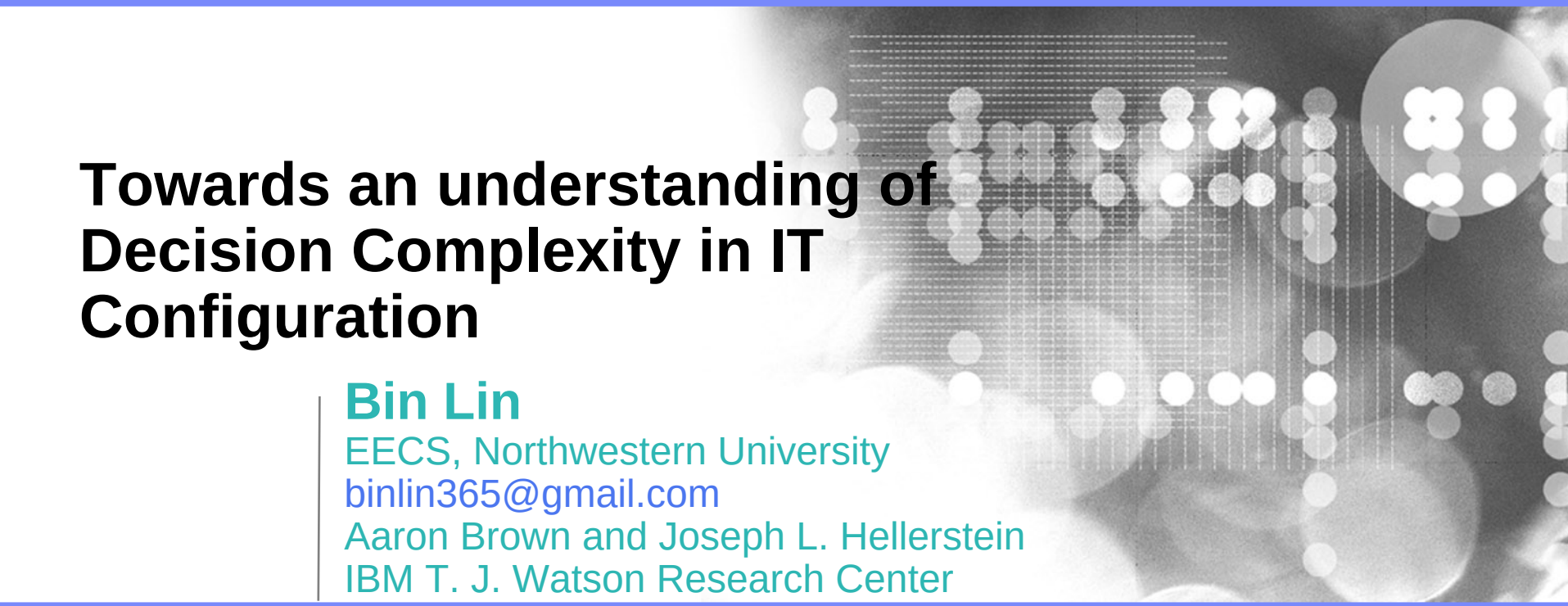


Towards an understanding of Decision Complexity in IT Configuration

Bin Lin
EECS, Northwestern University
binlin365@gmail.com
Aaron Brown and Joseph L. Hellerstein
IBM T. J. Watson Research Center

Take-away points

- **We investigated decision complexity in IT configuration procedures**
  - Developed an initial model of decision complexity
  - Used an carefully-mapped analogous domain to explore complexity space
  - Conduct an extensive user study
  - Quantitative results showing the key factors
  - Next steps are to explore further in simulated IT environment

Outline

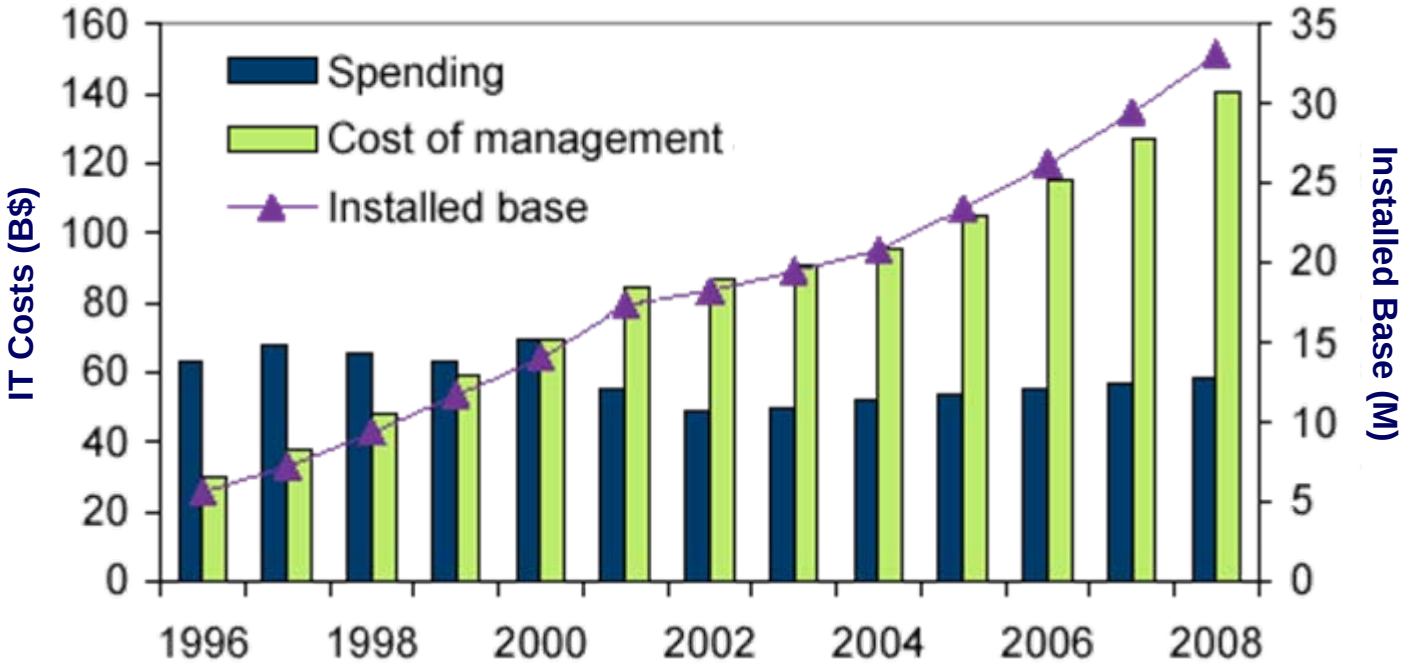
- **Context and previous work**
- **Motivation**
- **Model**
- **User study**
- **Conclusion**
- **Future work**

Outline

- **Context and previous work**
- **Motivation**
- **Model**
- **User study**
- **Conclusion**
- **Future work**

