Presented at the 2007 ISPA/SCEA Joint Annual International Conference and Workshop - www.iceaaonline.com

Myth Buster: Do Engineers Trust Parametric Models Over Their Own Intuition?

Ricardo Valerdi
MIT
Cambridge, MA 02139
rvalerdi@mit.edu

Abstract

This paper explores the abilities of engineers to estimate everyday tasks and their reliance on their own intuition when performing cost estimates. The approach to answering these questions is similar to that of the popular television show MythBusters which aims to separate truth from urban legend using controlled experiments. In MythBusters, methods for testing myths and urban legends are usually planned and executed in a manner to produce the most visually dramatic results possible, which generally involves explosions, fires, or vehicle crashes. While the question of parametric models versus intuition is not as exciting, we provide an interesting result that demonstrates the difference between what is real and what is fiction in the world of cost estimation.

Two heuristics, representativeness and anchoring, are explored in two experiments involving psychology students, engineering students, and engineering practitioners. The first experiment, designed to determine if there is a difference in estimating ability in everyday quantities, demonstrates that the three groups estimate with relatively equal accuracy. The results shed light on the distribution of estimates and the process of subjective judgment. The second experiment, designed to explore abilities for estimating the cost of software-intensive systems given incomplete information, shows that predictions by engineering students and practitioners are within 3-12% of each other. Results also show that engineers rely more on their intuition than on parametric models to make decisions.

The value of this work is in helping better understand how software engineers make decisions based on limited information. Implications for the development of software cost estimation models are discussed in light of the findings from the two experiments.

1. Introduction

The process of estimating the cost of software has been of interest to researchers for decades. Some have developed sophisticated algorithms calibrated with historical data to improve the estimation process (Bailey & Basili 1981; Boehm et al 2000; Putnam & Myers 2003). Others have found ways to combine different estimation methods such as bottoms up and analogy to arrive at estimates with a high degree of confidence (Jorgensen et al 2003; Jorgensen 2004). While this research has helped shift the field of software cost estimation from an art to more of a science, the process of estimation remains prone to human errors and biases. These can be especially problematic when there is little information available about the people, technologies, development environment, and process used for developing software.

Even in the face of missing information, humans make assumptions that help them develop software cost estimates. While these assumptions are not always justified, they have a strong influence on the outcome and accuracy of software cost estimates. The fields of human decision making and cognitive science help to further inform this issue.

Tversky and Kahneman (1974) proposed that many human decisions are based on beliefs concerning the likelihood of uncertain events. Occasionally, beliefs concerning uncertain events are expressed in numerical form as odds or subjective probabilities. Their work showed that people rely on a limited number of heuristic principles which reduce the complex task of assessing probabilities and predicting values to simpler judgmental operations. Many heuristics exist in software engineering (Endres & Rombach 2003); arguably the most popular one in software cost estimation is the cube root law (Cook & Leishman 2004) which contends that the software development time in calendar months is roughly three times the cube root of the estimated effort in person-months provided by a model like COCOMO II. This paper does not focus on technology-based heuristics, rather on decision making heuristics that rely heavily on subjective assessments by software engineers.

The subjective assessment of probabilities resembles the subjective assessment of physical quantities such as distance or size. For example, the apparent distance of an object is determined in part by its clarity. The more sharply the object is seen, the closer it appears to be. Similarly, in software engineering, the cost of developing software often depends on the intuitive judgments by the stakeholders involved relative to their point of view.

It is proposed that two heuristics developed by Tversky and Kahneman (1974) have an application in software cost estimation. The first is *representativeness* which is based on the concept that people are concerned with the degree to which *A* is representative of *B*. The symbol *A* could represent a completed software project and *B* could be a new project being estimated. The experiments described in this paper explore this heuristic in the context of predictions of every day values and software-intensive systems.

A second heuristic proposed by Tversky and Kahneman is called *anchoring* which is concerned with the ability for people to make an estimate by starting from an initial value that is adjusted to yield the final answer. The initial value, or starting point, may be suggested by the formulation of the problem, or it may be the result of a partial computation. In the case of this paper, the initial value will be related to the progress of a software-intensive project as it approaches completion. The second experiment described in this paper will explore the application of this heuristic in the context of software cost estimation.

1.1 Research Questions

In light of the current theories of human cognition and decision making, the interest in this paper is to explore how software engineers make decisions on the basis of limited information. The research questions of interest are:

How accurately can software engineers estimate future events given limited information?

