

## The Impact of Irrelevant and Misleading Information on Software Development Effort Estimation:

A Randomized, Controlled Field Experiment of Forty-Six Outsourcing Companies

Magne Jørgensen,  
[magnej@simula.no](mailto:magnej@simula.no)  
København, 22 May, 2008

### “Clouds Make Nerds Look Better”



- Sunshine increases tipping, impacts stock-market, and, increases happiness.
- Study of university applicants:
  - 12% higher chance when sunshine compared to worst cloudcover.
  - Nerds had significantly higher chance compared to non-nerds on cloudy days.
    - Nerd-factor measured as academic rating divided by social rating (e.g., leadership).

## Irrelevant information is everywhere ...

- Requirement specifications and other information provided in an estimation situation typically include
  - some misleading information (on purpose or accidentally)
  - much estimation irrelevant information
  - much information of low importance for the estimation work
- There are good (and not so good) reasons for this, e.g.,
  - information may be relevant for other purposes than effort estimation,
  - "copy-paste" of general information about the clients' processes and organization from previous specifications,
  - lack of competence in how to write a good requirement specification
- Are we more rational than stock investors and university applicant assessors, or do we get impacted by irrelevant information?

## The impact of the # of pages ...

- Context: Software engineering students, experimental context.
- Task: Estimated the effort of the same programming task.
  - Group A: Received the original specification, which was one page long.
  - Group B: Received a version of the specification that had exactly the same text, but was seven pages long. The increased length was achieved through double line space, wide margins, larger font size and more space between paragraphs.

## The impact of the # of pages ...

- Group A (1 page spec.) and Group B (7 pages spec.) estimates were, on average, 117 and 173 work-hours, respectively.
  - Longer specification → higher estimates.

## The impact of irrelevant context information

- Context: Software professionals participating on an estimation seminar.
- Group A received the original programming task specification.
- Group B received the same specification, with added work-effort irrelevant information.
  - Information about the end users desktop applications, other systems' web design, work effort irrelevant information about user passwords.

## The impact of irrelevant context information

- Results:
  - Group A (no irrelevant information) average: 20 work-hours
  - Group B (a lot of work-effort irrelevant information added) average: 39 work-hours
- More information → Higher estimates.

## Anchoring experiment ...

- Context: Software professionals participating in a estimation seminar.
- Three groups: High anchor, Control group, Low anchor
- HIGH group:
  - *“The customer has indicated that he believes that **1000** work-hours is a reasonable effort estimate for the specified system.”*
- LOW group:
  - *“The customer has indicated that he believes that **50** work-hours is a reasonable effort estimate for the specified system.”*
- HIGH and LOW group
  - *“However, the customer knows very little about the implications of his specification on the development effort and you shall not let the customer’s expectations impact your estimate.”*
  - *“Your task is to provide a realistic effort estimate of a system that meets the requirements specification and has a sufficient quality.”*
- CONTROL group: No information about the customer expectation.

## Anchoring experiment ...

- *Results:*
  - *HIGH anchor (1000) group average: 555 work-hours*
  - *CONTROL group (no anchor) average: 456 work-hours*
  - *LOW anchor (50) group average: 99 work-hours*
- *None felt they had been much impacted, and most of the software professionals claimed that they had not been impacted at all.*

## How Much Can we Trust Such Experiments in Artificial Settings?

- Short time for estimation work
  - More use of surface indicators?
- Only to some extent following their ordinary estimation processes
- Sufficient expertise not always there
- Individual estimates (no group work)

## A randomized, controlled trial in field settings

- Forty-six companies from various countries estimated the same five projects: Russia (15 companies), Ukraine (5), India (7), Bulgaria (4), Romania (3), Pakistan (5), Belarus (2), Moldova (1), Poland (1), Serbia (1), Slovakia (1), and Vietnam (1). T
- We accepted only estimators with professional experience from projects similar to those to be estimated, i.e., we allowed only reasonably experienced estimators.
- The companies were hired and paid for their estimation work, i.e., they did not (seen from their point of view) participate in an experiment.
  - The companies were on average paid about 1500 USD for the estimation work, ranging from 400 to 4000 USD.
  - The effort a company estimated to spend on the estimation of the five projects varied from about 40 work-hours to about 200 work-hours.
  - They were told that they would not be invited to develop the systems, but that their job was to provide realistic effort estimates.
- Random allocation to “manipulations” of requirement specification, similar to those previously done in artificial settings.

