Developing Theories in Design Science

Prof. Dr. Roel Wieringa
University of Twente,
The Netherlands
www.cs.utwente.nl/~roelw
Outline

• Theories in the design cycle
• Theories
  – Explanation
  – Generalization
  – Scientific inference
• Empirical research to support theories
Engineering cycle

Implementation evaluation = Problem investigation

- Development of problem theories about stakeholders and their problems, or of design theories about artifacts and their real-world behavior, based on real-world observations

Treatment design
- Design some artifact

Treatment validation
- Development of design theories about artifacts and their real-world behavior, based on simulations

Treatment implementation
- Specify requirements!
  - Requirements contribute to goals?
  - Available treatments?
    - Design new ones!

- Context & Artifact → Effects?
  - Effects satisfy Requirements?
  - Trade-offs for different artifacts?
  - Sensitivity for different Contexts?

Implementation evaluation = Problem investigation

- Stakeholders' Goals?
- Conceptual problem framework?
- Phenomena? Causes, mechanisms, reasons?

This is a checklist. See appendix A in the book & on my web site

Vitoria, 29 Sept 2016

© R.J. Wieringa
• Prototypes will be lost, design theories should stay
Empirical knowledge questions

• **Descriptive** knowledge questions:
  – What happened?
  – How much? How often?
  – When? Where?
  – What components were involved?
  – Who was involved?
  – Etc. etc.

• **Explanatory** knowledge questions:
  – Why?
    1. What has *caused* the phenomena?
    2. Which *mechanisms* produced the phenomena?
    3. For what *reasons* did people do this?

**Journalistic questions.**
Yield facts. Can be generalized into a descriptive theory.

**Beyond the facts.**
Can be generalized into an explanatory theory.
Example questions

• Design software to estimate Direction of Arrival of plane waves, to be used in satellite TV receivers in cars
  • Is the DoA estimation accurate enough in this context?
    – Description: Measurements
    – Explanation: Signal theory, algorithm properties
  • Is it fast enough?
    – Description: Measurements
    – Explanation: Signal theory, algorithm properties

• Design a Multi-Agent Route Planning system to be used for aircraft taxi route planning
  • Is this routing algorithm deadlock-free on airports?
    – Explanation: mathematical proof
  • How much delay does it produce?
    – Description: Measurements
    – Explanation: MARP properties, taxiing mechanisms on airports
Example questions

• Design a data location regulation auditing method

• Is the method usable and useful for consultants?
  – Description: Measurements, opinions
  – Explanation: Theory about interaction between method and cognitive or social mechanisms of auditors during auditing

Answering descriptive or explanatory questions about artifact in context

Descriptive or explanatory theory about interaction between artifact and context
Outline

• Theories in the design cycle
• **Theories**
  – Explanation
  – Generalization
  – Scientific inference
• Empirical research to support theories
What is a theory?

• A **theory** is a belief that there is a pattern in phenomena.
  – Idealizations: “Merging two faculties reduces cost.” “This works in theory, but not in practice.”
  – Speculations: “The NSA is monitoring all my email.”
  – Opinions: “The Dutch lost the soccer competition because they are not a team.”
  – Wishful thinking: “My technique works better than the others.”
  – **Scientific theories:** Theory of electromagnetism
Scientific theories

• A **scientific** theory is a belief that there is a pattern in phenomena, that has survived
  – Tests against experience:
    • Observation, measurement
    • Possibly: experiment, simulation, trials
  – Criticism by critical peers:
    • Anonymous peer review
    • Publication
    • Replication

• Examples
  – *Theory of electromagnetism*
  – *Technology acceptance model*
  – *Theory of the UML*

• Non-examples
  • *Religious beliefs*
  • *Political ideology*
  • *Marketing messages*
  • *Most social network discussions*
Scientific design theories

• A **scientific design theory** is a belief that there is a pattern in the interaction between an artifact and its context

• Examples:
  - *Theory of the UML in software engineering projects*
  - *Theory about accuracy and speed of DOA algorithms in a context of plane waves and white noise*
  - *Theory about delays in routes planned by MARP on airports*
  - *Theory about usability of data location auditing method for auditors*
The structure of scientific theories

1. Conceptual framework
   - Definitions of concepts. Ontology: concepts and their logical relations.

2. Generalizations
   - Express beliefs about patterns in phenomena.
   - In theories of natural phenomena: Laws of nature
   - In design theories: generalizations about interactions between artifact and context.
Theory of electromagnetism

• Conceptual framework (concepts defined to describe and explain the relevant phenomena):
  – Definitions of electric current, electric charge, potential difference, electric resistance, electric power, capacitance, electric field, magnetic field, magnetic flux density, inductance, ..., ... and their units.