How much do engineers rely on their intuition to perform cost estimates?

The exploration of these questions can help inform the field of software cost estimation on many fronts. First, they provide empirical evidence to help better understand the way software

engineers make decisions based on limited information. Second, they shed light on the cognitive limits of software engineers under controlled scenarios which allows for comparison to other populations; technical vs. non-technical as well as student vs. practitioner. This helps determine whether software engineers are necessarily better or worse at estimating certain phenomena. Third, they help determine to what degree software engineers rely on the *representativeness* and *anchoring* heuristics for purposes of decision making.

2. Method

Following the lead of the popular British television show MythBusters, which aims to separate truth from urban legend, two experiments were conducted to test the research questions. In this case, the urban legend is that engineers trust parametric models more than they trust their own intuition. The two experiments were conducted to assess the ability of participants to estimate common quantities as well as the duration of development for a software-intensive system given an elapsed period of time. The first experiment was inspired by previous work on optimal predictions in everyday cognition (Griffiths & Tenenbaum 2006) but was extended to the area of cost estimation by applying the idea of cognitive estimation limits. The original set of questions remained the same so that data from previous studies could be compared to newly obtained data. Results were obtained for this experiment through the use of a survey instrument provided in Appendix A. The second experiment involved only engineering students and practitioners since it was intended to assess the ability of participants to estimate the duration, in person months, of the development of a software-intensive system and reliance on intuition over a parametric model.

2.1 Participants

Participants were tested in three groups, with each group making predictions about different phenomena. The first group, made up of 142 undergraduate students, participated in the experiment as part of a psychology class and is referred to as *psychology students* throughout the paper. The second group, made up of 36 graduate-level engineering students, participated in the experiment as part of a lecture in a project management class and is referred to as *engineering students* throughout the paper. The third group, made up of 49 software and system cost estimation professionals, participated in the experiment as part of a day-long workshop on cost estimation and is referred to as *practitioners* throughout the paper. The engineering students had anywhere between 0-2 years of work experience in cost estimation whereas the practitioners have an average of 12 years of experience and were familiar with advanced cost estimation principles.

2.2 Description of Experiment #1

The first experiment was conducted by giving individual pieces of information to each of the participants in the study, and asking them to draw a general conclusion. For example, many of the participants were told the amount of money that a film had supposedly earned since its release, and asked to estimate what its total "gross" would be, even though they were not told for how long it had been playing. In other words, participants were asked to predict t_{total} given t_{past}.

No additional information was given about the film such as the genre, country of origin, actors, or production studio.

In addition to the returns on films, the participants were asked about things as diverse as the number of lines in a poem (given how far into the poem a single line is), an individual's life span (given his current age), the duration of a Pharaoh's reign (given he had reigned for a certain time), the run-time of a film (given an already elapsed time), the total length of the term that would be served by an American congressman (given how long he has already been in the House of Representatives), the time it takes to bake a cake (given how long it has already been in the oven), and the amount of time spent on hold in a telephone queuing system (given an already elapsed time). All of these items have known values and well-established probability distributions. The intent of the experiment was to determine whether the individual participants were able to provide an estimate from the lone pieces of data and, as a group, derive the expected distribution of answers for each item. The eight questions are provided in Appendix A, Part I.

2.3 Description of Experiment #2

The second experiment was conducted in a similar fashion except it only involved the engineering students and practitioners because of the technical content. The focus was to capture the estimation limits of participants given a limited amount of information and the reliance of intuition when performing cost estimates. The first part of the experiment contained questions about the expected duration of a software-intensive project given an elapsed period of time. Participants were given four system life cycle phases to use as their mental model: conceptualize, develop, operational test & evaluation, and transition to operation. Similar to experiment 1, no additional information was given about the project such as application domain, development organization, or historical performance. Participants were asked to predict the total effort needed for a project, t_{total}, given a certain amount of effort had already been expended on one or more life cycle phases, t_{past}. In the first question, t_{past} = 300 person months for the Conceptualize phase. In the second question, t_{past} = 300 person months in the Conceptualize and Develop phases. In the third question, t_{past} = 300 person months for the Conceptualize, Develop, and Operational Test & Evaluation phases. The three questions are provided in Appendix A, Part II.

