

Innledning

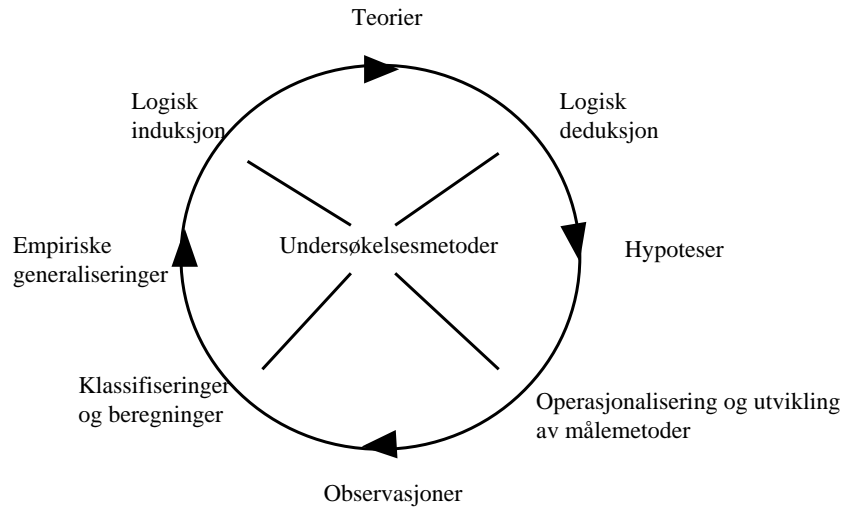
- Planlegging og kontroll av prosessforbedringer krever at **ført**ilstand (evt **nå**tilstand) og forandring studeres.
- Undersøkelsemetodikk er et negligjert tema i de fleste prosessforbedringsrammeverkene. F eks, i ISO 9000-3 står det at man skal måle kvalitet, ikke hvordan kvalitet skal måles eller hvordan man skal utforme undersøkelsen som målingene skal inngå i.
- Undersøkelse av effekter av tiltak er et svært komplisert område.

Spørsmål: Under hvilke rammevilkår må stort sett undersøkelser i forbindelse med systemutvikling gjøre sine studier av nåsituasjon og effekter av tiltak?

INF5180 2



Undersøkellesmetoder (Wallace's modell)



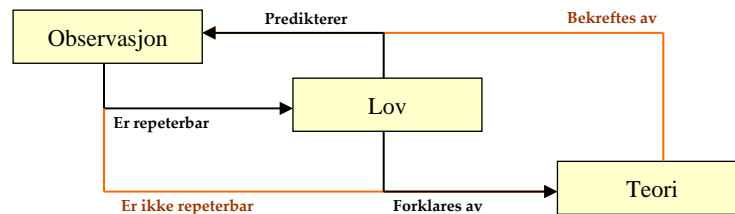
INF5180

3



Teorier, lover og observasjoner

- Teorier kan komme "hvor som helst fra" og være basert på logisk deduksjon eller flyktige observasjoner
- En lov er basert på gjentatte observasjoner under noenlunde tilsvarende forhold
- Observasjoner kan være foretatt velstrukturerte eksperimenter eller være basert på case-studier og brukes til å bekrefte teorier.



Lover forklarer *hvordan* ting skjer, men ikke *hvorfor*. Det gjør derimot **teorien**.

Ref Endres, Rombach: A Handbook of Software and Systems Engineering

INF5180

4



Metoder og observasjoner (I)

Grader av objektivitet:

- **Introspeksjon:** Eksaminering av ens egne tanker, følelser og inntrykk.
- **Observasjon:**
 - Observasjon fra utsiden/innsiden (deltagende => "action research")
 - Skjulte/åpne observasjoner
 - Strukturert/eksplorativ observasjon
 - Ofte mangel på repeterbarhet
 - Ofte case-studium ("science of the particular")
- **Eksperiment:** "Vitenskapelig metode", holde alle variable foruten de som undersøkes konstant (ceteris paribus).
 - Repeterbarhet
 - Generaliserbarhet

INF5180

5



Metoder og observasjoner (II)

Grader av generalitet:

