

Analyzing Awareness, Decision, and Outcome Sequences of Project Design Groups

A Platform for Instrumentation of Workshop-based Experiments

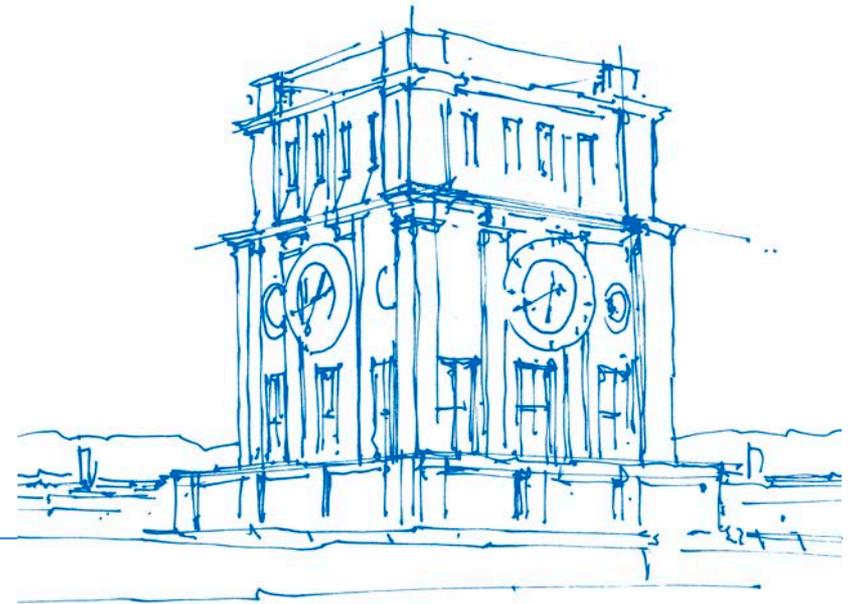
Bryan R. Moser (Massachusetts Institute of Technology)

Carl O. R. Frühling (Technical University of Munich)

Dec 18th, 2018



Product Development



Uhrenturm der TUM

Agenda

- **Research domain and method**
- Data analysis:
 - Predictor and outcome sequences
 - Clustering and similarities
- Predictability and findings
- Summary and outlook

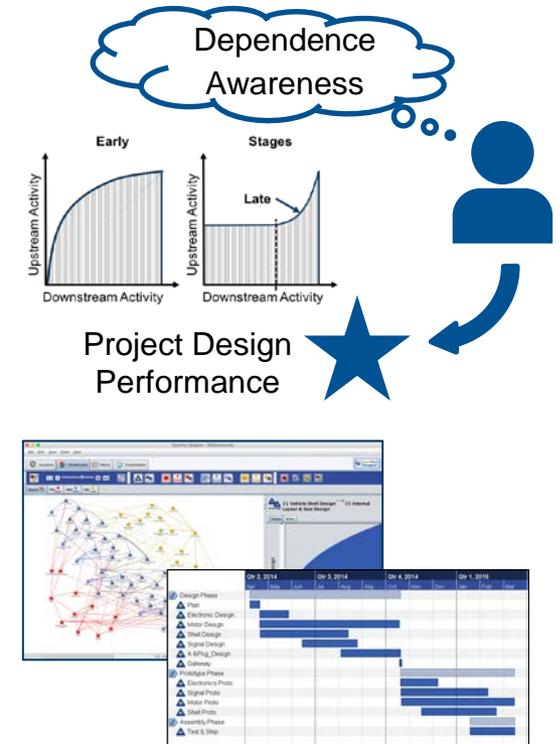
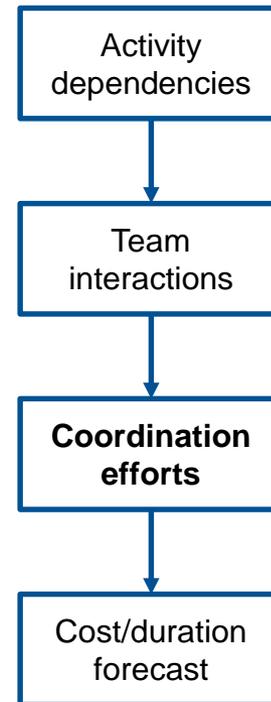
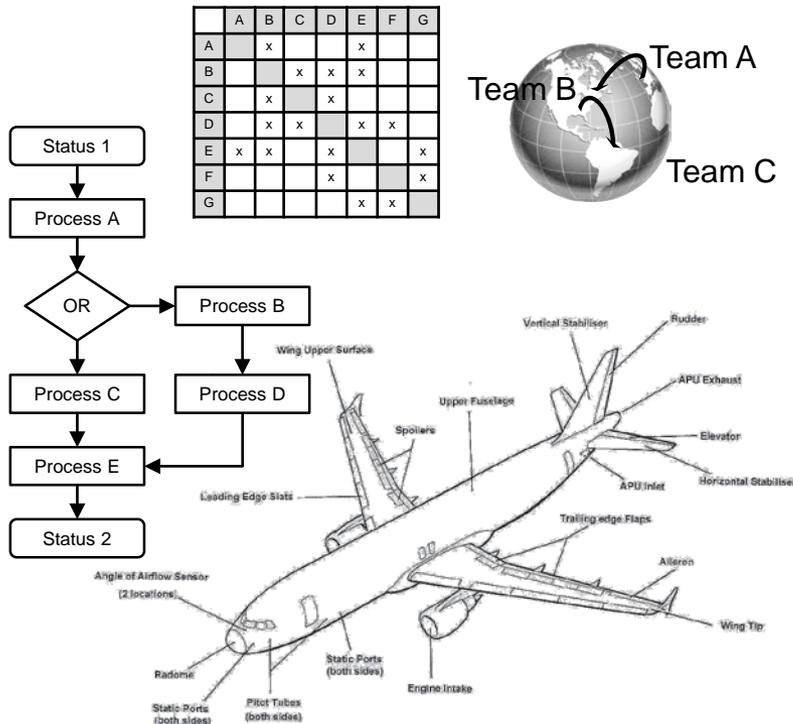
Project planner should be aware of activity dependencies to account for coordination efforts

