Applying the Experimental Paradigm to Software Engineering

Natalia Juristo

Universidad Politécnica de Madrid

Spain

8th European Computer Science Summit

Current situation

- **16.3%** of software projects are successful

  The project is completed on time and within budget, and has all the features and functions specified at the start

- **52.7%** of software projects cost more, take longer or do less

  The project is completed and operational, but it cost more than budgeted (189% more), took longer than estimated and offers fewer features and functions than originally specified (42%)

- **31%** are cancelled

  The project is called off at some point during development before the system is put into operation

Do we know why?

Current situation

- Knowledge
  - Today the results of applying software development methods are unpredictable
  - There is no evidence to support most of the beliefs on which software systems development is based

- Practice
  - Method selection for and decision making on software production is based on suppositions and subjective opinions
    - When, by chance (or thanks to practitioners' personal and non-transferable know-how), the right methods are used, the software construction projects run smoothly and output the desired product
    - When the wrong methods are applied, the project develops haphazardly and the output product tends to be of poor quality

Content

1. The Scientific Method
   1. What is science?
   2. Scientific laws
   3. Predicting & understanding
2. Experimental Software Engineering
   1. Is the scientific method applicable to SE?
   2. Experiment & laboratory
   3. Designing experiments
   4. Challenges in applying the scientific method to SE

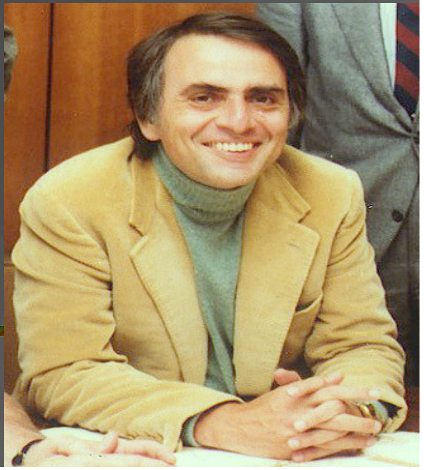
The Scientific Method

A process called science

- Science is a process of understanding the world

- *Science is a way of thinking much more than it is a body of knowledge*

Carl Sagan



Explaining the world

- Science looks for explanations about how a phenomenon works and why it works as we perceive it

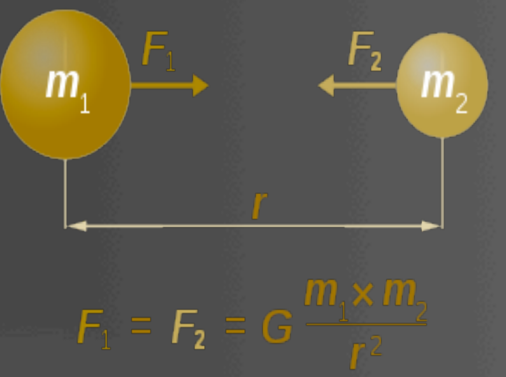
- These explanations are known as laws or theories
  - Nature generally acts regularly enough to be described by laws

Scientific laws

- Are patterns of behaviour
- Describe cause-effect relationships
- Explain
  - why some events are related
  - how the mechanism linking the events behaves

We only perceive nexus

- We cannot perceive laws directly through our senses
  - Anyone can see an apple fall, but Newton's inverse-square law of gravitation only becomes apparent through special systematic measurement



- Two activities are necessary
  - Systematic objective observation
  - Inference of links between cause & effect

The scientific method

- Collection of empirical **data**
  - Systematic observation to appreciate nexus

- Theoretical **interpretation** of data
  - Formation of hypotheses (right or wrong) about the mechanism relating the events
- Collection of empirical **data**
  - Hypotheses must be tested against reality to find out whether or not they are true