## Large variance in estimates!

Effort Estimation Distributions

Project	Minimum	Q1	Median	Q3	Maximum
RDinner	45	119	190	339	1320
DocAssist	61	186	330	438	1200
AA	160	316	509	715	2280
DES	17	134	192	347	1160
IMWOS	240	649	895	1316	3371

## Length of specification ...

- **H1:** A reduction in number of pages of the requirement specification leads to lower effort estimates, even when the written content is exactly the same.
  - Manipulation: Text identical. One version 3 pages, the other 12 pages.
  - Length of specification is clearly not relevant for the development effort, but will it be used as an indicator?

## Results: Length of specification (H1) [System: DocAssist]

### The Effect of the Reduced Length of Specification

Group	Median
Manipulated (3 pages spec.)	295 work-hours (n=24)
Ordinary (12 pages spec.)	330 work-hours (n=22)

A small effect – perhaps not even that ...  
Effect seems to be reduced (perhaps removed) with more time and expertise.

## Client expectation ...

- **H2:** Presenting the actual effort of the system to be replaced (a low numerical value in our case) early in the requirement specification leads to lower effort estimates.
  - The following text was included early in the manipulated requirement specifications: *“The preliminary budget of the new system is \$10 000 [corresponding to about 100 work-hours with typical pricing in the country in which it will be built]. The preliminary budget is not built on any knowledge about the actual cost of developing the new system, and will, if needed, be extended to cover the expenses necessary to build a quality system with the desired functionality.”*
  - 100 work-hours is a very low value for this project and the companies were instructed to not use this as input to their effort estimate, but they may use it unconsciously.

## Results: Client expectation (H2) [System: IMWOS]

### Numerical Anchor

Group	Median estimate
Manipulated (client's expectation)	724 work-hours (n=23)
Ordinary	956 work-hours (n=23)

A significant, large effect.  
However, lower effect than in our previous laboratory experiments.



## Time schedule pressure ...

- **H3:** Information about that the client requires a short development period leads to lower effort estimates.
  - The following text was included early in the manipulated requirement specifications: “[the client] *expects that the system development starts February 3, 2008 and can be launched February 23, 2008. This three week period should include all development and testing.*”
  - A short development period should lead to, if anything, more rather than less use of effort, but may also induce “wishful thinking”.

## Results: Time schedule pressure (H3) [System: DES]

The effect of time schedule pressure

Group	Median
Manipulated (Informed that the client expected the system to be developed during 3 weeks period.)	142 work-hours (n=24)
Ordinary	214 work-hours (n=21)

Very large, significant effect!  
Opposite effect of what I would considered as  
“normative” estimation behavior.

**The effects seem consequently to be important in industry settings? Why does this happen?**

- Hot topic among researchers. We do not know very much. It is difficult to study.
- The main reason is that brain activity is mainly unconscious, i.e., we are not in control of most of our thought processes and attention.

**Example: The Cocktail Party Effect**



## HELP! My brain is out of control ...



- The lack of brain control implies that it is hard to defend positions like:
  - “I know why I like what I like”
  - “My estimate is based on information X”
  - “I will not be impacted in my judgment by a dinner with one potential providers”
- This is, however, what most people seem to do.
- The reason for our unwillingness to accept the lack of control may be a strong desire to believe that we are rational individuals.
  - Ironically, the rational reaction to our lack of control is to admit irrationality.

[ **simula** . research laboratory ]

21

## We cannot be that irrational, or we would have been extinct ...



- Research studies tend to focus on judgmental biases, not when our judgments are good, i.e., the picture derived from the research is strongly biased towards demonstration of poor performance.
- The effect of irrelevant information is a consequence of high performance tailored (evolved) to other, much more important, situations (survival and reproduction) combined with the relatively slow speed of mental activities and neural speed:
- If the working memory (the conscious part of our brain) should do all processing work, we would not be able to walk and talk at the same time - probably not even walk or talk.