Research domain

Example

Complex projects with global teams

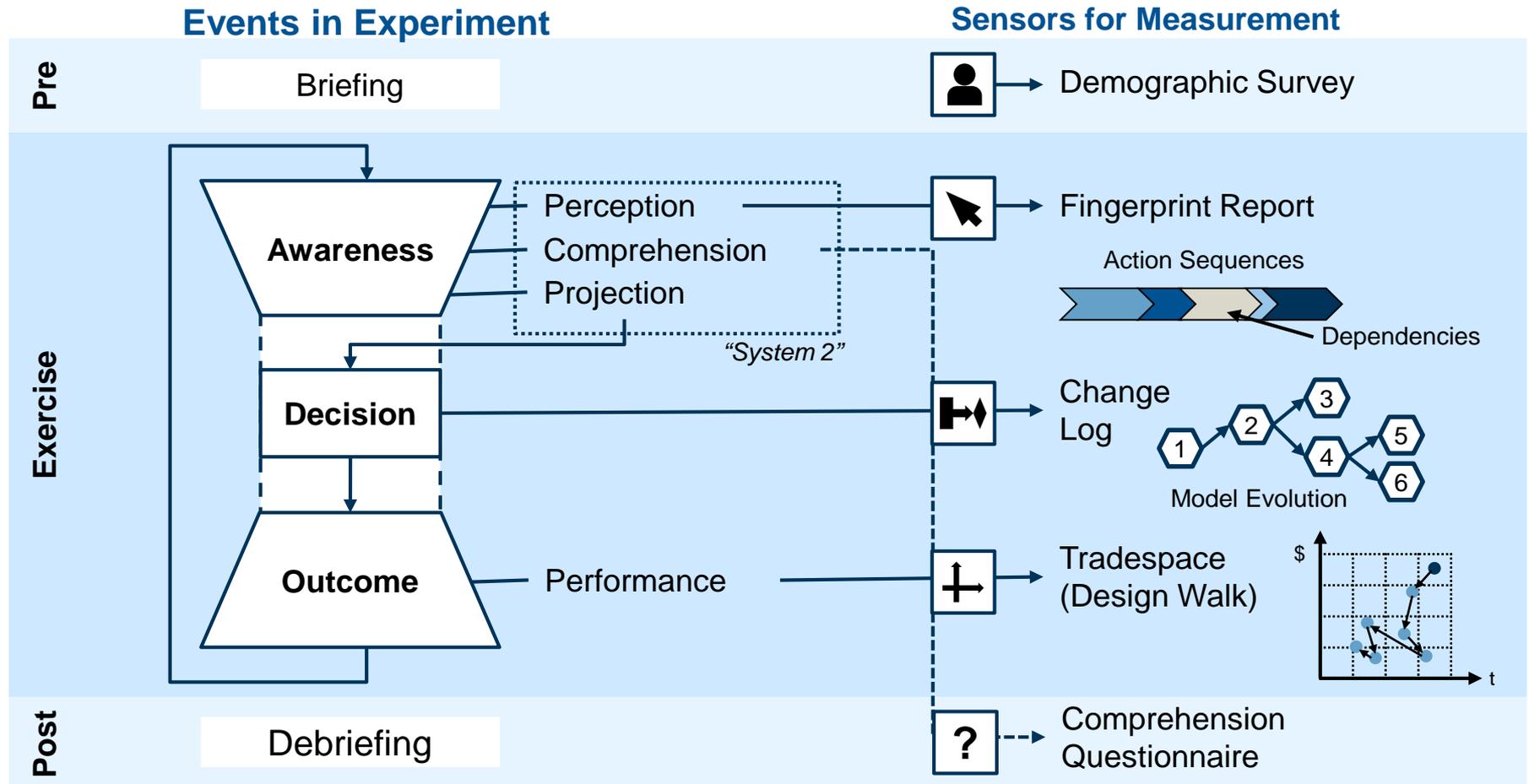
Project Design in TeamPort Software



We conduct workshop-based experiments to collect real-time awareness and performance data

Research method (1/2)