The second part of the experiment asked participants to predict the total systems engineering effort for a software-intensive system, t_{total}, given the predicted effort from a cost model, $t_{predicted}$, and a historical data point, $t_{historical}$, from a similar system of equivalent scope and complexity. A relatively new cost model, COSYSMO, was selected for this experiment to avoid any unbalanced expertise from practitioners. Moreover, both the engineering students and the practitioners received an initial tutorial on the use of COSYSMO and its definitions to ensure that there were no misinterpretations of the questions. In the first question, $t_{predicted}$ = 100 person months and $t_{historical}$, = 110 person months. In the second question, $t_{predicted}$ = 1,000 person months and $t_{historical}$, = 1,100 person months. The two questions are provided in Appendix A, Part III.

3. Results

People's predictions about everyday events were on the whole extremely accurate. The results of the responses from the psychology students are provided in Figure 1.

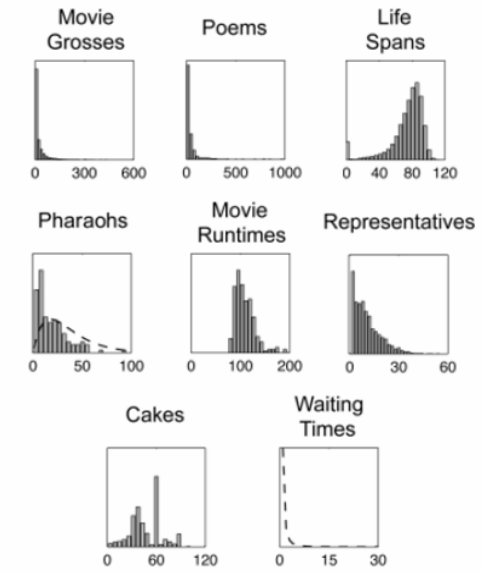


Figure 1. Relative Probabilities of t values for Psychology Students, n = 142
(Griffiths & Tenenbaum 2006)

The distributions for movie grosses and poems are approximately power-law which accurately indicates that the majority of movies gross very little money but there are a few which become blockbuster hits. For example, out of over 7,300 films worldwide from the period 1900-2006 only three films grossed over $1B. Similarly, the majority of poems are very short but there are a few which are very long.

The distribution of life spans approximately follows a Gaussian distribution which accurately indicates that most males, at least in the U.S., the distribution is centered around the average life expectancy of 75. Half of the male population dies before reaching the age of 75 and half of the population dies after but at a much sharper rate. The movie runtime also follows a quasi-Gaussian distribution since most movies run at least 90 minutes and some of them longer. The distribution of length of terms for representatives is approximately Erlang which accurately indicates that most representatives serve a small amount of two-year terms. Very few of them get re-elected despite the fact that they are eligible to get re-elected an unlimited number of times. The cake distribution is complex and irregular but can be described as a bimodal distribution that is Gaussian-like around the value of 45 and spikes at the value 60. This is consistent with recipes that indicate that most cakes take either 45 minutes or 60 minutes depending on the type of cake, ingredients, and altitude among other factors. The complete list of sources of data for the eight questions is provided in Appendix B.

Of particular interest is the similarity in the distribution of the answers across the three population types and the proximity in the mean values for t_{total}. The psychology students and engineering students were just as accurate in estimating t_{total} for the eight questions in the first experiment compared to the practitioners as shown in Table 1.

Table 1. Mean Values of Results for Experiment 1

	Psychology Students (n = 142)	Engineering Students (n = 36)	Practitioners (n = 49)
Movie Grosses (in Millions)	40	41	42
Poems (lines)	22	20	21
Life Spans (years)	76	73	78
Pharaohs (years)	30	23	23
Movie Runtimes (Minutes)	120	105	108
Representatives (years)	18	21	22
Cakes (minutes)	53	48	50
Waiting times (minutes)	10	7	9

When it came to estimating t_{total} for the scenarios presented in the second experiment, there was a negligible difference between engineering students and practitioners as shown in Table 2. Note that the standard deviation is shown in brackets below the mean value. It should be noted that the number of samples differs slightly from experiment 1 because of missing data from one participant.