[ **simula** . research laboratory ]

22

# What should we do to improve our judgments?

## What we definitely should avoid ...

- Exposure to obviously irrelevant information, e.g., customer expectations that will have the role as anchors in effort estimation situations.
- A belief that the impact from irrelevant information only happens to other than yourself.
  - This will effectively prevent actions to take place.
- Information that “dilutes” the impact from the most essential information.
  - Much evidence to support the claim that more information of lesser quality or relevance typically leads to too little emphasis on the most relevant information.

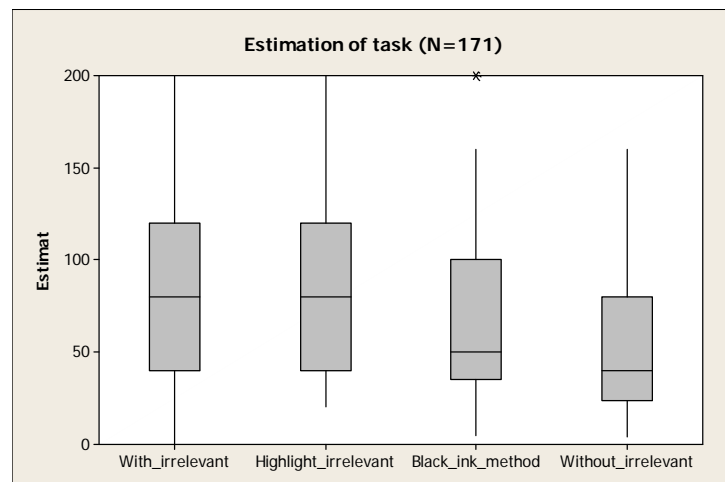
## Debiasing techniques ...

- Awareness of own biases does not help directly, but indirectly in that other less vulnerable judgment processes are chosen.
- Analytic, as opposed to intuition-based, estimation processes helps.
  - But, as long as they are not mechanical, there is room for impact from irrelevant and misleading information.
- The “black-ink method” (see next slide) may help.
- Debiasing techniques are typically the second best option.
- **The only really effective method is to remove the irrelevant and misleading information.**

[ **simula** . research laboratory ]

25

## The “Black Ink”-method (experiment with Java developers)



[ **simula** . research laboratory ]

26

**A possible approach  
(published in May/June 2008 issue of IEEE  
Software)**

**Step 1:** Let another person than those estimating the effort develop a “package” of the requirement specification and other estimation relevant information where misleading, irrelevant and non-essential information have been removed.

**What we should do ...**

**Step 2:** Estimate the most likely effort based on the filtered estimation package

**NB 1:** Ensure that everybody involved clearly understands that the purpose of the estimation work is to derive the most likely use of effort, and not something else.

**NB 2:** If a person knows anything about the desired outcome of the estimation process or other biasing information, this person should be excluded from the estimation work.

## What we should do ...

**Step 3:** Read through the less relevant, but not essential information to examine the need for adjustment.

**NB 1:** No adjustment should be allowed unless **very good** argumentation.

**NB 2:** If the estimate is “too high” (e.g., for the budget of the client), the solution is to remove functionality or simplify solution, **not** to adjust the estimate. The process of removing and simplifying should follow the same principles as outlined for the first estimate.

## Final comment on research method

Different types of studies have their role in software engineering research:

- Exploratory field observations (e.g., case studies, action research) of software organizations are useful to generate hypotheses and provide knowledge about highly inter-connected relationships.
- Experiments with students and software professionals in artificial settings are useful as pilot studies and “proof” of **existence** of effects or relationships.
- Experiments in field settings are not much in use, but can be the best method to provide knowledge about the **size** of effects. This method is, I believe, under-utilized.
- *More on this in: Hanney, J. and M. Jørgensen. "The Role of Deliberate Artificial Design Elements in Software Engineering Experiments." IEEE Transactions on Software Engineering, 2008.*