• Generalizations
  – Electric charges attract or repel one another with a force inversely proportional to the square of their distance.
  – Magnetic poles attract or repel one another in a similar way and always come in North-South pairs.
  – An electric current inside a wire creates a corresponding circular magnetic field outside the wire.
  – A current is induced in a loop of wire when it is moved towards or away from a magnetic field
Theory of cognitive dissonance

- **Conceptual framework**: beliefs, dissonance, resolution
- **Generalization**: People seek consistency among their cognitions. They resolve this by creating comfortable beliefs.
- **Scope**: all human beings
Theory of the UML

• Concepts: UML concepts, definitions of software project, of software error, project effort, definition of concept of domain, understandability

• Descriptive generalization: (UML) \( X \ (SE \ project) \rightarrow \) (Less errors, less effort than similar non-UML projects)

• Explanatory generalizations:
  - UML models resemble the domain more than other kinds of models;
  - they are easier to understand for software engineers;
  - So they they make less errors and there is less rework (implying less effort).
Facts versus theories

**Facts**

- Observed sample of cases
  - Cases are prototypes, people, projects, etc.
  - Facts are what is measured.

**Descriptive theory of the population**

**Explanatory theory of the case/sample**

- Why?

**Generalize**

**Unobserved population**

**Explanatory theory of the population**

- Why?
Outline

• Theories in the design cycle
• Theories
  – Explanation
  – Generalization
  – Scientific inference
• Empirical research to support theories
Three kinds of explanation

Facts

- Cases are prototypes, people, projects, etc.
- Facts are what is measured.

Explanatory theory of the case/sample

- Explain by
  - Causes
  - Mechanisms
  - Reasons
  - Why?

Descriptive theory of the population

- Explain by
  - Causes
  - Mechanisms
  - Reasons
  - Why?

Unobserved population

Observed sample of cases

Generalize
Example: light

- **Descriptive question:** Is the light on?
  - Based on observation: Yes.
  - When? Now.
  - Where? Here.

- **Explanatory question:** Why is it on?
  1. **Cause:** because someone turned the light switch, it is on (and not off). Explains difference with off-state.
  2. Why does this cause the light to switch on? **Mechanism:** because the switch and light bulbs are connected by wires to an electricity source, in this architecture ..., and these components have these capabilities ..... Explains how on-state is produced.
  3. By why did someone turn the light on? **Reasons:** Because we wanted sufficient light to be able to read, and it was too dark to read. Explains which stakeholder goal is contributed to.
Example: coffee

• **Causal explanation:** effect attributed to a cause. Explain difference in outcomes by difference in interventions. Causation is difference-making.
  – *The coffee made me stay awake late.*

• **Architectural explanation:** Outcome produced by interaction among components. Explain capability of system in terms of capabilities of components
  – *Mechanism of action: Caffeine has a psychostimulant effect because it antagonizes adenosine, which normally inhibits neurotransmitters such as dopamine.*

• **Rational explanation:** Outcome contributes to a goal. Explain outcome in terms of rational takeholder choices.
  – *I worked late because I wanted to finish the paper before the deadline.*
Example: software

- **Descriptive question:** What is the performance of this program?
  - Execution time for different classes of inputs?
  - Memory usage?
  - Accuracy?
  - Etc. etc.

- **Explanatory question:** Why does this program have this performance (compared to others)?
  1. **Cause:** Variation in execution time is caused by variation in input; etc.
  2. **Mechanism:** Execution time varies this way because it has this architecture with these components
  3. **Reasons:** Observed execution time varies this way because users want to be on-line all the time, and therefore provide these inputs
Another example: method

• **Descriptive question:** What is the performance of this method for developing software?
  – Understandability for practitioners
  – Learnability
  – Quality of the result
  – Perceived utility
  – Etc. etc.

