities than did the technicians. This leads to the conclusion that managers believed things could be done faster than the technicians believed they could be done.

When comparing personality traits, Table 5-8 showed the Level of Formal Education (LoE) and Years of Experience (YoE) demographics began to play a larger role. The conclusion is that, while certain personality traits tend to respond in certain ways, the primary driver in scheduling practices is the management/technical divide. There appear to be some weak correlations between personality traits and scheduling practices. Given all of this, if a Decision Maker learns the demographic of a stakeholder, she can update her belief about the expected opinions of that stakeholder, and based on the correlations seen in this research, have an idea of what to expect from that stakeholder's duration estimates.

Outside the scheduling estimates, Section 4.1 summarized several thoughts subjects had regarding general reasons why projects fall behind. One of the factors mentioned frequently by members of both the management and technical

demographics was the inability of project staff to focus on one particular project.

When staff were re-directed to other projects, it affected the original duration

estimates which were made assuming a certain level of availability of project

personnel (Gould 2005, 251). Other factors mentioned included aggressive schedules

and poor planning, which correlate to Kahneman and Tversky's planning fallacy

(Kahneman 2011, 246, 249).

From a separate survey described in Section 4.2, results showed that

management and technical subjects generally agreed that enough people were

assigned to the projects, but once again pointed out that those assigned must be

allowed to focus on the project at hand if schedules are to be met. Results from this

same survey showed that managers and technicians were also in relative agreement,

when provided a list of activities, that each of those activities were necessary to

successfully complete a project. The major disagreement between the two groups

resulted from a debate as to whether or not all activities required were provided in the

list.

The COA Survey described in Section 5.1 revealed a further disagreement

between the two groups regarding what constituted "necessary" work. Technicians

polled were more likely to regard extra troubleshooting as risk mitigations whereas

managers were divided as to whether or not extra troubleshooting constituted risk

mitigation or gold plating.

Based on these results, it is concluded that stakeholder perception plays a

major role in determining the time allotted for a given activity. During the planning

process, these perceptions may be driving some of the aggressive schedules that were

mentioned by project subjects and in some GAO reports (GAO 1980a, 52,37, 2014, 10) as well as some of the debates on how long an activity should take.

Table 5-8 showed that technicians typically provided longer estimates than the managers. Table 5-14 showed technicians also tended to provide estimates that fell into the "pessimist" pattern, where the prior distribution was right-skewed. Based on the results of the COA Survey in Table 5-2, they also preferred to have extra time to ensure their systems were operating at full capacity. Table 5-8 showed that managers, on the other hand, provided smaller estimates overall and their variances were smaller, meaning that their extreme estimates were clustered more tightly around their "most likely" estimate. Table 5-14 showed their prior distributions, although still favoring the "pessimist" model, were more likely to result in either the "optimist" or "neutral" models, where the distributions were left-skewed or did not have skew. Given that managers typically develop and drive the schedules, these results show that scheduling tendencies among the managers, coupled with the planning fallacy may be creating the aggressive schedules mentioned by many subjects in Section 4.1. Based on the results of the COA survey in Table 5-1 and Table 5-2, matters are further complicated by the fact that a manager may not perceive an activity as critical to the success of the project, while a technician may have the opposite view.

If a schedule is only developed to the milestone level, which, according to the GAO reports in Section 2.1.3 (GAO 2014, 25) and the survey results in Section 4.1, tends to happen, a technician may be trying to pack extra activities into the milestone which the manager did not account for, activities which the technician believes are

critical to the success of the mission. Combining these results leads to the conclusion

that, because time was not allotted for these activities and the manager may not

realize they are being accomplished, it may appear that the technician is including too

much contingency in the estimate. The technician, on the other hand, believes the

activities are critical and may not realize that the manager has no knowledge of the

activity. This leads to a perception that the schedule is too compressed. If this

happens frequently, the confirmation bias may also be reinforcing the opinion that the

schedule is always too compressed, as memories of projects completed early fade into

the background.

Managers, on the other hand, may be holding on to memories of successfully

completed projects which reduce their perceived contingency required to successfully

complete a project on time (Kahneman 2011, 80–81). Ultimately, it is critical for

project stakeholders to maintain open lines of communication to ensure that plans are

communicated clearly (Kremer 2017b). This can help ensure that perceptions of

required tasks are in-line with one another which will lead to overall better

scheduling.

8.1.2 Aggregating Estimates

Even with good communication, there will still be different opinions about

how long an activity should take. Given the diversity of project stakeholders, Section

3.5.3 described a specific application of Morris' method to aggregate expert opinion

as applied to the PERT methodology. This allowed for better incorporation of those

diverse opinions by calculating a single posterior distribution that encompassed the

assessments of all stakeholders (Morris 1977, 680–86). The procedure begins by

obtaining the three standard duration estimates, "best case", "worst case", and "most likely", used in the PERT method from both the DM and the Experts (Malcolm et al. 1959, 650–52). In the PERT methodology, these estimates would be modeled by a beta distribution (Clark 1962, 406). The new method, as described in Section 3.5.1, models these estimates according to one of three different distributions: the GEV Max distribution for the pessimists, the GEV Min distribution of the optimists, and the Normal distribution for the neutrals. In Section 3.5.1, equations were developed to convert the estimates provided into the parameters that would eventually describe the shape of the prior distribution once the overall model was selected. In this case, the estimate distribution was treated not as a frequency distribution of potential activity durations, but as a model of the stakeholder's beliefs about duration (Dennis V. Lindley 1983, 4).

Once the priors were established, Section 3.5.2 described a method of calibrating the Expert priors, again based on the work of Morris (Morris 1977, 682–84). The beta distribution was chosen as a descriptive model of the calibration filter. This allowed the mode of the calibrated prior to remain the same, but modified the variance to reflect the DM's belief regarding whether or not the Expert overstated, understated, or correctly stated his knowledge of the situation. Tables in Appendix A.10 – Appendix A.12 relate the beta parameters to the likelihood of surprise for the three distributions used to model the Expert's prior duration assessments.

Once the prior distributions are established and calibrated, a posterior distribution is calculated, using Equation 3-28, which is the final step in Morris' method (Morris 1977, 686). Given the algebraic complexities of determining the

mean and variance of this posterior distribution, Section 3.5.3 described the method

used to approximate the results of the posterior curve as calculated by Equation 3-28.

The three approximation models chosen were the GEV Max distribution, the GEV

Min distribution, or the Normal distribution. Section 3.5.3 describes the equations

used to calculate the parameters of these approximating distributions. These

approximations allowed for the quick calculation of expected value and variance,

which are required for the development of a network schedule.

Even with good communication, project stakeholders will have different

opinions about how long an activity should take. The posterior distribution derived

using this method is reflective of the collective knowledge of all participating project

stakeholders. Biases are more likely to cancel each other out (Surowiecki 2005, 10)

and each stakeholder will know that their input was included in the derivation of the

final schedule.

8.2 Future Work

The methods described above provide a synopsis of the diversity of project

stakeholders and a new method for aggregating those diverse opinions. There are

still, however, several areas ripe for further research and development. The sections

below provide a summary of those areas of improvement.

8.2.1: Participant Dependence

For the purposes of this research, it was assumed that the Experts were all

statistically independent when calculating the posterior. Research has shown,

however, that it is extremely difficult to have true independence given how humans

with similar backgrounds tend to cluster together (Winkler 1968, B-65, 1981, 480; Clemen 1987, 373). An avenue of further research is to determine the dependency of the Experts and develop a method to incorporate this dependency into the assessment of the posterior distribution.

8.2.2 Research Expansion and Refinement

This study gathered data from a very particular niche within the workforce. One area of further research is to expand beyond this particular group into other projects, not only at Wallops Flight Facility, but NASA at large to determine whether or not the trends noted here are local or global to NASA. Beyond that, expansion into other career fields is another area of exploration. Applying the same processes, personnel in areas such as construction, healthcare, and software development could be studied. For example, are there differences in the perception of how long it takes to care for a patient between hospital administrators and the doctors who provide the care (Dr. Jeffrey Neely, personal communication)? Do personnel in the construction industry feel the same way about the project constraints as was suggested in this research by personnel at NASA?

Beyond the potential organizational culture differences, there may be global cultural differences that should be considered. This research was conducted in the United States, but if the assessments completed in this research were to be conducted elsewhere in the world, the results may differ based on cultural priorities and norms (Grisham 2010, 104). Knowing these priorities and scheduling tendencies in advance could allow cross-cultural project teams to account for these differences when developing project schedules. Although the results may differ from those found in

this research, if other organizations and industries around the world struggle with disagreements about how long a project should take, this research provides some suggestions for understanding the "why" behind those differences and a method to merge those differences into a single estimate a project manager can use in the development of a schedule.

Additionally, further refinements to the methods used in this research could be used when polling a larger group. For example, the constraints questions could be further specified to determine if there are consistent "break points" among the preferences based on specific project inputs, where one preference gives way to another. A collection of actual activity durations could be compared to the estimates to help determine a method for accuracy calibration among the participants.

8.2.3 Data for the Decision Maker

While the two parts of this research (as described in Section 1.2) were independent of one another in that the model used to update the estimates did not account for the traits/estimating trends of the subjects providing the estimates, one area of future work is to merge the two parts together to develop a method to incorporate an estimator's traits as data from which a decision maker can update her beliefs. For example, if the decision maker knows that one group tends to estimate longer durations than another, this information could be used as part of the Bayesian updating process or perhaps as a method to calibrate various stakeholders to account for differences in project constraint priorities.

For companies that maintain a knowledge repository, one possible area of study is the development of a method for incorporation of this knowledge into the estimating process. Using the Bayesian model described in this research, a repository of project completion times could serve as a base-rate from which the project manager could start her assessment. Details specific to her project would then allow her to update her prior to more accurately reflect her beliefs about her specific project given the current project requirements and constraints. Maintaining a record of different expert's estimates versus actual completion times would also aid in the efforts to calibrate the experts. Once this data was collected, another area of future study could be to compare project managers who use the knowledge repository as compared to those who do not to determine whether or not the data improve the accuracy of the schedule estimates.

8.2.4 Communication of Assumptions

For the scheduling estimates collected in this research, "project completion" was defined as the completion of all tasks listed in the scheduling survey. It has been pointed out, however, that different team members will have different definitions of what constitutes the completion of the project. These different definitions could affect resource allocation if an organization is involved in more than one project, as one stakeholder perceives a project to be complete while another may still have tasks to accomplish. This, then, can also lead to project team members continuing to ask for assistance from other members who believed they were complete with the project and could allocate their time elsewhere. One area of future research is to study the definition of project completion among different stakeholders and determine how best to incorporate those definitions when developing the project schedule and advertising the completion date.

One finding of this research suggested that one reason for differences in activity estimates was due to differing assumptions about what actions were required to successfully complete the activity. An area for future research would be to ask participants to once again provide estimates, but to explicitly state all assumptions about the sub-activities that comprised each activity (Moder and Rodgers 1968, B-82). This would help ensure that any differences seen in the estimates were driven by beliefs about how long an activity should take as opposed to a lack of communication regarding the sub-activities that constituted the activity listed in the schedule.

Successful completion of a project includes all aspects of project management, both technical and programmatic. Based on statements made in an IG Report (Martin 2012, 13–14) and the results seen in this study regarding project constraint preferences, it appears that technical and programmatic success are treated as two separate entities and that technical success takes precedence over programmatic success. One are of future research is to determine why this perception exists, if it exists in organizations outside of NASA, and how to change the perception such that management of all project constraints is considered when evaluating project success.

8.2.5 Dominating Outliers

It was noted on several of the example estimates that, in a group of estimates, an extreme outlier would often dominate the mode of the posterior distribution due to the differing density levels in the tails of the priors as described in Section 7.3.5. If, however, multiple experts are in agreement in their opinion on a duration estimate with only one outlier expressing disagreement, the posterior distribution should reflect this. One area of further research is to determine a method that accounts for

the agreement among the experts and lessens the impact of the outlier on the posterior distribution.

8.2.6 Confidence and Risk

While not statistically correct, it is clear from Section 5.2.4 that a confidence rating on a single number has a meaning attached to it for many people. One recommended avenue of study is to look more closely into this interpretation of confidence and better explore the meaning stakeholders assign to the value. After requesting this information, subjects would be asked to explain exactly what that confidence value means and how they interpret the value. The same could be said for the "Risk" project constraint discussed in Section 5.2.1. Although not clearly defined, most stakeholders appeared to have some idea of a concept of risk independent of any other constraint (e.g. schedule risk or cost risk or quality risk). While it is believed that subjects perceived risk as a threat to mission success or personnel safety, further study into these perceptions are warranted.

8.2.7 Approximations and Direct Calculation

In an effort to simplify the calculations required to determine the expected value and variance of the posterior distribution, approximations of the posterior curve resulting from Equation 3-28 were used. This method involves calculating the PDF of the posterior at set intervals across the range of predicted duration estimates and then matching that curve with an approximation with a known equation for the mean and variance. The approximations used in this case were defined over the interval from ($-\infty$ to $+\infty$) to accommodate and encompass a wide range of expert duration

266

estimates ("Generalized Extreme Value Distribution - Wikipedia" 2016, "Normal Distribution" 2016). In some cases, this allowed a non-zero probability that the duration of an activity could be negative, which is a physical impossibility. One recommendation to remedy this is to use a gamma distribution for the "pessimists" and a Weibull distribution for the "optimists". These distributions are both defined from [0, +∞) and would ensure no duration estimate would be less than zero ("Gamma Distribution - Wikipedia" 2016, "Weibull Distribution" 2016). This would require recalculating many of the equations in Sections 3.5.1 and 3.5.3 to reflect the new distributions with their defining parameters. A distribution would also need to be found that roughly modeled the Normal distribution, but was only defined on the positive number line. Perhaps a better solution would be to develop an equation that described the posterior curve, where the parameters were defined by the estimates provided by the stakeholders. Once a method for calculating the mean and variance of that curve was determined, then the approximation would no longer be necessary.

8.2.8 Filter settings

As part of the Bayesian updating process, Decision Maker's (DM's) are asked to assess the likelihood of the Expert being surprised by the actual duration. In Section 3.5.2, surprise was defined as the actual duration falling below the 0.1 fractile or above the 0.9 fractile. For the purposes of this research, tables were developed which defined the α and β parameters required for each likelihood assessment from 0.01 to 0.99 (see Appendix A.10 – Appendix A.12). Based on the current procedure, once the DM decides on a likelihood value, she would need to reference these tables

in order to choose the appropriate values for α and β. It may be possible, however, to

calculate these values knowing only the likelihood assessment of the DM.

From Figure 8-1, it can be seen the relationship of α and β is linear.
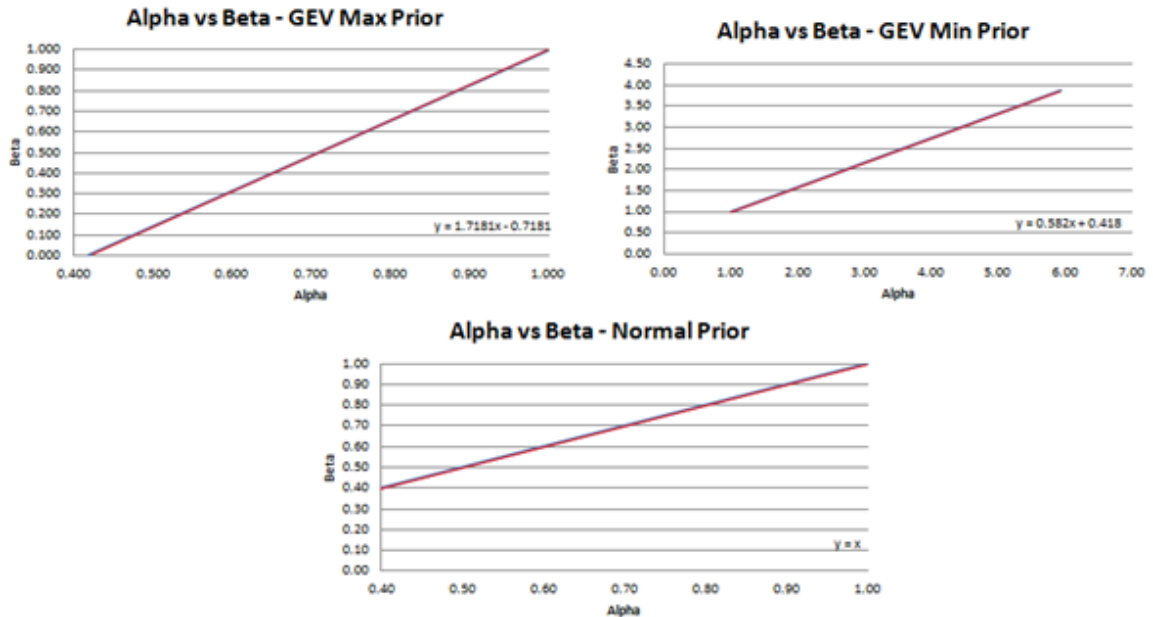
Figure 8-1: Relationship of α and β for the Beta Filter

Once α is determined, β can be easily calculated using one of the three

equations in Table 8-1, based on the form of the prior distribution:

Prior Distribution:	**Relationship of α and β:**
GEV Max	$\beta = 1.7181\alpha - 0.7181$
GEV Min	$\beta = 0.582\alpha - 0.418$
Normal	$\alpha = \beta$