Context

- **Rapidly-rising costs of IT system management**



Source: IDC 2004

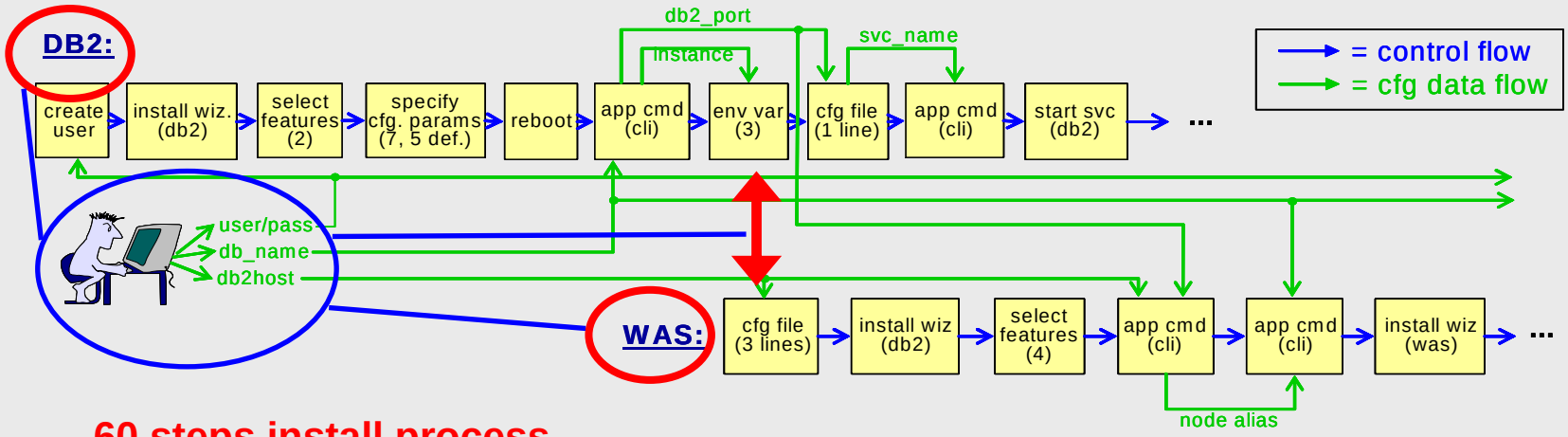
The System Management Problem

- **System management costs are driven by labor**
  - Labor costs are affected by required skill, time, and error rate
  - These factors are directly related to the complexity of management tasks
- ***Reducing management cost means reducing management complexity***

Context: Low-level IT Configuration Procedure

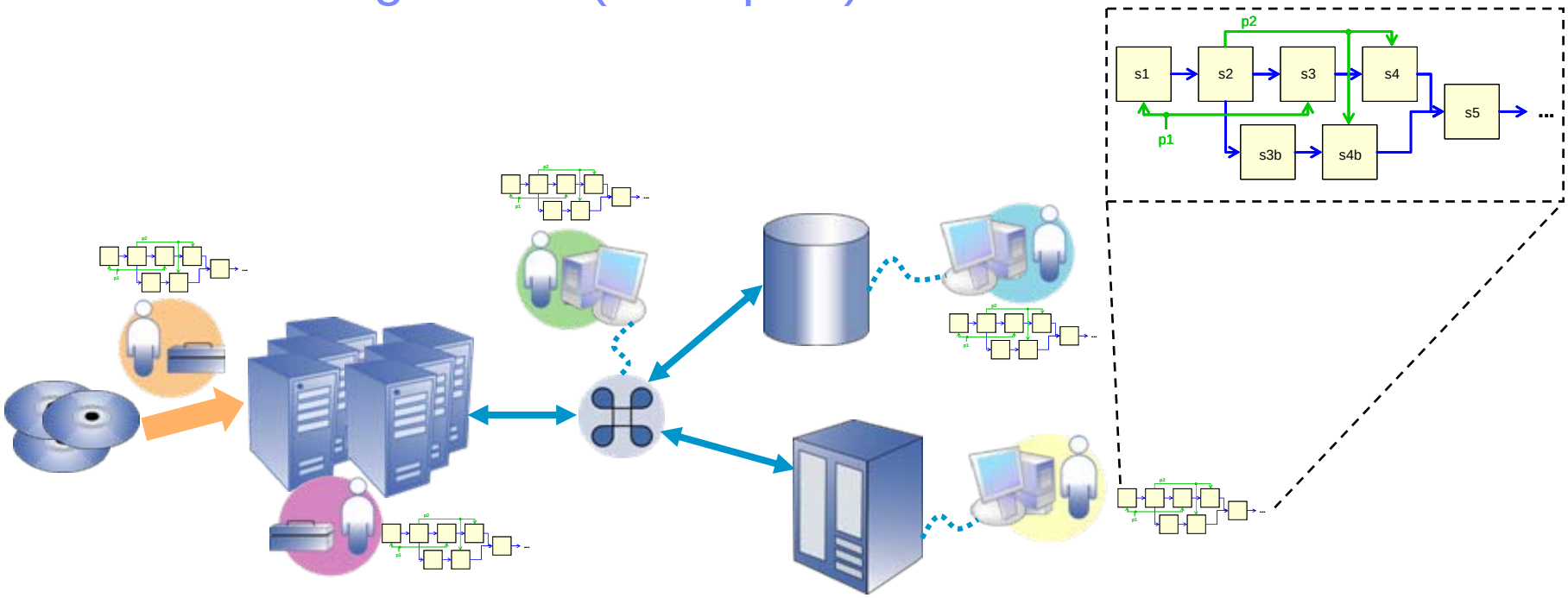
- **A significant source of complexity and error**
- **An interactive series of steps carried out by the system and its human administrator/operator to construct a working solution**
- **Human-driven configuration procedures occur at every stage.**

- Example: complexity of J2EE provisioning



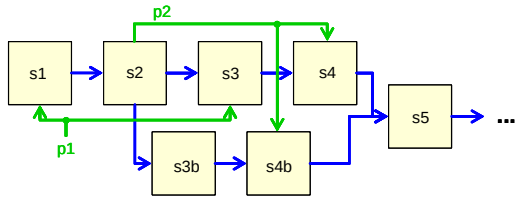
60 steps install process

Context: IT Configuration (examples)



- **Examples of configuration procedures in enterprise systems:**
  - Installing, provisioning, upgrading, or decommissioning software or hardware
  - Configuring multiple systems to work together
    - E.g., connecting a database to a web server
  - Adjusting system parameters to alter system performance or availability
  - Restoring a system damaged by failure or hacker attack

Context: Quantifying IT Process Complexity

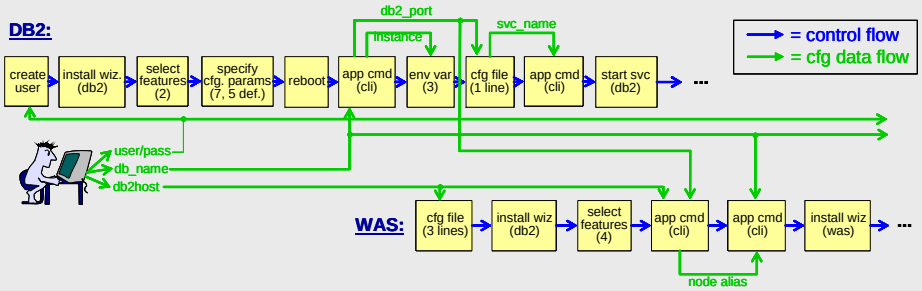


- **Technical problem**
  - Identify metrics and develop benchmarks for quantifying the exposed operational complexity of IT processes
- **Importance**
  - Complexity of systems management processes drives labor cost
  - Labor cost reductions are extremely important to services organizations (SO) and customers
  - A quantitative framework for complexity can guide process improvements to reduce labor cost

Previous work by IBM researchers

- Create an initial model of configuration complexity and demonstrates its value for a change management system.
- Metrics that indicate some configuration complexity, including execution complexity, parameter complexity, and memory complexity.
- [Brown et al, IM'05; HOTOS'05]

