Prof. Dr. Harald Störrle
Danmarks Tekniske Universitet (DTU)

# The Good, the Bad, and the Ugly:
## UML Diagram Layout Matters

# Background



- **Experience suggests that the appearance (and in particular, the layout) of SE diagrams (e.g., UML, BPMN, EPC, …) greatly depends on its appearance.**
  - "Prettiness" is not just superficial add-on, it carries m…
- **Several om…**
- **…inconclusive or weakly significant results**
  **…very small subsets of UML with questionable validity.**

**Attention!**
**DEF: A DIAGRAM is a visualization of a projection of a MODEL.**
Large numbers of model metric publications have "Class Diagram" or similar in the title, but are really addressing "Class Models".

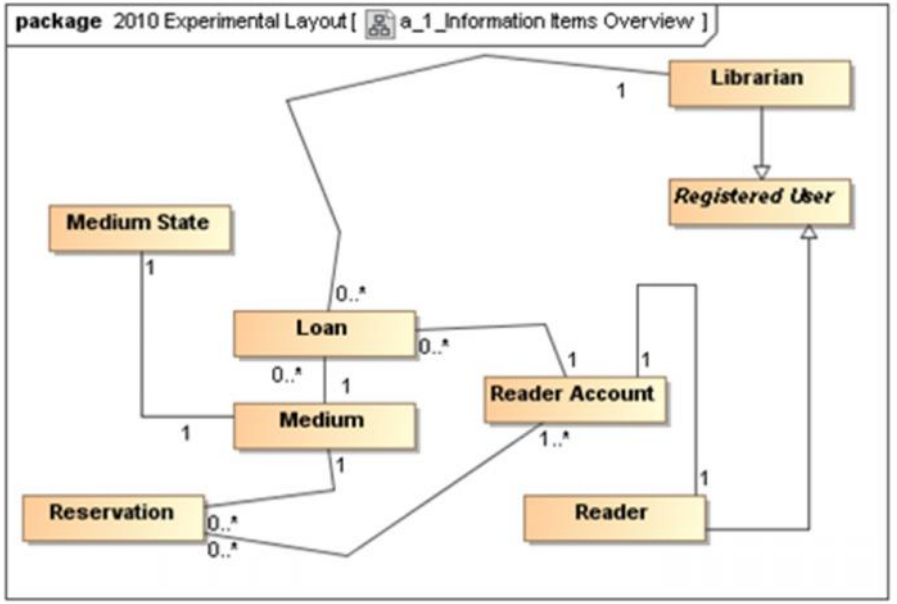
*"We could not identify a statistically significant relation between diagram quality and [understandability]."*

[Eichelberger & Schmid, J. Information & Software Technology 51 (2009) p. 1696]

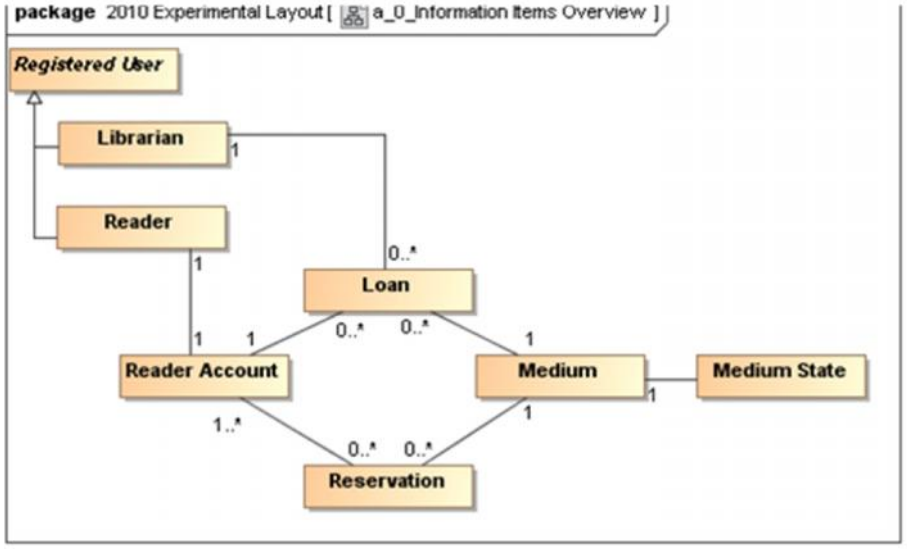
# Good vs. Bad (UML Diagram) Layout



- **Elements of bad layout**
  - Edge crossings and bends
  - Overlaping/obscuring elements
  - Varying colors/sizes
  - Varying text orientation