Table 8-1: Relationship of α and β for the Beta Filter

There is also a relationship between the value of α and the likelihood of

surprise assessed by the DM. This relationship is not linear, but it does follow a clear

pattern that could be described by a formula.
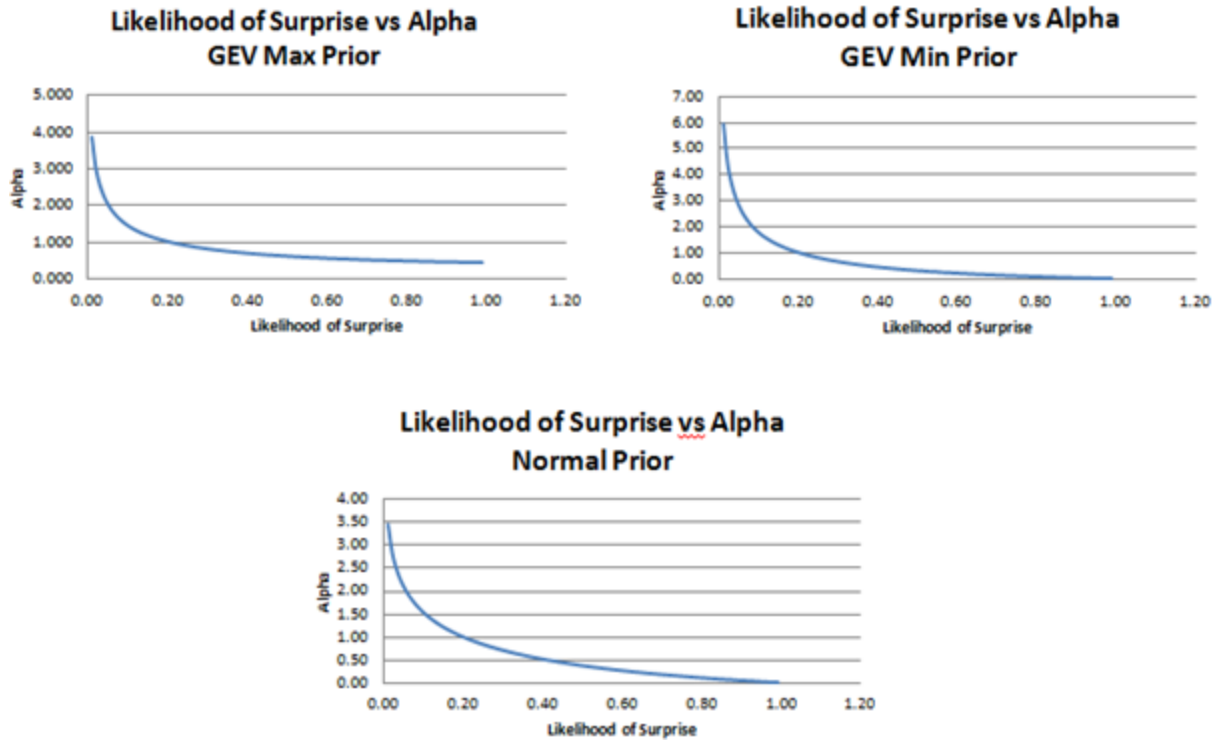
Figure 8-2: Relationship of Likelihood of Surprise and α for the Beta Filter

Although the Excel™ Trendline function provided a close approximating formula, when compared to the results from the table, the formula began to break down and some values of the likelihood were skipped altogether due to rounding. Further study into the relationship between α and the likelihood of surprise would render the tables created for this research unnecessary. Once the DM had assessed her belief in the likelihood of surprise, she could use a formula to determine α, and then go on to use the formulas provided in Table 8-1 to compute the β parameter.

For this research, the assessment of the likelihood of surprise was left entirely to the discretion of the DM. It was a subjective assessment based on the DM's knowledge of different Experts and their propensity to over- or under-exaggerate their level of knowledge. Given that objective data can be difficult to come by, one further area of study is to provide the DM a point of reference from which to base her

likelihood assessment. One recommendation is to use the Δ value calculated in Equation 3-39 Equation 3-40 as a point of reference. If, for example, the density between the Expert's "best case" and "worst case" estimates is less than 0.997 (i.e. the 3σ value recommended in Section 3.5.1), then the DM should use the "overstated" model. If it is above 0.997, she should use the "understated" model. An Expert density of exactly 0.997 would not require any calibration. The challenge would be to determine an acceptable value of Δ to use as the cut-off point and also to determine an appropriate way to relate the likelihood value to the value of Δ. Another recommendation is to record the Expert's self-assessed confidence intervals on the "best case" and "worst case" estimates and then compare those self-assessed values to the density between those values as calculated by Δ. The challenge in this case would again be how to correlate the differences noted with a likelihood assessment.

Appendices

This section includes the surveys provided to each of the subjects, as well as the responses gathered from each of the surveys. This is the raw data that was used determine the results and conclusions presented here. Below is a list describing the content of each of the appendices.

Appendix #	Title	Description
A.1	Recruitment Email	Email used to contact prospective participants and determine their interest/availability to participate in the study
A.2	Traits/ Opinions Survey	Gathered the required demographic information from the subject. Also gathered the data required to study the questions regarding constraint preferences and risk tolerances
A.3	Scheduling Survey	Collected data on duration estimates for specific projects at Wallops Flight Facility. Also gathered data on opinion of the provided task list and project personnel availability
A.4	Follow-On Survey	Intended to gather data to compare actual durations with the estimated durations. Information provided was not usable, so results were not included in this research
A.5	"Course of Action" Survey	Collected data on the perception managers and technicians of what constitutes "necessary" work
A.6	Participant List	Provides a summary of participants who consented to participate in the study identified by their random identifier and demographic category
A.7	Utility results	Consolidate results of the Risk Tolerance questions found in the Traits/Opinions survey
A.8	AHP results	Consolidates results of the Project Constraint Preferences questions found in the Traits/Opinions survey
A.9	Scheduling Survey – Estimation Results and Calculations	Summarizes the results of the duration estimations, confidences and calculated means and variances for the project surveys provided
A.10	GEV Max Beta Filters	Table providing the required hyper-parameters to calibrate an Expert with a GEV Max prior
A.11	GEV Min Beta Filters	Table providing the required hyper-parameters to calibrate an Expert with a GEV Min prior
A.12	Normal Beta Filters	Table providing the required hyper-parameters to calibrate an Expert with a Normal prior
A.13	DesignExpert™ Experiment Settings	Tables describing the configuration settings needed to re-create the experiments conducted here using the DesignExpert™ program

All,

      Good morning! I was wondering if I could get your help with something. Some already know about this, but if you don't, I've started a research project for a degree I'm working on that's looking at how we here at Wallops schedule projects. Some have helped me out with some data gathering last winter, but I found out some requirements the University of Maryland has regarding research, so I need to re-do a few things. All that to say:

This research will take a look at how different people schedule project activities. The basic plan of the research is to get several estimates of how long people *think* a project will take and then compare it to the actual project to see how long it *actually* took. Hopefully, it will show who we should be listening to when it comes to scheduling a project. I'll be tracking several projects over the course of the next year or so and basically (for projects that are selected and that you have been assigned to work) I would just need you to fill out a survey for each project on how long you think an activity should take (bonus: since these are real-world projects, your input may be used in building the actual schedule for the project). The scheduling surveys should take about 10 minutes to complete…maybe up to 30 minutes if you want to be really thorough on some of the questions. It will basically be like when a Range Services Manager (RSM) would ask you how long you think something should take, just with a few more details. There's also a "Follow-On" survey where you'll track how long it took to do different activities in each project along with a few other questions about how the project played itself out. If you fill out your hours as the project goes along, it should take 5-10 minutes to complete the survey. If you wait to the end, it may take a little longer since you'll need time to go back and figure what you did (maybe closer to 30-60 minutes, but that's worst-case-scenario). There's one other survey that will only be completed once which will gather some basic demographic information (education, years of service at Wallops, etc.). That survey should take maybe 30 minutes, again depending on how thorough you want to be. Participation would be entirely voluntary and if you decide not to participate, it won't affect anything. All final data will be masked so your responses won't be tied back specifically to you. <Name removed> and <Name removed> have both given me permission to press forward with this research, so no worries there either.

If you're willing to help me out, shoot me an e-mail to let me know, and I'll get in contact with you from there on the next few steps.

Thanks!

Lauren

P.S. It would be best to only contact me by e-mail (as opposed to in person) for questions and for these next steps to help maintain anonymity.

All of the following questions are optional. While all answers will be useful to this research, if any particular question or group of questions make you uncomfortable, you are free to leave the question blank.

Demographics

Please enter the 3-digit code given to you when you signed the consent form: _____

1. Which category best describes your current position with the company
(Senior Management would be anyone not directly involved with the project.
Focused more on the program level)
- Technical
- Management

2. How long have you been in your current position? (if you have worked in a similar position at a different company, include that time)

3. What is your completed level of education?
- High School Diploma
- Associates Degree (or equivalent)
- Bachelors
- Masters

Risk Tolerance

The following 5 questions will be used to determine your risk tolerance. Your answers will form points on a curve, the shape of which is a good indicator of whether you are risk tolerant or risk averse.

For these questions, consider the following scenario. You have been given a chance to play a "lottery" to win $5000. The person offering this lottery has given you a choice. You can either play the game, or he will trade you your chance at $5000 for straight cash that he will pay on the spot. The questions below ask you to determine the lowest monetary value for which you will trade your chance at winning. For example, if you had a 5% chance of winning $5000, would you trade that 5% chance for $15 on the spot? $20? Remember to use the LOWEST dollar amount you would be willing to trade.

1. You have a 10% chance of winning $5000. How much money would you trade this chance for if you knew you could walk away with cash in-hand?

2. You have a 35% chance of winning $5000. How much money would you trade this chance for if you knew you could walk away with cash in-hand?

3. You have a 50% chance of winning $5000. How much money would you trade this chance for if you knew you could walk away with cash in-hand?

4. You have a 68% chance of winning $5000. How much money would you trade this chance for if you knew you could walk away with cash in-hand?

5. You have a 87% chance of winning $5000. How much money would you trade this chance for if you knew you could walk away with cash in-hand?

Preference Analysis

The following questions deal with what's important to you with respect to different aspects of project completion. Each question asks your preference between two options and then using a scale of 1-9 (where 1 means no preference and 9 means a strong preference) to indicate how strong your preference is for your chosen option. As we all know, projects encounter a variety of problems as they progress. The questions below seek to determine which project constraints you consider more important to manage. To answer the question, type in the preference followed by the strength of your preference. For example, if you have a moderate preference of dealing with an increase in cost over an increase in schedule, answer the question "Cost, 3". Below is a table describing the meaning behind the 1-9 values.

1- A and B equally important
3- A slightly more important than B
5- A strongly more important than B
7- A very strongly more important than B
9- A absolutely more important than B
2,4,6, and 8- intermediate values between the defined numbers

Examples:
1. Increased Cost indicates that your overall project cost will increase, but you may have the advantage of decreased schedule, increased quality, etc.
2. Increased Schedule indicates that your overall schedule will increase, but you may have the advantage of increased quality or decreased risk
3. Increased Risk indicates that your overall risk of failing to meet project objectives will increase, but you may have the advantage of less cost or a shorter schedule
4. Decreased Quality indicates that your project quality may decrease (example: antenna has a jitter, but it still "works"), but you might finish early or under budget or schedule
5. Decreased Resources indicates that your project may have fewer resources, but you may finish under budget

1. Would you prefer an Increased Cost or an Increased Schedule? On a scale of 1-9, by how much do you prefer your chosen option?

2. Would you prefer an Increase Cost or Increased Risk? On a scale of 1-9, by how much do you prefer your chosen option?

3. Would you prefer Increased Cost or Decreased Quality? On a scale of 1-9, by how much would you prefer your chosen option?

4. Would you prefer Increased Schedule or Increased Risk? On a scale of 1-9, by how much do you prefer your chosen option?

5. Would you prefer Increased Schedule or Decreased Quality? On a scale of 1-9, by how much do you prefer your chosen option?

6. Would you prefer Increased Risk or Decreased Quality? On a scale of 1-9, by how much do you prefer your chosen option?

1. Below is a list of activities that must be completed in order to finish the
_____ Project. For each of the activities below, please write down how long
you believe each activity will take in the "Most Likely" column based on your
experiences at Wallops and elsewhere. Next to your estimate, please list how
confident you are (one a scale from 0% - 100%) Please also write down how long you
think the activity will take in a "best case" scenario (everything goes right) and a
worst case scenario (everything goes wrong).
Estimates can be in either hours or days (example: 4 hours, .5 days, 3 days, .5 hours,
etc.)
Assume a work week is ___ hour days, Monday - _____
Assume a team of _____ people
Assume personnel are working on the project _____ % of their time

Activity Name	**Most Likely/Confidence?**	**Best Case**	**Worst Case**
1. Activity 1			
2. Activity 2			
3. Activity 3			
4. Activity 4			

2. Do you believe more people need to be added to the project? If so, how many
more?

3. Do you believe that each of the activities listed above needs to be completed in
order to successfully finish this project? If not, which ones can be removed (write
down activity number or name)

4. Do you believe there are any additional activities required to successfully
complete this project that are not listed above? If so, list them below (use a second
page if more space is required).

5. Given everything you know about the above project (team assigned, activities required) and operations at Wallops:

For Engineering Projects: What date do you believe this project will be completed?

For Deployments: What date to you believe the team needs to deploy in order to be ready for mission support

In both cases, please provide your "most likely" estimate with a confidence level (from 0% - 100%) your "best case" estimate and your "worst case" estimate

	Most Likely/Confidence?	**Best Case**	**Worst Case**
Project Completion Date			

If you would like to make any comments about the reasoning behind your estimates or any of your other answers, please provide them on the next page.

<u>Comments:</u>

1. Now that you've completed the project, please list how long each of the activities actually took. If you did not work on all of the activities, fill out the sheet based on the activities you did work on.
Work week was: ___ hour days, Monday - _____
Team consisted of: _____ people
You/Team was able to work on the project: _____ % of their time

Note: These "actual times" can be in hours or days (example: 4 days, 3 hrs, .5 hours, etc.)

Activity Name	Actual Completion Time
1. Activity 1	
2. Activity 2	
3. Activity 3	
4. Activity 4	
5. Activity 5	

2. Were there any unexpected challenges? What were they?

3. Did you feel like you had enough time to complete each activity to your satisfaction or did you feel rushed to meet deadlines?

If you would like to make any comments about the reasoning behind your estimates or any of your other answers, please provide them on the next page.
Comments:

1. Management or Technical?

 Pick whichever option you picked for the previous survey

- o Management
- o Technical

<depending on what the participant chooses, it will direct them to one of two pages as described below>

■■

<If the Participant chooses Management>

For Management

Your team has been testing a piece of equipment for 2 weeks. Right now, it meets requirements, but is just barely within the specification. They've worked with equipment like this before and have stated its performance could be better. They think they know where the issue is and they believe an extra week of testing will get the system up to its full capability. The project is currently right on schedule and this extra week will mean a schedule slip that will increase the overall cost of the project and delay the readiness review by one week.

1. Do you leave the system "as is" and press forward with the readiness review or take the extra week to get the system up to its full capability?

- o Leave "as is" and press forward
- o Take the extra week to work on the system

2. Given that the system currently meets the requirements, would you consider the extra week gold-plating (i.e. going unnecessarily above and beyond the requirement) or risk mitigation (i.e. less chance the equipment could fail)?

- o Gold-plating
- o Risk Mitigation

3. Given your experiences throughout your professional career, why do you believe projects fall behind schedule?
Please keep your responses generic (don't identify specific people or projects) and remember to also keep them respectful!

<If the Participant chooses Technician>

For Technicians

You've been testing a piece of equipment for 2 weeks. Right now it meets requirements, but is just barely within the specification. You've worked with equipment like this before and know its performance could be better. You think you know where the issue is and you believe an extra week of testing will get the system up to its full capability. The project is currently right on schedule and this extra week will mean a schedule slip that will increase the overall cost of the project and delay the readiness review by one week.

1. Do you leave the system "as is" and press forward with the readiness review or take the extra week to get the system up to its full capability?

- o Leave "as is" and press forward
- o Take the extra week to work on the system

2. Given that the system currently meets the requirements, would you consider he extra week gold-plating (i.e. going unnecessarily above and beyond the requirement) or risk mitigation (i.e. less chance the equipment could fail)?

- o Gold-plating
- o Risk Mitigation

3. Given your experiences throughout your professional career, why do you believe projects fall behind schedule?
Please keep your responses generic (don't identify specific people or projects) and remember to also keep them respectful!

A.6 Participant List

Participant ID	Demographic Designator
408	M1B
481	M1M
498	M1M
164	M1M
548	M1M
969	M1M
858	M1T
838	M2B
380	M2B
458	M2M
518	M2M
148	M2T
157	M2T
222	M4B

Participant ID	Demographic Designator
127	T1B
191	T1B
739	T1B
912	T1B
399	T1H
824	T1M
670	T1T
203	T2B
712	T2B
819	T2B
158	T2H
538	T2M
661	T2T
434	T2T
493	T2T
441	T2T
857	T3B
315	T3M
396	T3T
774	T4H
619	T4H
798	T4H
424	T4H
463	T4T

For the charts above:

Position Demogrpahic:

M = Management

 T = Technical

Years of Experience (YoE) Demographic:

1 = 0 - 7

2 = 8 - 14

3 = 15 - 23

4 = 24+

Level of Formal Education (LoE) Demographic:

M = Master's

B = Bachelor's

T = Technical/Associates degree

H = High School

A.7 Utility results

In the chart below, the bold/underlined 3-digit numbers on the top row represent subject designators. The first column tells the subject their probability of winning $5000. The second column represents CME for the lottery (P[win] * 5000). Values below the subject designators represent each participant's "cash on the table" value for which they would trade their chance at winning $5000.