- Example: complexity of J2EE provisioning



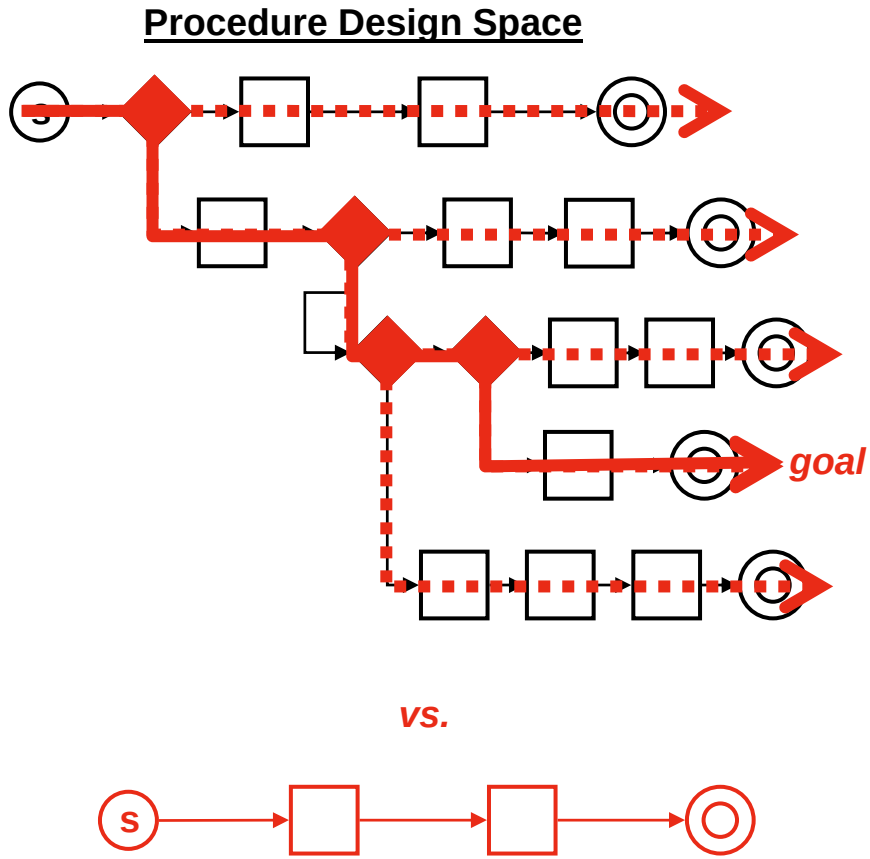
- Process complexity: ~~manual~~ **automated**
  - Execution
    - ~~59~~ **5** steps, ~~27~~ **1** context switches
  - Parameter
    - ~~32~~ **17** parameters used ~~61~~ **17** times, ~~18~~ **0** outside of source context
    - Source score: ~~125~~ **94**
  - Memory (LIFO stack model)
    - Size: max ~~8~~ **0**, avg ~~4.4~~ **0**

Outline

- Context and previous work
- **Motivation**
- Model
- User study
- Conclusion
- Future work

Next Step: Decision Complexity

- **Previous metrics assume expert skill**
  - Do not consider complexity arising from decision-making by non-experts
- **Capturing complexity impact of decisions along a specific procedure's path**
  - Parameterized by skill level
- **Understanding the overall complexity across all possible procedures**
- **Quantifying the tradeoff between flexibility and simplicity**



Decision Complexity

- **The complexity faced by a non-expert system administrator**
  - The person providing IT support in a small-business environment, who is confronted by decisions during the configuration process.
- **A measure of difficulty of identifying appropriate sequence of configuration actions to perform in order to achieve a specified configuration goal.**

Outline

- Context and previous work
- Motivation
- **Model**
- User study
- Conclusion
- Future work

Decision Complexity (An initial model & methodology)

- **Factors that affect complexity**
  - **constraints**

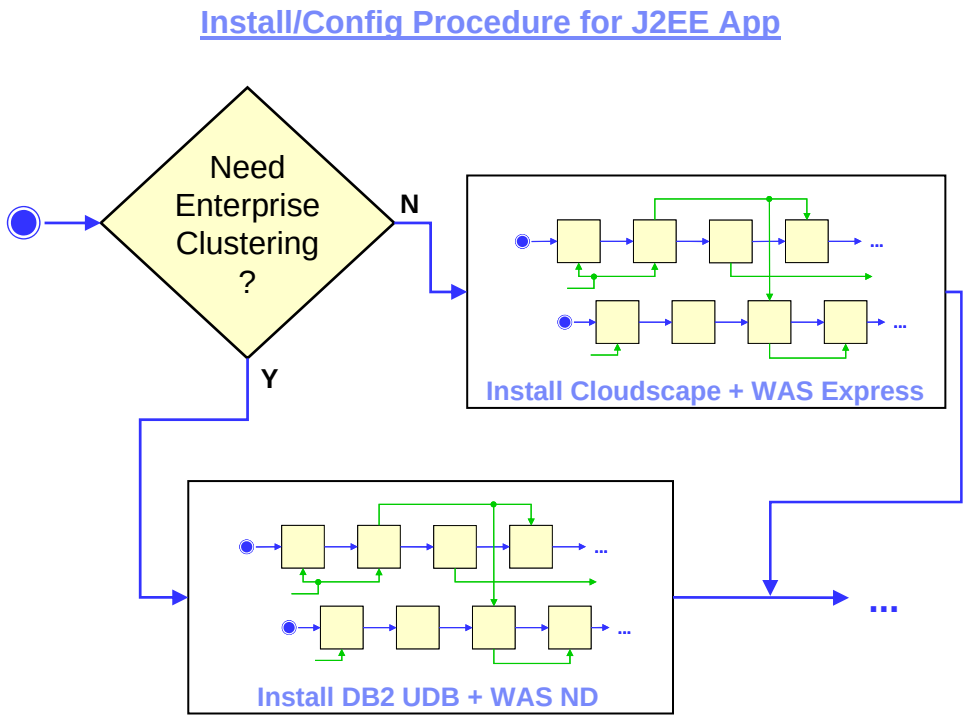
    e.g. compatibility between software products, capabilities of a machine

    consequences

    e.g. functionality, performance

    levels of guidance

    e.g. documentation, previous configuration experience
- **Manifestation**
  - task time, user-perceived difficulty, error probability
- **A starting point to drive data collection (user study)**
- **After we have the real-world data, refine the model**



Model details: levels of guidance

- **Global information**
  - E.g. documentation, design guide, deployment patterns
- **Short-term goal-oriented information**
  - E.g. wizard-based prompts indicating the appropriate next step
- **Confounding information**
  - E.g. alternate configuration instructions for a different platform than the target
- **Position information**
  - E.g. feedback on the current state of the system and the effect of the previous action

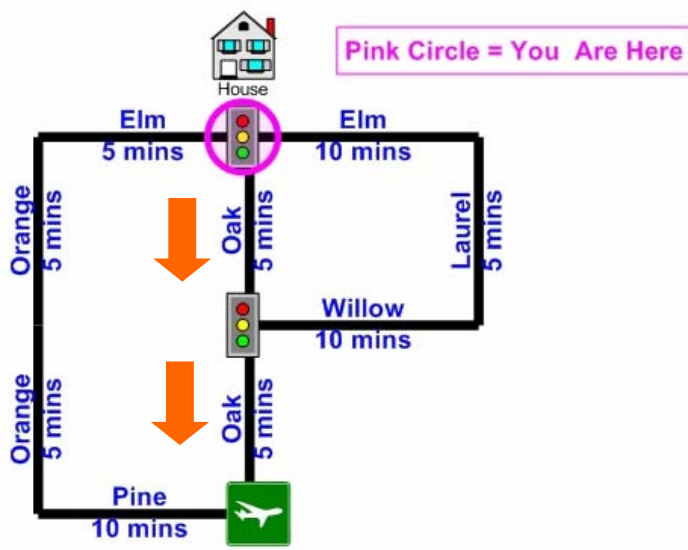
Decision Complexity (challenge & solution)