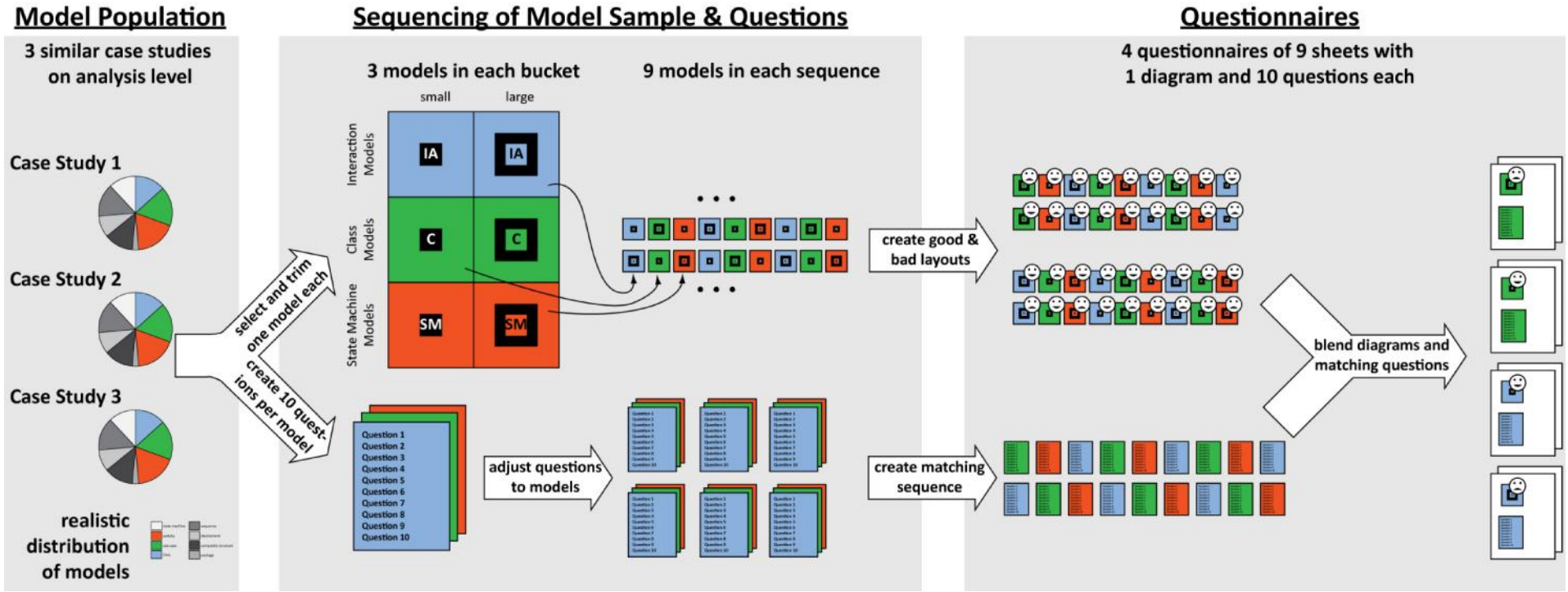



- **Elements of good layout**
  - Join similar edges
  - Cluster similar elements
  - Orthogonal arrangement
  - Place elements in flow



# Study Design: Models & Diagrams

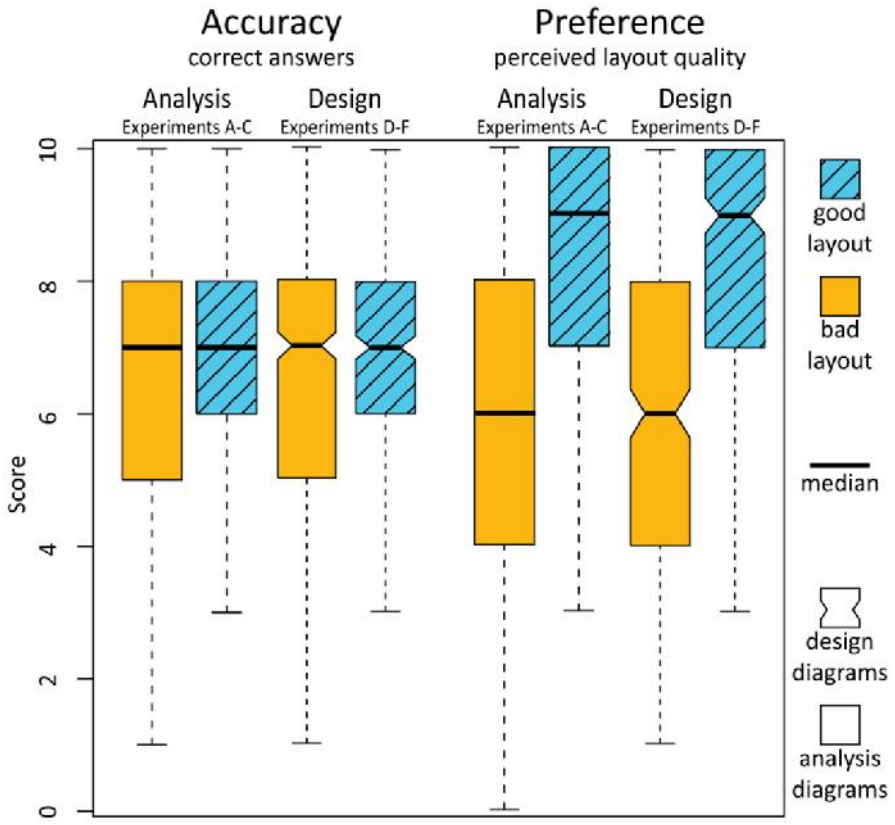





- **Within-Subjects design reduces impact of individual variance.**
- **Models from different case studies reduce semantic inferences.**
- **Systematic variation of independent variables cancels out learning.**