P(win)	Utility	458	408	481	838	157	498	164	548	148	380	969	858	222	518
0.1	500	750	500	500	1000	100	500	20	300	5	500	501	750	350	250
0.35	1750	2000	2000	1750	2500	250	1750	40	1500	25	2500	1751	2000	1000	1000
0.5	2500	3000	2500	2500	5000	1000	2500	50	2000	100	3500	2501	2750	2000	2000
0.68	3400	4500	3200	3400	5000	1500	3400	75	3000	200	4000	3401	3750	3000	3500
0.87	4350	4950	4250	4350	5000	2000	4350	100	3500	250	0	4351	4500	4000	4000

P(win)	Utility	661	774	434	493	396	619	399	798	203	158	463	538	424	127
0.1	500	100	50	500	1000	50	200	499	500	10	1000	1000	400	1000	100
0.35	1750	350	100	1750	2500	50	350	1749	1750	50	1500	2000	1575	2000	200
0.5	2500	500	2000	2500	3000	100	500	2499	2500	250	2000	3500	2250	4000	1000
0.68	3400	2500	2500	3400	3500	100	1000	3300	3400	2000	3000	4000	3060	0	2500
0.87	4350	3500	4200	4350	4000	150	1500	4351	4350	4000	4000	4500	3915	0	4000

P(win)	Utility	670	824	191	739	441	712	819	857	315	912				
0.1	500	400	20	500	500	250	20	25	250	50	500				
0.35	1750	2000	25	1750	1750	1000	40	100	850	250	700				
0.5	2500	4000	100	2500	2500	2500	100	200	1250	500	2500				
0.68	3400	4500	200	3400	3400	3000	200	400	2500	1000	3500				
0.87	4350	4750	1000	4350	4350	4000	500	600	3500	2000	4250				

A.8 AHP Results

Each box represents the results from a single participant. The 3-digit number in the top row is the subject designator. The first row lists the constraints and the weight assigned to each constraint based on the results of the "Preference Analysis" section of the "Traits/Opinions" Survey (Appendix A.2) and Equation 3-1. The final column is the consistency among the weights for each subject calculated using Equation 3-2.

458	**Weights**
Cost	0.131
Schedule	0.604
Quality	0.035
Risk	0.230
Consistency	0.422

408	**Weights**
Cost	0.260
Schedule	0.183
Quality	0.052
Risk	0.505
Consistency	0.242

148	**Weights**
Cost	0.620
Schedule	0.270
Quality	0.066
Risk	0.044
Consistency	0.206

380	**Weights**
Cost	0.239
Schedule	0.572
Quality	0.049
Risk	0.140
Consistency	0.374

838	**Weights**
Cost	0.334
Schedule	0.452
Quality	0.106
Risk	0.108
Consistency	0.110

518	**Weights**
Cost	0.549
Schedule	0.266
Quality	0.140
Risk	0.044
Consistency	0.395

969	**Weights**
Cost	0.112
Schedule	0.062
Quality	0.191
Risk	0.635
Consistency	0.088

858	**Weights**
Cost	0.275
Schedule	0.519
Quality	0.119
Risk	0.088
Consistency	0.142

548	**Weights**
Cost	0.084
Schedule	0.136
Quality	0.469
Risk	0.311
Consistency	0.159

498	**Weights**
Cost	0.391
Schedule	0.208
Quality	0.236
Risk	0.165
Consistency	0.674

222	**Weights**
Cost	0.362
Schedule	0.427
Quality	0.069
Risk	0.142
Consistency	0.033

164	**Weights**
Cost	0.552
Schedule	0.265
Quality	0.114
Risk	0.069
Consistency	0.083

661	Weights
Cost	0.335
Schedule	0.502
Quality	0.037
Risk	0.126
Consistency	0.127

774	Weights
Cost	0.440
Schedule	0.404
Quality	0.075
Risk	0.081
Consistency	0.005

493	Weights
Cost	0.507
Schedule	0.289
Quality	0.050
Risk	0.154
Consistency	0.198

396	Weights
Cost	0.300
Schedule	0.602
Quality	0.049
Risk	0.049
Consistency	0.129

399	Weights
Cost	0.495
Schedule	0.313
Quality	0.108
Risk	0.083
Consistency	0.059

798	Weights
Cost	0.433
Schedule	0.433
Quality	0.070
Risk	0.064
Consistency	0.001

203	Weights
Cost	0.318
Schedule	0.534
Quality	0.046
Risk	0.101
Consistency	0.077

158	Weights
Cost	0.272
Schedule	0.600
Quality	0.063
Risk	0.065
Consistency	0.107

463	Weights
Cost	0.529
Schedule	0.315
Quality	0.051
Risk	0.105
Consistency	0.086

538	Weights
Cost	0.491
Schedule	0.350
Quality	0.097
Risk	0.062
Consistency	0.035

434	Weights
Cost	0.072
Schedule	0.041
Quality	0.444
Risk	0.444
Consistency	0.060

424	Weights
Cost	0.304
Schedule	0.514
Quality	0.065
Risk	0.116
Consistency	0.118

127	Weights
Cost	0.260
Schedule	0.541
Quality	0.059
Risk	0.140
Consistency	0.271

670	Weights
Cost	0.334
Schedule	0.376
Quality	0.041
Risk	0.249
Consistency	0.510

824	Weights
Cost	0.557
Schedule	0.248
Quality	0.138
Risk	0.057
Consistency	0.072

191	Weights
Cost	0.348
Schedule	0.498
Quality	0.118
Risk	0.035
Consistency	0.312

739	Weights
Cost	0.291
Schedule	0.491
Quality	0.151
Risk	0.067
Consistency	0.074

441	Weights
Cost	0.283
Schedule	0.581
Quality	0.037
Risk	0.099
Consistency	0.230

712	Weights
Cost	0.249
Schedule	0.619
Quality	0.045
Risk	0.088
Consistency	0.197

819	Weights
Cost	0.609
Schedule	0.247
Quality	0.036
Risk	0.108
Consistency	0.407

857	Weights
Cost	0.161
Schedule	0.554
Quality	0.044
Risk	0.241
Consistency	0.188

315	Weights
Cost	0.284
Schedule	0.494
Quality	0.109
Risk	0.114
Consistency	0.186

619	Weights
Cost	0.262
Schedule	0.582
Quality	0.114
Risk	0.043
Consistency	0.386

912	Weights
Cost	0.514
Schedule	0.299
Quality	0.044
Risk	0.143
Consistency	0.197

A.9 Scheduling Survey – Estimation Results and Calculations

The following pages in Appendix A.9 provide the results of the Scheduling Surveys. The chart below gives an example of how to interpret the surveys. The Demographic Designator corresponds to the designators described in Section 3.2.1. For each value, the number in parenthesis represents the 3-digit participant designator. For example, ML (481) is the row listing the "most likely" estimates for Activity 1 – Activity 11 for participant 481. Participant 481 is a manager with 0-7 years of experience and a Master's degree (M1M). Cells with "BLNK" show where the participant did not provide an estimate or the value could not be calculated due to missing data. Cells highlighted in grey indicate that responses were provided shortly after the project started, but before any real durations could be reported. If the PERT duration of an activity could not be calculated for one person within a survey, that activity was removed from the summation for all participants for that activity and replaced with "BLNK". Confidence estimates were also set as "BLNK" for any participant that did not provide a ML estimate (even if they provided a confidence estimate). If the BC estimate was larger than the ML estimate, estimates for that activity were replaced with "BLNK". Units are in hours. Note that given the small sample size and close working quarters at Wallops Flight Facility, there are five "dummy projects" inserted in this appendix to help protect the anonymity of the subjects. These are fictitious projects with fictitious estimates made to resemble the actual projects. Data from these projects was not used in the analyses completed in this study.

Project designator   Activity Designator

Demographic Designator

ML = Most Likely
BC = Best Case
WC = Worst Case
(XXX) = Participant Designator
PERT = (BC +(4*ML)+WC)/6

Te = ∑PERT

VAR = (WC-BC)/6

Conf= Confidence in ML estimate

Survey 1

		A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11
M1M	ML (481)	2	5	18	8	5	2	9	5	BLNK	1	9
M1B	ML (408)	2	1	8	4	1	1	8	1	1	4	40
	BC (481)	2	4	9	8	4	1	8	4	BLNK	1	6
	BC (408)	1	0.5	6	2	0.5	0.5	6	0.5	0.5	2	24
	WC (481)	4	9	27	12	9	4	18	9	BLNK	2	12
	WC (408)	4	2	12	6	2	2	12	2	2	6	56
	PERT (481)	2.33	5.50	18.00	8.67	5.50	2.17	10.33	5.50	BLNK	1.17	9.00
	PERT (408)	2.17	1.08	8.33	4.00	1.08	1.08	8.33	1.08	1.08	4.00	40.00
	Te(481)	68.17										
	Te(408)	72.25										
	VAR (481)	0.11	0.69	9.00	0.44	0.69	0.25	2.78	0.69	BLNK	0.03	1.00
	Conf (481)	0.95	0.80	0.70	0.70	0.80	0.95	0.90	0.80	BLNK	0.70	0.90
	VAR (408)	0.25	0.06	1.00	0.44	0.06	0.06	1.00	0.06	0.06	0.44	28.44
	Conf (408)	0.75	0.85	0.50	0.50	0.85	0.85	0.50	0.85	0.85	0.85	0.60

286

		A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11
	Survey 1											
	-	**A1**	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**	**A9**	**A10**	**A11**
M1M	ML (481)	2	5	18	8	5	2	9	5	BLNK	1	9
M1B	ML (408)	2	1	8	4	1	1	8	1	1	4	40
	-											
	BC (481)	2	4	9	8	4	1	8	4	BLNK	1	6
	BC (408)	1	0.5	6	2	0.5	0.5	6	0.5	0.5	2	24
	-											
	WC (481)	4	9	27	12	9	4	18	9	BLNK	2	12
	WC (408)	4	2	12	6	2	2	12	2	2	6	56
	PERT (481)	2.33	5.50	18.00	8.67	5.50	2.17	10.33	5.50	BLNK	1.17	9.00
	PERT (408)	2.17	1.08	8.33	4.00	1.08	1.08	8.33	1.08	BLNK	4.00	40.00
	-											
	Te(481)	68.17										
	Te(408)	71.17										
	-											
	VAR (481)	0.11	0.69	9.00	0.44	0.69	0.25	2.78	0.69	BLNK	0.03	1.00
	Conf (481)	0.95	0.80	0.70	0.70	0.80	0.95	0.90	0.80	BLNK	0.70	0.90
	VAR (408)	0.25	0.06	1.00	0.44	0.06	0.06	1.00	0.06	0.06	0.44	28.44
	Conf (408)	0.75	0.85	0.50	0.50	0.85	0.85	0.50	0.85	0.85	0.85	0.60

		A1	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**	**A9**	**A10**	**A11**	**A12**	**A13**	**A14**	**A15**	**A16**	**A17**	**A18**
	Survey 2																		
T2H	ML (158)	0.25	0.25	0.75	0.25	1	0.5	2	0.25	1	0.5	3	6	0.25	2	0.5	0.6	1	0.5
M1M	ML (164)	2	0.4	1.5	0.5	0.5	0.5	1	0.25	0.5	0.25	2	2	1	1	0.25	0.25	0.5	0.5
	BC (158)	0.17	0.17	0.5	0.17	1	0.33	1.5	0.17	0.75	0.33	3	5	0.25	2	0.33	0.33	0.83	0.33
	BC (164)	1.5	0.3	1.25	0.3	0.3	0.3	0.75	0.2	0.3	0.2	1.5	1.5	0.75	0.75	0.2	0.2	0.4	0.4
	WC (158)	0.5	0.5	1	0.5	1.5	0.75	2.5	0.5	1.5	0.75	4	8	0.5	2.5	0.75	0.75	1.5	0.75
	WC (164)	3	0.75	2.5	0.75	0.75	0.75	2	0.75	0.7	0.5	3	3	1.5	1.5	0.5	0.5	1	1
	PERT (158)	0.28	0.28	0.75	0.28	1.08	0.51	2	0.28	1.04	0.51	3.17	6.17	0.29	2.08	0.51	0.58	1.06	0.51
	PERT (164)	2.08	0.44	1.63	0.51	0.51	0.51	1.13	0.33	0.5	0.28	2.08	2.08	1.04	1.04	0.28	0.28	0.57	0.57
	Te(158)	21.4																	
	Te(164)	15.9																	
	VAR (158)	0	0	0.01	0	0.01	0	0.03	0	0.02	0	0.03	0.25	0	0.01	0	0	0.01	0
	Conf (158)	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.9	0.95	0.95	0.95	0.95	0.95	0.95
	VAR (164)	0.06	0.01	0.04	0.01	0.01	0.01	0.04	0.01	0	0	0.06	0.06	0.02	0.02	0	0	0.01	0.01
	Conf (164)	0.9	0.95	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.95	0.9	0.9	0.9	0.9

		A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
	Survey 3										
	-	**A1**	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**	**A9**	**A10**
T2B	ML (819)	3	2	2	1	1	1	1	2	0.25	2
M1M	ML (548)	2	3	2	2	1	2	1	1	4	2
	-										
	BC (819)	3	2	2	1	1	1	1	2	0.25	2
	BC (548)	1	1	1	1	0.5	1	0.5	0.5	2	1
	-										
	WC (819)	6	4	4	3	3	3	3	4	0.5	4
	WC (548)	4	5	4	4	2	4	2	2	5	4
	PERT (819)	3.5	2.33	2.33	1.33	1.33	1.33	1.33	2.33	0.29	2.33
	PERT (548)	2.17	3	2.17	2.17	1.08	2.17	1.08	1.08	3.83	2.17
	-										
	Te (819)	18.5									
	Te (548)	20.9									
	-										
	VAR(819)	0.25	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0	0.11
	CONF (819)	BLNK ->									
	-										
	VAR (548)	0.25	0.44	0.25	0.25	0.06	0.25	0.06	0.06	0.25	0.25
	CONF (548)	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7

	Survey 4	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17
	_	**A1**	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**	**A9**	**A10**	**A11**	**A12**	**A13**	**A14**	**A15**	**A16**	**A17**
M1M	ML (164)	6	1	1	16	2	2	2	2	2	3	1	1	1	1	2	1	1
T2H	ML (158)	4	1	1	1	8	4	6	1.5	4	8	1	6	4	2	8	2	2
	_																	
	BC (164)	4	0.5	0.5	12	1.5	1.5	1	1.5	1.5	2	0.5	0.5	0.5	0.5	1	0.5	0.5
	BC (158)	4	0.5	0.75	0.75	6	3	4	1	2	7	1	4	3	1	6	1	2
	_																	
	WC (164)	8	2	2	27	2.5	3	3	3	3	4	2	2	2	2	3	2	2
	WC (158)	8	2	2	8	16	8	8	3	8	16	4	8	5	4	16	4	4
	_																	
	PERT (164)	6	1.08	1.08	17.2	2	2.08	2	2.08	2.08	3	1.08	1.08	1.08	1.08	2	1.08	1.08
	PERT (158)	4.67	1.08	1.13	2.13	9	4.5	6	1.67	4.33	9.17	1.5	6	4	2.17	9	2.17	2.33
	_																	
	Te (164)	47.1																
	Te (158)	70.8																
	VAR (164)	0.44	0.06	0.06	6.25	0.03	0.06	0.11	0.06	0.06	0.11	0.06	0.06	0.06	0.06	0.11	0.06	0.06
	CONF (164)	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
	_																	
	VAR (158)	0.44	0.06	0.04	1.46	2.78	0.69	0.44	0.11	1	2.25	0.25	0.44	0.11	0.25	2.78	0.25	0.11
	CONF (158)	BLNK ->																

	Survey 5	A1	A2	A3	A4	A5
M1B	ML (408)	12	6	6	6	6
T1H	ML (399)	8	16	4	4	8
	-					
	BC (408)	8	4	4	4	4
	BC (399)	4	4	2	2	4
	-					
	WC (408)	29	12	12	12	12
	WC (399)	16	24	8	8	16
	-					
	PERT (408)	14.2	6.67	6.67	6.67	6.67
	PERT (399)	8.67	15.3	4.33	4.33	8.67
	-					
	Te (408)	40.8				
	Te (399)	41.3				
	-					
	VAR (408)	12.3	1.78	1.78	1.78	1.78
	CONF (408)	0.5	0.8	0.8	0.8	0.7
	-					
	VAR (399)	4	11.1	1	1	4
	CONF (399)	0.5	0.75	0.7	0.5	BLNK

	Survey 6	A1	A2	A3	A4
T4H	ML (424)	1	7	8	8
M1M	ML (548)	6	6	3	3
	-				
	BC (424)	1	4	6	6
	BC (548)	3	3	2	2
	-				
	WC (424)	2	8	10	12
	WC (548)	8	8	4	4
	-				
	PERT (424)	1.17	6.67	8	8.33
	PERT (548)	5.83	5.83	3	3
	-				
	Te (424)	24.2			
	Te (548)	17.7			
	-				
	VAR (424)	0.03	0.44	0.44	1
	CONF (424)	BLNK ->			
	-				
	VAR (548)	0.69	0.69	0.11	0.11
	CONF (548)	0.7	0.7	0.7	0.7

		A1	A2	A3	A4	A5	A6
	Survey 7						
		A1	**A2**	**A3**	**A4**	**A5**	**A6**
T4T	ML (463)	BLNK ->					
M2T	ML (148)	8	1	8	8	10	10
	BC (463)	4	2	4	4	4	4
	BC (148)	6	0.5	6	6	8	8
	WC (463)	9	9	9	13.5	9	13.5
	WC (148)	9	1.5	9	9	12	12
	PERT (463)	BLNK ->					
	PERT (148)	7.83	1	7.83	7.83	10	10
	Te (463)	BLNK					
	Te (148)	44.5					
	VAR (463)	0.69	1.36	0.69	2.51	0.69	2.51
	CONF (463)	BLNK->					
	VAR (148)	0.25	0.03	0.25	0.25	0.44	0.44
	CONF (148)	0.9	0.9	0.85	0.85	0.85	0.85

	Survey 8	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14
T2T	ML (441)	4	4	4	6	9	2	0.5	1	2	1	1	2	2	1
T2B	ML (712)	1.5	1	1.5	1	3	6	1.5	2	2.5	0.5	0.5	2	0.5	0.5
	BC (441)	2	1	2	4	6	1	0.5	0.5	1	0.5	0.5	1	1	0.5
	BC (712)	0.5	0.5	1	0.5	2.5	4	0.5	1.5	2	0.5	0.5	1.5	0.5	0.5
	WC (441)	18	18	9	18	18	4	9	9	18	2	2	4	9	2
	WC (712)	3	1.5	4	1.5	4	9	4	3	4	1	1	3	2	1
	PERT (441)	6	5.83	4.5	7.67	10	2.17	1.92	2.25	4.5	1.08	1.08	2.17	3	1.08
	PERT (712)	1.58	1	1.83	1	3.08	6.17	1.75	2.08	2.67	0.58	0.58	2.08	0.75	0.58
	Te (441)	53.3													
	Te (712)	25.8													
	VAR (441)	7.11	8.03	1.36	5.44	4	0.25	2.01	2.01	8.03	0.06	0.06	0.25	1.78	0.06
	CONF (441)	0.8	0.8	0.8	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.85	0.85	0.95
	VAR(712)	0.17	0.03	0.25	0.03	0.06	0.69	0.34	0.06	0.11	0.01	0.01	0.06	0.06	0.01
	CONF (712)	BLNK ->													