- Hard to conduct a full user study to validate the model (constraints, consequences, levels of guidance) using real IT processes with practicing system administrators
- First step: measuring decision complexity in a simplified domain:

  Route-planning

  – navigating a car from one point to another

Factors	Route planning domain
Constraints	Traffic
Guidance (Global info)	Map, Expert path
Guidance (Goal-oriented info)	GPS
Guidance (Position info)	Current position indicator
Consequence	Reach the destination or not



Analogy between route planning and general IT configuration domain

- **Traffic**
  - Static update
  - Dynamic update
    - Road close
    - Travel time update
- **Expert path**

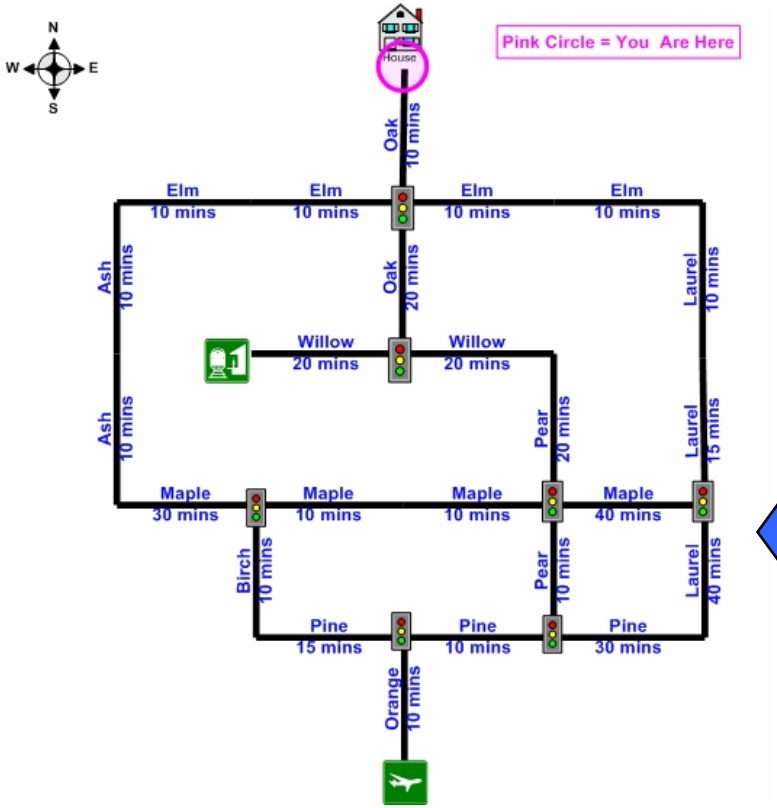
- **Constraints** (e.g. version compatibility)
  - Pre-specified
  - Unexpected
    - Constraints that eliminate the viability of one installation path
    - Constraints that change the resulting performance
- **Previous experience, or info in a "how-to" guide**

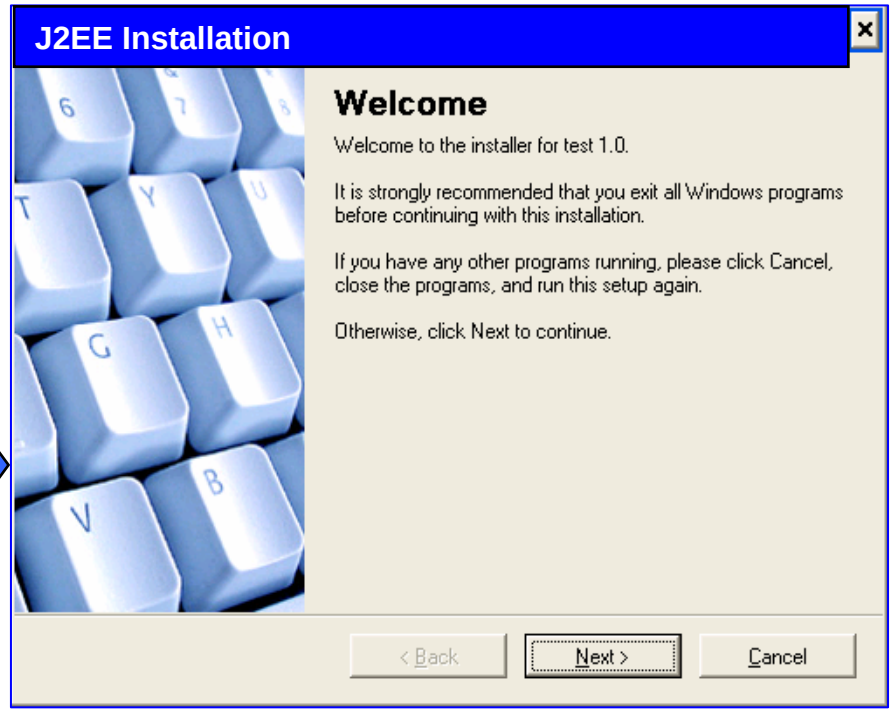
Analogy between route planning and general IT configuration domain

- **GPS**
- **Position indicator**
- **Path difference**

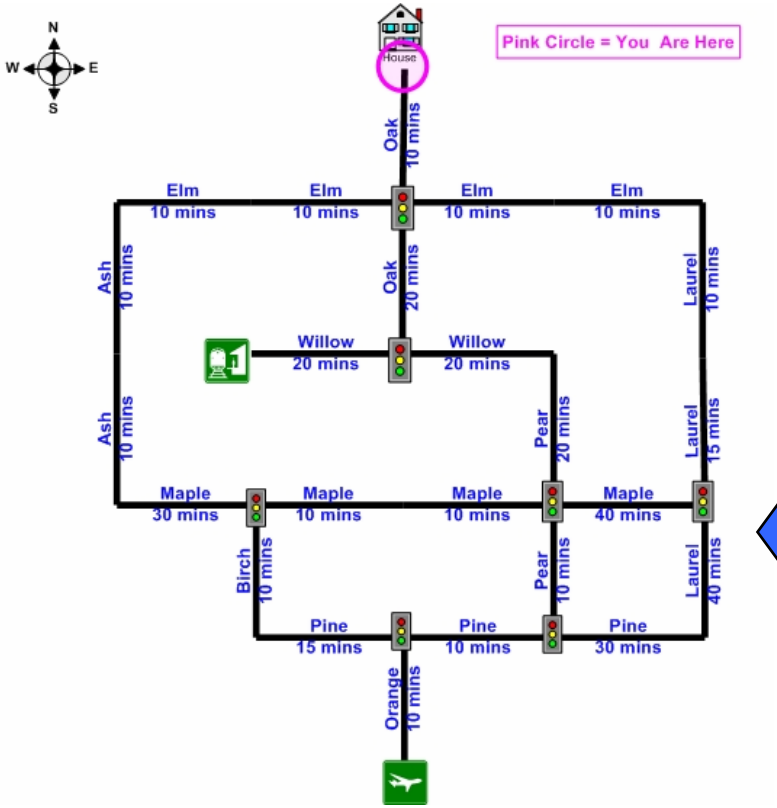
- **An omniscient expert**
- **Feedback info in IT context**
- **Different consequences resulted from configuration decisions**

