views.

The Future

The preceding parts of the tutorial dealt with well-established empirical paradigms, here we look at new and experimental paradigms such as visualization, simulation, various kinds of sampling techniques, as well as ways of leveraging from academic and industrial collaboration.

A BIBLIOGRAPHY

Note that this bibliography is not meant to be complete. What we have attempted to do is identified those areas important to credible empirical studies where both people in large organizations and technology are present. The books or articles marked with an asterisk (*) are required reading for interested in these kinds of empirical studies.

Case Study Methods

*Allen S. Lee, "A Scientific Methodology for MIS Case Studies", MIS Quarterly, March 1989, pp. 33 - 50. This is a fundamental paper on the logic and rational describing scientific case studies. Sloan school at MIT uses this paper as the principle template for how to do case studies in organizational settings.

Data Analysis and Statistics

*G. E. P. Box, W. G. Hunter, and J. S. Hunter, Statistics for Experimenters, John Wiley & Sons, 1978. This book is a good all round statistics book with many examples and good theoretical treatment.

George G. Roussas, A First Course in Mathematical Statistics, Addison-Wesley Publishing Co., 1973. Don't be fooled by the title, this is not a statistics book for the mathematically faint of heart. The author gives an excellent theoretical treatment of statistics.

*Sidney Siegel and John N. Castellan, Jr. Nonparametric Statistics for the Behavioral Sciences, McGraw-Hill Inc., second edition, 1988. This is an excellent book that list most important nonparametric test and shows their uses with many examples. The authors discuss experimental design in the introductory section.

Empirical Studies of an Individual

E. Soloway and S. Iyengar (editors), Empirical Studies of Programmers, Ablex Publishing Corp., 1986. This series has many good solid studies of individuals and what they are doing when they are developing software. The Curtis' paper is important because it addresses directly the external threat to validity of taking the results of student experiments and applying them directly to professional software developers.

Foundation Papers for Understanding People Effects in Experiments

*A.R. Dennis and J.S. Valacich, "Computer Brainstorms: More Heads Are Better Than One", Journal of Applied Psychology, 78:4, pp. 531—537, 1993.

*G.A. Miller, "The Magical Number 7, Plus or Minus 2: Some Limits on Our Capacity for Processing Information," Psychological Review, 63:2, March 1956, pp. 81—97.

*H.M. Parson, "What Happened at Hawthorne?" Science, Vol 183, March 1974, pp. 922—932.

Large Engineering Organizations

*Thomas .J. Allen. Managing the Flow of Technology. MIT Press, Cambridge, MA, 1977. Tom summarizes his work from his first 20 years with an excellent treatment of how engineering organizations work and how individuals in them communicate and exchange technology.

*Jeffery K. Liker and Walton M. Hancock, "Organizational Systems Barriers to Engineering Effectiveness", IEEE Transactions on Engineering Management, EM-33:2, May, 1986, pp. 82—91. One of the best descriptions of what can go wrong when organizations and processes are not matched with the reward system for individuals. The domain is automobile manufacture; but the results apply to many large software organizations.

Process Improvement

W. Edwards Deming. Out of the Crisis. MIT Press, Cambridge, MA. 1982. As corny as this may sound, we think this is one of the best books a manager could ever read about managing human resources in an organization. Every time we refer to this book, we are more amazed at how right most of the advice and observations are.

H. James Harrington. Business Process Improvement. McGraw Hill Inc., New York, 1991. Domain is not software production; however, we can all learn from his experiences. The author covers many topics including approaches to process improvement that 5ESS has tested with much success.

L.C. Briand, C. M. Differding, and H. D. Rombach. "Practical Guidelines for Measurement-Based Process Improvement", Software Process: Improvement & Practice, to appear. Lessons learned in using the GQM (goal, question, metric) approach to process improvement.

Social Experimental Design

Samuel D. Conte, H. E. Dunsmore, and V. Y. Shen. Software Engineering Metrics and Models. Benjamin/Cummings Publishing Company, Menlo Park, CA., 1985. This is not really about social experimental design; however, Chapter 3, Measurement and Analysis, gives an excellent introduction to controlled experiments in the domain of software engineering. The irony here is the book is out of print. The authors integrate the statistical analysis with the discussion of the experimental design; both topics are treated in depth — making it an unique contribution.