	Survey 9	A1	A2	A3	A4	A5	A6	A7	A8	A9
T1B	ML (912)	2	2	4	8	6	1	4	8	2
T4H	ML (619)	1	1	1	0.5	2	0.17	2.5	6	1
T2T	ML (661)	2	2	18	27	8	BLNK	9	54	2
	BC (912)	1	1	1	2	4	0.5	3	6	1
	BC (619)	0.5	0.5	0.5	0.25	1.5	0.17	1	4	0.5
	BC (661)	1	1	5	8	6	BLNK	9	27	1
	WC (912)	4	4	8	12	10	2	8	10	3
	WC (619)	2	2	2	2	3	0.5	5	9	2
	WC (661)	8	8	63	108	18	BLNK	27	90	8
	PERT (912)	2.17	2.17	4.17	7.67	6.33	BLNK	4.5	8	2
	PERT (619)	1.08	1.08	1.08	0.71	2.08	BLNK	2.67	6.17	1.08
	PERT (661)	2.83	2.83	23.3	37.3	9.33	BLNK	12	55.5	2.83
	Te (912)	37								
	Te (619)	16								
	Te (661)	146								
	VAR(912)	0.25	0.25	1.36	2.78	1	0.06	0.69	0.44	0.11
	CONF (912)	1	1	0.5	0.5	0.9	1	0.75	0.8	1
	VAR (619)	0.06	0.06	0.06	0.09	0.06	0	0.44	0.69	0.06
	CONF (619)	1	1	0.9	0.75	1	1	0.75	0.75	1
	VAR (661)	1.36	1.36	93.4	278	4	BLNK	9	110	1.36
	CONF (661)	0.75	0.75	0.6	0.85	0.9	BLNK	0.95	0.7	0.9

	Survey 10	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
T1B	ML (191)	4	16	8	4	4	16	8	4	4	32
T3M	ML (315)	20	30	20	7	10	30	20	7	20	60
M1M	ML (548)	20	10	5	7	5	10	5	7	5	15
	BC (191)	2	8	4	2	2	8	4	2	2	16
	BC (315)	15	25	15	0.5	5	25	15	0.5	20	30
	BC (548)	15	7	3	5	3	7	3	5	3	12
	WC (191)	12	48	24	12	12	48	24	12	12	96
	WC (315)	40	50	50	15	15	40	50	15	50	180
	WC (548)	30	20	7	10	10	20	7	10	7	20
	PERT (191)	5	20	10	5	5	20	10	5	5	40
	PERT (315)	22.5	32.5	24.2	7.25	10	30.8	24.2	7.25	25	75
	PERT (548)	20.8	11.2	5	7.17	5.5	11.2	5	7.17	5	15.3
	Te (191)	125									
	Te (315)	259									
	Te (548)	93.3									
	VAR (191)	2.78	44.4	11.1	2.78	2.78	44.4	11.1	2.78	2.78	178
	CONF (191)	0.9	0.9	0.9	0.9	0.9	0.8	0.8	0.8	0.8	0.9
	VAR (315)	17.4	17.4	34	5.84	2.78	6.25	34	5.84	25	625
	CONF (315)	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
	VAR (548)	6.25	4.69	0.44	0.69	1.36	4.69	0.44	0.69	0.44	1.78
	CONF (548)	0.4	0.4	0.3	0.4	0.3	0.4	0.3	0.4	0.3	0.4

	Survey 11	A1	A2	A3	A4	A5	A6	A7
T2T	ML (441)	18	18	3	4	18	13.5	4
T3T	ML (396)	16	16	6	6	45	16	8
M2T	ML (148)	16	16	8	8	40	16	8
	BC (441)	13.5	13.5	2	3	9	9	3
	BC (396)	12	12	4	4	36	12	6
	BC (148)	10	10	6	6	30	12	6
	WC (441)	22.5	27	9	9	36	18	9
	WC (396)	24	24	8	8	54	24	16
	WC (148)	26	26	16	16	60	24	16
	PERT (441)	18	18.8	3.83	4.67	19.5	13.5	4.67
	PERT (396)	16.7	16.7	6	6	45	16.7	9
	PERT (148)	16.7	16.7	9	9	41.7	16.7	9
	Te (441)	82.9						
	Te (396)	116						
	Te (148)	119						
	VAR (441)	2.25	5.06	1.36	1	20.3	2.25	1
	CONF(441)	0.75	0.75	0.8	0.8	0.8	BLNK	BLNK
	VAR (396)	4	4	0.44	0.44	9	4	2.78
	CONF (396)	0.85	0.85	0.85	0.85	BLNK	0.75	0.75
	VAR (148)	7.11	7.11	2.78	2.78	25	4	2.78
	CONF (148)	0.75	0.75	0.85	0.85	0.66	0.75	0.75

	Survey 12	A1	A2	A3
T1H	ML (399)	4	27	6
M1M	ML (481)	6	5	5
M1B	ML (408)	12	16	16
	BC (399)	3	16	2
	BC (481)	4	2	2
	BC (408)	8	12	12
	WC (399)	8	40	16
	WC (481)	8	9	9
	WC (408)	16	24	24
	PERT (399)	4.5	27.3	7
	PERT (481)	6	5.17	5.17
	PERT (408)	12	16.7	16.7
	Te (399)	38.8		
	Te (481)	16.3		
	Te (408)	45.3		
	VAR (399)	0.69	16	5.44
	CONF (399)	0.7	0.5	0.5
	VAR (481)	0.44	1.36	1.36
	CONF (481)	0.95	0.5	0.5
	VAR (408)	1.78	4	4
	CONF(408)	0.8	0.85	0.85

	Survey 13	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15
M2M	ML (518)	4	4	2	4	8	4	8	2	1	4	2	4	6	6	12
M1M	ML (498)	16	16	0.5	1	BLNK	1	8	4	0.2	2	BLNK	4	3	8	8
M4B	ML (222)	13.5	2	7	3	13.5	5	3	2.5	2	3	2	1.5	0.75	1.5	BLNK
M1M	ML (481)	6	6	1	3	4	1	2	16	4	2	BLNK	6	2	2	6
	BC (518)	2	3	1.5	2	6	3	4	1.5	0.5	2	1	2	4	4	8
	BC (498)	3	8	0.2	0.25	BLNK	0.25	4	2	0.1	1.5	BLNK	3	2	4	4
	BC (222)	9	1.5	4	2	7	2	2	1.5	1	1.5	1	0.75	0.42	0.75	BLNK
	BC (481)	3	4	1	3	4	1	1	10	3	2	BLNK	5	1	1	6
	WC (518)	6	5	4	6	10	5	10	4	1.5	5	3	5	8	8	16
	WC (498)	40	24	2	4	BLNK	16	16	8	0.5	4	BLNK	8	4	12	12
	WC (222)	27	3	9	5	18	7	4	4	2.5	5	2.5	3	1	2	BLNK
	WC (481)	8	8	3	5	8	3	3	20	6	3	BLNK	8	4	3	10
	PERT (518)	4	4	2.25	4	BLNK	4	7.67	2.25	1	3.83	BLNK	3.83	6	6	BLNK
	PERT (498)	17.8	16	0.7	1.38	BLNK	3.38	8.67	4.33	0.23	2.25	BLNK	4.5	3	8	BLNK
	PERT (222)	15	2.08	6.83	3.17	BLNK	4.83	3	2.58	1.92	3.08	BLNK	1.63	0.74	1.46	BLNK
	PERT (481)	5.83	6	1.33	3.33	BLNK	1.33	2	15.7	4.17	2.17	BLNK	6.17	2.17	2	BLNK
	Te (518)	48.8														
	Te (498)	70.3														
	Te (222)	46.3														
	Te (481)	52.2														
	VAR (518)	0.44	0.11	0.17	0.44	0.44	0.11	1	0.17	0.03	0.25	0.11	0.25	0.44	0.44	1.78
	CONF (518)	BLNK ->														

Survey 13 (cont.)															
	A1	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**	**A9**	**A10**	**A11**	**A12**	**A13**	**A14**	**A15**
VAR (498)	38	7.11	0.09	0.39	BLNK	6.89	4	1	0	0.17	BLNK	0.69	0.11	1.78	1.78
CONF (498)	0.95	0.9	0.95	0.5	BLNK	0.5	0.6	0.95	0.95	BLNK	BLNK	0.95	0.95	0.6	0.7
VAR (222)	9	0.06	0.69	0.25	3.36	0.69	0.11	0.17	0.06	0.34	0.06	0.14	0.01	0.04	BLNK
CONF (222)	0.75	0.75	0.75	0.8	0.9	0.9	0.7	0.7	0.8	0.85	0.8	0.9	0.9	0.8	BLNK
VAR (481)	0.69	0.44	0.11	0.11	0.44	0.11	0.11	2.78	0.25	0.03	BLNK	0.25	0.25	0.11	0.44
CONF (481)	0.9	1	0.9	0.9	0.8	0.8	0.8	0.9	0.8	0.9	BLNK	0.9	0.9	0.8	0.8

	Survey 14														
		A1	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**	**A9**	**A10**	**A11**	**A12**	**A13**	**A14**	**A15**
T2T	ML (661)	2	1.5	8	0.5	5	8	40	20	50	110	20	20	30	50	50
T4H	ML (619)	0.5	0.5	1	1	2	1	2	4	2	BLNK	BLNK	4	4	1	3
	BC (661)	1.5	0.75	5	0.25	2	3	20	6	30	110	20	20	10	30	30
	BC (619)	0.5	0.5	0.5	0.5	1	0	1	2	1	BLNK	BLNK	3	2.5	1	2
	WC (661)	8	3	12	1	8	24	70	30	100	450	50	40	60	100	100
	WC (619)	1	1	2	2	4	10	4	10	3	BLNK	BLNK	6	8	3	5
	PERT (661)	2.92	1.63	8.17	0.54	5	9.83	41.7	19.3	55	BLNK	BLNK	23.3	31.7	55	55
	PERT (619)	0.58	0.58	1.08	1.08	2.17	2.33	2.17	4.67	2	BLNK	BLNK	4.17	4.42	1.33	3.17
	Te (661)	309														
	Te (619)	29.8														

Survey 14 (cont.)															
	A1	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**	**A9**	**A10**	**A11**	**A12**	**A13**	**A14**	**A15**
VAR (661)	1.17	0.14	1.36	0.02	1	12.3	69.4	16	136	3211	25	11.1	69.4	136	136
CONF (661)	0.8	0.9	1	1	0.75	0.25	0.6	0.6	0.9	0.85	0.85	0.85	0.6	0.9	0.9
VAR (619)	0.01	0.01	0.06	0.06	0.25	2.78	0.25	1.78	0.11	BLNK	BLNK	0.25	0.84	0.11	0.25
CONF (619)	0.9	0.9	0.75	0.8	0.8	0.3	0.8	0.9	1	BLNK	BLNK	0.8	1	1	0.9

	Survey 15									
		A1	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**	**A9**	**A10**
M1M	ML (548)	8	5	8	5	7	10	15	7	7	10
T2T	ML (493)	120	30	60	30	5	10	10	5	5	5
T2B	ML (203)	42	15	42	15	10	63	21	5	10	3
M1B	ML (408)	13	7	20	10	2	13	15	2	3	10
M2B	ML (838)	15	10	15	10	1	BLNK	5	1	BLNK	BLNK
	BC (548)	5	2	5	2	5	7	10	5	5	7
	BC (493)	90	20	30	15	3	6	5	3	3	3
	BC (203)	21	10	21	10	5	42	15	3	5	1
	BC (408)	10	5	15	7	1	10	10	1	1	5
	BC (838)	10	5	10	5	0.5	BLNK	3	0.5	BLNK	BLNK
	WC (548)	10	10	10	10	10	15	20	10	10	15
	WC (493)	180	60	90	60	10	20	14	10	10	10
	WC (203)	63	21	63	42	15	126	42	10	21	5
	WC (408)	23	15	30	20	5	17	17	3	5	20
	WC (838)	22.5	15	25	15	3	BLNK	10	3	BLNK	BLNK

Survey 15 (cont.)										
	A1	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**	**A9**	**A10**
PERT (548)	7.83	5.33	7.83	5.33	7.17	BLNK	15	7.17	BLNK	BLNK
PERT (493)	125	33.3	60	32.5	5.5	BLNK	9.83	5.5	BLNK	BLNK
PERT (203)	42	15.2	42	18.7	10	BLNK	23.5	5.5	BLNK	BLNK
PERT (408)	14.2	8	20.8	11.2	2.33	BLNK	14.5	2	BLNK	BLNK
PERT (838)	15.4	10	15.8	10	1.25	BLNK	5.5	1.25	BLNK	BLNK
Te (548)	55.7									
Te (493)	272									
Te (203)	157									
Te (408)	73									
Te (838)	59.3									
VAR (548)	0.69	1.78	0.69	1.78	0.69	1.78	2.78	0.69	0.69	1.78
CONF (548)	0.7	0.6	0.6	0.7	0.6	0.6	0.7	0.6	0.5	0.5
VAR (493)	225	44.4	100	56.3	1.36	5.44	2.25	1.36	1.36	1.36
CONF (493)	0.5	0.7	0.5	0.5	0.8	0.8	0.6	0.8	0.8	0.8
VAR (203)	49	3.36	49	28.4	2.78	196	20.3	1.36	7.11	0.44
CONF (203)	0.7	0.8	0.6	0.5	0.8	0.5	0.6	0.7	0.7	0.8
VAR (408)	4.69	2.78	6.25	4.69	0.44	1.36	1.36	0.11	0.44	6.25
CONF (408)	0.6	0.4	0.4	0.8	0.9	0.75	0.5	0.9	0.8	0.6
VAR (838)	4.34	2.78	6.25	2.78	0.17	BLNK	1.36	0.17	BLNK	BLNK
CONF (838)	0.75	0.9	0.65	0.9	0.85	BLNK	0.8	0.85	BLNK	BLNK

	Survey 16	A1	A2	A3	A4	A5	A6
M2M	ML (518)	3	2	1	2	1	4
T2B	ML (819)	45	4	4	4	4	8
M1M	ML (498)	4	4	2	1	2	6
M4B	ML (222)	3.5	1.5	0.5	0.33	0.17	5
M1M	ML (164)	16	2	1	2	0.5	4.5
M1M	ML (481)	4	1	1	2	1	4
	BC (518)	2	1	0.5	1	0.5	2
	BC (819)	45	4	4	4	4	8
	BC (498)	2	1	1	1	1	4
	BC (222)	2.25	1	0.17	0.17	0.08	3
	BC (164)	12	1.5	0.75	1.5	0.3	4
	BC (481)	1	1	1	1	1	1
	WC (518)	4	3	2	3	1.5	6
	WC (819)	72	6	8	6	8	12
	WC (498)	6	12	6	4	4	10
	WC (222)	5	4	1	0.6	0.33	8
	WC (164)	24	3	2	3	1	8
	WC (481)	6	4	2	2	2	8
	PERT (518)	3	2	1.08	2	1	4
	PERT (819)	49.5	4.33	4.67	4.33	4.67	8.67
	PERT (498)	4	4.83	2.5	1.5	2.17	6.33
	PERT (222)	3.54	1.83	0.53	0.35	0.18	5.17
	PERT (164)	16.7	2.08	1.13	2.08	0.55	5
	PERT (481)	3.83	1.5	1.17	1.83	1.17	4.17

	Survey 16 (cont)	A1	A2	A3	A4	A5	A6
	Te (518)	13.1					
	Te (819)	76.2					
	Te (498)	21.3					
	Te (222)	11.6					
	Te (164)	27.5					
	Te (481)	13.7					
	VAR (518)	0.11	0.11	0.06	0.11	0.03	0.44
	CONF (518)	BLNK ->					
	VAR (819)	20.3	0.11	0.44	0.11	0.44	0.44
	CONF (819)	BLNK ->					
	VAR (498)	0.44	3.36	0.69	0.25	0.25	1
	CONF (498)	0.8	0.05	0.1	0.05	0.5	0.95
	VAR (222)	0.21	0.25	0.02	0.01	0	0.69
	CONF (222)	0.85	0.9	0.8	0.9	0.9	0.9
	VAR (164)	4	0.06	0.04	0.06	0.01	0.44
	CONF (164)	0.9	0.9	0.85	0.9	0.95	0.95
	VAR (481)	0.69	0.25	0.03	0.03	0.03	1.36
	CONF (481)	0.8	0.9	0.9	0.5	0.5	0.9

	Survey 17	**A1**	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**
T4H	ML (424)	2	2	2	2	4	2	6
M1M	ML (481)	5	5	5	5	5	1	2
M1B	ML (408)	1	1	1	1	BLNK	6	6
	BC (424)	1	1	1	1	2	1	4
	BC (481)	4	4	4	4	4	1	1
	BC (408)	0.5	0.5	0.5	0.5	BLNK	4	4
	WC (424)	4	4	4	4	8	4	6
	WC (481)	9	9	9	9	10	2	4
	WC (408)	2	2	2	2	BLNK	8	8
	PERT (424)	2.17	2.17	2.17	2.17	BLNK	2.17	5.67
	PERT (481)	5.5	5.5	5.5	5.5	BLNK	1.17	2.17
	PERT (408)	1.08	1.08	1.08	1.08	BLNK	6	6
	Te (424)	16.5						
	Te (481)	25.3						
	Te (408)	16.3						
	VAR (424)	0.25	0.25	0.25	0.25	1	0.25	0.11
	CONF (424)	1	1	1	1	1	1	1
	VAR (481)	0.69	0.69	0.69	0.69	1	0.03	0.25
	CONF (481)	0.8	0.8	0.8	0.8	0.8	0.95	0.95
	VAR (408)	0.06	0.06	0.06	0.06	BLNK	0.44	0.44
	CONF (408)	0.85	0.85	0.85	0.85	BLNK	0.75	0.75