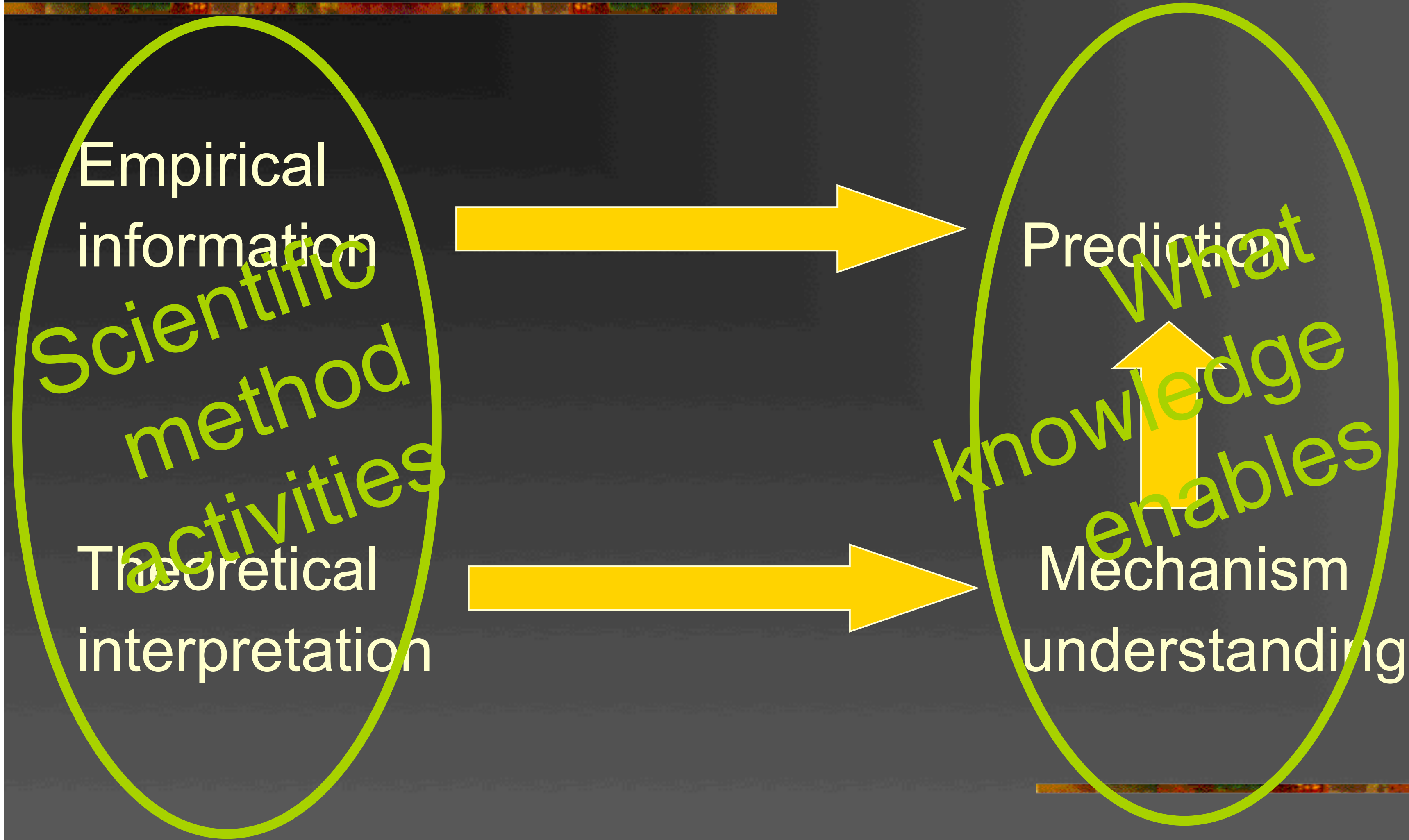
Building vs. understanding

- Humans are able to build interesting artefacts without scientific knowledge

- The builders do not necessarily understand the mechanisms governing the observed behaviour

Building without understanding

Predicting vs. understanding



Predicting without understanding

Data ⟷ theory

- Without an inferential leap to theory, the accumulation of data will never lead to an understanding of the mechanism

- In the absence of empirical data, theories move in the realm of speculation

Scientific research method

- A rigorous process for properly developing and evaluating explanations for observable phenomena based on reliable empirical evidence and neutral, unbiased independent verification

- Not based on arguments from authority or popular preferences

Useful for researching software development?

Experimental Software Engineering

What does all this have to do with software?

Experimental SE

- Generates scientific statements about building software through experimentation
  - This knowledge should help to identify the applicability conditions, strengths and weaknesses of the different software development technologies
- All engineering disciplines have taken a similar step
  - Achieve predictable results moving from beliefs, speculations and lucky guesses to scientific knowledge

Gaining SE knowledge with experiments

- Identify and understand
  - the variables that play a role in software development
  - the connections between variables
- Learn cause-effect relationships between the development process and the resulting products
- Establish laws and theories about software construction that explain development behaviour

Experimental Software Engineering

Experiment & Laboratory

Experiment

- Experiments
  - Model key characteristics of a reality in a controlled environment by manipulating them iteratively to investigate the impact of such variations and get a better understanding of a phenomenon
  - Are a formal, rigorous and controlled investigation in which the variables under study are given different values to find out what effect each value has
- The properties of a complex system are explained by analysing the behaviour of its parts

SE experiment

- Development decomposed into its parts
- Manipulated variables
  - Techniques (design, testing, etc.),
  - Developers (experience, knowledge, etc.)
  - Variables that can be assigned during development
- Investigated impacts
  - Effectiveness, efficiency, productivity, quality
  - Examples of instances
    - number of detected defects, number of code lines, etc.
  - Interesting characteristics obtained as a result of development

The laboratory

- Laboratory
  - Simplified and controllable reality where the phenomenon under study can be manipulated and studied
- Chemistry laboratory
  - Flasks and pipettes where temperatures and pressures are controlled
  - Real world: real substances with temperature and pressures
- Economics laboratory
  - Sets of individuals playing games to earn toy benefits
  - Real-world: markets (composed of thousands of agents) where real rewards are pursued
- What is a SE laboratory like??