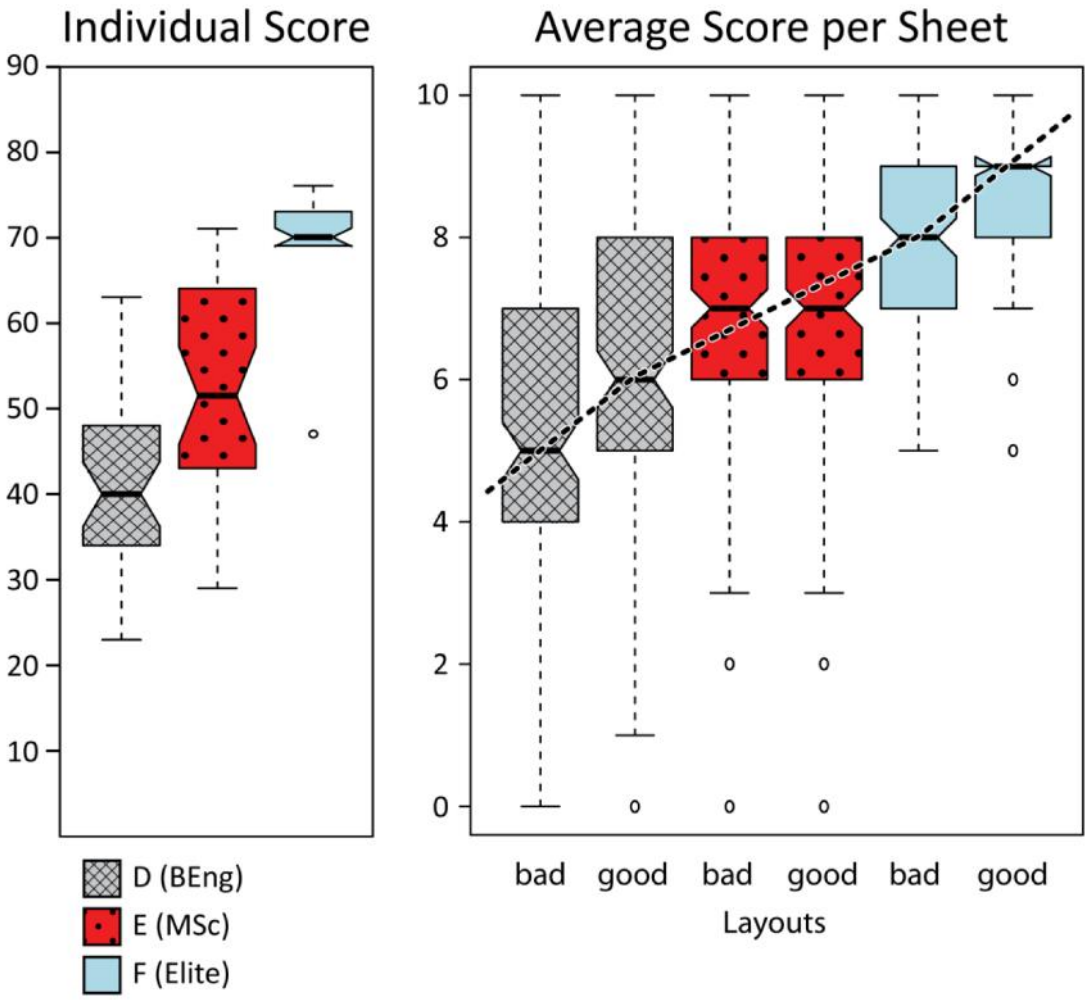
# Diagram Type ~ Impact

- **In contrast to previous studies, we observe a comparatively large effect.**
  - ...though not necessarily in user performance...
- **Cognitive load seems to benefit much more from good layout than objective performance indicators.**
  - This might be due to subjective coping strategies.
  - Dual stimulus experiments might shed light on this hypothesis.



# Expertise Level ~ Impact

- In the first experiment, we found no noticeable differences between novice modelers and advanced modelers, to our surprise.
- Maybe, there was no (large enough) expertise difference in our population?
- In a second experiment, we targeted different audiences (and different diagram types).



# Diagram Size ~ Impact (1/3)



- **Probably the simplest conceivable definition of diagram size is to simply count the number of elements in a diagram.**
  - While easy to implement, it obviously treats simple and complex elements alike, although they intuitively should contribute differently to diagram size.
  - Consider simple lines vs. multi-segment lines, rectangle vs. stick person.

- **A more refined definition of diagram size should introduce a weight factor for different elements.**
  - Consider line segments as separate elements, distinguish between simple/ medium/complex shapes and labels, with weights 1, 1.5, and 2, respectively.
  - However, different notations have different vocabulary sizes. So, a single grapheme in one notation (with a large vocabulary) can represent a greater amount of information than in another notation (with a smaller vocabulary).

- **Our third attempt adds the logarithm of the vocabulary size of the notation as another weight factor.**

# Diagram Size ~ Impact (2/3)



- **The three metrics were progressively more difficult to implement.**
  - We computed all three metrics manually for our sample of 38 diagrams.
  - We used the same diagrams as in previous studies.