	Survey 18	A1	A2	A3	A4	A5	A6	A7
M2T	ML (148)	1.5	1	2	1.5	2	80	27
M1M	ML (548)	1	0.5	2	3	2	2	2
	BC (148)	1	0.75	1	1	1	40	18
	BC (548)	1	0.25	1	2	1	1	1
	WC (148)	4	4	3	3	4	80	40
	WC (548)	3	1	3	4	3	4	4
	PERT (148)	1.83	1.46	2	1.67	2.17	73.3	27.7
	PERT (548)	1.33	0.54	2	3	2	2.17	2.17
	Te (148)	110						
	Te (548)	13.2						
	VAR (148)	0.25	0.29	0.11	0.11	0.25	44.4	13.4
	CONF (148)	0.9	0.9	0.9	0.8	0.8	0.9	0.75
	VAR (548)	0.11	0.02	0.11	0.11	0.11	0.25	0.25
	CONF (548)	0.7	0.7	0.7	0.7	0.7	0.7	0.7

	Survey 19	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14
T3T	ML (396)	BLNK ->													
T4H	ML (774)	BLNK ->													
M2T	ML (148)	8	5	6	1	1	4	4	50	8	2	3	6	2	8
M1T	ML (858)	6	2	2	2	1.5	6	16	20	1.5	1	16	4	5	5
M2T	ML (157)	BLNK ->													
	BC (396)	6	3	3	2	2	3	8	8	5	3	BLNK	2	2	3
	BC (774)	8	4	4	2	2	4	10	48	8	8	8	4	4	4
	BC (148)	6	4	5	0.5	0.75	3	2	49	6	1.5	2	5	1	7
	BC (858)	4	1.5	1.5	1	1	4	12	16	1	0.5	12	2	4	4
	BC (157)	4	2	1	2	0.5	1.5	4	2	1	2	1	0.7	1	1.5
	WC (396)	24	10	10	4	4	6	24	16	12	6	BLNK	8	8	9
	WC (774)	14	8	8	4	4	6	16	BLNK	10	10	10	6	6	6
	WC (148)	16	7	16	2	2	8	8	58	16	8	5	10	8	12
	WC (858)	8	4	4	3	3	8	20	24	3	3	20	5	6	6
	WC (157)	8	4	1.5	4	1	2	6	3	2	3	1.5	1	2	2
	PERT (396)	BLNK ->													
	PERT (774)	BLNK ->													
	PERT (148)	9	5.17	7.5	1.08	1.13	4.5	4.33	51.2	9	2.92	3.17	6.5	2.83	8.5
	PERT (858)	6	2.25	2.25	2	1.67	6	16	20	1.67	1.25	16	3.83	5	5
	PERT (157)	BLNK ->													
	Te (396)	BLNK													
	Te (774)	BLNK													
	Te (148)	117													
	Te (858)	88.9													

Survey 19 (cont.)														
	A1	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**	**A9**	**A10**	**A11**	**A12**	**A13**	**A14**
Te (157)	BLNK													
VAR (396)	9	1.36	1.36	0.11	0.11	0.25	7.11	1.78	1.36	0.25	BLNK	1	1	1
CONF (396)	BLNK->													
VAR (774)	1	0.44	0.44	0.11	0.11	0.11	1	BLNK	0.11	0.11	0.11	0.11	0.11	0.11
CONF (774)	BLNK->													
VAR (148)	2.78	0.25	3.36	0.06	0.04	0.69	1	2.25	2.78	1.17	0.25	0.69	1.36	0.69
CONF (148)	0.7	0.7	0.7	0.9	0.9	0.7	0.5	0.7	0.7	0.7	0.9	0.9	0.66	BLNK
VAR (858)	0.44	0.17	0.17	0.11	0.11	0.44	1.78	1.78	0.11	0.17	1.78	0.25	0.11	0.11
CONF (858)	0.7	0.5	0.5	0.75	0.75	0.75	0.75	0.5	0.5	0.25	0.5	0.75	0.75	0.75
VAR (157)	0.44	0.11	0.01	0.11	0.01	0.01	0.11	0.03	0.03	0.03	0.01	0	0.03	0.01
CONF (157)	BLNK->													

305

	Survey 20	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13
T2B	ML (203)	10	30	42	10	126	42	84	10	63	15	20	15	10
T2T	ML (493)	60	300	120	30	90	30	90	30	90	30	90	30	10
M1M	ML (969)	20	45	150	10	10	5	150	20	15	5	10	5	2
	BC (203)	5	20	30	5	105	21	63	5	42	10	10	5	5
	BC (493)	30	180	60	10	30	14	45	14	30	14	60	20	5
	BC (969)	15	25	100	5	5	2.5	100	10	10	2.5	5	2.5	1
	WC (203)	21	63	84	21	168	84	126	30	84	21	42	21	15
	WC (493)	90	400	240	30	180	60	180	45	120	60	180	45	14
	WC (969)	40	75	200	12.5	12.5	10	300	25	20	10	15	10	4
	PERT (203)	11	33.8	47	11	130	45.5	87.5	12.5	63	15.2	22	14.3	10
	PERT (493)	60	297	130	26.7	95	32.3	97.5	29.8	85	32.3	100	30.8	9.83
	PERT (969)	22.5	46.7	150	9.58	9.58	5.42	167	19.2	15	5.42	10	5.42	2.17
	Te (203)	502												
	Te (493)	1026												
	Te (969)	468												
	VAR (203)	7.11	51.4	81	7.11	110	110	110	17.4	49	3.36	28.4	7.11	2.78
	CONF (203)	0.8	0.7	0.7	0.5	0.6	0.5	0.6	0.75	0.7	0.7	0.7	0.8	0.8
	VAR (493)	100	1344	900	11.1	625	58.8	506	26.7	225	58.8	400	17.4	2.25
	CONF (493)	0.5	0.5	0.5	0.7	0.5	0.7	0.6	0.6	0.6	0.7	0.6	0.6	0.75
	VAR (969)	17.4	69.4	278	1.56	1.56	1.56	1111	6.25	2.78	1.56	2.78	1.56	0.25
	CONF (969)	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8

	Survey 21	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14
T2H	ML (158)	0.5	0.5	0.25	0.75	2	0.25	1	0.75	2	0.25	2	4	0.5	1
M1M	ML (548)	1	0.25	2	0.25	0.25	0.25	2	0.5	0.25	0.75	1	1	2	2
T4H	ML (424)	1.5	.75	1.5	0.5	1	0.5	1.25	1	BLNK	1	0.5	3	2	3
	BC (158)	0.25	0.25	0.17	0.5	1.5	0.17	0.75	0.5	1.5	0.17	1.5	2.5	0.5	0.5
	BC (548)	0.5	0.17	1.25	0.17	0.17	0.17	1.25	0.5	0.17	0.5	0.5	0.5	1.25	1.25
	BC (424)	1.5	0.25	1	0.25	0.5	0.25	1	0.5	BLNK	0.5	0.25	2	0.5	2
	WC (158)	1	1	0.5	1.5	3.5	0.5	1.5	1.5	3.5	0.5	3.5	7.5	1	2
	WC (548)	2	1	3.5	0.5	0.5	0.5	3.5	1	0.5	1.25	1.5	1.5	3.25	3.25
	WC (424)	2.5	2	3	1.5	1.5	1	3	1.25	BLNK	4	2	4	4	5
	PERT (158)	0.54	0.54	0.28	0.83	2.17	0.28	1.04	0.83	BLNK	0.28	2.17	4.33	0.58	1.08
	PERT (548)	1.08	0.36	2.13	0.28	0.28	0.28	2.13	0.58	BLNK	0.79	1	1	2.08	2.08
	PERT (424)	1.67	0.88	1.67	0.63	1.00	0.54	1.50	0.96	BLNK	1.42	0.71	3.00	2.08	3.17
	Te (158)	21.3													
	Te (548)	16.1													
	Te (424)	19.2													
	VAR (158)	0.13	0.13	0.06	0.17	0.33	0.06	0.13	0.17	0.33	0.06	0.33	0.83	0.08	0.25
	CONF (158)	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.9	0.95	0.95
	VAR (548)	1.5	0.83	2.25	0.33	0.33	0.33	2.25	0.5	0.33	0.75	1	1	2	2
	CONF (548)	0.9	0.95	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.95
	VAR(424)	0.17	0.29	0.33	0.21	0.17	0.13	0.33	0.13	BLNK	0.58	0.29	0.33	0.58	0.50
	CONF(424)	1	1	1	1	1	1	1	1	BLNK	1	1	1	1	1

	Survey 22	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15
T2T	ML (441)	3	3	3	5	8	1	0.25	2	1	2	2	1	1	2	1
T2B	ML (819)	2	0.5	2	0.5	2	5	0.75	1	1.5	0.75	1	1	0.75	0.75	2
M2T	ML (157)	0.5	2	1	0.5	5	5	1	1	2	1.5	2	0.75	1	2	3
T4H	ML (774)	1	1	2	3	3	2	0.74	1	1.4	0.4	2	0.4	2	1	3
	BC (441)	2	2	2	3.25	5	0.5	0.17	1.25	0.5	1	1	0.5	0.5	1	0.5
	BC (819)	1.25	0.5	1.25	0.17	1.5	3	0.5	0.5	1	0.5	0.5	0.5	0.5	0.5	1.25
	BC (157)	0.17	1.25	0.5	0.25	3	3	0.5	0.5	1.25	1	1.25	0.5	0.5	1.25	2
	BC(774)	0.5	0.5	1.25	2	2	1.5	0.5	0.5	1	0.17	1.5	0.25	1.25	0.5	2
	WC (441)	5.5	5.5	5.5	9	15	2	0.5	4	2	4	4	2	2	4	2
	WC (819)	4	1	4	1	4	9	2	2	3	2	2	2	2	2	4
	WC (157)	1	3	2	1	8	8	1.5	1.5	3	3	3	2	2	3	5
	WC (774)	2	2	4	5.5	5.5	4	2	2	3	1	4	1	4	2	5.5
	PERT (441)	3.25	3.25	3.25	5.38	8.67	1.08	0.28	2.21	1.08	2.17	2.17	1.08	1.08	2.17	1.08
	PERT (819)	2.21	0.58	2.21	0.53	2.25	5.33	0.92	1.08	1.67	0.92	1.08	1.08	0.92	0.92	2.21
	PERT (157)	0.53	2.04	1.08	0.54	5.17	5.17	1	1	2.04	1.67	2.04	0.92	1.08	2.04	3.17
	PERT (774)	1.08	1.08	2.21	3.25	3.25	2.25	0.91	1.08	1.6	0.46	2.25	0.48	2.21	1.08	3.25
	Te (441)	38.2														
	Te (819)	23.9														
	Te (157)	29.5														
	Te (774)	26.4														
	VAR (441)	0.58	0.58	0.58	0.96	1.67	0.25	0.06	0.46	0.25	0.5	0.5	0.25	0.25	0.5	0.25
	CONF (441)	0.8	0.8	0.8	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.85	0.85	0.95	0.85

308

Survey 22 (cont.)		A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15
	VAR (819)	0.46	0.08	0.46	0.14	0.42	1	0.25	0.25	0.33	0.25	0.25	0.25	0.25	0.25	0.46
	CONF (819)	0.95	0.9	0.95	0.5	0.8	0.5	0.6	0.95	0.95	0.95	0.95	0.95	0.95	0.6	0.7
	VAR (157)	0.14	0.29	0.25	0.13	0.83	0.83	0.17	0.17	0.29	0.33	0.29	0.25	0.25	0.29	0.5
	CONF(157)	0.8	0.9	1	1	0.75	0.25	0.6	0.6	0.9	0.85	0.85	0.85	0.6	0.9	0.9
	VAR (774)	0.25	0.25	0.46	0.58	0.58	0.42	0.25	0.25	0.33	0.14	0.42	0.13	0.46	0.25	0.58
	CONF (774)	0.9	0.95	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.95	0.9

	Survey 23	A1	A2	A3
T1H	ML (399)	3	24	5
M2B	ML (838)	4	3	3
M1M	ML (498)	10	15	15
	BC (399)	2	15.5	3
	BC (838)	2	2	2
	BC (498)	6	9	10
	WC (399)	5.5	44	9
	WC (838)	6.5	5	4.5
	WC (498)	16	23.5	24
	PERT (399)	3.25	25.9	5.33
	PERT (838)	4.08	3.17	3.08
	PERT (498)	10.3	15.4	15.7
	Te (399)	34.5		
	Te (838)	10.3		
	Te (498)	41.4		
	VAR (399)	0.58	4.76	1
	CONF (399)	0.7	0.5	0.5
	VAR (838)	0.42	0.33	0.25
	CONF (838)	0.95	0.5	0.5
	VAR (498)	1	1.42	1.5
	CONF (498)	0.8	0.85	0.85

	Survey 24	A1	A2	A3	A4	A5	A6
T3T	ML (396)	4	2	4	4	4	4
M2T	ML (157)	7	2	7	7	9	9
	BC (396)	2.5	1.25	2.5	2.5	2.5	2.5
	BC (157)	4.5	1	5	4	5.5	5.5
	WC (396)	11	3	11	11	14	14
	WC (157)	13	4	13	13	16.5	16.5
	PERT (396)	4.92	2.04	4.92	4.92	5.42	5.42
	PERT (157)	7.58	2.17	7.67	7.5	9.67	9.67
	Te (396)	27.6					
	Te (157)	44.3					
	VAR (396)	1.42	0.29	1.42	1.42	1.92	1.92
	CONF (396)	1	0.9	0.9	0.9	0.9	0.9
	VAR (157)	1.42	0.5	1.33	1.5	1.83	1.83
	CONF (157)	0.9	0.9	0.85	0.85	0.85	0.85

	Survey 25	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11
T4H	ML (798)	0.5	2	0.5	6	BLNK	27	3.5	2	4	27	BLNK
	BC (798)	0.25	1	0.25	4	BLNK	18	1.5	1	2	18	BLNK
	WC (798)	1	2.5	1	12	BLNK	36	8	4	9	36	BLNK
	PERT (798)	0.54	1.92	0.54	6.67	BLNK	27	3.92	2.17	4.5	27	BLNK
	Te (798)	74.3										
	VAR (798)	0.02	0.06	0.02	1.78	BLNK	9	1.17	0.25	1.36	9	BLNK
	CONF (798)	0.9	0.9	0.9	0.9	BLNK	0.6	0.9	0.9	0.9	0.9	BLNK

	Survey 26	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14
T4H	ML (774)	8	4	4	2	2	4	BLNK	BLNK	BLNK	4	8	5	5	6
	BC (774)	7	3	3	1	1	3	BLNK	BLNK	BLNK	3	7	4	4	5
	WC (774)	14	8	8	4	4	6	BLNK	BLNK	BLNK	6	10	6	6	7
	PERT (774)	8.83	4.5	4.5	2.17	2.17	4.17	BLNK	BLNK	BLNK	4.17	8.17	5	5	6
	Te (774)	54.7													
	VAR (774)	1.36	0.69	0.69	0.25	0.25	0.25	BLNK	BLNK	BLNK	0.25	0.25	0.11	0.11	0.11
	CONF (774)	0.85	1	1	1	0.8	1	BLNK	BLNK	BLNK	0.9	0.9	0.9	0.9	0.85

	Survey 27					
	-	**A1**	**A2**	**A3**	**A4**	**A5**
T1B	BC (739)	2	3	2.5	1	4
	WC (739)	54	18	18	54	6
	-					
	VAR (739)	75.1	6.25	6.67	78	0.11

	Survey 28															
		A1	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**	**A9**	**A10**	**A11**	**A12**	**A13**	**A14**	**A15**
T2M	ML (538)	1.6	4.3	BLNK	2.5	0.5	4	3	11.3	6.2	1.5	101	46.2	5	10	15
	BC (538)	1	3	BLNK	2	0.4	3	2	9	3	1	63	21	2	5	10
	WC (538)	2	6	BLNK	4	1	6	4	16	12	2	168	84	10	20	30
	-															
	PERT (538)	1.57	4.37	BLNK	2.67	0.57	4.17	3	11.7	6.63	1.5	106	48.3	5.33	10.8	16.7
	-															
	Te (538)	223														
	-															
	VAR (538)	0.03	0.25	BLNK	0.11	0.01	0.25	0.11	1.36	2.25	0.03	306	110	1.78	6.25	11.1
	CONF (538)	0.7	0.7	BLNK	0.7	0.7	0.7	0.7	0.7	0.5	0.5	0.5	0.5	0.5	0.5	0.5

	Survey 29																
	-	**A1**	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**	**A9**	**A10**	**A11**	**A12**	**A13**	**A14**	**A15**	**A16**
T4H	ML (619)	9	9	5	1	4	0.5	9	9	2	18	2.5	5.5	7	4	9	2.5
	BC (619)	6	8	3	1	2	0.5	5	5	1	9	2	4	6	3	5	1.5
	WC (619)	20	20	9	3	8	1	10	10	4	60	6	9	10	8	15	4
	-																
	PERT (619)	10.3	10.7	5.33	1.33	4.33	0.58	8.5	8.5	2.17	23.5	3	5.83	7.33	4.5	9.33	2.58
	-																
	Te (619)	108															
	-																
	VAR (619)	5.44	4	1	0.11	1	0.01	0.69	0.69	0.25	72.3	0.44	0.69	0.44	0.69	2.78	0.17
	CONF (619)	0.8	0.9	0.8	0.9	0.8	1	0.8	0.8	0.9	0.5	0.75	0.8	0.9	0.9	0.75	0.9