Table 2. Mean and Standard Deviation of Results for Experiment 2

	Engineering Students (n = 36)	Practitioners (n = 48)
Through one phase (PM)	1516 *[1011]*	1386 *[758]*
Through two phases (PM)	666 *[266]*	594 *[241]*
Through three phases (PM)	401 *[129]*	390 *[145]*
Project X (PM)	112 *[7]*	110 *[9]*
Project Y (PM)	1140 *[128]*	1122 *[111]*

The difference in estimates for t_{total} between engineering students and practitioners was 9%, 12%, and 3% for the three scenarios, respectively. Interestingly, engineering students estimated consistently higher than the practitioners in all three scenarios. However, the mean values of

their estimates were very close considering the small amount of information provided to both groups in the survey. The dispersion of the distributions of the estimates for the three scenarios, or coefficients of variation, were 0.66, 0.39, and 0.32 for the engineering students and 0.55, 0.41, and 0.37 for the practitioners. This indicates that both groups followed a similar pattern of increased intra-group agreement indicated by a reduction of the standard deviation of the distribution of their answers relative to the mean of the distribution.

The results from the second part of experiment #2, also displayed in Table 2, show that the difference in estimates for t_{total} between engineering students and practitioners was 2% for both scenarios. Engineering students again estimated consistently higher than the practitioners but this was relatively negligible considering the amount of information that was provided in the questionnaire. Coefficients of variation were 0.06 and 0.11 for the engineering students and 0.08 and 0.09 for the practitioners which demonstrate a relatively balanced set of responses from both groups.

4. Analysis

The two experiments performed shed light on the estimation accuracy of the three populations of participants. The psychology students served as a control group for comparing engineering student's and practitioner's ability to estimate every day values. As the results from the first experiment show, all three groups predicted values of every day events with relative accuracy with the exception of the Pharaoh question. Both the magnitude of errors and the variance in judgments across participants were substantially greater for this question than for our other questions. A Pharaoh is a title used to refer to any ruler, usually male, of the Egyptian kingdom in the pre-Christian, pre-Islamic period. Compared to other questions in the survey, which were of more contemporary tone, participants would typically not be aware of the typical rule of Egyptian rulers thousands of years ago. Therefore, they must depend on their judgment of present day events to produce an estimate.

Despite the lack of direct experience, people's predictions were not completely off the mark: Their judgments were consistent with having implicit knowledge of the correct form of the underlying distribution but making incorrect assumptions about how this form should be parameterized (i.e., its mean value). The predictions for the reigns of Pharaohs suggest a general strategy people might employ to make predictions about unfamiliar kinds of events, which is surely an important prediction problem faced in everyday life. Given an unfamiliar prediction task, people might be able to identify the appropriate form of the distribution by making an analogy to more familiar phenomena in the same broad class, even if they do not have sufficient direct experience to set the parameters of that distribution accurately.

For instance, participants might have been familiar with the length of time that various modern monarchs have spent in their positions, as well as with the causes (e.g., succession, death) responsible for curtailing those times, and it is not unreasonable to think that a similar thought process could have governed the durations of Pharaohs' reigns in ancient Egypt. Yet most people might not be aware of, or might not remember, just how short life spans typically were in ancient Egypt compared with modern expectations, even if they know life spans were somewhat shorter. The survey question regarding present-day life spans appeared immediately before the Pharaoh question which may have skewed their opinion based on current life expectancy norms.

If participants predicted the reign of the pharaoh by drawing an analogy to modern monarchs and adjusting the mean reign duration downward by some uncertain but insufficient factor, that

would be entirely consistent with the pattern of errors observed. Such a strategy of prediction by analogy could be an adaptive way of making judgments that would otherwise lie beyond people's limited base of knowledge and experience. This phenomenon is what is precisely described by the *representativeness* heuristic. By estimating by analogy, participants were able to approximately guess the mean length of a Pharaoh's reign. However, the analogy method is inaccurate when knowledge and experience are obstacles to the process as is often the case with software cost estimation. Large databases of historical projects may be available for use in estimation by analogy method but, when the context of the projects is not known, the utility of the projects may be overestimated and may actually lead to inaccurate conclusions about the applicability of the current project.

The results from the second experiment demonstrate an application of the *anchoring* heuristic but seen from a longitudinal perspective. The distribution of predictions of t_{total} by engineering students and practitioners decreased as the system life cycle progressed. In other words, as more of the project was complete, the smaller the standard deviation of responses for t_{total}. These results confirm previous hypotheses about a phenomenon referred to as the cone of uncertainty in cost estimation (McConnell 2006; Little 2006). Responses from the three stages are plotted and rotated ninety degrees to visually demonstrate the convergence of results. The responses from engineering students, shown in Figure 2, have a higher variance compared to the responses from practitioners, shown in Figure 3.