- **Case-studium**
 - Typisk tilfelle: Oppførselen til én systemutviklingsgruppe studeres i dybden.
 - Viktig: Kunnskap om studert case og å beskrive detaljert kontekst.
 - Styrke: Realisme. Dybde.
 - Svakheter: Kontroll over variable. Årsaksanalyser.
- **Representativt utvalg ("multi-case")**
 - Typisk tilfelle: Tilfeldig utvalg fra en gruppe (populasjon), f eks systemutviklingsgrupper som utvikler administrativ programvare vha Cobol.
 - Viktig: Gode statistikk-kunnskaper.
 - Styrke: Muliggjør en balanse mellom dybde og kontroll
 - Svakheter: Antagelser om utvalges representativitet og fordelinger
- **Studium av universelle sammenhenger (naturvitenskap)**
 - Typisk tilfelle: Eksperiment mhp hvordan det menneskelige øye reagerer på visse stimuli, der det kan antas at det resultatene er generelle (naturlover).
 - Viktig: Gode kunnskaper om eksperimentell metode.
 - Styrke: Kontroll
 - Svakheter: Realisme

INF5180

6



Introspeksjon (selv-evaluering)

For:

- *"Psychoanalysis is learned first of all on oneself, through the study of one's own personality ... There is whole series of very common and well-known mental phenomena which can be taken as material for self-analysis when one has acquired some knowledge of the method"* (Freud, A general introduction to psychoanalysis)
- enkel start på et studium og ofte eneste mulige metode for å få tak i viktig informasjon

Mot:

- *".. although subjects showed a clear preference for indentation (innrykk i programvarekoden), no statistically significant differences were found on performance measures"* (Schneiderman, Measuring computer program quality and comprehension)
- Vanskelig for andre å interpretere. Hva mener person A med "bra"? Lite generaliserbart.



Case-studium

- I dialogen "Menon" (Platon) spør Sokrates Menon om hva "dyd" (virtue) er. Menon svarer ved å gi eksempler (cases) og Sokrates reagerer med:
 - *".. forstår du ikke at jeg er ute etter det som er felles egenskaper ("simile in multis") for dyd ... oppfyll løftet ditt, og gi meg en generell beskrivelse av dyd"*
- Synspunktet at case-studier ("eksempler") ikke gir generell og kontekst-fri kunnskap og dermed ikke har noen vesentlig vitenskapelig verdi, har siden oldtiden vært vitenskapens holdning. Humanistiske fag, psykologi og andre "myke" disipliner har mye av den grunn ikke fått full aksept som vitenskaper. (*I middelalderen var f eks eksperimenter og observasjoner sett på som svært mye laverestående aktivitet enn den "rene tenkningen", bl a pga mangel på generalitet i resultatene.*)
- I nyere tid har case-studium imidlertid fått stadig større aksept som bidragsyter til dybdekunnskap som ofte kan generaliseres. Dette kanskje som en følge av at de i utgangspunktet mer vitenskapelige "korrelasjonsstudiene" ofte blir "tall-magi" uten forståelse av det bakenforliggende. (se f eks Flyvebjerg, "Rasjonalitet og Magt")



Representativt utvalg

For:

- Representerer en større populasjon enn enkelt «case»
- Mulig å bruke statistiske metoder til å teste hypoteser etc.

Mot:

- Ofte mange spekulative antakelser for å generalisere/teste hypoteser
- Mye misbruk av statistikk
- Bredder går ofte på bekostning av dybde
- Ofte vanskelig å si noe om årsak-virkning



Ren observasjon

For:

- generell metode
- kan utføres i realistiske (ikke-eksperimentelle) omgivelser

Mot:

- vanskelig å kontrollere effekten av omgivelsene (ref. antropologenes vanskeligheter med å ikke påvirke det samfunnet de skal observere)
- vanskeligere å generalisere enn eksperimenter



Hypotetisk-deduktiv metode

HDM:

- Påstander i de teorier, som fremsettes og etterprøves ved hjelp av metoden (HDM), betraktes ikke som absolutt sikre, men som hypoteser.
- Påstandene etterprøves og begrunnes ved at man utleder (deduserer) konsekvenser av dem og ser, hvordan disse henger sammen innbyrdes og med vår øvrige meninger og hvordan de passer med erfaringen.

HDM kan åpenbart brukes på det aller meste og ikke kun egnet for naturvitenskaplig aktivitet.