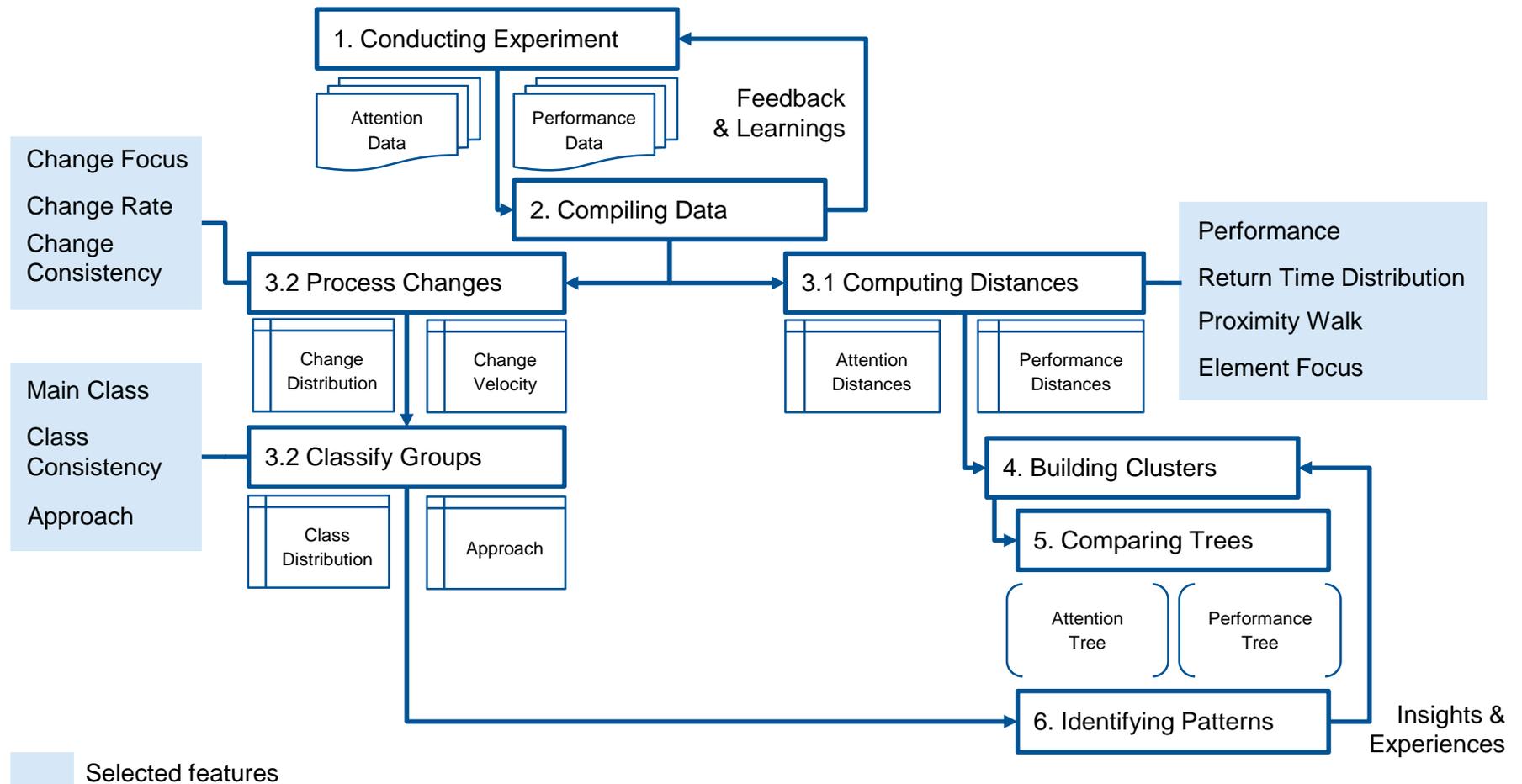
Illustrative



Features are selected from collected data for sequence and clustering analysis

Research method (2/2)

Illustrative



Agenda

- Research domain and method
- **Data analysis:**
 - Predictor and outcome sequences
 - Clustering and similarities
- Predictability and findings
- Summary and outlook

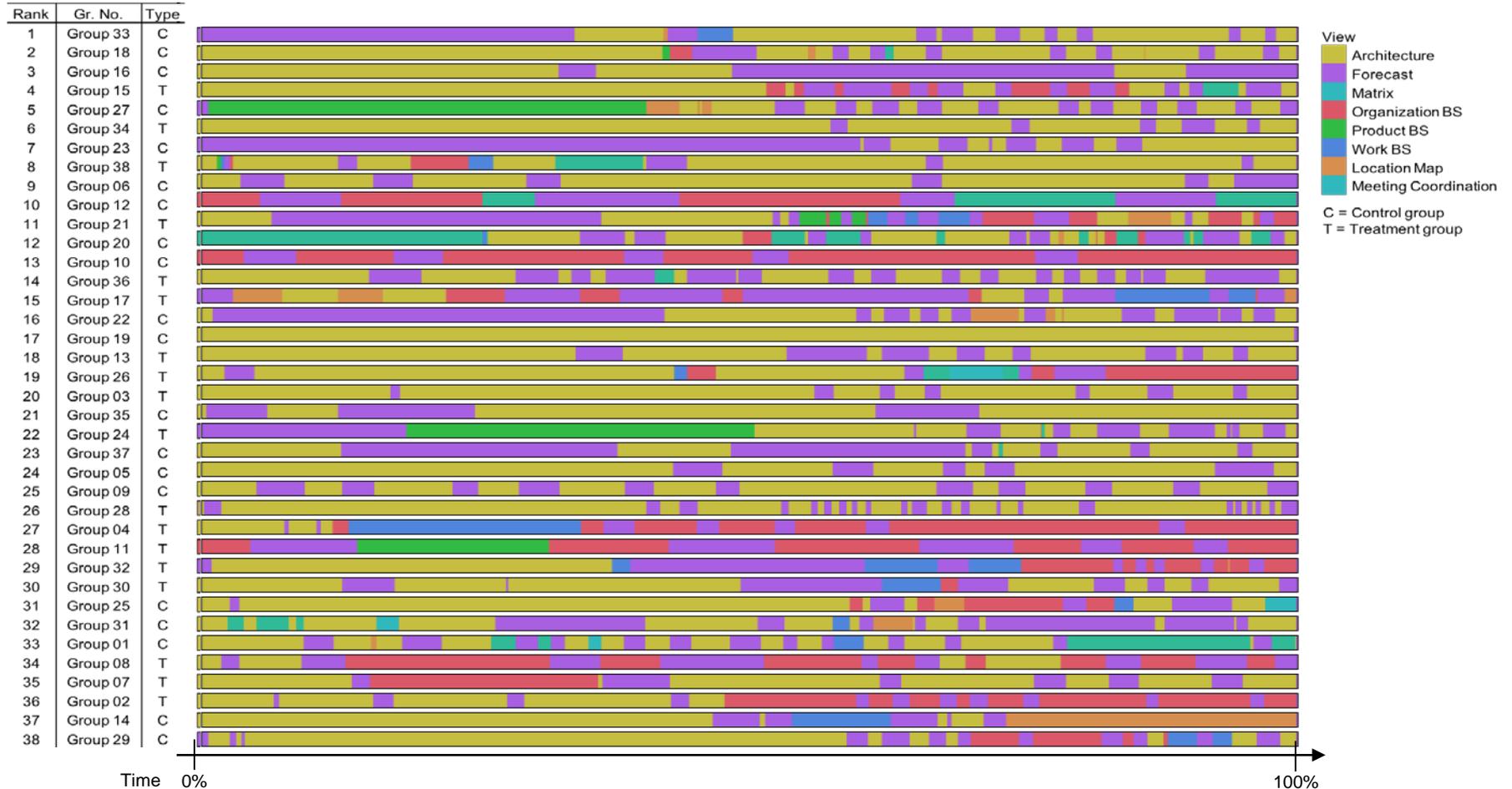
The “design walk” shows subject performance during experiment