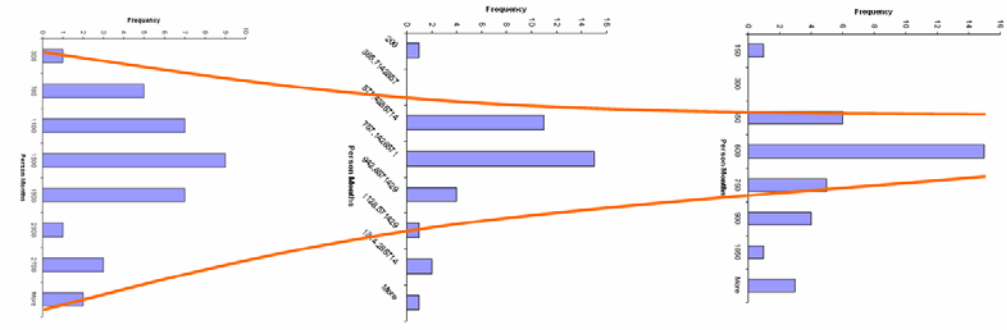


Figure 2. Engineering Student Estimates for Three Scenarios, n = 36

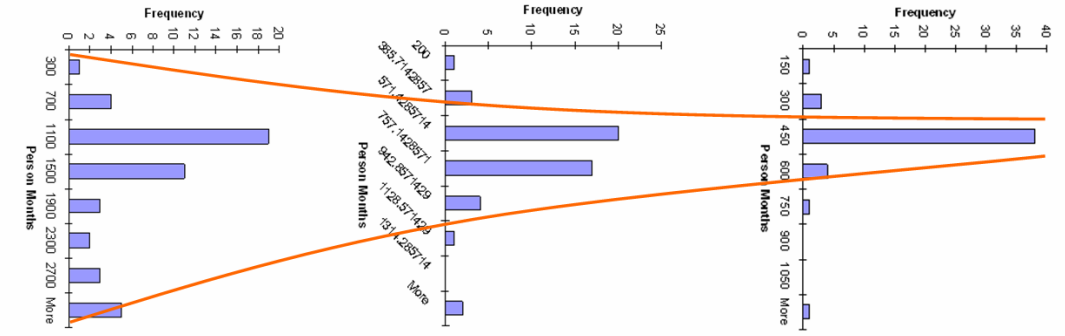


Figure 3. Practitioner Estimates for Three Scenarios, n = 48

Results from the second experiment also confirm that engineering students slightly overestimate compared to practitioners. The overestimation is even more apparent when the distributions of responses are visually compared. Even though the distributions are

approximately Gaussian and the mean values are within 3-12% of each other, the variance of responses from the engineering students is slightly higher.

Results from the final part of the second experiment, where two scenarios are provided and participants are asked to estimate t_{total}, is provided in Figures 4 and 5. From inspection of these results, it is evident that individuals trust historical data more than they trust cost models.