• **Explanatory question:** Why does this method have this performance?
  1. **Cause:** Difference in understanding of methods by software engineers is attributed to differences in the methods, and not to differences in people, software systems, etc. (cf. Testing of a medicine)
  2. **Mechanism:** These differences are explained by the structure of the method and/or the structure of cognition. (cf. Mechanism of action of a medicine)
  3. **Reasons:** Developers are rewarded if they use the method well
Fallibility and validity of theories

• All scientific theory are fallible
  – May turn out to be false
  – I.e. they are improvable

• Validity is degree of support
  – Never total

• Every step in a scientific argument must be assessed by the author on validity:
  – Design of a research setup
  – Descriptions of observations
  – Explanations
  – Generalizations

Beliefs in marketing, politics and religion are usually considered fallible by their defenders
Internal validity of an explanation

- **Internal validity** = degree of support for an explanation
- Three kinds of internal validity
  - Of causal explanations
  - Of architectural explanations
  - Of rational explanations
- Customarily stated in terms of threats that decrease support.
Threats to internal validity of causal explanations

• **Ambiguous relationship**: ambiguous covariation, ambiguous temporal ordering, ambiguous spatial connection?

• **Object of Study dynamics**: could there be interaction among OoSs? Could there be historical events, maturation, dropout of OoSs?

• **Sampling influence**: could the selection mechanism influence the OoSs? Could there be a regression effect?

• **Treatment control**: what other factors than the treatment could influence the OoSs? The treatment allocation mechanism, the experimental setup, the experimenters and their expectations, the novelty of the treatment, compensation by the researcher, resentment about the allocation?

• **Treatment instrument validity**: do the treatment instruments have the effect on the OoS that you claim they have?

• **Measurement influence**: will measurement influence the OoSs?
Threats to internal validity of architectural explanations

- **Analysis**: the analysis of the architecture may not support its conclusions with mathematical certainty.
  - Are components fully specified?
  - Are interactions fully specified?

- **Variation**: do the real-world case components match the architectural components?
  - Do they have the same capabilities?
  - Are all architectural components present in the real-world case?

- **Abstraction**: does the architectural model used for explanation omit relevant elements of real-world cases?
  - Are the mechanisms in the architectural model interfered with by other mechanisms, absent from the model but present in the real world case?
Threats to the internal validity of rational explanations

- **Goals:** The actor may not have the goals that the explanation says it has
- **Motivation:** a goal may not motivate an actor as much as the explanation says it did
Outline

• Theories in the design cycle
• Theories
  – Explanation
  – Generalization
  – Scientific inference
• Empirical research to support theories
Two kinds of generalization

Facts

Observed sample

- What happens in these cases?
- What average, variance in this sample?

Explanatory theory of the case/sample

Explain by
- Causes
- Mechanisms
- Reasons
- Why?

Descriptive theory of the population

Unobserved population

- What happens in all cases?
- What average, variance in this population?

Explanatory theory of the population

Explain by
- Causes
- Mechanisms
- Reasons
- Why?
Case-based generalization

• Precedes sample-based generalization in history.
• Examples
  – Newton’s prism experiment
  – Pascal’s experiment with air pressure
  – Lavoisier’s experiments with phosphorus
  – Oersted’s experiment with magnetic needle
• “If you build the same research setup, it will exhibit the same phenomena.”
  – Repeatability
  – Architecture
  – Similarity
Generalization by analogy: example

• **Observation:**
  – **Artifact:** This prototype implementation of the MUSIC algorithm,
  – **Context:** when used to recognize direction of arrival of plane waves received by an antenna array, in the presence of only white noise, running on a Montium 2 processor,
  – **Effect:** has execution speed less than 7.2 ms and accuracy of at least 1 degree.

• **Explanations:**
  – algorithm theory and signal theory

• **Generalization by analogy:**
  – All similar implementations
  – Running in similar contexts
  – Will show similar performance
Generalization by analogy: example

• **Observations:**
  – **Artifact:** This version of the UML
  – **Context:** Used in this software project
  – **Effect:** Produces software with less errors and less effort than in similar projects without the UML,

• **Explanation:**
  – **UML models are easier to understand for software engineers because they resemble the domain more than other kinds of models, and so the software engineers make less errors and there is less rework.**

• **Generalization**
  – **In similar projects,**
  – **UML**
  – **will have similar effects**
Generalization by analogy: general pattern

• Generalization:
  – All artifacts with similar architecture
  – Used in contexts with similar architecture
  – Will show similar effects

• Must be supported by architectural explanation.