Outcome analysis



The “fingerprints” show subject’s attention allocation sequences

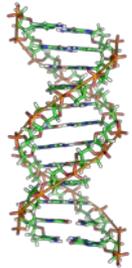
Predictor analysis



Bioinformatics use sequence similarity analyses to detect genetic relations

Sequence analysis in bioinformatics

1.) Genetic codes in DNA sequences



```

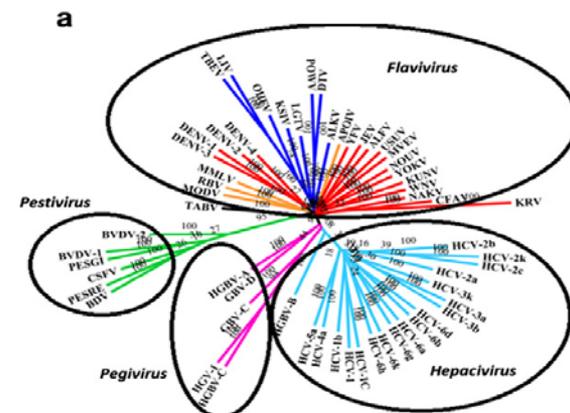
A5ASC3.1 14 SIKLWPPSGTTRLLVERMANNLST..ESIFTRK..YGLSKEEARENAKQTEEVACSTANQ.....HYEKEPDG DGS AVQLYAKECSKRLILEVLK 101
B4F917.1 13 SIKLWPPSESTRIMLVDRMTNNLST..ESIFSRK..YRLGKQEAHENAKTLEELCFALADE.....HFREEPDGDSSAVQLYAKETSMMLEVLK 100
A9S1V2.1 23 VFKLWPPSGTRERAVRQKMAKLLSS..ACFESQS..FARIELADAEHARALEEVAFGAQE.....ADSGGDKTBSAVMVMYAKHASKMLLETLR 109
B9GSN7.1 13 SVKLWPPGSTRMLVERMTKNFIT..PSFISRK..YGLLSKEEAEDAKKLEEVAFAAANQ.....HYEKQPDGDSSAVQIYAKESSRLMLEVLK 100
Q8H056.1 30 SFSIWPPQTQRTRDAVVRRLVDTLGG..DTILCKR..YGAVPAADAEPAARGIEAEAFDAASA..SGEAAATASVEEGIKALQLYSKEVSRRLDFVK 120
Q0D423.2 44 SLSIWPPSQTRDAVVRRLVQTLVA..PSILSKR..YGAVPEAEAGRAAAVEAEYAVTES..SSAAAAPASVEDGIEVLQAYSKEVSRRLLELAK 135
B9MWJ8.1 56 SFSIWPPQTQRTRDAIISRLLIETLST..TSVLSKR..YGTIPKEEASESRRIEEAFSGAST.....VASSEKDGLEVLQLYSKEISKRMLLETVK 141
Q01YC5.1 29 SFAVWPPTRRTRDAVVRRLVAVLSGDTTALRKRKYR..YGAVPAADAEARARAVEAQAADAASA.....SSSSSSSVDGIETLQLYSREVSNRLAFVR 121
A9NM46.1 13 SIKLWPPSESTRMLVERMTDNLSS..VSFFSRK..YGLLSKEEAENAKRIETATFLAND.....HEAKEPNLDSSVVFYAREASKLMLEALK 100
Q9C500.1 57 SLRIWPPTKQRTRDAVLRNLIETLST..ESILSKR..YGTLSDDATTVAKLIEEAYGVASN.....AVSSDDDGKILELYSKEISKRMLLESVK 142
Q2HR17.1 25 NYSIWPPKQRTRDAVKNRLLIETLST..PSVLTKR..YGTMSADEASAARIQIEEAFSVANA.....SSSTSNQNVITILEVYSKEISKRMIETVK 110
Q9M7N3.1 28 SFKIWPPQTQRTRDAVVRRLVETLTS..QSVLSKR..YGVIPSEEDATSAARIIEEAFSVASV..ASAASTGGRPEDEWIEVLHIYSQEIQRVVESAK 119
Q9M7N6.1 25 SFSIWPPQTQRTRDAVLRNLIETLST..PSILSKR..YGTLPQDEASETARLIEEAFSAAGS.....TASDADDGIEILQVYSKEISKRMIETVK 110
Q9LE82.1 14 SVKMWPPSKSTRMLVERMTKNITT..PSIFSRK..YGLLSVEEAEQDAKRIEDLAFATANK.....HFQNEPDGDTSAVHVYAKESSKMLLDVVK 101
Q9M651.2 13 SIKLWPPSLPTRKALIERITNFFSS..KTIFTEK..YGLSLTKDQATENAKRIEDIAFSTANQ.....QFEREPDGDGS AVQLYAKECSKRLILEVLK 100
B9R748.1 48 SLSIWPPQTQRTRDAVITRLLIETLSS..PSVLSKR..YGTIISHDEAESARRLEDEAFGVANT.....ATSAEDDGLLEILQLYSKEISRRMLDTVK 133
    
```