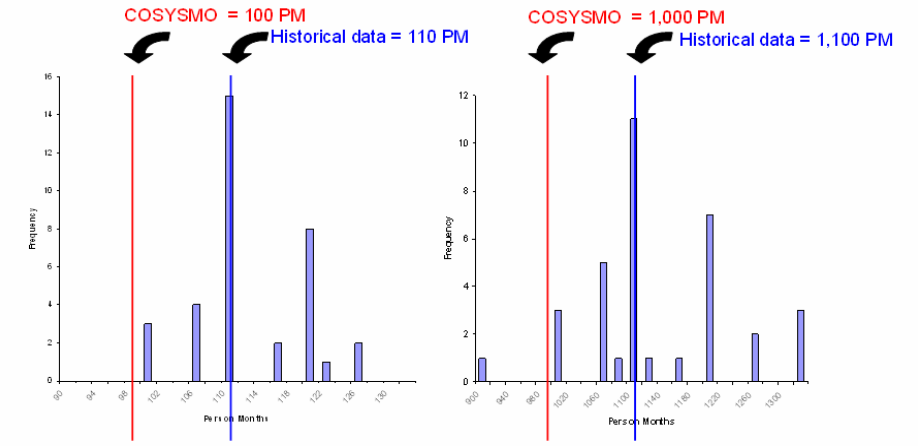


Figure 4. Engineering Student Estimates for Three Scenarios, n = 36

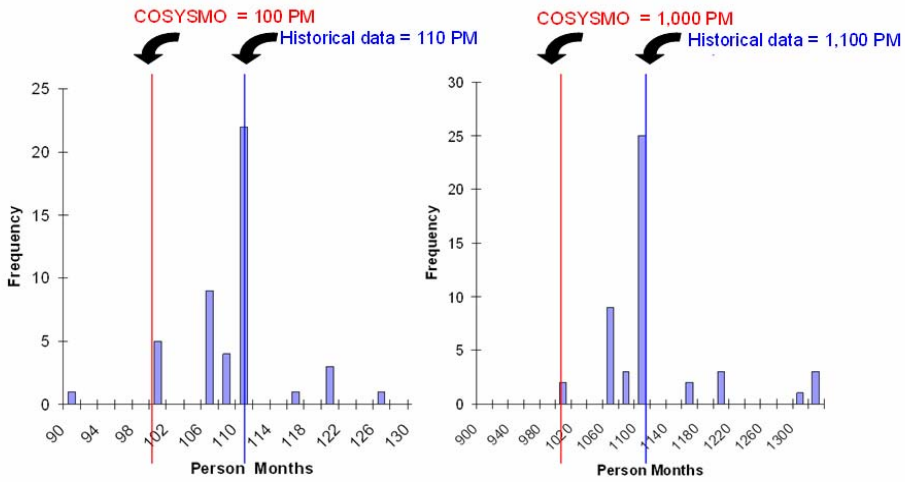


Figure 5. Practitioner Estimates for Three Scenarios, n = 48

Both the engineering students (Figure 4) and the practitioners (Figure 5) demonstrated a bias towards historical data and even overestimated the effort needed to develop the system despite the information provided. The responses from the engineering students had a higher variance than the practitioners as observed in other sections of the experiment.

The bias towards historical data is the most significant result from this experiment because it shows that estimators are more likely to base their estimate on their intuition, explained by the *representativeness* heuristic (*A* is representative of *B*), than on parametric models.

4.1 Threats to Validity

As discussed in other empirical software engineering studies (Jedlitschka & Ciolkowski 2006), it is necessary to identify possible threats to validity that could bring into question the experiment and its results.

The execution of the experiment itself could affect the internal validity of this study. Namely, the survey administration for the psychology students was performed by one set of researchers while the survey administration for the engineering students and practitioners was performed by another. While this was not done deliberately it could affect consistencies in survey administration and potentially affect the quality of the results due to the difference in experimental setting.

Another experimental threat could be that the survey participants, when given the set of questions to answer, were trying hard to find the right answer because they may have perceived this as a test of intelligence. This is a well known effect in educational measurement and is often referred to as the Pygmalion effect.

One aspect that would make this experiment feel quite different that a real world situation is that motivation for participating is very different than in a real project. Therefore, the biases may not be as visible in the experiment, especially for the practitioners. This could also explain the chronic overestimation by the participants.

Despite the healthy sample size, the survey was not distributed to a representative population of software engineers. Quite the contrary, the practitioners that participated are known to be involved in several process improvement initiatives. They are also employed by organizations which have traditionally motivated their employees to follow a high degree of process maturity. This could be considered a biased sample because of the tendency to be familiar with mature practices. This could severely affect the external validity of the results.

The sample of students was also not random. The students that participated in the experiments were undergraduate psychology students and graduate engineering students. Both are considered to be highly motivated and educated compared to the normal population and therefore could have know the correct answer to the questions being asked. It is less likely that they did not know the answer since they could have an “educated guess” was likely to be relatively accurate.

Even with these known issues of internal and external validity, it is believed that the results of the experiment are informative to the research questions since the pool of participants are likely to behave like decision makers in software organizations.

5. Discussion

Empirical data has been provided to explore the estimation accuracy of software engineers compared to two student populations. On the whole, judgments of everyday quantities such as movie times and life expectancy were quite accurate and exhibited known distribution profiles. Other everyday quantities, such as the reign of Pharaohs, were not as precise but nevertheless provided insight into the heuristics used by people to arrive at quantities of unfamiliar topics.