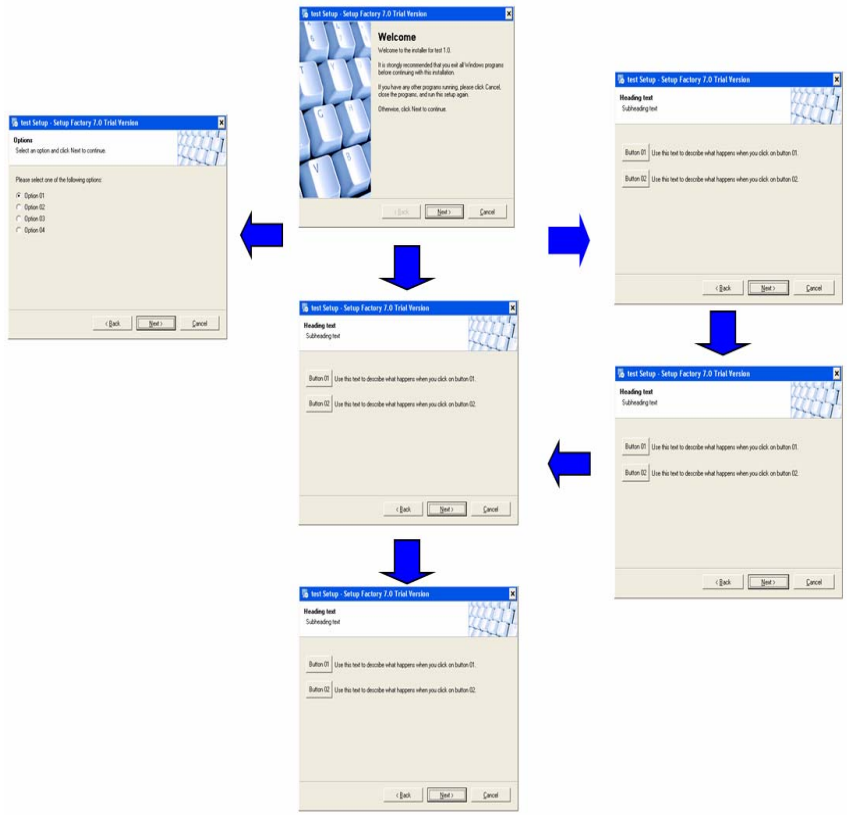
Analogy between route planning and specific IT installation process





Analogy between route planning and specific IT installation process



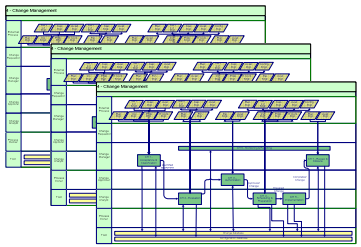


- Driving time per segment
- Global map
- Traffic
- Goal (reach the destination)

- Number of features achieved per step
- Flowchart of the overall process (text)
- Soft compatibility / machine capacity limit
- Achieve the max number of features

Decision Complexity Model

Existing IT conf Processes



Decision Complexity Model

(Constraints, Guidance, Consequence, …)

AvgTimePerStep

Rating (User perceived complexity)

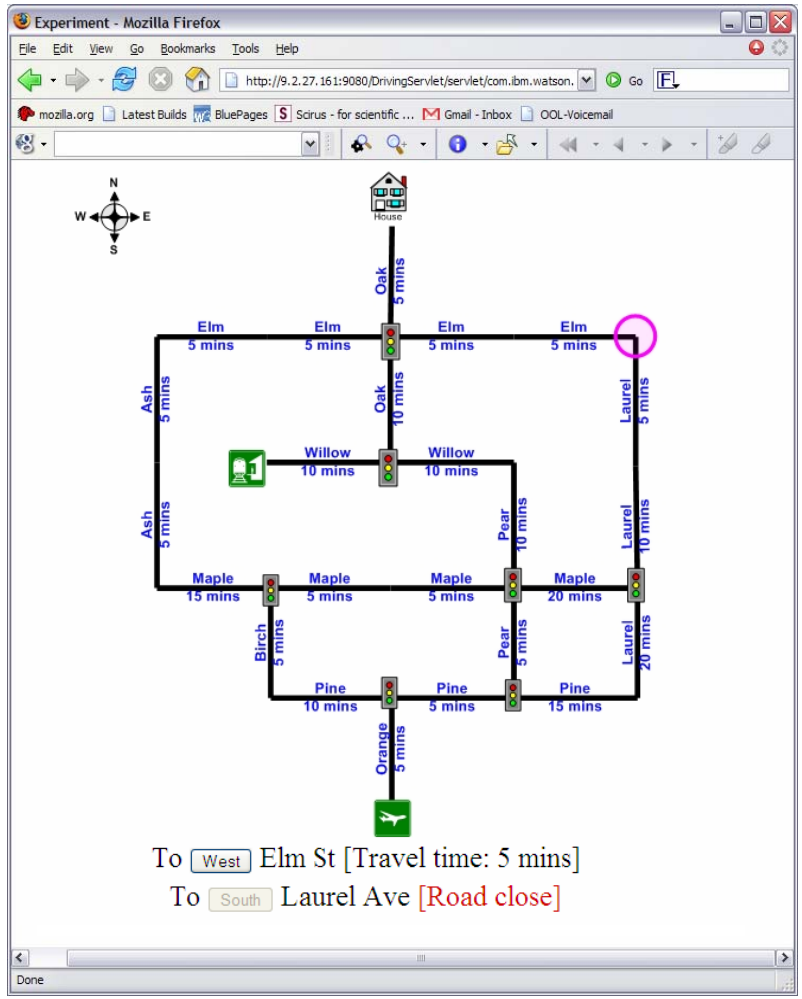
Error Rate

Outline

- Context and previous work
- Motivation
- Model
- **User study**
- Conclusion
- Future work

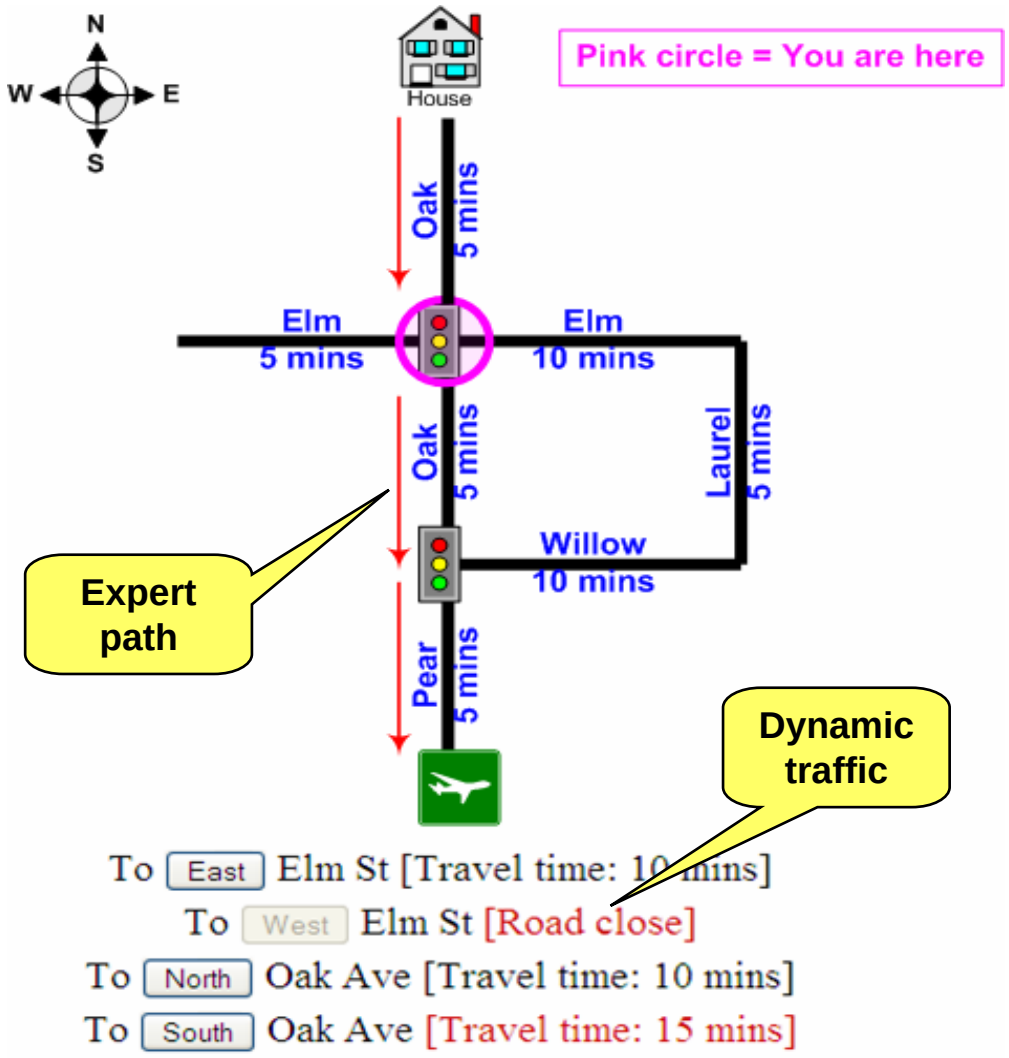
Decision Complexity (user study design)