	Survey 30								
	-	**A1**	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**
T4H	ML (798)	0.5	2	0.5	4	BLNK	2	3.5	2
	BC (798)	0.25	1	0.25	2	BLNK	1	1.5	1
	WC (798)	1	2.5	1	6	BLNK	4	8	4
	-								
	PERT (798)	0.54	1.92	0.54	4	BLNK	2.17	3.92	2.17
	-								
	Te (798)	15.3							
	-								
	VAR (798)	0.02	0.06	0.02	0.44	BLNK	0.25	1.17	0.25
	CONF (798)	0.9	0.9	0.9	0.9	BLNK	0.9	0.9	0.9

	Survey 31									
		A1	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**	**A9**
M1B	ML (408)	4	2	2	2	2	1	4	4	8
	BC (408)	2	1	1	1	1	0.5	2	2	6
	WC (408)	6	4	4	4	4	4	6	6	12
	PERT (408)	4	2.17	2.17	2.17	2.17	1.42	4	4	8.33
	Te (408)	30.4								
	VAR (408)	0.44	0.25	0.25	0.25	0.25	0.34	0.44	0.44	1
	CONF (408)	0.9	0.5	0.5	0.5	0.5	0.75	0.75	0.75	0.5

	Survey 32		
		A1	**A2**
T4T	ML (463)	4	2
	BC (463)	3	2
	WC (463)	7	4
	PERT (463)	4.33	2.33
	Te (463)	6.67	
	VAR (463)	0.44	0.11
	CONF (463)	0.8	0.9

			A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
	Survey 33											
	.	**A1**	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**	**A9**	**A10**
M1M	ML (548)	2	2	10	2	2	2	4	8	4	3
	BC (548)	1	1	3	1	1	1	2	4	2	1
	WC (548)	4	4	18	4	4	4	5	10	5	4
	.										
	PERT (548)	2.17	2.17	10.2	2.17	2.17	2.17	3.83	7.67	3.83	2.83
	.										
	Te (548)	39.2									
	.										
	VAR (548)	0.25	0.25	6.25	0.25	0.25	0.25	0.25	1	0.25	0.25
	CONF (548)	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7

		A1	A2	A3	A4	A5	A6	A7	A8
	Survey 34								
		A1	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**
T2T	ML (661)	4	8	18	2	6	4	BLNK	9
	BC (661)	1	6	9	0.5	4	3	BLNK	6.75
	WC (661)	8	18	36	4	18	8	BLNK	13.5
	.								
	PERT (661)	4.17	9.33	19.5	2.08	7.67	4.5	BLNK	9.38
	.								
	Te (661)	56.6							
	.								
	VAR (661)	1.36	4	20.3	0.34	5.44	0.69	BLNK	1.27
	CONF (661)	0.9	0.9	0.85	0.9	0.9	0.9	BLNK	0.9

	Survey 35										
		A1	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**	**A9**	**A10**
T3M	ML (315)	5	3	9	10	5	3	2	4	6	15
	BC(315)	3	2	6	6	3	2	1	2.5	4	10
	WC(315)	9	5.5	16.5	18	9	5.5	3.5	8	11	27.5
	PERT (315)	5.33	3.25	9.75	10.7	5.33	3.25	2.08	4.42	6.5	16.3
	Te (315)	66.8									
	VAR (315)	1	0.58	1.75	2	1	0.58	0.42	0.92	1.17	2.92
	CONF (315)	0.95	0.9	0.8	0.7	0.95	0.9	0.9	0.8	0.75	0.9

A.10 GEV Max Beta Filters

$\frac{1}{B(\alpha,\beta)}$	α	β	LoS	$\frac{1}{B(\alpha,\beta)}$	α	β	LoS	$\frac{1}{B(\alpha,\beta)}$	α	β	LoS
423.037	3.866	5.925	0.01	0.458	0.744	0.561	0.34	0.131	0.509	0.156	0.67
80.943	3.021	4.473	0.02	0.439	0.732	0.540	0.35	0.125	0.505	0.149	0.68
31.982	2.558	3.676	0.03	0.421	0.721	0.520	0.36	0.121	0.501	0.143	0.69
17.070	2.251	3.149	0.04	0.403	0.709	0.501	0.37	0.115	0.497	0.136	0.70
10.727	2.028	2.767	0.05	0.387	0.699	0.482	0.38	0.111	0.493	0.130	0.71
7.469	1.858	2.474	0.06	0.371	0.688	0.465	0.39	0.106	0.490	0.124	0.72
5.576	1.723	2.242	0.07	0.357	0.679	0.448	0.40	0.101	0.486	0.118	0.73
4.369	1.612	2.052	0.08	0.343	0.669	0.432	0.41	0.097	0.483	0.112	0.74
3.546	1.519	1.892	0.09	0.330	0.660	0.416	0.42	0.092	0.480	0.106	0.75
2.961	1.440	1.756	0.10	0.318	0.651	0.401	0.43	0.087	0.476	0.100	0.76
2.527	1.372	1.638	0.11	0.306	0.643	0.386	0.44	0.083	0.473	0.095	0.77
2.192	1.311	1.535	0.12	0.295	0.635	0.372	0.45	0.079	0.470	0.090	0.78
1.929	1.258	1.443	0.13	0.284	0.627	0.359	0.46	0.075	0.467	0.085	0.79
1.717	1.210	1.361	0.14	0.274	0.619	0.346	0.47	0.071	0.464	0.080	0.80
1.542	1.167	1.286	0.15	0.264	0.612	0.333	0.48	0.067	0.461	0.075	0.81
1.396	1.127	1.219	0.16	0.255	0.605	0.321	0.49	0.063	0.459	0.070	0.82
1.274	1.092	1.157	0.17	0.245	0.598	0.309	0.50	0.059	0.456	0.065	0.83
1.170	1.059	1.101	0.18	0.237	0.591	0.298	0.51	0.056	0.453	0.061	0.84
1.078	1.028	1.048	0.19	0.229	0.585	0.287	0.52	0.051	0.451	0.056	0.85
1.00	**1.00**	**1.00**	**0.20**	0.220	0.579	0.276	0.53	0.048	0.448	0.052	0.86
0.931	0.974	0.955	0.21	0.213	0.573	0.266	0.54	0.044	0.446	0.048	0.87
0.869	0.949	0.913	0.22	0.205	0.567	0.256	0.55	0.040	0.443	0.043	0.88
0.815	0.926	0.874	0.23	0.198	0.561	0.246	0.56	0.037	0.441	0.039	0.89
0.766	0.905	0.837	0.24	0.190	0.556	0.236	0.57	0.033	0.439	0.035	0.90
0.721	0.885	0.802	0.25	0.184	0.550	0.227	0.58	0.029	0.436	0.031	0.91
0.681	0.866	0.769	0.26	0.177	0.545	0.218	0.59	0.027	0.434	0.028	0.92
0.645	0.848	0.738	0.27	0.171	0.540	0.210	0.60	0.023	0.432	0.024	0.93
0.612	0.831	0.709	0.28	0.165	0.535	0.202	0.61	0.019	0.430	0.020	0.94
0.580	0.814	0.681	0.29	0.158	0.531	0.193	0.62	0.017	0.428	0.017	0.95
0.553	0.799	0.655	0.30	0.153	0.526	0.186	0.63	0.013	0.426	0.013	0.96
0.526	0.784	0.629	0.31	0.147	0.521	0.178	0.64	0.010	0.424	0.010	0.97
0.501	0.770	0.605	0.32	0.141	0.517	0.170	0.65	0.006	0.422	0.006	0.98
0.480	0.757	0.583	0.33	0.136	0.513	0.163	0.66	0.003	0.420	0.003	0.99

317

$\frac{1}{B(\alpha,\beta)}$	α	β	LoS	$\frac{1}{B(\alpha,\beta)}$	α	β	LoS	$\frac{1}{B(\alpha,\beta)}$	A	β	LoS
423.037	5.925	3.866	0.01	0.458	0.561	0.744	0.34	0.131	0.156	0.509	0.67
80.943	4.473	3.021	0.02	0.439	0.540	0.732	0.35	0.125	0.149	0.505	0.68
31.982	3.676	2.558	0.03	0.421	0.520	0.721	0.36	0.121	0.143	0.501	0.69
17.070	3.149	2.251	0.04	0.403	0.501	0.709	0.37	0.115	0.136	0.497	0.70
10.727	2.767	2.028	0.05	0.387	0.482	0.699	0.38	0.111	0.130	0.493	0.71
7.469	2.474	1.858	0.06	0.371	0.465	0.688	0.39	0.106	0.124	0.490	0.72
5.576	2.242	1.723	0.07	0.357	0.448	0.679	0.40	0.101	0.118	0.486	0.73
4.369	2.052	1.612	0.08	0.343	0.432	0.669	0.41	0.097	0.112	0.483	0.74
3.546	1.892	1.519	0.09	0.330	0.416	0.660	0.42	0.092	0.106	0.480	0.75
2.961	1.756	1.440	0.10	0.318	0.401	0.651	0.43	0.087	0.100	0.476	0.76
2.527	1.638	1.372	0.11	0.306	0.386	0.643	0.44	0.083	0.095	0.473	0.77
2.192	1.535	1.311	0.12	0.295	0.372	0.635	0.45	0.079	0.090	0.470	0.78
1.929	1.443	1.258	0.13	0.284	0.359	0.627	0.46	0.075	0.085	0.467	0.79
1.717	1.361	1.210	0.14	0.274	0.346	0.619	0.47	0.071	0.080	0.464	0.80
1.542	1.286	1.167	0.15	0.264	0.333	0.612	0.48	0.067	0.075	0.461	0.81
1.396	1.219	1.127	0.16	0.255	0.321	0.605	0.49	0.063	0.070	0.459	0.82
1.274	1.157	1.092	0.17	0.245	0.309	0.598	0.50	0.059	0.065	0.456	0.83
1.170	1.101	1.059	0.18	0.237	0.298	0.591	0.51	0.056	0.061	0.453	0.84
1.078	1.048	1.028	0.19	0.229	0.287	0.585	0.52	0.051	0.056	0.451	0.85
1.00	**1.00**	**1.00**	**0.20**	0.220	0.276	0.579	0.53	0.048	0.052	0.448	0.86
0.931	0.955	0.974	0.21	0.213	0.266	0.573	0.54	0.044	0.048	0.446	0.87
0.869	0.913	0.949	0.22	0.205	0.256	0.567	0.55	0.040	0.043	0.443	0.88
0.815	0.874	0.926	0.23	0.198	0.246	0.561	0.56	0.037	0.039	0.441	0.89
0.766	0.837	0.905	0.24	0.190	0.236	0.556	0.57	0.033	0.035	0.439	0.90
0.721	0.802	0.885	0.25	0.184	0.227	0.550	0.58	0.029	0.031	0.436	0.91
0.681	0.769	0.866	0.26	0.177	0.218	0.545	0.59	0.027	0.028	0.434	0.92
0.645	0.738	0.848	0.27	0.171	0.210	0.540	0.60	0.023	0.024	0.432	0.93
0.612	0.709	0.831	0.28	0.165	0.202	0.535	0.61	0.019	0.020	0.430	0.94
0.580	0.681	0.814	0.29	0.158	0.193	0.531	0.62	0.017	0.017	0.428	0.95
0.553	0.655	0.799	0.30	0.153	0.186	0.526	0.63	0.013	0.013	0.426	0.96
0.526	0.629	0.784	0.31	0.147	0.178	0.521	0.64	0.010	0.010	0.424	0.97
0.501	0.605	0.770	0.32	0.141	0.170	0.517	0.65	0.006	0.006	0.422	0.98
0.480	0.583	0.757	0.33	0.136	0.163	0.513	0.66	0.003	0.003	0.420	0.99

A.12 Normal Beta Filters

$\dfrac{1}{B(\alpha,\beta)}$	α	B	LoS
61.960	3.467	3.467	0.01
24.305	2.866	2.866	0.02
14.047	2.521	2.521	0.03
9.500	2.279	2.279	0.04
7.001	2.093	2.093	0.05
5.455	1.943	1.943	0.06
4.418	1.818	1.818	0.07
3.674	1.710	1.710	0.08
3.117	1.615	1.615	0.09
2.695	1.532	1.532	0.10
2.355	1.456	1.456	0.11
2.084	1.388	1.388	0.12
1.862	1.326	1.326	0.13
1.676	1.269	1.269	0.14
1.518	1.216	1.216	0.15
1.384	1.167	1.167	0.16
1.268	1.121	1.121	0.17
1.166	1.078	1.078	0.18
1.078	1.038	1.038	0.19
1.00	**1.00**	**1.00**	**0.20**
0.930	0.964	0.964	0.21
0.868	0.930	0.930	0.22
0.812	0.898	0.898	0.23
0.761	0.867	0.867	0.24
0.716	0.838	0.838	0.25
0.674	0.810	0.810	0.26
0.635	0.783	0.783	0.27
0.599	0.757	0.757	0.28
0.568	0.733	0.733	0.29
0.538	0.709	0.709	0.30
0.511	0.687	0.687	0.31
0.485	0.665	0.665	0.32
0.461	0.644	0.644	0.33

$\dfrac{1}{B(\alpha,\beta)}$	α	β	LoS
0.439	0.624	0.624	0.34
0.418	0.604	0.604	0.35
0.399	0.585	0.585	0.36
0.381	0.567	0.567	0.37
0.363	0.549	0.549	0.38
0.347	0.532	0.532	0.39
0.333	0.516	0.516	0.40
0.317	0.499	0.499	0.41
0.304	0.484	0.484	0.42
0.292	0.469	0.469	0.43
0.279	0.454	0.454	0.44
0.268	0.440	0.440	0.45
0.257	0.426	0.426	0.46
0.246	0.412	0.412	0.47
0.236	0.399	0.399	0.48
0.227	0.387	0.387	0.49
0.217	0.374	0.374	0.50
0.209	0.362	0.362	0.51
0.200	0.350	0.350	0.52
0.193	0.339	0.339	0.53
0.184	0.327	0.327	0.54
0.177	0.316	0.316	0.55
0.170	0.306	0.306	0.56
0.163	0.295	0.295	0.57
0.157	0.285	0.285	0.58
0.150	0.275	0.275	0.59
0.144	0.265	0.265	0.60
0.139	0.256	0.256	0.61
0.132	0.246	0.246	0.62
0.127	0.237	0.237	0.63
0.122	0.228	0.228	0.64
0.117	0.220	0.220	0.65
0.112	0.211	0.211	0.66

$\dfrac{1}{B(\alpha,\beta)}$	α	B	LoS
0.107	0.203	0.203	0.67
0.102	0.195	0.195	0.68
0.098	0.187	0.187	0.69
0.093	0.179	0.179	0.70
0.089	0.171	0.171	0.71
0.084	0.163	0.163	0.72
0.081	0.156	0.156	0.73
0.077	0.149	0.149	0.74
0.073	0.142	0.142	0.75
0.069	0.135	0.135	0.76
0.065	0.128	0.128	0.77
0.062	0.121	0.121	0.78
0.059	0.115	0.115	0.79
0.055	0.108	0.108	0.80
0.052	0.102	0.102	0.81
0.049	0.096	0.096	0.82
0.045	0.089	0.089	0.83
0.042	0.083	0.083	0.84
0.039	0.077	0.077	0.85
0.036	0.072	0.072	0.86
0.033	0.066	0.066	0.87
0.030	0.060	0.060	0.88
0.028	0.055	0.055	0.89
0.025	0.049	0.049	0.90
0.022	0.044	0.044	0.91
0.020	0.039	0.039	0.92
0.017	0.034	0.034	0.93
0.015	0.029	0.029	0.94
0.012	0.024	0.024	0.95
0.010	0.019	0.019	0.96
0.007	0.014	0.014	0.97
0.005	0.009	0.009	0.98
0.003	0.005	0.005	0.99

A.13 DesignExpert™ Experiment Settings

The tables below show the configurations used to set up the experiment runs in the
DesignExpert™ software. After selecting the "Optimal (Custom)" analysis option
and setting the number of factors and their levels, the information in the tables below
can be used to configure the experiment as was done in this research. In these tables,
delta represents the smallest change detected by the software, sigma is the standard
deviation among the collected weights, and power is a measure of the probability of
successfully detecting whether or not an effect is significant. Recommended power is
80%

Note that the Power levels shown for constraints may not match the values provided
here below and may change based on the final samples used in the design matrix.
These were the values the program calculated when the experiment was completed
for this research.

When populating the run-sheet, the runs will need to be adjusted to match the data
actually collected in this research. The runs suggested by DesignExpert™ are based
on the D-optimality criteria and do not match the demographics of the subjects who
provided information. The ANOVA completed on the data is based on the run-sheet
(i.e. the actual data collected from the subjects).