2.) Distance Matrix

$$D_{ij} = \sqrt{\sum (\mu_{ir} - \mu_{jr})^2 + \sum (\sigma_{ir} - \sigma_{jr})^2}$$

| 1037.067 | Group 01 | Group 02 | Group 03 | Group 04 | Group 05 | Group 06 | Group 07 | Group 08 | Group 09 | Group 10 | Group 11 | Group 12 | Group 13 |
|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| Group 01 | 0.000 | 0.047 | 0.413 | 0.046 | 0.075 | 0.054 | 0.057 | 0.053 | 0.074 | 0.062 | 0.056 | 0.052 | 0.189 |
| Group 02 | 0.047 | 0.000 | 0.377 | 0.070 | 0.052 | 0.081 | 0.047 | 0.080 | 0.092 | 0.093 | 0.085 | 0.084 | 0.155 |
| Group 03 | 0.413 | 0.377 | 0.000 | 0.436 | 0.353 | 0.443 | 0.382 | 0.441 | 0.438 | 0.455 | 0.446 | 0.446 | 0.233 |
| Group 04 | 0.046 | 0.070 | 0.436 | 0.000 | 0.085 | 0.016 | 0.058 | 0.015 | 0.057 | 0.025 | 0.018 | 0.019 | 0.217 |
| Group 05 | 0.075 | 0.052 | 0.353 | 0.085 | 0.000 | 0.090 | 0.033 | 0.089 | 0.103 | 0.103 | 0.094 | 0.094 | 0.144 |
| Group 06 | 0.054 | 0.081 | 0.443 | 0.016 | 0.090 | 0.000 | 0.064 | 0.010 | 0.061 | 0.013 | 0.006 | 0.007 | 0.276 |
| Group 07 | 0.057 | 0.047 | 0.382 | 0.058 | 0.033 | 0.064 | 0.000 | 0.061 | 0.085 | 0.077 | 0.068 | 0.068 | 0.169 |
| Group 08 | 0.053 | 0.080 | 0.441 | 0.015 | 0.089 | 0.010 | 0.061 | 0.000 | 0.059 | 0.019 | 0.013 | 0.012 | 0.224 |
| Group 09 | 0.074 | 0.092 | 0.438 | 0.057 | 0.103 | 0.061 | 0.085 | 0.059 | 0.000 | 0.063 | 0.059 | 0.060 | 0.223 |
| Group 10 | 0.062 | 0.093 | 0.455 | 0.025 | 0.103 | 0.013 | 0.077 | 0.019 | 0.063 | 0.000 | 0.010 | 0.011 | 0.238 |
| Group 11 | 0.056 | 0.085 | 0.446 | 0.018 | 0.094 | 0.006 | 0.068 | 0.013 | 0.059 | 0.010 | 0.000 | 0.006 | 0.229 |
| Group 12 | 0.052 | 0.084 | 0.446 | 0.019 | 0.094 | 0.007 | 0.068 | 0.012 | 0.060 | 0.011 | 0.006 | 0.000 | 0.228 |
| Group 13 | 0.189 | 0.155 | 0.233 | 0.217 | 0.144 | 0.226 | 0.169 | 0.224 | 0.223 | 0.238 | 0.229 | 0.228 | 0.000 |

3.) Hierarchical Clustering



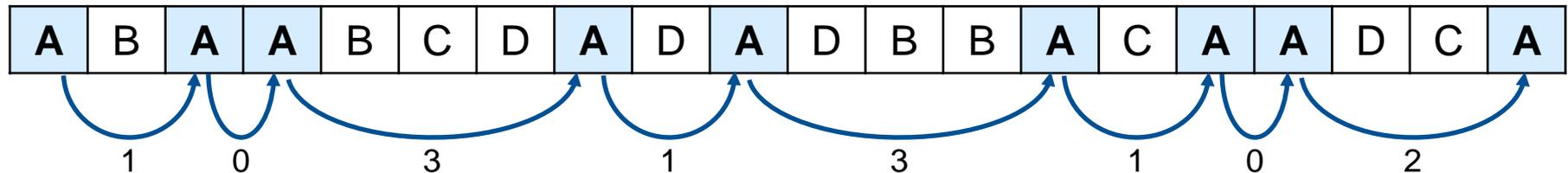
We adopted this method to analyze our sequential experiment data

The Return Time Distribution in a mouse click sequence represents subject's attention allocation

Feature calculation

Example

Example Sequence with alphabet size 4 $\rightarrow R = \{A, B, C, D\}$



| Return Time (A) | Frequency |
|-----------------|-----------|
| 0 | 2 |
| 1 | 3 |
| 2 | 1 |
| 3 | 2 |

Distance between two sequences (i, j)

$$D_{ij} = \sqrt{\sum (\mu_{ir} - \mu_{jr})^2 + \sum (\sigma_{ir} - \sigma_{jr})^2}$$

$$\mu_r = \frac{\sum (\text{Return Time} \times \text{Frequency})}{\sum \text{Frequency}}$$

$$\mu_A = \frac{0 \times 2 + 1 \times 3 + 2 \times 1 + 3 \times 2}{2 + 3 + 1 + 2} = 1.57$$

$$\sigma_r = \sqrt{\frac{\sum (\text{Frequency} - \mu_i)^2}{\sum \text{Frequency}}}$$

$$\sigma_A = \sqrt{\frac{(2 - 1.57)^2 + (3 - 1.57)^2 + (1 - 1.57)^2 + (2 - 1.57)^2}{2 + 3 + 1 + 2}} = 0.59$$

Return Time Distribution is an *alignment-free* sequence analysis method – the order of elements is not respected in comparison of two sequences

Calculated feature distances are hierarchically clustered with “neighbor-joining” method

Hierarchical clustering – predictor

Example

Sequence-based Return Time Distribution (RTD)