- **Web-based study**
  - larger subject pool
  - accurate timing data
  - standardized information
- **Questionnaire to collect user background**
- **Recording user interaction**
  - time spent, each decision point
  - comparison b/w user path & optimal path
  - user ranking of the complexity for testcases

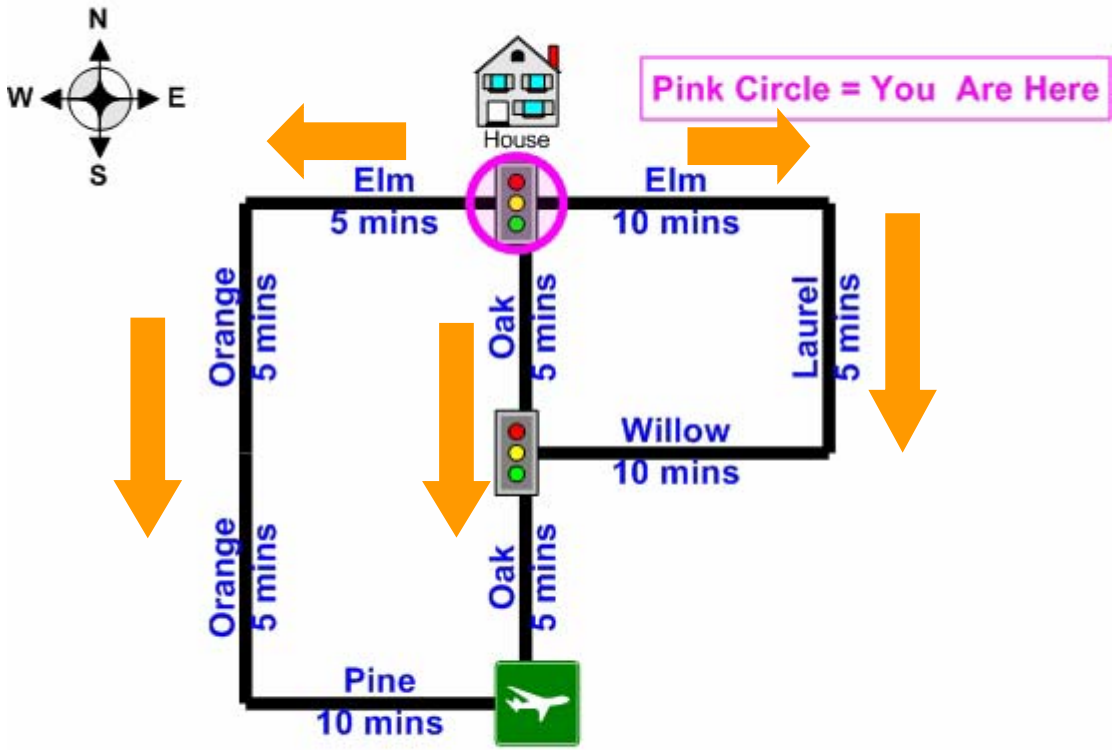


Testcase selection

- **Testcases**
  - Different combinations of factors
    - Static traffic
    - Dynamic traffic
    - Expert path
    - GPS
    - Difference in travel times
    - Position information
  - Selected 10 most relevant testcases
  - Example: dynamic traffic (road close, speed update) + expert path



Perspective of the user

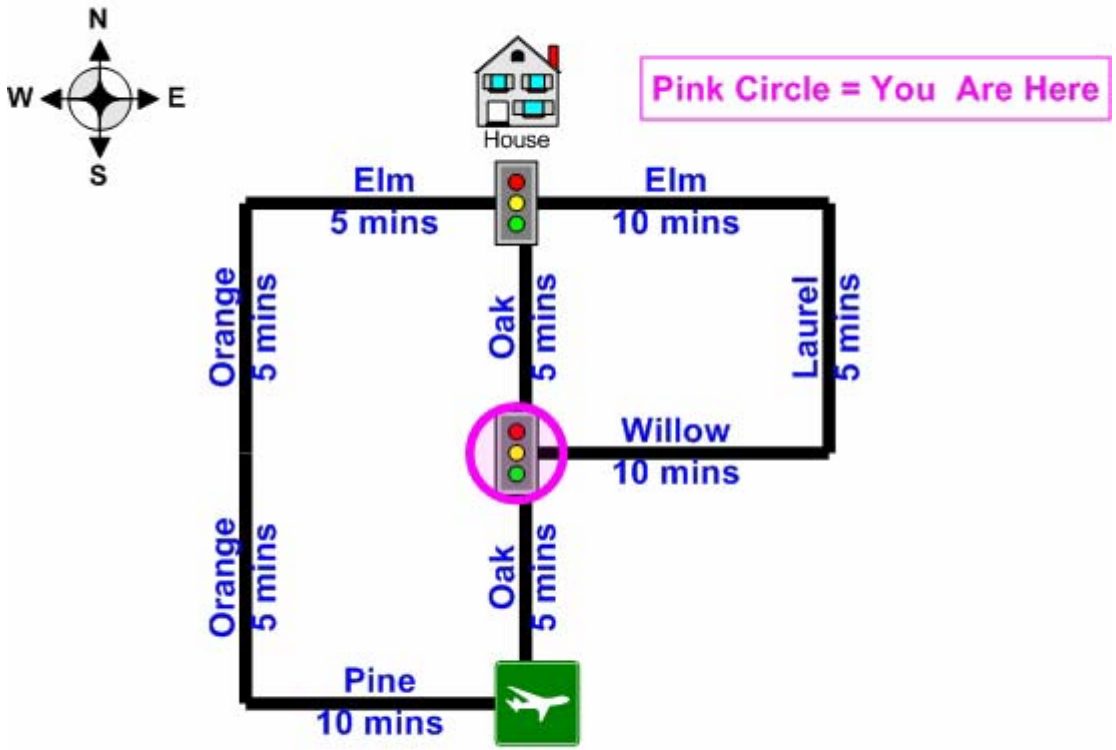


To **East** Elm St [Travel time: 10 mins]

To **West** Elm St [Travel time: 5 mins]

To **South** Oak Ave [Travel time: 5 mins]