I praksis er det likevel store forskjeller på gjennomføringen når metoden brukes på:

- gjenstander som i stor grad har "bevissthet" og påvirkes av teorier og observasjoner (f eks systemutviklingsorganisasjoner)
- materiale uten eller som i liten grad har "bevissthet" (f eks studium av værsystemer eller akvariefisk)

Fra "Argumentasjonsteori, språk og vitenskapsfilosofi" Føllesdal, Wallø og Elster

INF5180

11



Øvelse

- Angi problemer med å bruke undersøkelsesmetoder basert på "naturvitenskapelige prinsipper", dvs utviklet for studium av gjenstander uten bevissthet, for studium av systemutviklingsorganisasjoner.

INF5180

12



Ulike "effekter" å være klar over

- **Hawthorne-effekten**
 - Oppmerksomheten som ble rettet mot arbeiderne i "Hawthorne Plant of the Western Electric Company" syntes å ha større effekt enn de endringer i miljø som ble studert.
- **Heisenberg-effekten** (svært misbrukt begrep fra kvante-fysikken)
 - Umulig å måle uten å påvirke målingene. (Hvordan måle hvordan noe er når du ikke måler det, dvs normalsituasjonen?)
- **Weinberg-effekten**
 - Undersøkelse viste at systemutviklere som visste at de ble målt, reduserte ytelsen på de fleste parametre som **ikke** ble målt.
- **Spørsmålets utforming-effekten**
 - "Synes du forelesningen i dag var A: bra, B: meget bra, C: strålende, D: enestående"
- **Teoriladet observasjon-effekten**
 - Det man (tror man) observerer, er ikke upåvirket av hva man tror man kommer til å observere. (ref. mannen som skjøt etter en traktor under elgjakta)



Eksperiment

For:

- kontrollerbare omgivelser gir større muligheter til å generalisere
- relativt objektiv metode

Mot:

- som oftest urealistiske omgivelser (laboratorium), og dermed ofte objektive svar på spørsmål som ikke er stilt av noen
- objektiviteten på eksperimentnivå utfylles ofte med en enorm grad av subjektivitet når resultatene skal interpreteres (millimeter-nøyaktighet noen steder etterfulgt av synsing på km-nivå) til å gjelde i realistiske omgivelser.
- ofte ikke gjennomførbart med eksperimenter (av praktiske eller etiske årsaker)



Valg av metode

- Vanskelig å gi generelle retningslinjer mhp metodevalg.
- Metodevalg avhenger bl a av:
 - forhåndskunnskaper om studieemnet
 - muligheter til å kontrollere omgivelsene
 - størrelsen på studiet
 - behovet for generalitet i resultatene
 - tilgjengelighet til informasjon/data og andre ressurser
 - hva som er målet med undersøkelsen (prediksjon, årsaks-forståelse, intern bruk i en organisasjon,)
- **Øvelse:** Gi eksempler på hvordan disse punktene påvirker valg av metode. F eks dersom formålet er å måle effektiviteten til systemutviklerne.



Øvelse: Hvor viktig er tidlig bruk av PC?

(Dette er en historie fra virkeligheten, blant mange lignende historier.)

Firmaet "Intelligente barn AS" hevdet i en avisannonse at *"barn som tidlig begynner å bruke PC-er vil få bedre skolerestulater enn barn som begynner sent å bruke PC"*.

Dette hevdet "Intelligente barn AS" var konklusjonen fra en studie utført ved en del norske skoler. Studien hadde vist at *"barn som hadde brukt PC hjemme minst en gang i uka som 6-åringer, fikk bedre evalueringer fra lærerne i fag som norsk og matematikk i de første skoleårene."*

Anta at studien er korrekt utført. Er det da samsvar mellom det "Intelligente barn AS" hevder og konklusjonen fra studien?



Argumentasjon - Definisjoner

(Mye av stoffet om argumentasjon er hentet fra boka "Advocacy and opposition" av Rybacki og Rybacki.)

- *"Argumentation is a form of instrumental communication relying on reasoning and proof to influence belief or behavior through the use of spoken or written messages."*
- *"Persuasion is an attempt to move an audience to accept or identify with a particular point of view"*
- **Øvelse 1:** Hva er forskjellen mellom argumentasjon og overtalelse?
- **Øvelse 2:** Hva er sammenhengen mellom vitenskapelige undersøkelsesmetoder og argumentasjon?