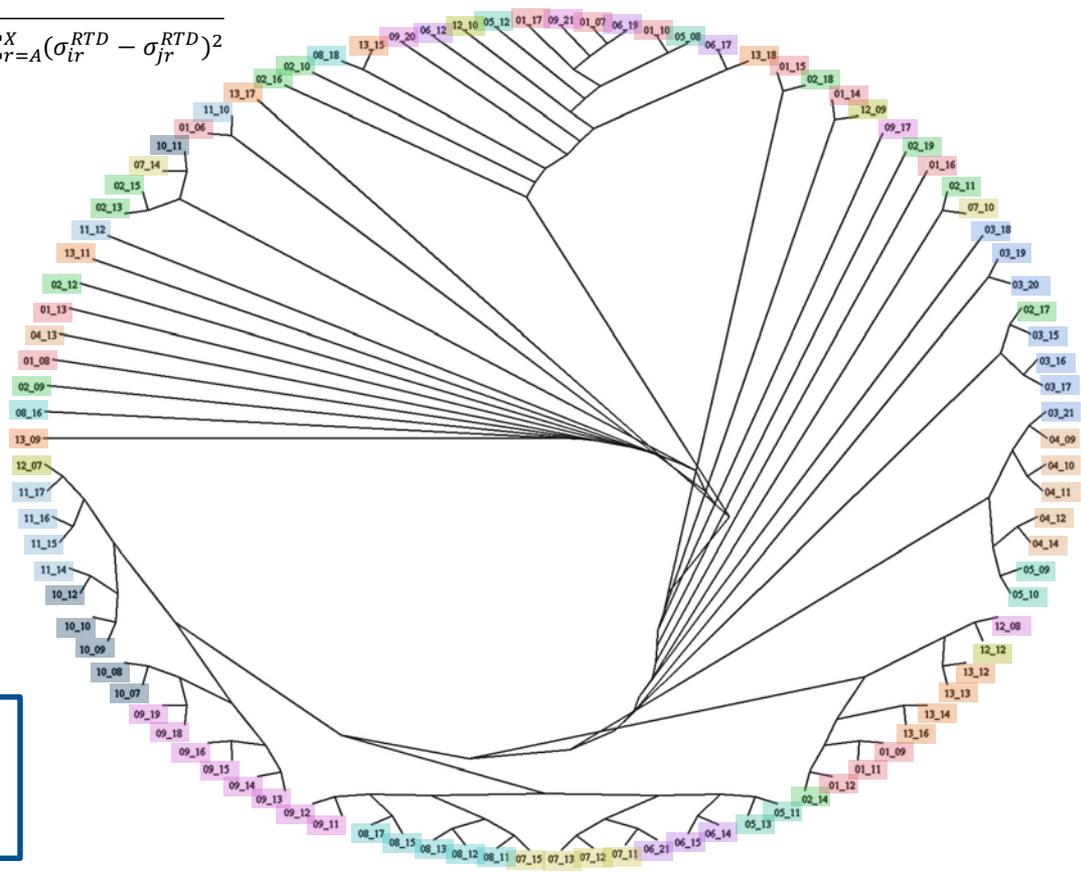
$$d_{ij}^{RTD} = \sqrt{\sum_{r=A}^X (\mu_{ir}^{RTD} - \mu_{jr}^{RTD})^2 + \sum_{r=A}^X (\sigma_{ir}^{RTD} - \sigma_{jr}^{RTD})^2}$$

$$\mu_r^{RTD} = \frac{\sum(\text{Return Time} \times \text{Frequency})}{\sum \text{Frequency}}$$

$$\sigma_r^{RTD} = \sqrt{\frac{\sum(\text{Frequency} - \mu_r^{RTD})^2}{\sum \text{Frequency}}}$$

n = 98 sequences

- Group 01
- Group 02
- Group 03
- Group 04
- Group 05
- Group 06
- Group 07
- Group 08
- Group 09
- Group 10
- Group 11
- Group 12
- Group 13



The RTD tree is very little distinct due to many zero distances

Clustering needs to be compared to find predictor and outcome correlation

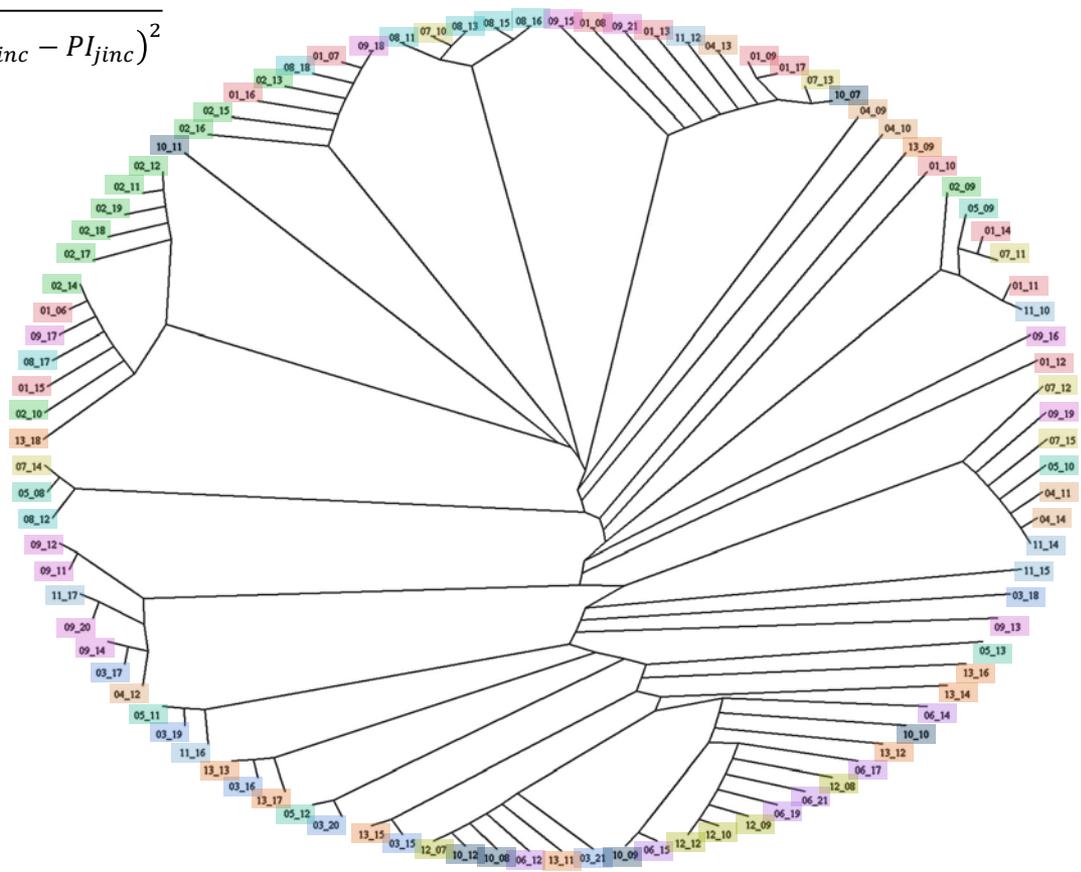
Hierarchical clustering – outcome

Example

Sequence-based Performance Impact (PI)

$$d_{ij}^{PI} = \sqrt{(PI_{itot} - PI_{jtot})^2 + (PI_{iinc} - PI_{jinc})^2}$$

n = 98 sequences



- Group 01
- Group 02
- Group 03
- Group 04
- Group 05
- Group 06
- Group 07
- Group 08
- Group 09
- Group 10
- Group 11
- Group 12
- Group 13