Much work is left to do in understanding the underlying reasons why people can turn observed coincidences into heuristics. Somehow, the human mind is capable of acquiring useful knowledge about the world and employing rational statistical mechanisms to make predictions about future occurrences. Moreover, the process of inferring hidden structure from observed

data seems to be more of an unconscious process as opposed to a deliberate one. The exploration of these concepts in software engineering can lead to future theories and hypotheses that will further inform how people use their cognitive abilities to make judgments.

5.1 Implications

Two main implications result from the results discussed. First, it was shown that students are pretty good estimators compared to practitioners, but they tend to overestimate. This support the argument others have made about the suitability of students for software engineering experiments (Host et al 2000; Carver, Jaccheri et al 2003; Carver, Shull et al 2003; Berander 2004). While students are not ideal for all types of experiments, they have proven to be adequate participants for experiments in cost estimation.

Another important implication of this work is that all survey participants were influenced more by historical information than by the answer provided by the cost model, even though the estimate provided by each was 10% apart. Furthermore, participants in the second experiment overestimated the effort needed to develop a system despite the historical data provided. Cost modeling research should continue to work towards the development of sophisticated models but should note the fact that software engineers will not depend on the answer provided by the models alone, contrary to urban myth. They will incorporate historical data, their own heuristics based on past observations, and personal biases regarding the specific situation. These heuristics and biases need to be considered not only from a technological standpoint as done in cost estimation (Peeters & Dewey 2000) but also from a cognitive standpoint in order to fully understand them. Myth uncovered, thanks to the MythBusters approach.

6. References

Bailey, J., Basili, V., "A Meta-Model for Software Development Resource Expenditures," *Proceedings of the Fifth International Conference on Software Engineering*, March 1981, pp. 107-116.

Berander, P., "Using Students as Subjects in Requirements Prioritization", *International Symposium on Software Engineering*, 2004, pp. 167-176.

Boehm, B. W., Abts, C., Brown, A. W., Chulani, S., Clark, B., Horowitz, E., Madachy, R., Reifer, D. J. and Steece, B., *Software Cost Estimation With COCOMO II*, Prentice Hall, 2000.

Carver, J., Jaccheri, L., Morasca, S., and Shull, F., "Issues in Using Students in Empirical Studies in Software Engineering Education", *International Software Metrics Symposium*, 2003, pp. 239-249.

Carver, J., Shull, F., and Basili, V., "Observational Studies to Accelerate Process Experience in Classroom Studies: An Evaluation", *International Symposium on Empirical Software Engineering*, 2003, pp. 72-79.

Cook, D. A., Leishman, T. R., "Lessons Learned from Software Engineering Consulting", *Journal of Defense Software Engineering*, February 2004.

Endres, A., Rombach, D. H., *A Handbook of Software and Systems Engineering: Empirical Observations, Laws, and Theories*, Pearson Addison Wesley, 2003.

Griffiths, T. L., Tenenbaum, J. B., "Optimal Predictions in Everyday Cognition", *Psychological Science*, Vol. 17, No. 9, 2006, pp. 767-773.

Höst, M., Regnell, B., and Wohlin, C., "Using Students as Subjects - A Comparative Study of Students and Professionals in Lead-Time Impact Assessment", *Empirical Software Engineering*, Vol. 5, No. 3, 2000, pp. 201-214.

Jedlitschka, A., Ciolkowski, M., "Reporting Experiments in Software Engineering", *Fraunhofer Institute for Experimental Software Engineering, Technical Report* ISERN-06-01, 2006.

Jorgensen, M., Indahl, U., Sjoberg, D. I. K., "Software Effort Estimation by Analogy and Regression Toward the Mean", *Journal of Systems and Software*, Vol 68, No. 3, 2003, pp. 253-262.

Jorgensen, M., "Top-Down and Bottom-Up Expert Estimation of Software Development Effort", *Information and Software Technology*, Vol. 46, No. 1, 2004, pp. 3-16.

Little, T., "Schedule Estimation and Uncertainty Surrounding the Cone of Uncertainty", *IEEE Software*, Vol. 23, No. 3, 2006, pp. 48-54.

McConnell, S., *Software Estimation: Demystifying the Black Art*, Microsoft Press, 2006.

Peeters, D., Dewey, G., "Reducing Bias in Software Cost Estimates", *Journal of Defense Software Engineering*, April 2000.

Putnam, L. H., Myers, W., *Five Core Metrics: The Intelligence Behind Successful Software Management*, Dorset House, 2003.

Tversky, A., Kahneman, D., "Judgment Under Uncertainty: Heuristics and Biases", *Science*, Vol. 185, 1974, pp. 1124-1131.