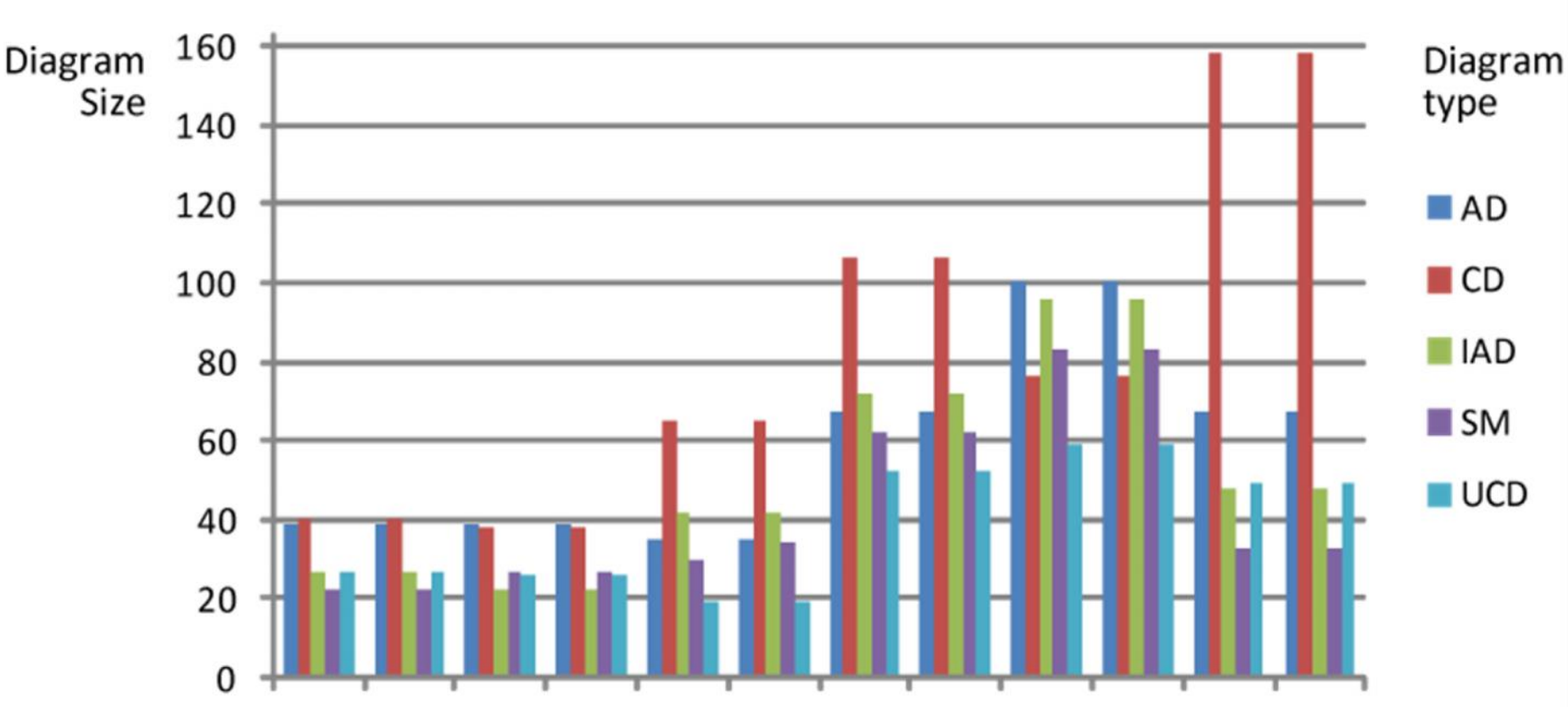
- **We expected them to also be progressively more accurate in correlating diagram size and modeler performance.**

- **We compared the outcome of the three metrics with each other and found that they correlate extremely well.**
  - 0.967…0.992, Pearson‘s product-moment correlation, $p<10^{-15}$

- **By Occam‘s law, we eliminated the second and third metric proposal and define diagram size as number of diagram elements.**

# Samples

- **We used the same samples reported in our previous study on layout quality impact.**
  - 78 participants (mostly students at different education levels).
  - 60 diagrams (30 models with a good/bad layout) of the 5 most commonly used UML diagram types.

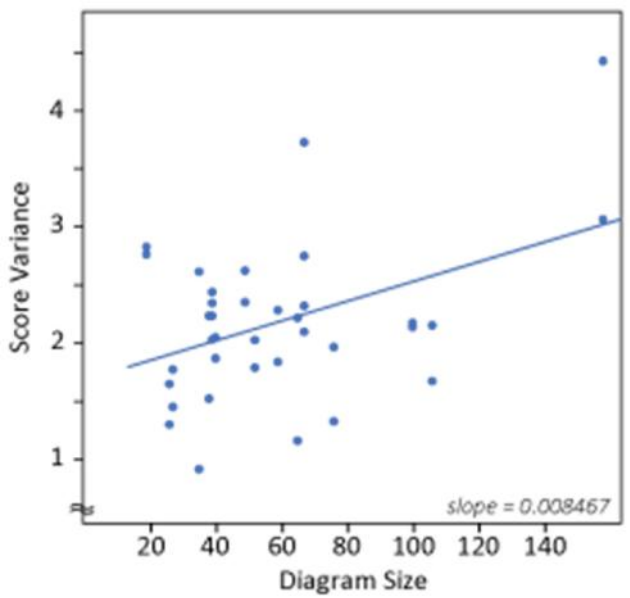


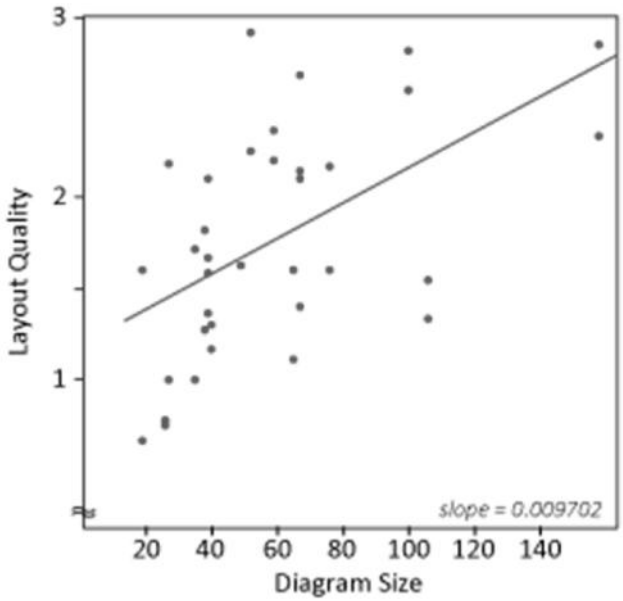
# Diagram Size ~ Impact (3/3)