Project Constraint Analysis – by Demographic	
Design Parameter	**Selected Setting**

Effects Analyzed	Main Effects
	    A: Position
	    B: Years of Experience
	    C: Level of Formal Education
	Interaction
	    AB: Management|Years of Experience

Exchange:	Coordinate
Optimality	D
Blocks	1
Model Points	11
Additional Model Points	2
Lack-of-Fit points	6
Replicate Points	17

Constraint	Delta	Sigma	Delta/Sigma	Power A	Power B	Power C
Cost	0.27	0.15	1.80	99.9%	83.1%	81.8%
Schedule	0.29	0.16	1.81	99.9%	83.6%	82.4%
Quality	0.19	0.1	1.90	99.9%	86.8%	85.7%
Risk	0.24	0.13	1.85	99.9%	84.9%	83.7%

Table A-1: DOE Experiment Set-up – Project Constraints

Risk Aversion	
Design Parameter	**Selected Setting**

Effects Analyzed	Main Effects
	  A: Position
	  B: Years of Experience
	  C: Level of Formal Education
	Interaction
	  AB: Management|Years of Experience

Exchange:	Coordinate
Optimality	D
Blocks	1
Model Points	11
Additional Model Points	3
Lack-of-Fit points	6
Replicate Points	18

Constraint	Delta	Sigma	Delta/Sigma	Power A	Power B	Power C
Utility	2150	1280	1.680	99.9%	81.3%	83.9%

Table A-2: DOE Experiment Set-up –Risk Aversion

Confidence	
Design Parameter	**Selected Setting**

Effects Analyzed	Main Effects
	  A: Position
	  B: Years of Experience
	  C: Level of Formal Education

Exchange:	Coordinate
Optimality	D
Blocks	1
Model Points	8
Additional Model Points	2
Lack-of-Fit points	5
Replicate Points	11

Constraint	Delta	Sigma	Delta/Sigma	Power A	Power B	Power C
Confidence	0.23	0.115	2	99.9%	81.8%	81.8%

Table A-3: DOE Experiment Set-up – Confidence Analysis

Skew Analysis						
Design Parameter		**Selected Setting**				
Effects Analyzed		Main Effects A: Position B: Years of Experience C: Level of Formal Education				
Exchange:		Coordinate				
Optimality		D				
Blocks		1				
Model Points		8				
Additional Model Points		2				
Lack-of-Fit points		7				
Replicate Points		12				
Constraint	**Delta**	**Sigma**	**Delta/Sigma**	**Power A**	**Power B**	**Power C**
$\dfrac{(ML-BC)}{(ML-BC)+(WC-ML)}$	0.19	0.105	1.80952	99.9%	83.2%	83.2%

Table A-4: DOE Experiment Set-up – Duration Estimate Skew

Outlying Estimate Analysis						
Design Parameter		**Selected Setting**				
Effects Analyzed		Main Effects A: Position B: Years of Experience C: Level of Formal Education				
Exchange:		Coordinate				
Optimality		D				
Blocks		1				
Model Points		8				
Additional Model Points		3				
Lack-of-Fit points		6				
Replicate Points		12				
Constraint	**Delta**	**Sigma**	**Delta/Sigma**	**Power A**	**Power B**	**Power C**
BC/(ML+BC)	0.07	0.0396	1.7677	99.8%	80.8%	80.8%
WC/(ML+WC)	0.18	0.0985	1.8274	99.9%	82.5%	82.5%

Table A-5: DOE Experiment Set-up – Outlying Estimate Analysis

Bibliography

"About GAO." 2015. Accessed February 17. http://www.gao.gov/about/index.html.

Alpert, Marc, and Howard Raiffa. 1982. "A Progress Report on the Training of Probability Assessors." In *Judgment Under Uncertainty: Heuristics and Biases*. New York, NY: Cambridge University Press.

Ariely, Dan. 2009a. *Upside of Irrationality Unexpected Benefits of Defying Logic at Work & at Home*. 1 edition. Harper.

———. 2009b. *Predictably Irrational, Revised and Expanded Edition: The Hidden Forces That Shape Our Decisions*. 1 Exp Rev edition. HarperCollins e-books.

Arkes, Hal R. 1985. "The Psychology of Sunk Cost." *The Psychology of Sunk Cost* 35 (1): 124–40. doi:10.1016/0749-5978(85)90049-4.

Baecher, Gregory. 1999. "Expert Elicitation in Geotechnical Risk Assessments." USACE Draft Report. College Park, MD: Department of Civil Engineering, University of Maryland.

"Bayes' Theorem." 2017. *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Bayes%27_theorem&oldid=768607734.

Bennett, F., M. Lu, and S. AbouRizk. 2001. "Simplified CPM/PERT Simulation Model." *Journal of Construction Engineering and Management* 127 (6): 513–14. doi:10.1061/(ASCE)0733-9364(2001)127:6(513).

Benson, P. George, Shawn P. Curley, and Gerald F. Smith. 1995. "Belief Assessment: An Underdeveloped Phase of Probability Elicitation." *Management Science* 41 (10): 1639–53.

Berlin, Isaiah, Henry Hardy, and Michael Ignatieff. 2013. *The Hedgehog and the Fox : An Essay on Tolstoy's View of History*. 2nd ed. Princeton: Princeton University Press.

"Beta Distribution." 2016. *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Beta_distribution&oldid=753618637.

"Beta Function." 2016. *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Beta_function&oldid=749020939.

"Binomial Distribution." 2016. *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Binomial_distribution&oldid=753619524.

Bram, Uri. 2011. *Thinking Statistically*. 3 edition. Capara Books.

Brenner, Lyle A., Derek J. Koehler, Varda Liberman, and Amos Tversky. 1996. "Overconfidence in Probability and Frequency Judgments: A Critical Examination." *Organizational Behavior and Human Decision Processes* 65 (3): 212–19. doi:10.1006/obhd.1996.0021.

Budescu, David V., and Adrian K. Rantilla. 2000. "Confidence in Aggregation of Expert Opinions." *Acta Psychologica* 104 (3): 371–98. doi:10.1016/S0001-6918(00)00037-8.

Buehler, Roger, Dale Griffin, and Michael Ross. 1994. "Exploring the 'Planning Fallacy': Why People Underestimate Their Task Completion Times." *Journal of Personality and Social Psychology* 67 (3): 366–81. doi:10.1037/0022-3514.67.3.366.

Chaloner, Kathryn M., and George T. Duncan. 1983. "Assessment of a Beta Prior Distribution: PM Elicitation." *Journal of the Royal Statistical Society. Series D (The Statistician)* 32 (1/2): 174–80. doi:10.2307/2987609.

Clark, Charles E. 1962. "The PERT Model for the Distribution of an Activity Time." *Operations Research* 10 (3): 405–6.

Clemen, Robert T. 1986. "Calibration and the Aggregation of Probabilities." *Management Science* 32 (3): 312–14.

———. 1987. "Combining Overlapping Information." *Management Science* 33 (3): 373–80.

Davidson, Lynn B., and Dale O. Cooper. 1980. "Implementing Effective Risk Analysis at Getty Oil Company." *Interfaces* 10 (6): 62–75.

Dawes, Robyn M. 1979. "The Robust Beauty of Improper Linear Models in Decision Making." *American Psychologist* 34 (7): 571–82. doi:10.1037/0003-066X.34.7.571.

Dawes, Robyn M., and Bernard Corrigan. 1974. "Linear Models in Decision Making." *Psychological Bulletin* 81 (2): 95–106. doi:10.1037/h0037613.

DeGroot, Morris H., and Stephen E. Fienberg. 1983. "The Comparison and Evaluation of Forecasters." *Journal of the Royal Statistical Society. Series D (The Statistician)* 32 (1/2): 12–22. doi:10.2307/2987588.

DesignExpert (version 9.0.6.2). 2015. Stat-Ease, Inc.

Einhorn, Hillel J. 1974. "Expert Judgment: Some Necessary Conditions and an Example." *Journal of Applied Psychology* 59 (5): 562–71. doi:10.1037/h0037164.

Einhorn, Hillel J., and Hogarth. 1978. "Confidence in Judgment: Persistence of the Illusion of Validity." *Confidence in Judgment: Persistence of the Illusion of Validity.* 85 (5): 395–416. doi:10.1037/0033-295X.85.5.395.

"Euler–Mascheroni Constant." 2016. *Wikipedia.* https://en.wikipedia.org/w/index.php?title=Euler%E2%80%93Mascheroni_constant&oldid=745226377.

Farr, Michael. 2012. "PMP Examp Power Prep: Course Slides and Practice Exams." CMF Solutions and ESI.

French, S. 1986. "Calibration and the Expert Problem." *Management Science* 32 (3): 315–21.

French, Simon. 1980. "Updating of Belief in the Light of Someone Else's Opinion." *Journal of the Royal Statistical Society. Series A (General)* 143 (1): 43–48. doi:10.2307/2981768.

———. 1985. "Group Consensus Probability Distributions: A Critical Survey." In *Bayesian Statistics 2.* New York, NY: Elsevier Science Publishes.

"Gamma Distribution - Wikipedia." 2016. Accessed May 16. https://en.wikipedia.org/wiki/Gamma_distribution.

GAO. 1976. "Space: Acquisition and Utilization of Wind Tunnels by the National Aeronautics and Space Administration." PSAD-76-133. Washington, D.C. http://www.gao.gov/products/PSAD-76-133.

———. 1977a. "Space: NASA's Resource Data Base and Techniques for Supporting, Planning, and Controlling Programs Need Improvement." PSAD-77-78. Washington, D.C. http://www.gao.gov/products/PSAD-77-78.

———. 1977b. "Space: National Aeronautics and Space Administration Should Provide the Congress with More Information on the Pioneer Venus Project." PSAD-77-65. Washington, D.C. http://www.gao.gov/products/PSAD-77-65.

———. 1977c. "Space: Status and Issues Pertaining to the Proposed Development of the Space Telescope Project." PSAD-77-98. Washington, D.C. http://www.gao.gov/products/PSAD-77-98.

———. 1977d. "Space Transportation System: Past, Present, Future." PSAD-77-113. Washington, D.C. http://www.gao.gov/products/PSAD-77-113.

———. 1980a. "Space: A Look at NASA's Aircraft Energy Efficiency Program." PSAD-80-50. Washington, D.C. http://www.gao.gov/products/PSAD-80-50.

———. 1980b. "Space: The Federal Weather Program Must Have Stronger Central Direction." LCD-80-10. Washington, D.C. http://www.gao.gov/products/LCD-80-10.

———. 1982. "Government Operations: GAO Position on Several Issues Pertaining to Air Force Consolidated Space Operations Center Development." Fo/MASAD-82-45. Washington, D.C. http://www.gao.gov/products/MASAD-82-45.

———. 1988a. "Space Exploration: NASA's Deep Space Missions Are Experiencing Long Delays." GAO/NSIAD-88-128BR. Washington, D.C. http://www.gao.gov/products/NSIAD-88-128BR.

———. 1988b. "Space Station: NASA Efforts To Establish a Design-To-Life-Cycle Cost Process." GAO/NSIAD-88-147. Washington, D.C. http://www.gao.gov/products/NSIAD-88-147.

———. 1989. "Weather Satellites: Cost Growth and Development Delays Jeopardize U.S. Forecasting Ability." GAO/NSIAD-89-169. Washington, D.C. http://www.gao.gov/products/NSIAD-89-169.

———. 1991a. "Space Station: NASA's Search for Design, Cost, and Schedule Stability Continues." GAO/NSIAD-91-125. Washington, D.C. http://www.gao.gov/products/NSIAD-91-125.

———. 1991b. "Weather Satellites: Action Needed to Resolve Status of the U.S. Geostationary Satellite Program." GAO/NSIAD-91-252. Washington, D.C. http://www.gao.gov/products/NSIAD-91-252.

———. 1991c. "Weather Satellites: The U.S. Geostationary Satellite Program Is at a Crossroad." GAO/T-NSIAD-91-49. Washington, D.C. http://www.gao.gov/products/T-NSIAD-91-49.

———. 1992a. "Space: NASA's Development of EOSDIS." GAO/IMTEC-92-42R. Washington, D.C. http://www.gao.gov/products/IMTEC-92-42R.

———. 1992b. "Weather Forecasting: Cost Growth and Delays in Billion-Dollar Weather Service Modernization." GAO/IMTEC-92-12FS. Washington, D.C. http://www.gao.gov/products/IMTEC-92-12FS.

———. 1993a. "NASA Program Costs: Space Missions Require Substantially More Funding Than Initially Estimated." GAO/NSIAD-93-97. Washington, D.C. http://www.gao.gov/products/NSIAD-93-97.

———. 1993b. "Space Station: Program Instability and Cost Growth Continue Pending Redesign." GAO/NSIAD-93-187. Washington, D.C. http://www.gao.gov/products/NSIAD-93-187.

———. 1994a. "NASA: Major Challenges for Management." GAO/T-NSIAD-94-18. Washington, D.C. http://www.gao.gov/products/T-NSIAD-94-18.

———. 1994b. "Space Shuttle: NASA's Plans for Repairing or Replacing a Damaged or Destroyed Orbiter." GAO/NSIAD-94-197. Washington, D.C. http://www.gao.gov/products/NSIAD-94-197.

———. 1997. "NASA: Major Management Challenges." GAO/T-NSIAD-97-178. Washington, D.C. http://www.gao.gov/products/T-NSIAD-97-178.

———. 1998. "Space Surveillance: DOD and NASA Need Consolidated Requirements and a Coordinated Plan." GAO/NSIAD-98-42. Washington, D.C. http://www.gao.gov/products/NSIAD-98-42.

———. 2001. "Space Station: Inadequate Planning and Design Led to Propulsion Module Project Failure." GAO-01-633. Washington, D.C. http://www.gao.gov/products/GAO-01-633.

———. 2002a. "Space Station: Actions Under Way to Manage Cost, but Significant Challenges Remain." GAO-02-735. Washington, D.C. http://www.gao.gov/products/GAO-02-735.

———. 2002b. "Space Transportation: Challenges Facing NASA's Space Launch Initiative." GAO-02-1020. Washington, DC. http://www.gao.gov/products/GAO-02-1020.

———. 2003. "NASA: Major Management Challenges and Program Risks." GAO-03-849T. Washington, D.C. http://www.gao.gov/products/GAO-03-849T.

———. 2004. "NASA: Lack of Disciplined Cost-Estimating Processes Hinders Effective Program Management." GAO-04-642. Washington, D.C. http://www.gao.gov/products/GAO-04-642.

———. 2006a. "NASA: Implementing a Knowledge-Based Acquisition Framework Could Lead to Better Investment Decisions and Project Outcomes." GAO-06-218. Washington, D.C. http://www.gao.gov/products/GAO-06-218.

———. 2006b. "NASA: Sound Management and Oversight Key to Addressing Crew Exploration Vehicle Project Risks." GAO-06-1127T. Washington, D.C. http://www.gao.gov/products/GAO-06-1127T.

———. 2006c. "NASA's James Webb Space Telescope: Knowledge-Based Acquisition Approach Key to Addressing Program Challenges." GAO-06-634. Washington, D.C. http://www.gao.gov/products/GAO-06-634.

———. 2006d. "National Aeronautics and Space Administration: Long-Standing Financial Management Challenges Threaten the Agency's Ability to Manage Its Programs." GAO-06-216T. Washington, D.C. http://www.gao.gov/products/GAO-06-216T.

———. 2006e. "Next Generation Air Transportation System: Preliminary Analysis of the Joint Planning and Development Office's Planning, Progress, and Challenges." GAO-06-574T. Washington, D.C. http://www.gao.gov/products/GAO-06-574T.

———. 2006f. "Polar-Orbiting Operational Environmental Satellites: Cost Increases Trigger Review and Place Program's Direction on Hold." GAO-06-573T. Washington, D.C. http://www.gao.gov/products/GAO-06-573T.

———. 2007. "NASA: Challenges in Completing and Sustaining the International Space Station." GAO-07-1121T. Washington, D.C. http://www.gao.gov/products/GAO-07-1121T.

———. 2008. "NASA: Ares I and Orion Project Risks and Key Indicators to Measure Progress." GAO-08-186T. Washington, D.C. http://www.gao.gov/products/GAO-08-186T.

———. 2009a. "Geostationary Operational Environmental Satellites: Acquisition Is Under Way, but Improvements Needed in Management and Oversight." GAO-09-323. Washington, D.C. http://www.gao.gov/products/GAO-09-323.

———. 2009b. "NASA: Assessments of Selected Large-Scale Projects." GAO-09-306SP. Washington, D.C. http://www.gao.gov/products/GAO-09-306SP.

———. 2009c. "NASA: Projects Need More Disciplined Oversight and Management to Address Key Challenges." GAO-09-436T. Washington, D.C. http://www.gao.gov/products/GAO-09-436T.

———. 2010. "NASA: Key Management and Program Challenges." GAO-10-387T. Washington, D.C. http://www.gao.gov/products/GAO-10-387T.

———. 2011. "NASA: Issues Implementing the NASA Authorization Act of 2010." GAO-11-216T. Washington, D.C. http://www.gao.gov/products/GAO-11-216T.

———. 2012. "NASA: Assessments of Selected Large-Scale Projects." GAO-12-207SP. Washington, D.C. http://www.gao.gov/products/GAO-12-207SP.

———. 2013. "James Webb Space Telescope: Actions Needed to Improve Cost Estimate and Oversight of Test and Integration." GAO-13-4. Washington, D.C. http://www.gao.gov/products/GAO-13-4.

———. 2014. "Space Launch System: Resources Need to Be Matched to Requirements to Decrease Risk and Support Long Term Affordability." GAO-14-631. Washington, D.C. http://www.gao.gov/products/GAO-14-631.

———. 2017. "NASA Commercial Crew Program: Schedule Pressure Increases as Contractors Delay Key Events." GAO-17-137, Washington DC. February 16. http://www.gao.gov/products/GAO-17-137.

Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. 2013. *Bayesian Data Analysis, Third Edition*. 3 edition. Chapman and Hall/CRC.

"Generalized Extreme Value Distribution - Wikipedia." 2016. Accessed August 23. https://en.wikipedia.org/wiki/Generalized_extreme_value_distribution.

Genest, Christian, and Mark J. Schervish. 1985. "Modeling Expert Judgments for Bayesian Updating." *The Annals of Statistics* 13 (3): 1198–1212.

Goldratt, Eliyahu-. 1997. *Critical Chain*. The North River Press Publishing Corporation. http://www.amazon.com/Critical-Chain-Eliyahu-M-Goldratt/dp/0884271536/ref=sr_1_1_twi_pap_1?ie=UTF8&qid=1458348235&sr=8-1&keywords=Critical+Chain.

Golenko-Ginzburg, Dimitri. 1988. "On the Distribution of Activity Time in PERT." *The Journal of the Operational Research Society* 39 (8): 767–71. doi:10.2307/2583772.

Gould, Frederick. 2005. *Managing the Construction Process: Estimating, Scheduling, and Project Control*. 3rd ed. Upper Saddle River, New Jersey: Pearson

Education In. https://www.amazon.com/Managing-Construction-Process-Estimating-Scheduling/dp/013113406X/ref=sr_1_1?ie=UTF8&qid=1491012734&sr=8-1&keywords=Managing+the+Construction+Process+Third+Edition.

Grisham, Thomas W. 2010. *International Project Management : Leadership in Complex Environments*. Hoboken, N.J. : Wiley,.