• Nonexample:
  – Wallnuts look like brains.
  – Brains can think.
  – Therefore .... Wallnuts can think

• This is only superficial similarity

• There is no mechanism that produces thinking in brains and wallnuts!
All real-world cases are different!

• All artifacts with similar architecture
  – ... a real-world artifact may not have a sufficiently close match
  – *E.g. they use the different version of UML correctly in a different way*

• Used in contexts with similar architecture
  – ... a real-world context may contain interfering mechanisms
  – *No deadline pressure, no personnel turnover, no coordination breakdown, ...*

• Will show similar effects
Two kinds of generalization, so two kinds of validity

- **External validity** = Degree of support for generalization by analogy
- **Conclusion validity** = Degree of support for a sample-based generalization
Threats to external validity

- **Object of study**
  - **Similarity:** Does the OoS satisfy the population predicate?
  - **Ambiguity:** Does the OoS satisfy other population predicates too?

- **Representative sampling**
  - **Sample-based research:** Study population represents theoretical population?
  - **Case-based research:** Selected cases be representative of the population?

- **Treatment**
  - **Treatment similarity:** Experimental treatment similar to real treatments?
  - **Compliance:** Is the treatment performed as specified?
  - **Treatment control:** Other factors that could influence the OoSs?

- **Measurement**
  - **Construct validity:** are the definitions of constructs to be measured valid?
  - **Measurement instrument validity**?
  - **Construct levels:** Representative measured range of values?
Analytic induction

• Improve external validity of an analogic generalization by analytic induction:

• Cases are studied one by one, theory updated in between
  1. Start with an initial theory about how mechanisms produce phenomena
  2. Select a confirming or falsifying case
  3. Do case study
  4. Update the theory (conceptual framework and/or generalization)
  5. Stop when budget is finished or theory appears stable

• As always, supports theory only fallibly!
Outline

• Theories in the design cycle
• Theories
  – Explanation
  – Generalization
    • Case-based generalization
    • Sample-based generalization
  – Scientific inference
• Empirical research to support theories
Two kinds of generalization

Facts
- Observed sample
  - What happens in these cases?
  - What average, variance in this sample?

Explanatory theory of the case/sample
- Explain by
  - Causes
  - Mechanisms
  - Reasons
  - Why?

Descriptive theory of the population
- Unobserved population
  - What happens in all cases?
  - What average, variance in this population?

Explanatory theory of the population
- Explain by
  - Causes
  - Mechanisms
  - Reasons
  - Why?
Descriptive statistics

• Summarize information in a sample
  – Sample mean, median, mode
  – Sample variance, standard deviation, max, min
  – Sample correlation
How can we generalize from descriptive statistics?

• “Similarly selected samples will have similar statistics.”
  – Treats sample as a case. Need to define relevant similarity.

• “If the sample is very large and randomly selected, then probably it has very similar statistics as the population.”
  – Law of large numbers. But: big data is usually not random.

• “In repeated random sampling from the same population, the sample averages will converge on the population average.”
  – Frequentist statistical inference: next slides

• “Use a sample of \((X, Y)\) values to estimate \(Y\) as a function of \(X\) in the population.”
  – Statistical learning, regression

• “Use a sample to update our assumed population distribution.”
  – Bayesian inference.
Inferential statistics

• Infer properties of a population from observations of a sample
• Sample-based generalization arose late in the 19th century
  – Fisher: test a null hypothesis
  – Neyman-Pearson: decide between alternative hypotheses
Statistics is John Lennon science

• No theory
• No causation
• No mechanisms
• No reasons
• No religion too

• Just averages, correlations and variance
  – Supported by repeatable empirical evidence
  – Sufficient for prediction in a nondeterministic world

• Statistical conclusions do not explain anything
  – “X explains 40% of the variance of Y” means “40% of the observed variation of X is related to the observed variation of Y.”
Frequentist statistical inference

• “Given N observations of variable X, infer some property of the probability distribution of X.”
  – Unobservable distribution of X
  – Finite number of observations of X
  – Sampling procedure

• Takes different forms:
  – Test a hypothesis about the population distribution of X, e.g. about its mean, by computing the probability of the data given the hypothesis.
  – Estimate a confidence interval for a parameter of the population distribution X
Sampling logic