Studier av Grupper

- Vanskelig å unngå Hawthorne-effekten
 - Kan innebære at man skjuler sin hensikt, enten ved å observere i smug, eller ved å fremstå med en annen hensikt enn å observere en bestemt effekt.
- Ulike klassifiseringsskjemaer er utviklet
 - SYMLOG, IPA klassifiserer individers oppførsel og responser i h.t. Ulike modeller.
- Egenrapportering baserer seg på strukturerte intervjuer eller utfylling av spørreskjemaer.
- Eksperimenter eller case-studier?
 - Eksperimenter gir forskerne frihet til å isolere en bestemt effekt og å holde andre ting konstant. Ofte vanskelig å unngå at mennesker responderer annerledes enn de ville ha gjort i et naturlig miljø.
 - Case-studiene har den fordelen at man observerer mennesker som gjør det de ville ha gjort uansett, i sitt naturlige miljø. Selvsagt begrenset hva man kan få anledning til å observere uten å forstyrre naturlig aktivitet.



Argumentasjon - etiske normer

- **Etiske** standarder for argumentasjon:
 - Kunnskap. Fortrinnsvis fra forskjellige synsvinkler.
 - Gode hensikter. Særlig, ingen "skjulte agendaer".
 - Rasjonalitet.
 - "Argumentasjonsfrihet". Respekter andres rett til å ha en annen argumentasjon. Ingen personangrep.



Argumentasjon - Elementer

Elementer i argumentasjon (Toulmin's modell):

PRIMÆR-elementer:

- Påstand (I noen tilfelle: Konklusjon)
- Momenter (Informasjon som styrker eller svekker påstanden)
- Logisk sammenheng mellom moment og påstand (Angivelse av hvorfor momentene styrker eller svekker påstanden.)

SEKUNDÆR-elementer:

- Støtteargumentasjon ("backing") for den logiske sammenhengen mellom moment og påstand.
- Angivelse av påstandens styrke.
- Angivelse av påstandens begrensinger.



Argumentasjon - Typer

Typer argumentasjoner:

- Argumentasjon for påstander om årsakssammenhenger
- Argumentasjon for påstander om samvariasjon
- Argumentasjon basert på generaliseringer
- Argumentasjon basert på helt eller nesten tilsvarende tilfeller
- Argumentasjon basert på autoriteter
- Argumentasjon basert på dilemma



Argumentasjon - Hva bør unngås

Hva bør **unngås** i en argumentasjon:

- Uholdbare generaliseringer
- Uholdbare overføringer av gyldighetsområde
- "At dytte folk en mening på, hvis vanvid alle kan forstå" - for så å argumentere mot denne
- Irrelevante argumenter
- Sirkelargumentasjon
- "Gå rundt grøten"
- "Snakke om noe annet"
- Personangrep
- Bruke definisjoner og begreper til å påvirke på en ikke-rasjonell måte
- Anta at noe er sant fordi det ikke er bevist at det er usant
- Appellering til fordommer
- Usaklig appell til følelser
- Manipulerende bruk av av autoriteter, tradisjon, humor, tvetydigheter, ladede ord, fremmedord,



Argumentasjon - Råd

Hva fører til en god argumentasjon?

I forberedelsesfasen:

- Innhenting av relevant informasjon fra ulike ståsteder
- Kritisk forhold til informasjonens kvalitet.
- Standpunkt (konklusjon) tas **etter** at informasjon er innhentet og analysert
- Ikke-relevante faktorer har ingen/liten påvirkning
- God innsikt i egen vurderingsevne og fordommer

Selve argumentasjonen:

- Klargjøring av rammene for argumentasjonen (begreper, ståsted, motivasjon, eget kunnskapsnivå, informasjonens kvalitet, målsetning, ...)
- Balansert fremstilling av momenter for **og i mot**
- Fokuseret og relevant argumentasjon
- God sammenheng mellom momenter og konklusjon
- Konklusjon (påstand/standpunkt) som er utledet fra momentene og logiske slutningsregler. **Motsetning:** Standpunkt som forsøkes støttes av momenter.



The Standish Group's 1994 CHAOS Report*



"The Standish Group research shows a staggering 31.1% of projects will be canceled before they ever get completed. Further results indicate 52.7% of projects will cost 189% of their original estimates."