Appendix A. Survey Instrument

Survey on Intuitive Judgments

Name _____ Years of work experience _____

Years of experience in cost estimation (of any kind) _____

What do you consider yourself to be (check all that apply)?

☐	☐	☐	☐	☐
Program Manager	software engineer	hardware engineer	systems engineer	Other_____

Each of the questions below asks you to predict either a duration or a quantity based on a single piece of information. Read each question and write your prediction on the line below it. We are interested in your intuitions, so please don't make complicated calculations. Just tell us what you think.

Part I: 8 questions

1. *Movie Grosses.* Imagine you hear about a movie that has taken in $10M at the box office, but don't know how long it has been running. What would you predict for the total amount of box office intake for that movie? _____

2. *Poems.* If your friend read you her favorite line of poetry and told you it was line 5 of a poem, what would you predict for the total length of the poem? _____

3. *Life Spans.* Insurance agencies employ actuaries to make predictions about people's life spans – the age at which they will die – based upon demographic information. If you were assessing an insurance case for an 18 year old man, what would you predict for his life span? _____

4. *Pharaohs.* If you opened a book about the history of ancient Egypt to a page listing the reigns of the pharaohs, and noticed that at 4000 BC a particular pharaoh had been ruling for 11 years, what would you predict for the total duration of his reign? _____

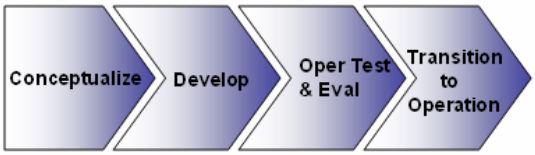
5. *Movie Runtimes.* If you made a surprise visit to a friend and found that they had been watching a movie for 30 minutes, what would you predict for the total length of the movie? _____

6. *Representatives.* If you heard a member of the House of Representatives had served for 15 years, what would you predict their total term in the House to be? _____

7. *Cakes.* Imagine you are in somebody's kitchen and notice that a cake is in the oven. The timer shows that it has been baking for 35 minutes. What would you predict for the total amount of time the cake needs to bake? _____

8. *Waiting times.* If you were calling a telephone box office to book tickets and had been on hold for 3 minutes, what would you predict for the total time you would be on hold? _____

Part II: 3 questions
These questions require you to be familiar with the four life cycle phases covered in COSYSMO. They are: (1) Conceptualize, (2) Develop, (3) Operational Test & Evaluation, and (4) Transition to Operation



1. *Through one phase of Systems Engineering.* Imagine that a project has taken 300 Person Months of systems engineering effort through the end of the **Conceptualize** phase. What is the total systems engineering effort you predict will be needed to deliver the system (i.e., through the completion of Transition to Operation)? _____

2. *Through two phases of Systems Engineering.* Imagine that a project has taken 300 Person Months of systems engineering effort through the end of the **Conceptualize & Develop** phases. What is the total systems engineering effort you predict will be needed to deliver the system? _____

3. *Through three phases of Systems Engineering.* Imagine that a project has taken 300 Person Months of systems engineering effort through the end of the **Conceptualize, Develop, and OT&E** phases. What is the total systems engineering effort you predict will be needed to deliver the system? _____

Part III: 2 questions
These questions assume that the COSYSMO was used to obtain systems engineering effort estimates.

1. The effort estimate for Project X provided by COSYSMO is **100** Person Months. Historical data from your organization shows that a similar system of equivalent scope & complexity took **110** Person Months to complete. What would you predict for the total systems engineering effort for Project X? _____

2. The effort estimate for Project Y provided by COSYSMO is **1,000** Person Months. Historical data from your organization shows that a similar system of equivalent scope & complexity took **1,100** Person Months to complete. What would you predict for the total systems engineering effort for Project Y? _____

END

Appendix B. Data Source for Estimating Everyday Predictions

Dataset	Source (# of data points)
Movie Grosses	http://www.worldwideboxoffice.com
Poems	http://www.emule.com/
Life Spans	http://www.ssa.gov/OACT/STATS/table4c6.html (Actuarial Life Table)
Pharaohs	http://www.touregypt.net
Movie Runtimes	http://us.imdb.com/chart/
Representatives	http://bioguide.congress.com/
Cakes	http://allrecipes.com/
Waiting Times	Griffiths & Tenenbaum (2006)