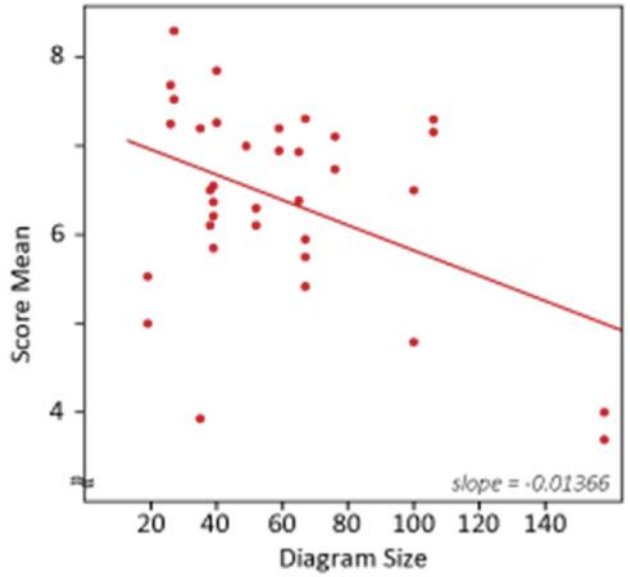
**With increasing diagram size, score mean decreases and score variance increases.**

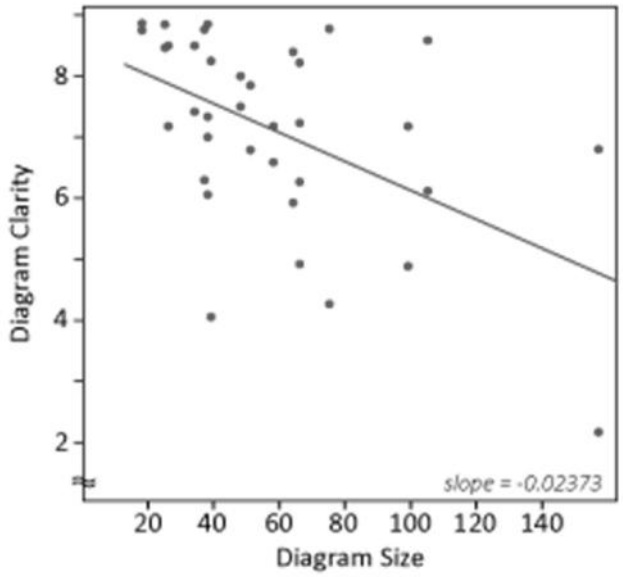
**Simultaneously, subjective assessments also decrease.**

**Interestingly, judgments of layout quality also increase.**



# Diagram Size ~ Impact



| Objective Performance | Score Mean $r$ | ES | $p$ | SIG | Score Variance $r$ | ES | $p$ | SIG |
|---|---|---|---|---|---|---|---|---|
| All Diagrams | −0.423 | L | 0.010 | ** | 0.424 | L | 0.010 | ** |
| Bad Layout | −0.491 | L | 0.039 | * | 0.534 | L | 0.023 | * |
| Good Layout | −0.396 | M | 0.104 | * | 0.303 | M | 0.222 | |

| Diagram Assessment | Layout Quality $r$ | ES | $p$ | SIG | Layout Clarity $r$ | ES | $p$ | SIG |
|---|---|---|---|---|---|---|---|---|
| All Diagrams | 0.538 | L | < 0.001 | *** | −0.508 | L | 0.002 | ** |
| Bad Layout | 0.521 | L | 0.027 | * | −0.563 | L | 0.015 | * |
| Good Layout | 0.573 | L | 0.013 | * | −0.766 | L | 0.0002 | *** |

| Cognitive Load | Diagram Understanding $r$ | ES | $p$ | SIG | Diagram Complexity $r$ | ES | $p$ | SIG |
|---|---|---|---|---|---|---|---|---|
| All Diagrams | −0.338 | M | 0.044 | ** | −0.081 | S | 0.640 | |
| Bad Layout | −0.452 | L | 0.060 | * | −0.313 | M | 0.207 | |
| Good Layout | −0.197 | S | 0.434 | | 0.152 | S | 0.548 | |

r: Pearson's r as computed by `corr.test` in R

S, M, L: Pearson's convention

- **Diagram size affects all performance indicators and assessments, good layouts are affected less: corroborates earlier findings.**
- **However, there is a negative correlation to complexity, and positive correlation to perceived quality – experimental artifacts?**

# Expertise Level + Diagram Size ~ Impact



**Controlling for expertise level shows much larger correlations for novices than experts.**

**We explain this as a lack of coping strategies in novices.**

**Or conversely:**
**the capacity to cope with size defines expertise.**

**And it gets worse with poor layouts.**

| | Novices | | | Experts | | |
|---|---|---|---|---|---|---|
| Objective Performance | Score Mean (low/high expertise) | | | | | |
| | $r$ | ES | $p$ SIG | $r$ | ES | $p$ SIG |
| All Diagrams | −0.494 | L | 0.002 ** | 0.018 | S | 0.917 |
| Bad Layout | −0.397 | M | 0.103 . | −0.173 | S | 0.493 |
| Good Layout | −0.615 | L | 0.007 ** | 0.243 | M | 0.331 |

| | Novices | | | Experts | | |
|---|---|---|---|---|---|---|
| Objective Score | Score Variance (low/high expertise) | | | | | |
| | $r$ | ES | $p$ SIG | $r$ | ES | $p$ SIG |
| All Diagrams | 0.290 | M | 0.086 . | 0.053 | S | 0.764 |
| Bad Layout | 0.254 | M | 0.309 | 0.204 | M | 0.432 |
| Good Layout | 0.343 | M | 0.163 | −0.085 | S | 0.736 |