- **Theoretical population.** E.g. all software projects.
- **Study population,** listed by a *sampling frame.*
  - E.g. all software projects listed in the project archives of three companies.
  - Or all projects currently running in these companies.
- **Sample** selected from the sampling frame
  - Random sample: random selection with replacement
  - Simple random sample: random selection without replacement
  - Convenience sample: opportunistic selection
Methodology of statistical inference

- Statistical model of distribution of $X$ in the $X$-box

Sample selection

Statistical inference

$X$-Box

- Starts with unobservable distribution of numbers,
- Observes a sample of numbers,
- Concludes something about the unobservable distribution of numbers
Central-Limit Theorem

• Let $X_1, \ldots, X_n$ be a sample from a distribution with mean $\mu$ and standard deviation $\sigma$. Then the sample mean $\bar{X}_n$ is approximately normally distributed with mean $\mu$ and standard deviation $\sigma^2 / n$. The approximation gets better as $n$ gets larger.

• Equivalently: $Z_n \equiv \frac{\sqrt{n} \left( \bar{X}_n - \mu \right)}{\sigma}$ is approximately distributed as $N(0, 1)$.

• If $\sigma$ is unknown: $T_n \equiv \frac{\sqrt{n} \left( \bar{X}_n - \mu \right)}{S_n}$ is approximately distributed as $N(0, 1)$. For small samples, it is more accurate to use the t-distribution itself.
• Population distribution does not need to be normal!
• Must have finite $\sigma$ and $\mu$.

• Repeated random sampling will give an approximately normal distribution of sampling means, centering on the population mean, with standard error of $\sigma / \sqrt{n}$.
• (Fluctuation due to random sampling of non-identical population elements)

Illustration of CLT

Population distribution does not need to be normal!

Samples of size 60

Narrower distribution of sampling means

Margin of error (MOE): 2 SEs on each side of the population mean.

**FIGURE 3.4**
The upper panel displays the population distribution, with lines marking SD units, showing $\sigma = 20$. Below is the mean heap. The superimposed curve is the sampling distribution of the mean, with lines marking SE units. In this example, $N = 15$, and 200 samples have been taken. The SE = $\sigma / \sqrt{N} = 20 / \sqrt{15} = 5.16$. 
95% confidence intervals

In the long run, 95% of these confidence intervals around the sample mean, will contain the population mean

Bars extend MOE either side of $M$

For this $M$, bars don’t capture $\mu$

For this $M$, bars don’t capture $\mu$
Confidence intervals using $\frac{s}{\sqrt{n}}$ to estimate $\frac{\sigma}{\sqrt{n}}$

For this $M$, bars don't capture $\mu$
Confidence—interval estimation

In the long run, approximately 95% of these confidence intervals around the sample mean, will contain the population mean.

FIGURE 3.9
All that the researcher knows: the data points of a single sample with $N = 15$, as shown in Figure 3.6, but now the 95% CI has been calculated, using $s$. 
Example: observational study


- Sample of 20 open source web applications selected randomly from the study population of OS web applications in the OSVDB that
  - had more than 1 release,
  - are larger than 3 KLOC,
  - have exploitable vulnerabilities and
  - of which source code is available

- Count the number of implementation vulnerabilities (security vulnerability caused by coding errors rather than by design flaws or configuration errors).
  - (= Chance model and measurement procedure)

- Observation: The average percentage of vulnerabilities caused by coding errors per OS web application in the sample is 73%.
  - (Descriptive statistic)
Example continued

• **Statistical inference:**
  – Assuming a random sample, and
  – assuming that the proportion $p$ of coding errors is constant and independent across web applications,
  – Vulnerabilities caused by coding errors are binomially distributed over the population, with variance $\frac{p(1-p)}{n}$
  – the average percentage of vulnerabilities caused by coding errors in any OS web application in the study population is roughly 73% ± 4% with roughly 95% confidence.
Confidence interval estimation is more informative than hypothesis testing, and can (and should) replace it.
Hypothesis test: if your hypothesis is outside the confidence interval, your data does not support the hypothesis.

**FIGURE 3.9**
All that the researcher knows: the data points of a single sample with $N = 15$, as shown in Figure 3.6, but now the 95% CI has been calculated, using $s$.

Repeat the research to deal with random fluctuations!
Null-hypothesis significance test (NHST): if your null hypothesis is outside a 95% confidence interval, reject the hypothesis at 5% level.