Grubbs, Frank E. 1962. "Attempts to Validate Certain PERT Statistics or 'Picking on PERT.'" *Operations Research* 10 (6): 912–15.

Hammond, Kenneth R. 1996. *Human Judgment and Social Policy : Irreducible Uncertainty, Inevitable Error, Unavoidable Injustice*. New York : Oxford University Press,.

Harrison, J. Michael. 1977. "Independence and Calibration in Decision Analysis." *Management Science* 24 (3): 320–28.

Heath, Chip, and Rich Gonzalez. 1995. "Interaction with Others Increases Decision Confidence but Not Decision Quality: Evidence against Information Collection Views of Interactive Decision Making." *Organizational Behavior and Human Decision Processes* 61 (3): 305–26. doi:10.1006/obhd.1995.1024.

Hogarth, Robin M. 1975. "Cognitive Processes and the Assessment of Subjective Probability Distributions." *Journal of the American Statistical Association* 70 (350): 271–89. doi:10.2307/2285808.

Howard, Ron. 1995. *Apollo 13*. Adventure, Drama, History.

Hubbard, Douglas W. 2009. *The Failure of Risk Management: Why It's Broken and How to Fix It*. Hoboken, New Jersey: John Wiley & Sons, Inc. http://www.amazon.com/Failure-Risk-Management-Why-Broken-ebook/dp/B0026LTMAU/ref=tmm_kin_swatch_0?_encoding=UTF8&qid=1456584143&sr=8-1.

———. 2010. *How to Measure Anything: Finding the Value of Intangibles in Business*. 2 edition. Hoboken, N.J: Wiley.

Jeffreys, Harold. 1983. *Theory of Probability*. Oxford [Oxfordshire] : Clarendon Press ;

Jenner, Lynn. 2015. "Sounding Rockets Overview." Text. *NASA*. March 6. http://www.nasa.gov/mission_pages/sounding-rockets/missions/index.html.

Johnson, D. 1998. "The Robustness of Mean and Variance Approximations in Risk Analysis." *The Journal of the Operational Research Society* 49 (3): 253–62. doi:10.2307/3010474.

———. 2002a. "Triangular Approximations for Continuous Random Variables in Risk Analysis." *The Journal of the Operational Research Society* 53 (4): 457–67.

———. 2002b. "Triangular Approximations for Continuous Random Variables in Risk Analysis." *The Journal of the Operational Research Society* 53 (4): 457–67.

Johnson, David. 1997. "The Triangular Distribution as a Proxy for the Beta Distribution in Risk Analysis." *Journal of the Royal Statistical Society. Series D (The Statistician)* 46 (3): 387–98.

Johnson, Timothy R., David V. Budescu, and Thomas S. Wallsten. 2001. "Averaging Probability Judgments: Monte Carlo Analyses of Asymptotic Diagnostic

Value." *Averaging Probability Judgments: Monte Carlo Analyses of Asymptotic Diagnostic Value* 14 (2): 123–40. doi:10.1002/bdm.369.

Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. Reprint edition. Farrar, Straus and Giroux.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47 (2): 263–91. doi:10.2307/1914185.

Kane, Robert L. 1995. "Creating Practice Guidelines: The Dangers of Over-Reliance on Expert Judgment." *Journal of Law, Medicine and Ethics* 23: 62.

Keefer, Donald L., and Samuel E. Bodily. 1983. "Three-Point Approximations for Continuous Random Variables." *Management Science* 29 (5): 595–609.

Keefer, Donald L., and William A. Verdini. 1993. "Better Estimation of PERT Activity Time Parameters." *Management Science* 39 (9): 1086–91.

Kremer, Steven. 2013a. "Research Range Services 2013 Annual Report." Annual Report. Wallops Flight Facility: NASA. http://www.nasa.gov/centers/wallops/home/#.U9wrXSiwXvc.

———. 2013b. "Wallops Range User's Handbook." 840-HDBK-0003. Wallops Flight Facility: NASA. http://sites.wff.nasa. gov/multimedia/docs/wffruh.pdf.

———. 2015. "Research Range Services 2015 Annual Report." Wallops Flight Facility: NASA.

———. 2017a. "Chapter 5 Comments," January 3.

———. 2017b. "Ch6 - RE: Research Project - Fighting down Panic :)," February 6.

———. 2017c. "RE: Research Project- Fighting down Panic :)," February 6.

Lichtenstein, Sarah, Baruch Fischhoff, and Lawerence Phillips. 1977. "Calibration of Probabilities: The State of the Art." In *Decision Making and Change in Human Affairs*. The Netherlands: D. Reidel Publishing Company.

Lindley, D. V. 1982. "The Improvement of Probability Judgements." *Journal of the Royal Statistical Society. Series A (General)* 145 (1): 117–26. doi:10.2307/2981425.

Lindley, D. V., A. Tversky, and R. V. Brown. 1979. "On the Reconciliation of Probability Assessments." *Journal of the Royal Statistical Society. Series A (General)* 142 (2): 146–80. doi:10.2307/2345078.

Lindley, Dennis V. 1983. "Theory and Practice of Bayesian Statistics." *Journal of the Royal Statistical Society. Series D (The Statistician)* 32 (1/2): 1–11. doi:10.2307/2987587.

Malcolm, D. G., J. H. Roseboom, C. E. Clark, and W. Fazar. 1959. "Application of a Technique for Research and Development Program Evaluation." *Operations Research* 7 (5): 646–69.

Mamet. 2015. "David Mamet Quotes at BrainyQuote.com." *BrainyQuote*. Accessed July 11. http://www.brainyquote.com/quotes/quotes/d/davidmamet478663.html.

Mantel Jr., Samuel J, Jack R Meredith, Scott M. Shafer, and Margaret M Sutton. 2004. *Core Concepts, with CD: Project Management in Practice*. 2 edition. Hoboken, NJ: Wiley.

Marquand, Richard. 1983. *Star Wars: Episode VI - Return of the Jedi*. Action, Adventure, Fantasy.

Martin, Paul K. 2012. "NASA's Challenges to Meeting Cost, Schedule, and Performance Goals." Audit IG-12-021. NASA. http://oig.nasa.gov/audits/reports/FY12/IG-12-021.pdf.

MATLAB (version 9.1.0.441655). 2016. Natick, MA: MathWorks, Inc.

Meehl, Paul E. 1954. *Clinical versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*. Minneapolis, MN: Jones Press, Inc.

Megill, Robert. 1971. *An Introduction to Risk Analysis*. Petroleum Publishing Company.

Microsoft. 2017. "Use a PERT Analysis to Estimate Task Durations - Project." Accessed April 13. https://support.office.com/en-us/article/Use-a-PERT-analysis-to-estimate-task-durations-864b5389-6ae2-40c6-aacc-0a6c6238e2eb.

"MinStableDistribution—Wolfram Language Documentation." 2017. Accessed March 7. https://reference.wolfram.com/language/ref/MinStableDistribution.html.

Moder, Joseph J., and E. G. Rodgers. 1968. "Judgment Estimates of the Moments of Pert Type Distributions." *Management Science* 15 (2): B76–83.

Montgomery, Douglas C. 2008. *Design and Analysis of Experiments*. 7 edition. Hoboken, NJ: Wiley.

Morris, Peter A. 1974. "Decision Analysis Expert Use." *Management Science* 20 (9): 1233–41.

———. 1977. "Combining Expert Judgments: A Bayesian Approach." *Management Science* 23 (7): 679–93.

———. 1983. "An Axiomatic Approach to Expert Resolution." *Management Science* 29 (1): 24–32.

———. 1986. "Observations on Expert Aggregation." *Management Science* 32 (3): 321–28.

Mosleh, A., V. M. Bier, and G. Apostolakis. 1988. "A Critique of Current Practice for the Use of Expert Opinions in Probabilistic Risk Assessment." *Reliability Engineering & System Safety* 20 (1): 63–85. doi:10.1016/0951-8320(88)90006-3.

Mumpower, Jeryl L.Stewart, Thomas R. 1996. "Expert Judgement and Expert Disagreement." *Thinking & Reasoning* 2 (2/3): 191–212. doi:10.1080/135467896394500.

Murphy, Allan H., and Robert L. Winkler. 1977. "Reliability of Subjective Probability Forecasts of Precipitation and Temperature." *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 26 (1): 41–47. doi:10.2307/2346866.

NASA. 2014. "NASA Space Flight Program and Project Management Handbook." NASA/SP-2014-3705. Washington, D.C. http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20150000400.pdf.

———. 2015a. "NASA Space Flight Program and Project Management Requirements W/Change 1-13." Accessed July 16. http://nodis3.gsfc.nasa.gov/npg_img/N_PR_7120_005E_/N_PR_7120_005E_.pdf.

———. 2015b. "NPR 7120.5C NASA Program and Project Management Processes and Requirements." Accessed August 1.

http://nodis3.gsfc.nasa.gov/displayCA.cfm?Internal_ID=N_PR_7120_005C_
&page_name=main.

"NASA Sounding Rockets Annual Report 2013." 2013. Annual Report NP-2013-11-
078-GSFC. Wallops Flight Facility: NASA.
http://sites.wff.nasa.gov/code810/files/Sounding%20Rockets%20Annual%20
Report%202013_sm.pdf.

NIST. 2017a. "1.3.6.7.1. Cumulative Distribution Function of the Standard Normal
Distribution." Accessed March 5.
http://www.itl.nist.gov/div898/handbook/eda/section3/eda3671.htm.

———. 2016b. "NIST/SEMATECH e_Handbook of Statistical Methods." Accessed
December 2.
http://www.itl.nist.gov/div898/handbook/eda/section3/eda366g.htm.

"NIST/SEMATECH E-Handbook of Statistical Methods." 2016. Accessed December
2. http://www.itl.nist.gov/div898/handbook/eda/section3/eda366h.htm.

"Normal Distribution." 2016. *Wikipedia*.
https://en.wikipedia.org/w/index.php?title=Normal_distribution&oldid=75291
7181.

Önkal, Dilek, J. Frank Yates, Can Simga-Mugan, and Şule Öztin. 2003. "Professional
vs. Amateur Judgment Accuracy: The Case of Foreign Exchange Rates."
Organizational Behavior and Human Decision Processes 91 (2): 169–85.
doi:10.1016/S0749-5978(03)00058-X.

Pearson, E. S., and J. W. Tukey. 1965. "Approximate Means and Standard Deviations
Based on Distances between Percentage Points of Frequency Curves."
Biometrika 52 (3/4): 533–46. doi:10.2307/2333703.

Pickard, William F. 2004. "Inverse Statistical Estimation via Order Statistics: A
Resolution of the Ill-Posed Inverse Problem of PERT Scheduling." *Inverse
Problems* 20 (5): 1565. doi:10.1088/0266-5611/20/5/014.

PMI. 2013. *A Guide to the Project Management Body of Knowledge (PMBOK®
Guide)*. Fifth Edition, Kindle Version. Newtown Square, Pa: Project
Management Institute.

R Core Team. 2014. *R: A Language and Environment for Statistical Computing*.
Vienna, Austria: R Foundation for Statistical Computing. http://www.R-
project.org/.

Raiffa, Howard. 1968. *Decision Analysis: Introductory Lectures on Choices Under
Uncertainty*. Reading, Mass.: Longman Higher Education.

Regnier, Eva. 2005a. "Hidden Assumptions in Project Management Tools," no. 11
(January): 1–4.

———. 2005b. "Activity Completion Times in PERT and Scheduling Network
Simulation, Part II." *DRMI Newletter*, no. 12 (April): 1,4-9.

Roberts, Harry V. 1965. "Probabilistic Prediction." *Journal of the American
Statistical Association* 60 (309): 50–62. doi:10.2307/2283136.

Roebber, Paul, and Lance Bosart. 2014. "The Complex Relationship between
Forecast Skill and Forecast Value: A Real-World Analysis: Weather and
Forecasting: Vol 11, No 4." Accessed February 23.
http://journals.ametsoc.org/doi/abs/10.1175/1520-
0434(1996)011%3C0544%3ATCRBFS%3E2.0.CO%3B2.

Rowe, Gene, and George Wright. 2001. "Differences in Expert and Lay Judgments of Risk: Myth or Reality?" *Risk Analysis* 21 (2): 341–56. doi:10.1111/0272-4332.212116.

Ruland, William. 1978. "The Accuracy of Forecasts by Management and by Financial Analysts." *The Accounting Review* 53 (2): 439–47.

Savage, Leonard J. 1971. "Elicitation of Personal Probabilities and Expectations." *Journal of the American Statistical Association* 66 (336): 783–801. doi:10.2307/2284229.

Schervish, Mark J. 1984. "Combining Expert Judgments." Technial Report 294. Pittsburgh, PA: Department of Statistics, Carnegie Mellon University.

———. 1986. "Comments on Some Axioms for Combining Expert Judgments." *Management Science* 32 (3): 306–12.

Selvidge, J. E. 1980. "Assessing the Extremes of Probability Distributions by the Fractile Method*." *Decision Sciences* 11 (3): 493–502. doi:10.1111/j.1540-5915.1980.tb01154.x.

Shanteau, James. 1992. "The Psychology of Experts: An Alternative View." In *Expertise and Decision Support*. New York : Plenum Press,.

Shih, N. -H. 2005. "Estimating Completion-Time Distribution in Stochastic Activity Networks." *The Journal of the Operational Research Society* 56 (6): 744–49.

Silver, Nate. 2012. *The Signal and the Noise: Why So Many Predictions Fail — but Some Don't*. 1 edition. New York: Penguin Press HC, The.

Sniezek, Janet A, and Rebecca Henry. 1990. "Revision, Weighting, and Commitment in Consensus Group Judgment." *Revision, Weighting, and Commitment in Consensus Group Judgment* 45 (1): 66–84. doi:10.1016/0749-5978(90)90005-T.

"Statistical Distributions." 2016. Accessed August 23. http://people.stern.nyu.edu/adamodar/New_Home_Page/StatFile/statdistns.htm.

Steyn, Herman. 2001. "An Investigation into the Fundamentals of Critical Chain Project Scheduling." *International Journal of Project Management* 19 (6): 363–69. doi:10.1016/S0263-7863(00)00026-0.

Surowiecki, James. 2005. *The Wisdom of Crowds*. Reprint edition. New York: Anchor.

Tetlock, Philip. 2005. *Expert Political Judgment*. Kindle Edition. Princeton, New Jersey: Princeton University Press. https://www.amazon.com/dp/B00C4UT1A4/ref=dp-kindle-redirect?_encoding=UTF8&btkr=1.

Trumbo, D, C Adams, M Milner, and L Schipper. 1962. "Reliability and Accuracy in the Inspection of Hard Red Winter Wheat." *Cereal Science Today* 7.

Tsai, Claire I., Joshua Klayman, and Reid Hastie. 2008. "Effects of Amount of Information on Judgment Accuracy and Confidence." *Organizational Behavior and Human Decision Processes* 107 (2): 97–105. doi:10.1016/j.obhdp.2008.01.005.

Tversky, Amos. 1974. "Assessing Uncertainty." *Journal of the Royal Statistical Society. Series B (Methodological)* 36 (2): 148–59.

———. 1975. "A Critique of Expected Utility Theory: Descriptive and Normative Considerations." *Erkenntnis (1975-)* 9 (2): 163–73.

Tversky, Amos, and Daniel Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science*, New Series, 185 (4157): 1124–31.

———. 1981. "The Framing of Decisions and the Psychology of Choice." *Science*, New Series, 211 (4481): 453–58.

Tversky, Amos, and Eldar Shafir. 1992. "Choice under Conflict: The Dynamics of Deferred Decision." *Psychological Science* 3 (6): 358–61.

Tversky, Amos, and Peter Wakker. 1995. "Risk Attitudes and Decision Weights." *Econometrica* 63 (6): 1255–80. doi:10.2307/2171769.

Ward, Dan. 2015. "Ward.pdf." Accessed July 9. http://www.dau.mil/pubscats/ATL%20Docs/Sep-Oct11/Ward.pdf.

waynehale. 2015. "Ten Years After Columbia: STS-112, the Harbinger." *Wayne Hale's Blog*. Accessed August 1. https://waynehale.wordpress.com/2012/12/03/ten-years-after-columbia-sts-112-the-harbinger/.

"Weibull Distribution." 2016. *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Weibull_distribution&oldid=757939623.

Weiss, David, and James Shanteau. 2014. "Empirical Assessment of Expertise (PDF Download Available)." Accessed February 23. https://www.researchgate.net/publication/10614553_Empirical_Assessment_of_Expertise.

West, Mike, and Jo Crosse. 1992. "Modelling Probabilistic Agent Opinion." *Journal of the Royal Statistical Society. Series B (Methodological)* 54 (1): 285–99.

Whittlesea, Bruce W.A. 1990. "Illusions of Immediate Memory: Evidence of an Attributional Basis for Feelings of Familiarity and Perceptual Quality." *Illusions of Immediate Memory: Evidence of an Attributional Basis for Feelings of Familiarity and Perceptual Quality* 29 (6): 716–32. doi:10.1016/0749-596X(90)90045-2.

Winkler, Robert L. 1968. "The Consensus of Subjective Probability Distributions." *Management Science* 15 (2): B61–75.

———. 1981. "Combining Probability Distributions from Dependent Information Sources." *Management Science* 27 (4): 479–88.

———. 1986. "Expert Resolution." *Management Science* 32 (3): 298–303.

Winston, Wayne L. 2003. *Operations Research: Applications and Algorithms*. 4 edition. Belmont, CA: Cengage Learning.

Yates, J. Frank. 1990. *Judgment and Decision Making*. Englewood Cliffs, N.J: Prentice Hall College Div.

Zajonc, Robert B. 1968. "Attitudinal Effects of Mere Exposure." *Attitudinal Effects of Mere Exposure.* 9 (2, Pt.2): 1–27. doi:10.1037/h0025848.

Zio, E. 1996. "On the Use of the Analytic Hierarchy Process in the Aggregation of Expert Judgments." *Reliability Engineering & System Safety* 53 (2): 127–38. doi:10.1016/0951-8320(96)00060-9.