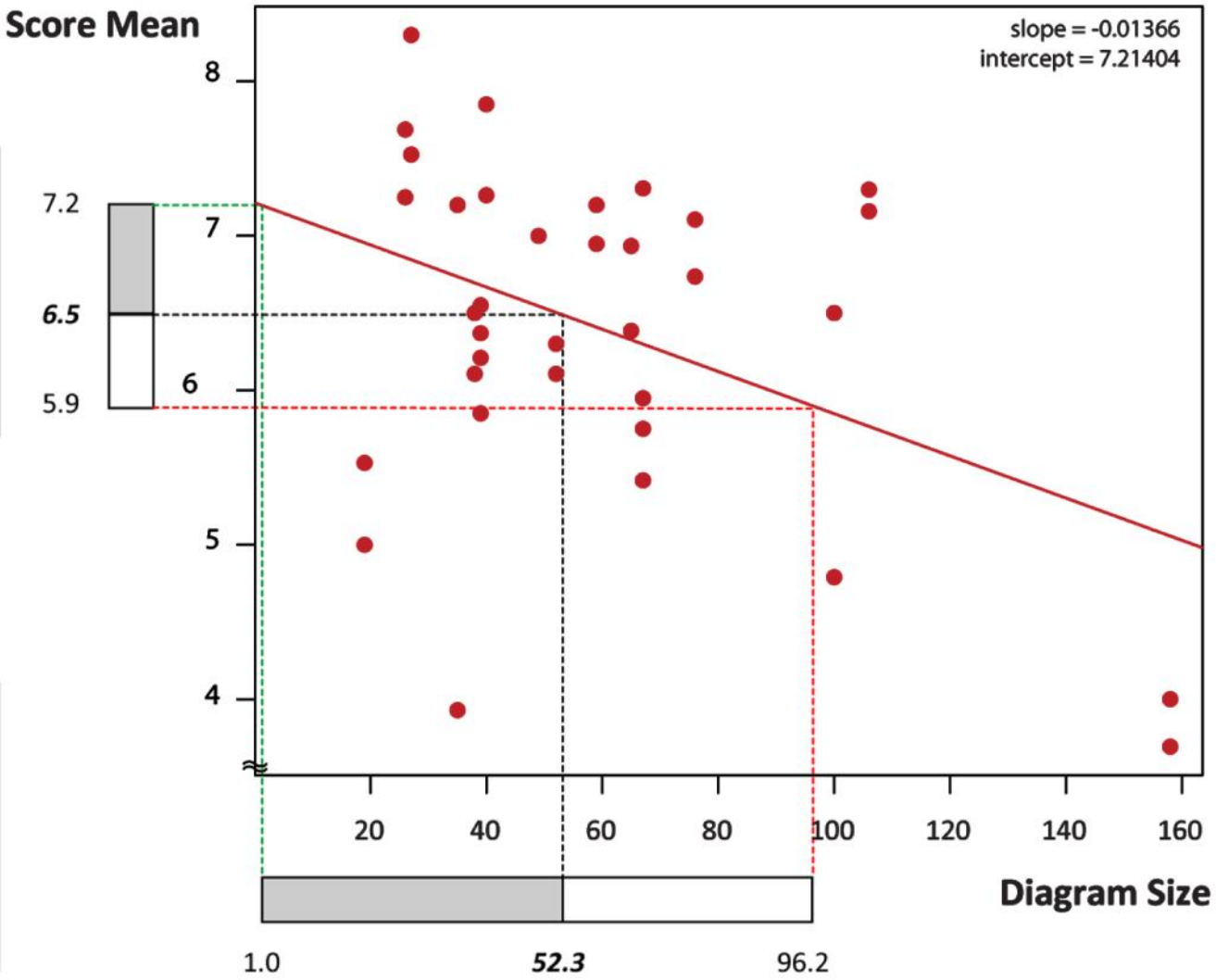
| | Novices | | | Experts | | |
|---|---|---|---|---|---|---|
| Diagram Assessment | Layout Quality (low/high expertise) | | | | | |
| | $r$ | ES | $p$ SIG | $r$ | ES | $p$ SIG |
| All Diagrams | 0.569 | L | 0.0003 *** | 0.484 | L | 0.003 ** |
| Bad Layout | 0.534 | L | 0.023 * | 0.516 | L | 0.028 * |
| Good Layout | 0.615 | L | 0.007 ** | 0.536 | L | 0.022 * |

| | Novices | | | Experts | | |
|---|---|---|---|---|---|---|
| Diagram Assessment | Layout Clarity (low/high expertise) | | | | | |
| | $r$ | ES | $p$ SIG | $r$ | ES | $p$ SIG |
| All Diagrams | −0.525 | L | 0.001 *** | −0.440 | L | 0.007 ** |
| Bad Layout | −0.742 | L | 0.0004 *** | −0.698 | L | 0.001 ** |
| Good Layout | −0.554 | L | 0.017 * | −0.570 | L | 0.014 * |

# Optimal Diagram Size

| Score | | Q1 | Median | Q3 |
|---|---|---|---|---|
| Layouts | Good | 6.2 | 6.9 | 7.2 |
| | All | 5.9 | ***6.5*** | 7.2 |
| | Bad | 5.5 | 6.4 | 7.2 |

| Size | | Q1 | Median | Q3 |
|---|---|---|---|---|
| Layouts | Bad | 1.0 | 23.0 | 74.2 |
| | All | 1.0 | ***52.3*** | 96.2 |
| | Good | 1.0 | 59.6 | 125.5 |



# Conclusions (1/2)



- **Diagram Layout has a highly significant effect on all aspects of modeler performance when reading/understanding diagrams.**
  - **Diagram type:** no correlation (i.e. holds for analysis and design phase)
  - **Layout quality:** high positive correlation (i.e. layout quality matters)
  - **Expertise level:** high negative correlation (i.e. novices benefit more)
  - **Diagram Size:** negative correlation (i.e. size matters)