Perspective of the user

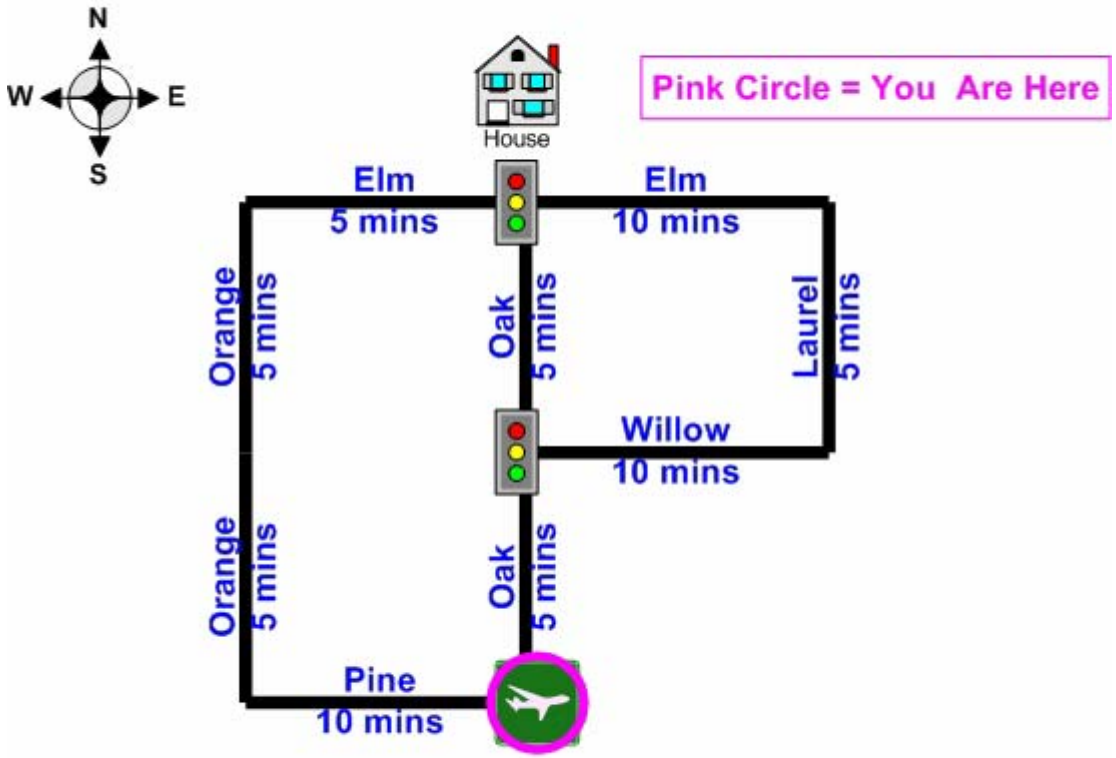


To **East** Willow St [Travel time: 10 mins]

To **North** Oak Ave [Travel time: 5 mins]

To **South** Oak Ave [Travel time: 5 mins]

Perspective of the user



Mission Complete !

Online user study overview

- **3 experiments (6 testcases each)**
  - an experiment randomly assigned after user logs in**.**
  - 10 different testcases with 1 warm-up
  - user ranks testcases difficulty on a scale of 1 (easiest) to 6 (most difficult)
- **1st stage, 35 users**
- *Experiment refined (only testcase order changed)*
- **2nd stage, 23 users**

Summary of testcases

No	Pos indicator	Traffic type	Update type	Path diff	Expert path	GPS
1	√	×	×		×	×
2	√	static	travel time		×	×
3	√	dynamic	road close		×	×
4	√	dynamic	travel time		×	×
5	√	×	×		√	×
6	√	dynamic	travel time		√	×
7	√	dynamic	travel time		×	√
8	√	dynamic	road close	bigger	×	×
9	√	dynamic	travel time	bigger	×	×
10	×	×	×		×	×

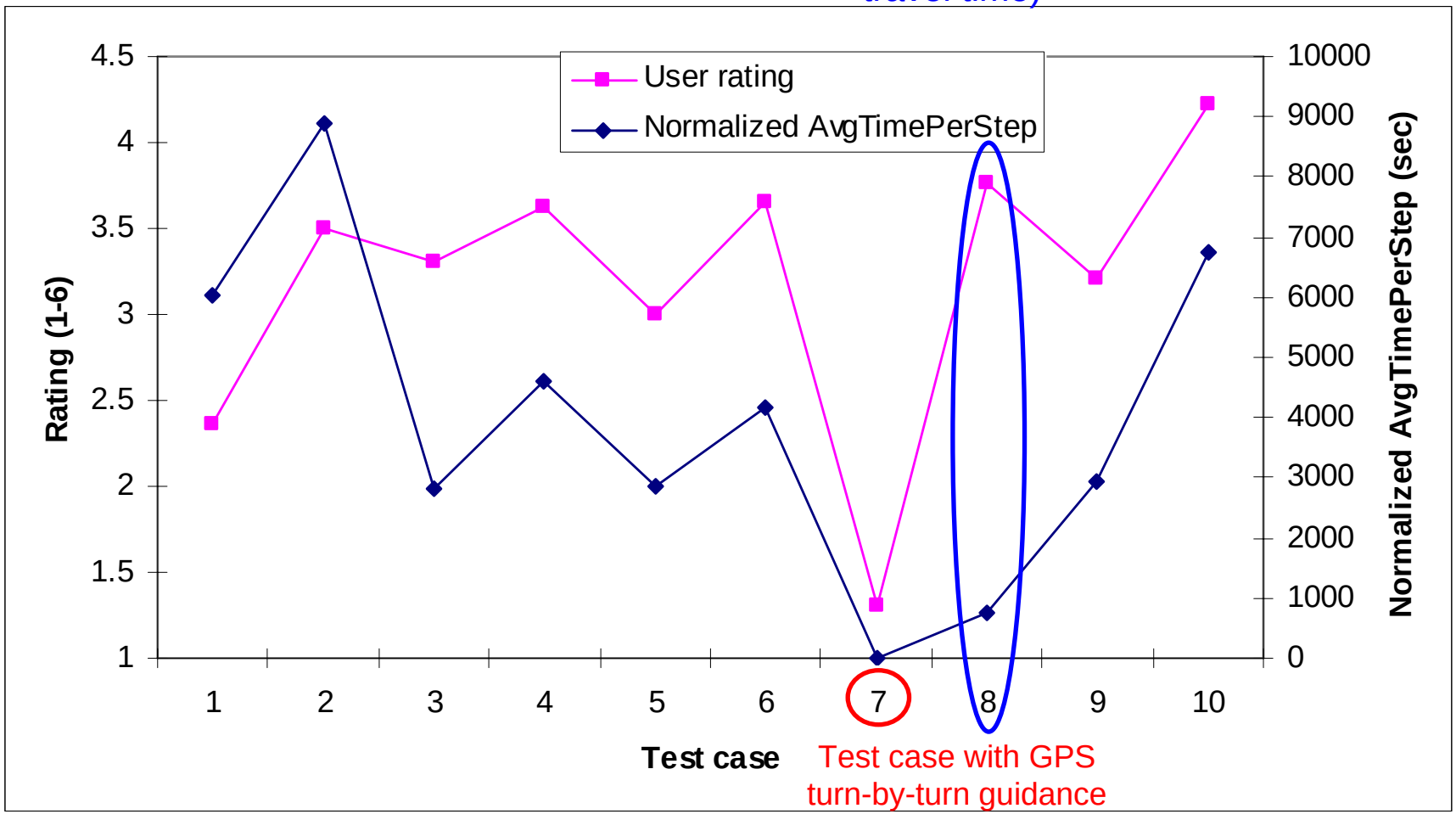
Note: dynamic traffic has two types - road close and travel time update