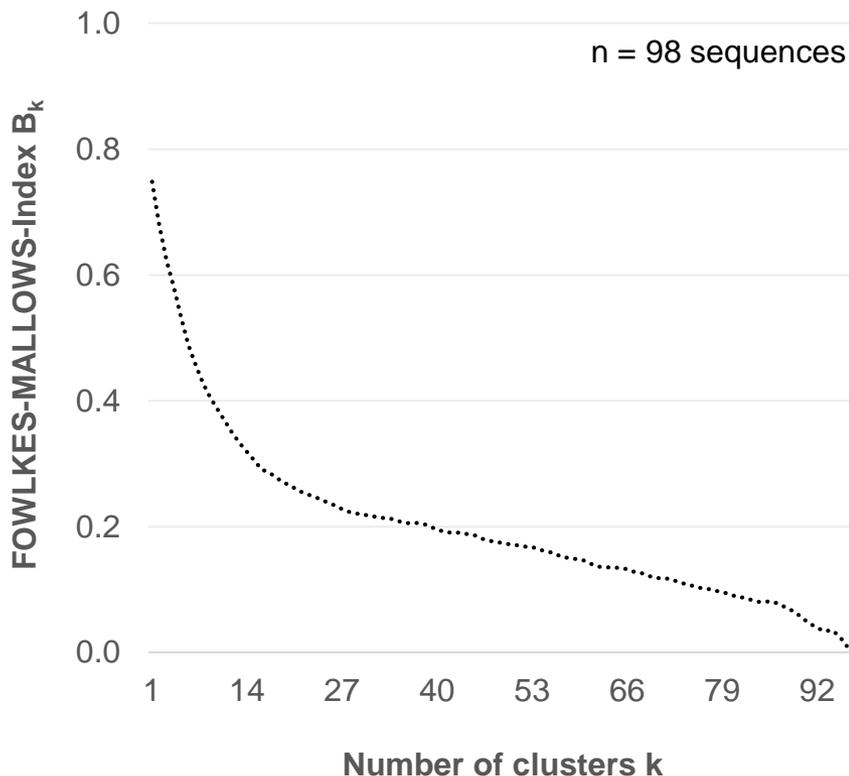
Agenda

- Research domain and method
- Data analysis:
 - Predictor and outcome sequences
 - Clustering and similarities
- **Predictability and findings**
- Summary and outlook

Predictability is represented by similarity between predictor and outcome clustering

Fowlkes-Mallows-Index (FMI)

FMI for randomized clusterings (baseline)



Explanation

- FMI is a **matching index** that determines the **similarity** of two hierarchical clusterings
 - FMI = 0**: no similarity at all
 - FMI = 1**: identical clusterings
- FMI is calculated over **matching matrix** ($M = [m_{ij}]$) holding number of common items between i^{th} and j^{th} cluster of the two clustering
- Rows and columns** of M are summed up for all possible **numbers of clusters** (k)

$$m_i = \sum_{j=1}^k m_{ij}; m_j = \sum_{i=1}^k m_{ij}$$

- Matching index B_k** is calculated

$$B_k = \frac{T_k}{\sqrt{P_k \times Q_k}}$$

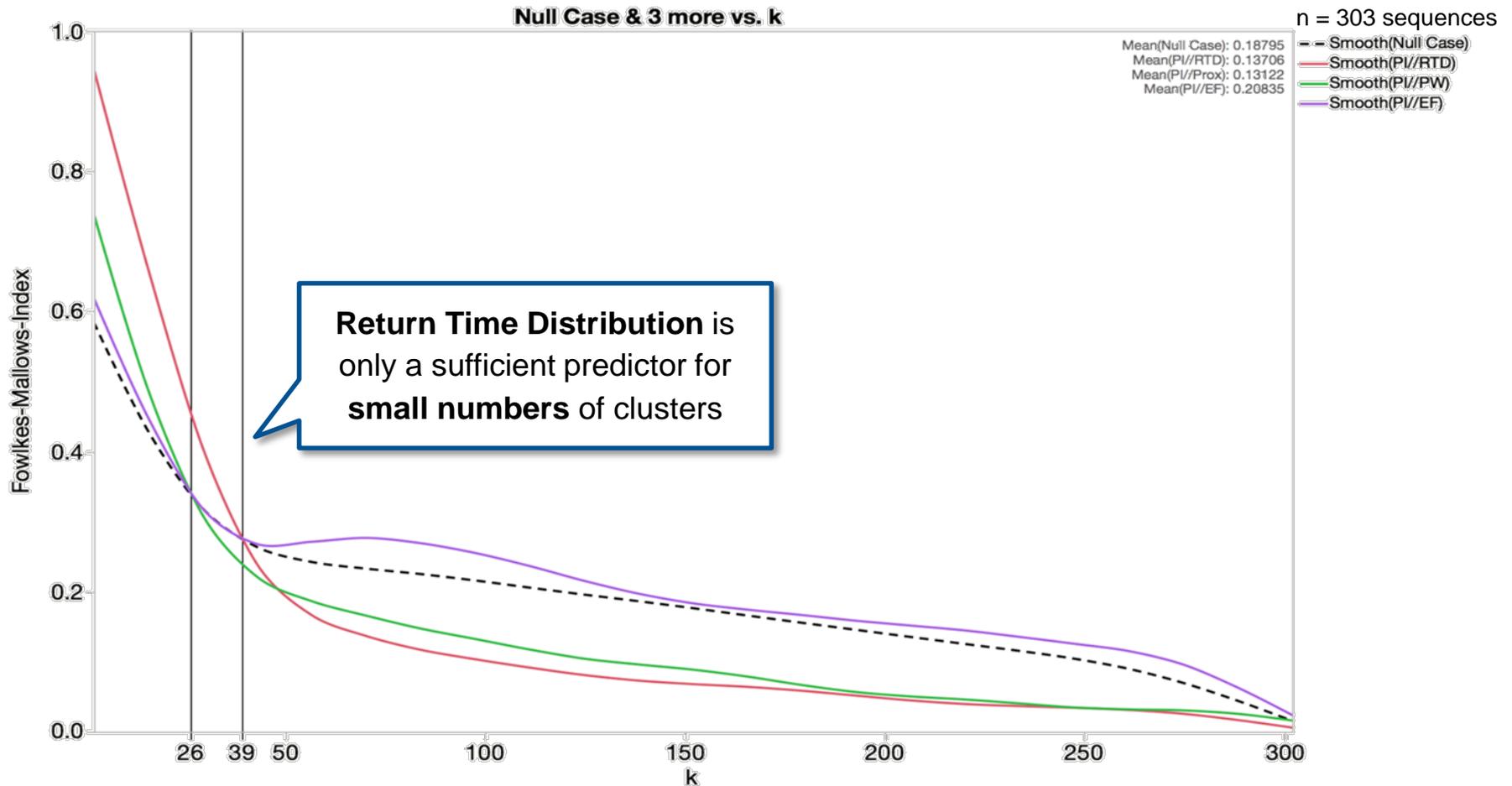
$$T_k = \sum_{i=1}^k \sum_{j=1}^k m_{ij}^2 - n$$