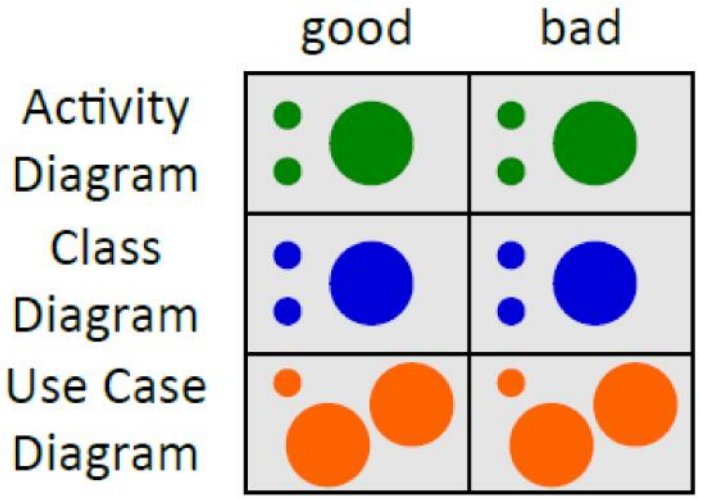
- **Insights**
  - Generality and Validity of our study far surpass previous studies (e.g., n=156).
  - Diagram size can be effectively measured as number of diagram elements.
  - A reasonable guideline for a maximum safe size of diagram is around 50 elements; above that limit, most modelers will perform less than average.
  - Diagram Sized is „irreducible“.

- **New questions arise.**
  - Can size be mitigated by medium (e.g. A0-sized posters, zooming, …)?
  - How do people actually read diagrams (starting point, strategy)?
  - How does diagram flow interact with reading strategies?

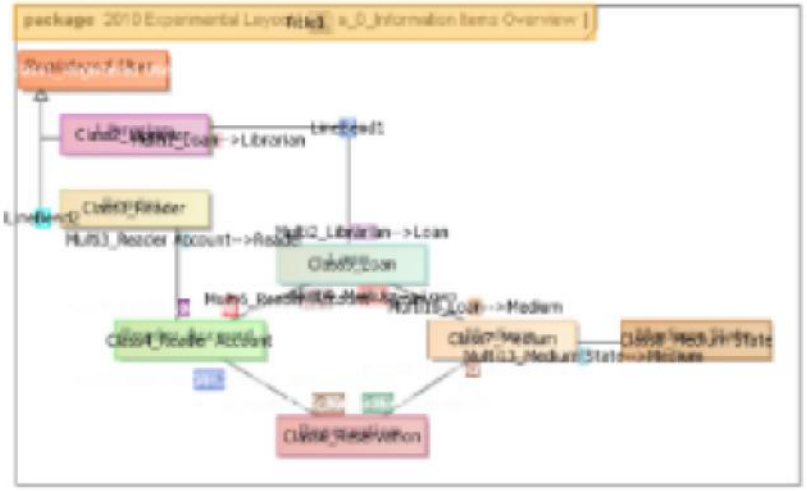
# Eye Tracking Study

## Participants

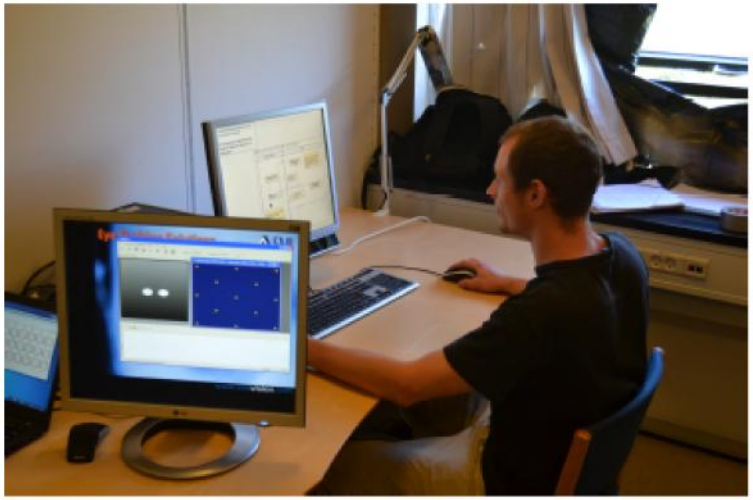

## Model Samples




## Instrumentation



## Setup



# Eye Tracking Study

Diagram Type
*[Class, Activity, Use Case]*

Diagram Size
*[Small, Large]*

Layout Quality
*[good, bad]*

Modeler Expertise
*[novicxe, experienced]*

**Modeler**

## Objective Performance

| Test Score | Time Needed |
|---|---|
| *[0..10]* | *[s]* |

## Subjective Experience

| Preference | Layout Clarity |
|---|---|
| *[1..5]* | *[1..5]* |
| Preference | Understandability |
| *[ordering]* | *[1..5]* |

## Cognitive Load

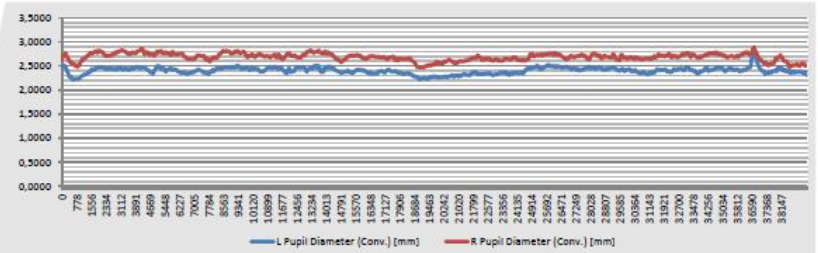
| Blink Rate | Perceived Difficulty |
|---|---|
| *[blinks/s]* | *[1..5]* |
| Pupil Dilation | Fixation Duration |
| *[mm]* | *[ms]* |