FIGURE 3.9
All that the researcher knows: the data points of a single sample with \( N = 15 \), as shown in Figure 3.6, but now the 95% CI has been calculated, using \( s \).
Statistical difference-making experiments: independent samples

• Random sample from a study population
• Randomly allocate treatments T1 and T2.
  – Gives random samples of the population treated by T1 and of the population treated by T2
• Apply treatments, measure sample outcomes
• Estimate difference in average population outcomes
  – Or test a hypothesis about the population difference

The two samples may coincidentally be very different, exaggerating the outcome difference. Replicate!
Another setup: repeated measures

• Random sample from a study population
• Apply T1, measure sample outcome
• Apply T2, measure sample outcome
• Estimate difference in average population outcomes
  – Or test a hypothesis about the population difference

Now the two samples are similar, but T1 may influence the outcome of T2. Mitigate by cross-over designs. Replicate anyway.
Statistical conclusion validity

- **Statistical conclusion validity** = Degree of support for a statistical inference
Threats to statistical conclusion validity

Assumptions of confidence interval estimation

• **Stable distribution.** Does X have a stable distribution, with fixed parameters?

• **Scale.** Does X have an interval or ratio scale?

• **Sampling.** Is sample selection random?

• **Sample size.** If the $z$ distribution is used, is the sample sufficiently large for the normal approximation to be used?

• **Normality.** If the $t$-distribution is used, is the distribution of X normal, or is the sample size larger than 100?

Assumption of statistical difference-making experiments

• **Treatment allocation.** Are treatments allocated randomly to sample elements?
Outline

• Theories in the design cycle
• Theories
  – Explanation
  – Generalization
  – Scientific inference
• Empirical research to support theories
• Analogic inference to similar cases must be based on architectural explanations (in terms of mechanisms or reasons)
Example (repeated & elaborated)

• **Observations:**
  – **Artifact:** This prototype implementation of the MUSIC algorithm,
  – **Context:** when used to recognize direction of arrival of plane waves received by an antenna array, in the presence of only white noise, running on a Montium 2 processor,
  – **Effect:** has execution speed less than 7.2 ms and accuracy of at least 1 degree.

• **Abduction (inference to the best explanation):**
  – algorithm theory and signal theory

• **Generalization by analogy:**
  – All *architecturally similar* implementations
  – Running in *architecturally similar* contexts
  – Will show *similar* performance produced by same mechanisms
Example (repeated)

• **Observations:**
  – Artifact: *this version of the UML*
  – Context: *Used in this software project*
  – Effect: *Produces software with less errors and less effort than in similar projects without the UML,*

• **Explanation: Abduction (inference to the best explanation):**
  – *UML models are easier to understand for software engineers because they resemble the domain more than other kinds of models, and so the software engineers make less errors and there is less rework.*

• **Generalization**
  – *In architecturally similar projects,*
  – *Using UML in a similar way*
  – *will produce similar effects by the same mechanisms*
Sample-based inference

1. Descriptive inference
   - Data from samples
   - Observations, sample statistics

2. Statistical inference
   - Generalizations over a population

3. Abductive inference
   - Explanations in terms of mechanisms, causes, reasons

4. Analogic inference
   - Architectural explanation facilitates further analogic generalization to similar populations

- Statistical inference yields descriptive generalization over a study population.
- Differences in outcome may be explainable by causes.
- Architectural explanation facilitates further analogic generalization to similar populations.
Example: observational study


• Sample of 20 open source web applications selected randomly from the OS web applications in the OSVDB that satisfies a number of conditions.

• Count the number of implementation vulnerabilities

• Observation: The average percentage of implementation vulnerabilities per OS web application in the sample is 73%.

• Statistical inference:
  – the average percentage of vulnerabilities caused by coding errors in any OS web application in the study population is roughly 73% ± 4% with roughly 95% confidence.
Example extended

• **Abduction (inference to the best explanation):**
  – Coding errors that cause implementation vulnerabilities are caused by cognitive imitations and project coordination mechanisms.
  – (But which ones?)

• **Analogic generalization from study population to theoretical population:**
  – The cognitive mechanisms that produce these coding errors (whatever they are) are common across all OS web application programmers.
  – If there are no interfering mechanisms, then it is plausible that 73% of all vulnerabilities in OS web applications are implementation vulnerabilities.