These results are still – in spite of their age - the most frequently quoted cost estimation figures. They are still used as:

- Documentation of how bad software projects are
- Input to governmental reports
- Excuse for large overruns (we are better than average!)
- Support for the belief that "we have improved a lot since 1994".
- A means to get research funding

**How Large Are Software Cost Overruns? A Review of the 1994 CHAOS Report*
Av Magne Jørgensen og Kjetil Moløkken Østvold, Simula research Laboratory

INF5180

24



Problem 1

- What does 189% average cost overrun mean?
 - "... 52.7% of projects (the so-called challenged projects) will overrun their initial cost estimates by 189%"
 - "The average cost overruns for combined challenged and cancelled projects is 189%."
 - "... 52.7% of projects will cost 189% of their original estimates"
- All formulations are used by the Standish Group



A web-search on the use of the numbers revealed a confusion ...

189% or 89% overrun?

- 50% of the documents described the result as "*189% cost overrun*",
- 40% as "*189% of original estimate*", and
- 10% as "*89% cost overrun*".

All projects, challenged projects, or challenged+cancelled projects?

- 70% of the documents related the result to "*53% of the projects*" (without explicitly pointing out that this 53% referred to challenged projects only),
- 16% to "*all projects*",
- 8% to "*challenged and cancelled projects*",
- only 6% explicitly pointed out that the average cost overrun is based on "*challenged projects*" only.



Problem 2: Lack of correspondance with other surveys

Study	Jenkins [9]	Phan [10]	Bergeron [11]
Year	1984	1988	1992
Respondents	23 software organizations	191 software projects	89 software projects
Country of Respondents	USA	USA	Canada
Average Cost Overrun	34%	33%	33%

INF5180

27



Problem 3: Lack of knowledge about research method

- No description of accuracy measures (!!!), selection process, or analysis method.
- We have sent several mails to them and received responses like:
 - *"Why would you be surprised that we would not want to give out detailed information about how we conduct our studies? This is how we make a living. There is absolutely no incentive for us to want anyone to replicate our study - that's like giving away our business for free."*
- After 3 mails they stopped responding.
 - Not even willing to respond on whether the 1994-study results should be interpreted as "89%" or "189%" cost overrun.
- They sell their CHAOS reports, but not to us ...
 - We had to order the report through SINTEF (our thanks to Tore Dybå)

INF5180

28



Potential reasons for the high cost accuracy numbers

- Incorrect interpretation of own results.
 - Seems as if the initial results were interpreted as 189% of initial estimate, i.e., 89% cost overrun. Later, however, this is by the Standish Group described as 189% cost overrun, i.e. 189% cost overrun.
- No category for cost underrun in data collection schema
 - Underruns described as 0% overrun?
 - Project challenged on time or functionality may still have effort underruns.
- Non-random sampling
 - *"We then called and mailed a number of confidential surveys to a random sample of top IT executives, asking them to share **failure stories** [!!!]. During September and October of that year, we collected the majority of the 365 surveys we needed to publish the CHAOS research."*



Evidence Based Software Engineering – EBSE*

- Utvalget av "best practices" innen systemutvikling øker hele tiden
- Kvaliteten på det som publiseres om disse er mildt sagt varierende
 - "Alle" undersøkelsesmetoder er representert
 - Selve undersøkelsen er sjelden dokumentert
 - Det viser seg at en del analyseinstitutter nekter å gi fra seg data om hvordan undersøkelsene er gjennomført (Standish Groups "Chaos Report")
- Tore Dybå ved Simula/SINTEF har sett på det som gjøres innen medisin og beskrevet en metode for å forsøke å finne best mulig evidence for valg av en teknologi eller metode - EBSE

****Evidence-based Software Engineering for Practitioners, published in IEEE Software 2004.***

Tore Dybå, Barbara A. Kitchenham, and Magne Jørgensen



Øvelsesoppgave

Skisser en undersøkelse for å finne ut om JAVA eller C++ bør være programmeringsspråket for begynneropplæringen i programmeringen. Gjør nødvendige antakelser ang. målsetning m.m.. **Spesielt viktig:** Begrunn valg av undersøkelsesmetode og målestørrelser.