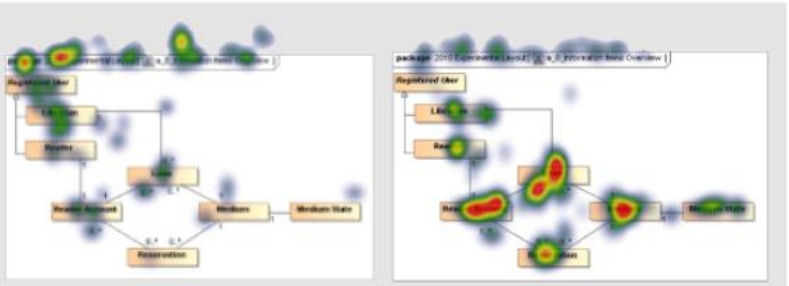
## Reading Strategy

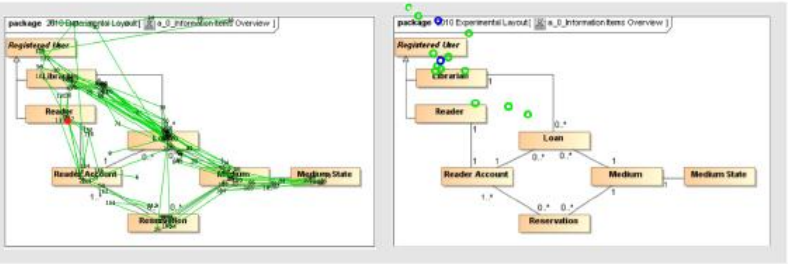
| Scan Start | Scan Path |
|---|---|
| *[AoI]* | *[AoI*]* |

**Point of reference / Replication**

Validation of previous studies by repeating the same experiment, with the same (subjective) measurements on a sub-sample of previously applied stimuli.





# Conclusions (2/2)



- **Modelers have reading strategies on UML diagrams.**
  - Reading strategies differ in starting point and exploration path; we saw the strategies “Document Reading”, “Graph Following”, and “Random Walk”.
- **Different behavior is found at different levels of expertise and layout quality.**
  - If a layout flow is present, it is used (i.e., flow is an effective guide).
  - If high expertise is present, systematic exploration is (mostly) observed.
- **Layout flaws (crossings, bends) trigger as much cognitive load as proper elements.**
  - The performance decrease of modelers seen with increasing diagram size can be mitigated by reducing layout flaws.
- **Our study design is biased in the sense that it did not contain instances of the center-out layout pattern.**
  - Different reading strategies may arise.

# The Good, the Bad, and the Ugly



- **The most important purposes and usages of models are related to communication.**
- **So, the value of diagrams as presentations of models is essential, not ephemeral.**
- **Diagrams are not just pretty pictures of/for models, they carry meaning and value.**
- **In “pretty” diagrams, good and bad (content) makes a difference – in “ugly” diagrams, it‘s all the same.**

# The Model Observatory

Why do people model and how do they use their models?
Are there any differences between different groups?
Does it pay to model, and if so: when and why?

Help us answer these questions and more by answering a few questions - it takes less than 5 minutes!

**http://tinyurl.com/MU-survey-2014**

Prof. Dr. Harald Störrle

Software Engineering Section
Informatics and Mathematical Modelling
Technical University of Denmark

Matematiktorvet
Building 303b, Room 056
DK-2800 Kgs. Lyngby

**Tel** **0045 4525 3757**
**EMail** **hsto@dtu.dk**
**Web** **www2.compute.dtu.dk/~hsto**

DTU

# References



[Mai14] Maier, Anja M. and Baltsen, Nick and Christoffersen, Henrik and Störrle, Harald. Towards Diagram Understanding: A Pilot-Study Measuring Cognitive Workload Through Eye-Tracking. In *Proc. Intl. Conf. Human Behavior in Design*, 2014.

[SBCM14] Harald Störrle, Nick Baltsen, Henrik Christoffersen, and Anja M. Maier. On the Impact of Diagram Layout: How Are Models Actually Read? In Stefan Sauer, Manuel Wimmer, Marcela Genero, and Shaz Qadeer, editors, *Joint Proc. MODELS 2014 Poster Session and ACM Student Research Competition*, volume 1258, pages 31–35. CEUR, 2014.

[Stö11] Harald Störrle. On the Impact of Layout Quality to Unterstanding UML Diagrams. In *Proc. IEEE Symp. Visual Languages and Human-Centric Computing (VL/HCC'11)*, pages 135–142. IEEE Computer Society, 2011.

[Stö12] Harald Störrle. On the Impact of Layout Quality to Unterstanding UML Diagrams: Diagram Type and Expertise. In Gennaro Costagliola, Andrew Ko, Allen Cypher, Jeffrey Nichols, Christopher Scaffidi, Caitlin Kelleher, and Brad Myers, editors, *Proc. IEEE Symp. Visual Languages and Human-Centric Computing (VL/HCC'12)*, pages 195–202. IEEE Computer Society, 2012.

[Stö14] Harald Störrle. On the Impact of Layout Quality to Understanding UML Diagrams: Size Matters. In Jürgen Dingel, Wolfram Schulte, Isidro Ramos, and Emilio Abrahao, Silviaand Insfran, editors, *Proc. 17th Intl. Conf. Model Driven Engineering Languages and Systems (MoDELS'14)*, number 8767 in LNCS, pages 518–534. Springer Verlag, 2014.