$$P_k = \sum_{i=1}^k m_i^2 - n$$

$$Q_k = \sum_{j=1}^k m_j^2 - n$$

A good performance predictor is larger than similarity index baseline

Clustering similarity analysis



Sequence analysis shows that project planners should focus on activity dependencies

Findings

| | Tested hypotheses | Validation |
|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------|
|  | High performing Project Design groups allocate their attention different from low performing Project Design groups. | Valid |
|  | High performing Project Design groups allocate their attention more to activities and dependencies than low performing Project Design groups. | Valid |
|  | High performing Project Design groups focus on the project architecture before making changes on the project model. | Not valid |
|  | Project Design groups become aware of activity dependencies through laying out the project architecture themselves. | Not valid |
|  | High performing Project Design groups follow similar action patterns which low performing Project Design groups do not follow. | Not tested |

Agenda

- Research domain and method
- Data analysis:
 - Predictor and outcome sequences
 - Clustering and similarities
- Predictability and findings
- **Summary and outlook**

Our research approach is continuously enhanced for global scalability

Key takeaways

- **Activity dependencies** have big impact in complex global projects
- Project designer benefit from **visualization of dependencies** in project planning software
- **Awareness** for activity dependencies increases **design performance**
- Research platform sensors in **workshop-based experiments** allow collection of data for **attention allocation, decision-making and design performance**
- **Sequence analysis** is an appropriate method to analyze **behavioral patterns**
- **Selection and clustering** of the right data features leads to insights about **successful design patterns**

Outlook

- Further **attention allocation features** could be considered for **clustering analysis**
- **Decision-making data** was collected but not entirely analyzed, yet
- Research platform allows to **add further sensors** and **scalability of experiments**

Key references

- **B. R. Moser, W. Grossmann, and P. Starke**, "Mechanism of Dependence in Engineering Projects as Sociotechnical Systems," presented at the 22nd ISPE Concurrent Engineering Conference (CE2015), 2015.
- **B. R. Moser and R. T. Wood**, "Design of Complex Programs as Sociotechnical Systems," in *Concurrent Engineering in the 21st Century*, J. Stjepandić, Ed., ed Switzerland: Springer International Publishing, 2015, pp. 197-220.
- **D. H. Rumsfeld**, "DoD News Briefing," ed. Washington, D.C.: Federal News Service Inc., 2002.
- **J. Luft and H. Ingham**, "The Johari Window: a graphic model of awareness in interpersonal relations," *Human relations training news*, vol. 5, pp. 6-7, 1961.
- **F. Marle and L.-A. Vidal**, *Managing Complex, High Risk Projects - A Guide to Basic and Advanced Project Management*. London: Springer-Verlag, 2016.
- **M. R. Endsley**, "Toward a Theory of Situational Awareness in Dynamic Systems," *Human Factors*, vol. 37, pp. 32-64, 1995.
- **D. Kahneman**, *Thinking, Fast and Slow*, 1st ed. New York: Farrar, Straus and Giroux, 2011.
- **N. Chucholowski, P. Starke, B. R. Moser, E. Rebentisch, and U. Lindemann**, "Characterizing and Measuring Activity Dependence in Engineering Projects," presented at the Portland International Conference on Management of Engineering & Technology 2016, Honolulu, Hawaii, USA, 2016.
- **P. Kolekar, M. Kale, and U. Kulkarni-Kale**, "Alignment-free distance measure based on return time distribution for sequence analysis: Applications to clustering, molecular phylogeny and subtyping," *Molecular Phylogenetics and Evolution*, vol. 65, pp. 510-522, 2012.
- **N. Saitou and M. Nei**, "The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees," *Molecular Biology and Evolution*, vol. 4, pp. 406-425, 1987.
- **E. B. Fowlkes and C. L. Mallows**, "A Method for Comparing Two Hierarchical Clusterings," *Journal of the American Statistical Association*, vol. 78, pp. 553-569, 1983.

Contact

Bryan R. Moser

Massachusetts Institute of Technology

Email: bry@mit.edu

Carl O. R. Fruehling

Technical University of Munich

Email: cfruehling@mytum.de