SE laboratory

- Students
  - rather than professionals
- Toy software
  - rather than real systems
- Exercises
  - rather than real projects
- Academic workshops or industrial tutorials
  - rather than real knowledge & experience in industry
- Phases, techniques
  - rather than whole projects

Weakness

- How representative is any lab finding of reality?

- Different levels of experimental studies
  - In vitro experiments
  - In vivo experiments (from mice to monkeys)
  - Field experiments (from volunteers to clinical trials)

External validity

- Is concerned with the extent to which the results can be generalized
  - from the unique and idiosyncratic experimental settings, procedures and participants
  - to other populations and conditions
- Generalizability of experimental results to
  - the target population of the study
  - the universe of other populations

Experimental Software Engineering

Designing Experiments

Cause-effect relationships

- The independent variable
  - is the variable that is thought to be the cause
  - must meet two requirements
    - be changeable
    - the change must be controllable

- The dependent variable is
  - the effect brought about by this cause
    - is not manipulated
  - measured to see how it is affected by the manipulation of the independent variable

Extraneous variables

- If extraneous variables also vary systematically with the independent variables
  - then conclusions regarding causality are not valid
  - the observations are "confounded"
- Experiment design involves controlling the influence of extraneous variables on the dependent variables
- Good design avoids confounding variables

Control strategies

- Control neutralizes variation of extraneous variables

- Control strategies
  - Constancy
    - Keeping extraneous variables constant
  - Blocking
    - Neutralizing known extraneous variables
    - Purposely assigning every value of the blocked variable to every group
  - Randomization
    - Neutralizing unknown extraneous variables
    - Random assignment of subjects to experiment conditions

Experiment design

- The validity of the design of experiments is a fundamental part of the scientific method
- The design of a controlled experiment is a set of strategies aiming to control
- Without a valid design, valid conclusions cannot be drawn
- Statistics cannot fix a badly designed experiment

Internal validity

- Is concerned with whether we can accurately infer that
  - the independent variable caused the effect on the dependent variable
- Certainty with which we can establish the cause of the variations in results
  - Were there any extraneous variables that could have caused the observed effect?

Experimental Software Engineering

Current Status

Journals, conferences, books

- A specialized journal
  - Empirical Software Engineering Journal
- A specialized conference
  - Empirical SE and Measurement Conference
- A couple of books
  - *Experimentation in SE: An Introduction*
    Wohlin, Runeson, Höst, Ohlsson, Regnell, Wesslén
    Springer 2012
  - *Basics of SE Experimentation*
    Juristo & Moreno
    Springer 2001

ESE: Origins

- The first experiments were run in the early 1980s by Victor Basili's group at the University of Maryland with NASA's Software Engineering Laboratory
- The use of experiments to examine the applicability of SE technologies has gradually gained in importance as a research methodology
- Empirical studies have finally become recognized as an important component of the SE discipline
  - Fraction of empirical studies is rising in the last 3-4 years

ESE: Evolution

- 1993-2003: Leading SE journals
  - TSE, TOSEM, JSS, EMSE, IST, IEEE Software, IEEE Computer, SP&E
  - 78 experiments
- 1977-2006: ICSE
  - 3.2% had some type of empirical evaluation. Of such
    - 0.9% case studies; 0% experiments; 0.7% quasi-experiments
  - 0% the contribution was pure empirical
- 2012: ICSE
  - 71% had an empirical evaluation
  - 20% the contribution was pure empirical
    - experiments, case studies & qualitative

SE Experimental paradigm is still immature

- SE experiments are mostly exploratory
  - They produce objective observation but cannot yet be explained
  - The scientific method’s inference step is not being exercised
  - Finding (statistical) patterns is not enough
  - Mechanisms have to be found that explain the patterns

SE Experimental paradigm is still immature

- SE experiments have flaws
  - Lack of thoroughly thought-out designs to rule out extraneous variables in each experiment
  - Proper analysis techniques are not always used
- Lack of replications
- Lack of field experiments

It is not Just Import the Paradigm

- ESE lays the foundations for carrying out experimental research into SE
- It is not enough just to apply experimental design and statistical data analysis to take experimental research in SE forward
  - A discipline's specific experimental methodology cannot be imported directly from others

THE ROLE OF SCIENTIFIC METHOD IN SOFTWARE DEVELOPMENT

Natalia Juristo
Universidad Politécnica de Madrid
Spain