*Charles M. Judd, Eliot R. Smith, and Louise H. Kidder, Research Methods in Social Relations, Holt, Rinehart and Winston, Inc., sixth edition, 1991. This is the best book that we have ever seen on the topic of social experimental design. They cover almost all aspects of the topic. The book is excellent for getting references to other work.

A Primer on Empirical Studies

D. E. Perry
Bell Laboratories
Murray Hill NJ 07974
dep@bell-labs.com

A. A. Porter*
University of Maryland
College Park, MD 20742
aporter@cs.umd.edu

L. G. Votta
Bell Laboratories
Naperville, IL 60566
votta@bell-labs.com

INTRODUCTION

We have created the tutorial for both practicing software engineers and software engineering researchers who want to understand the empirical bases of our field. In the former case, we expect that practitioners are seeking to improve the way they build, or manage the building of, software systems. In the latter case, we expect that researchers are seeking to improve there understanding of empirical work and their skill in undertaking empirical research.

While there are a variety of different ways the problems of empirical studies might be approached, we take a specific-focus, practical and concrete approach. By this we mean that instead of gathering a vast amount of general data and then trying to make some sense of it either by selection or statistical analysis, we advocate an approach that takes a narrower and deeper view in which specific questions are asked and data gathered and analyzed to answer those specific questions. Our approach is one which is more tuned to software engineering needs: we advocate a model in which there is more control involved resulting in more detailed information gathered. Rather than emphasize general phenomenology, we emphasize trying to find underlying mechanisms.

The primary goal, then, of the tutorial is to enable the attendees to assess the credibility of empirical work either as reported in the software engineering literature or as done by themselves and to apply the results to their own work. Within in this context, we discuss various models of empirical studies and the space of various kinds of empirical work. We note that good empirical science is the result of iterative experimentation – that is, we formulate hypotheses, test and evaluate them and then formulate new hypotheses or reformulate our original ones and continue the cycle again.

Having established the dimensions of empirical work, we then use this basis to establish criteria for evaluating both the experimental structures and the experimental results. We show how to use these criteria with published examples and show how to evaluate and apply these results – that is, how to make sense out of what is read, how to be able to separate the wheat from the chaff, and how to find the practical utility of the empirical work.

We then discuss various tricks of the trade and show how to exploit data that already exists locally and data reported in empirical studies. We discuss the position of statistics in our model and the importance of minimal manipulation of data. We then look to the future and present new techniques that are on the horizon suchs as simulation and sampling.

*This work is supported in part by a National Science Foundation Faculty Early Career Development Award, CCR-9501354.

STRUCTURE OF THE TUTORIAL

Introduction

We begin by characterizing various ways of doing empirical studies and indicate the scope of our approach and the layout of the tutorial.

Models of Empirical Studies

We first layout the space of empirical studies distinguishing between anecdotal studies, case studies, and experiments. Orthogonal to his dimension of the space is the context in which the studies are done: in vivo versus in vitro, independently and collaboratively.

We then present the anatomy of an experiment and consider the issues of construct validity, generalizability, and cost-benefit tradeoffs. In particular, we explicate our experimental logic.

An important aspect of empirical studies is the interplay between hypothesis and its test and the iterative and cyclic nature of empirical science. Particularly important are experiments designed to replicate or confirm as well as explicate or refute other experiments.

Evaluations of Empirical Studies

Having laid the foundations for the critical aspects of empirical studies, we now apply them to specific studies. For each study we discuss the goals, the empirical structure, and the results. We consider the problems of asking the right questions and using the right model to establish a particular hypothesis. We define the basis for evaluating empirical work and practice with well-chosen examples. We provide practical insights from our own experience and emphasize the utility of the structures and results as they apply practical problems.

Tricks of the Trade

Where the previous section emphasizes evaluation, this section emphasizes construction of empirical studies to gain specific information. These basic constructs show how to capitalize on existing information as well as constructing specific well-defined case studies and repeatable experiments. We further will show how to evolve your experimental structures as your understanding evolves and how to triangulate your empirical work to establish different but complimentary