• **NB the paper does not draw these conclusions.**
Validity of inferences

a) **Descriptive validity**: no information added in the descriptions

b) **Internal validity**: degree of support for explanations

c) **External validity**: degree of support for analogic generalizations

d) **Statistical conclusion validity**: degree of support for statistical inference
Descriptive validity

• **Data preparation**
  – Will the prepared data represent the same phenomena as the unprepared data?
  – If data may be removed, would this be defensible beyond reasonable doubt?
  – Would your scientific opponents produce the same descriptions from the data?

• **Data interpretation**
  – Will the interpretations that you produce be facts in your conceptual research framework?
  – Would your scientific peers produce the same interpretations?
  – Will the interpretations that you produce be facts in the conceptual framework of the subjects? Would subjects accept them as facts?

• **Descriptive statistics**
  – Is the chance model of the variables of interest defined in terms of the population elements?
Outline

• Theories in the design cycle
• Theories
  – Explanation
  – Generalization
  – Scientific inference
• Empirical research to support theories
### Different designs support different inferences

<table>
<thead>
<tr>
<th></th>
<th>Observational study (no treatment)</th>
<th>Experimental study (treatment)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Case-based:</strong></td>
<td>Observational case study</td>
<td>• <strong>Expert opinion</strong> (mental simulation by experts),</td>
</tr>
<tr>
<td></td>
<td><strong>Evidence for or against architectural explanations in similar cases</strong></td>
<td>• <strong>Mechanism experiments</strong> (simulations, prototyping),</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• <strong>Technical action research</strong> (experimental use of the artifact in the real world)</td>
</tr>
<tr>
<td><strong>Sample-based:</strong></td>
<td><strong>Survey:</strong> <em>Evidence for or against estimations of properties of population distributions</em></td>
<td>• <strong>Statistical difference-making experiment</strong> (treatment group – control group experiments):</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Evidence for or against causality</em></td>
</tr>
<tr>
<td></td>
<td>Investigate samples drawn from a population, look at averages and variation</td>
<td></td>
</tr>
</tbody>
</table>
Take-home

• Design theories are about the effects of an artifact in a context
• Theory consists of conceptual framework and generalizations
• Explanations can be causal, architectural, rational
• Generalization can be case-based (analogic) or sample-based (statistical)
• Theories are fallible and must be assessed on validity

Wieringa, R.J. (2014) *Design science methodology for information systems and software engineering*. Springer Verlag


Additional slides
Example: experimental study

- Four groups of 9 to 26 students made UML domain model from Use case model for two systems, with or without using System Sequence Diagrams (SSDs) and System operations contracts (SOCs). Four-group crossover design.
  - Theoretical population: all software engineers
  - Smaller theoretical population: all software engineering students
  - Study population: all participants in an SE class
  - Sample: Self-selected sample of volunteers
  - Groups within this sample: students randomly allocated to UML or to UML+SSD+SOC
Example continued

• Observation:
  – **In the observed samples**, when SSDs and SOCs were used, average correctness of models was higher, and effort to produce them was lower.

• Generalization by statistical inference:
  – Pairwise t-test, simple repeated measures ANOVA and mixed repeated measures ANOVA support the generalization that average correctness of models and effort to produce them is better when SSDs and SOCs are used in the population of all software engineering students. This conclusion is plausible but not always correct.

• Explanation:
  – By listing all possible causes, and assessing them on their plausibility, the use of SSDs and SOCs is the most plausible cause of these effects (and not the competence of the students or the positive expectation of the experiments, or …)

• Generalization by analogy to similar populations, e.g. the population of all SE students or of professional software engineers.
  – Need to discuss if the social or cognitive mechanisms that produce the results in the student population, are the same as those in the theoretical population of all SE students or of all professional software engineers.
An aside

- They did this ..... but unfortunately found hardly any support for a statistically significant difference.
Example extended

• **Causal explanation:**
  – By listing all possible causes, and assessing them on their plausibility, the use of SSDs and SOCs is the most plausible cause of these effects (and not the competence of the students or the positive expectation of the experiments, or ...)

• **Architectural explanation:**
  – The results are produced by social or cognitive mechanisms in the student population.
  – (But which ones?)

• **Generalization by analogy to similar populations, e.g. the population of professional software engineers.**
  – The mechanisms that produce the phenomena in the student population are the same as those in the population of professional software engineers.
  – If there no interfering mechanisms, then this phenomenon will probably also be produced in the population of professional software engineers.