Metrics we focus on

- **Average time spent per step (e.g. time / no. of steps)**
- **User rating (in the end of each experiment)**
- **Error rate (user picked non-optimal path)**

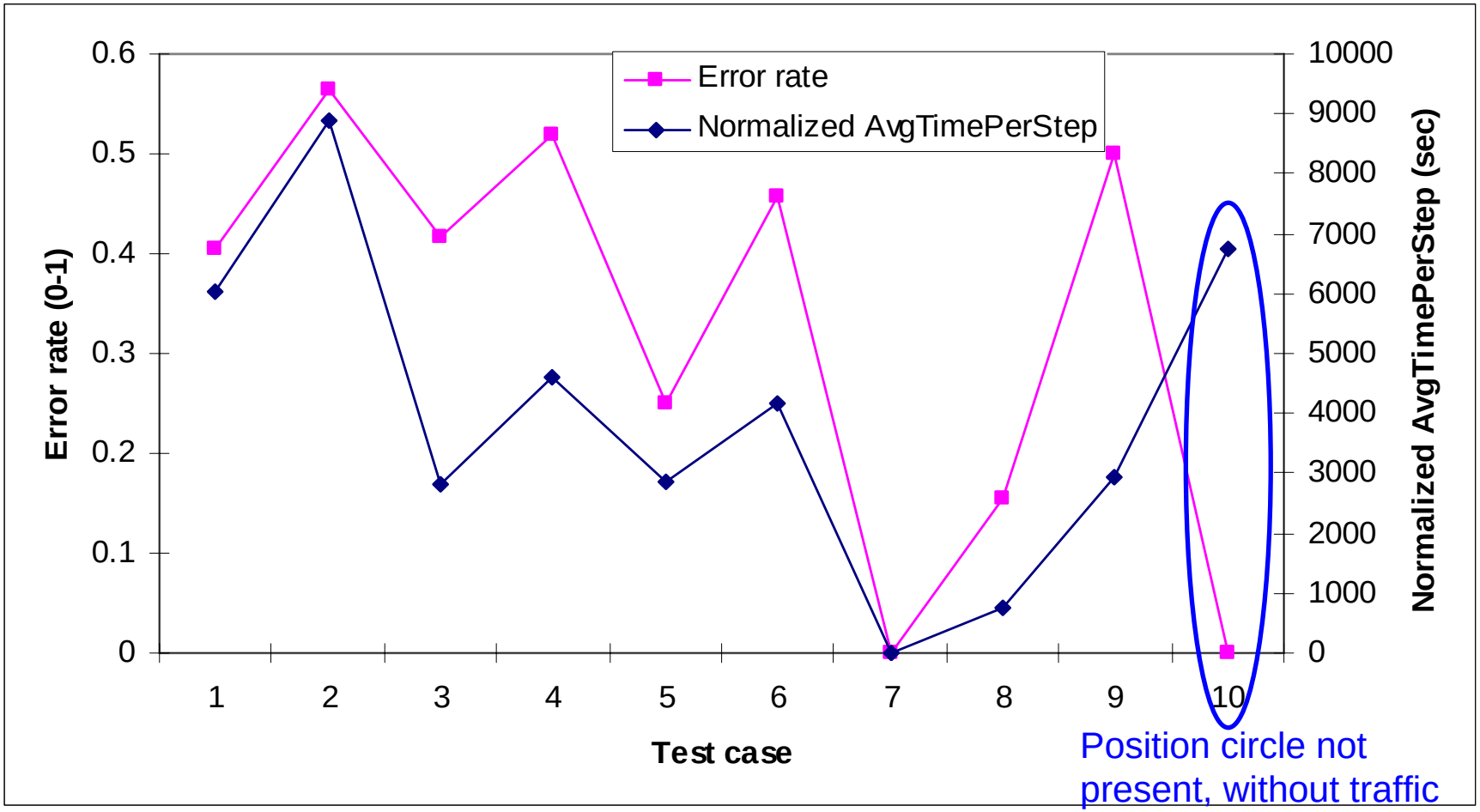
User rating & Time

Position indicator + Dynamic traffic (road close) + (bigger difference in travel time)



Avg Std for time over all testcases: **4368 sec**

Error rate & Time



Avg Std for time over all testcases: **4368 sec**

1st conclusion about decision complexity

- Lots of variance across users
- Confirm that decision complexity has different impacts on:
  - **User-perceived difficulty**
  - **Objective measures (time and error rate)**
- *With these assumption, further data analysis…*

Analysis approach

- **Step I: general statistical analysis of all data**
  - Each testcase measured as an independent data point
  - **Goal**: identify factors that explain the most variance
- **Step II: pair-wise testcase comparisons**
  - Get more insight into specific effects of factor value
  - **Goal**: remove inter-user variance

ANOVA: Time
(Analysis of Variance using linear-space regression model)

- **First: baseline analysis of variability**

Factor	Sum Squares
Testcase#	32.778
Residual	71.585

- ***Maximum variability that can be explained by complexity model factors: 32.778***

- **Next: analysis of complexity factors:**

Factors	Sum Squares
Constraints	16.764
Guidance (goal)	11.397

- **(16.764+ 11.397) / 32.778 = 86% of testcase variance is explained by Constraints & Guidance (short-term goal-oriented)**

Pair test: Time

	1st Study (95% CT)	2nd Study (95% CT)
Constraints	static traffic > dynamic traffic(road close) static traffic > without traffic	static traffic > dynamic traffic(travel time update) N/A
Guidance(goal)	without expert path > with expert path	N/A

ANOVA: User Rating
(Analysis of Variance using linear-space regression model)

- **First: baseline analysis of variability**

Factor	Sum Squares
Testcase#	51.671
Residual	79.653

- ***Maximum variability that can be explained by complexity factors: 51.671***

- **Next: analysis of complexity factors:**

Factors	Sum Squares
Guidance (goal)	42.272
Guidance (position)	5.668
Constraints	1.683

- **(42.272 + 5.668 + 1.683) / 51.671 = 96% of testcase variance is explained by Guidance (goal) & Guidance (position) & Constraints**

Pair test: Rating

	1st Study (95% CT)	2nd Study (95% CT)
Guidance(position)	without position indicator > with position indicator	without position indicator > with position indicator
Constraints	static traffic > dynamic traffic (road close)	N/A

Outline

- Context and previous work
- Motivation
- Model
- User study
- **Conclusion**
- Future work

Summary of results

- **Time is influenced by:**
  1. Constraints
     - static constraints > dynamic; static constraints > without constraints
  2. Guidance (short-term goal-oriented)
     - without guidance > with such guidance
- **Rating is influenced by:**
  1. Guidance (short-term goal-oriented)
  2. Guidance (position)
     - without guidance > with such guidance
  3. Constraints
     - static constraints > dynamic
- **Error rate: hard to say statistically, except**
  - error rate is reduced when guidance (short-term goal-oriented) is present
  - error rate is reduced when guidance (position) is not present

Summary of results (cont.)

- **Depending on its goal (user, time or error rate), optimization for less complexity will have different focus**

- ***Next, possible design approaches for reducing complexity…***

Advice to designers

- **An installation procedure with easily-located clear info (e.g. wizard-based prompts) for next step will**
  - reduce task time & user-perceived complexity,
  - not clear for error rate.

- **A procedure with feedback on current state of the system and effect of the previous action (e.g. message windows following a button press) will**
  - reduce perceived complexity,
  - unlikely to improve task time or error rate.

Advice to designers (cont.)

- **A procedure that auto adapts to diff software & hardware versions to reduce compatibility constraints will**
  - reduce task time,
  - possible small reduction in perceived complexity.

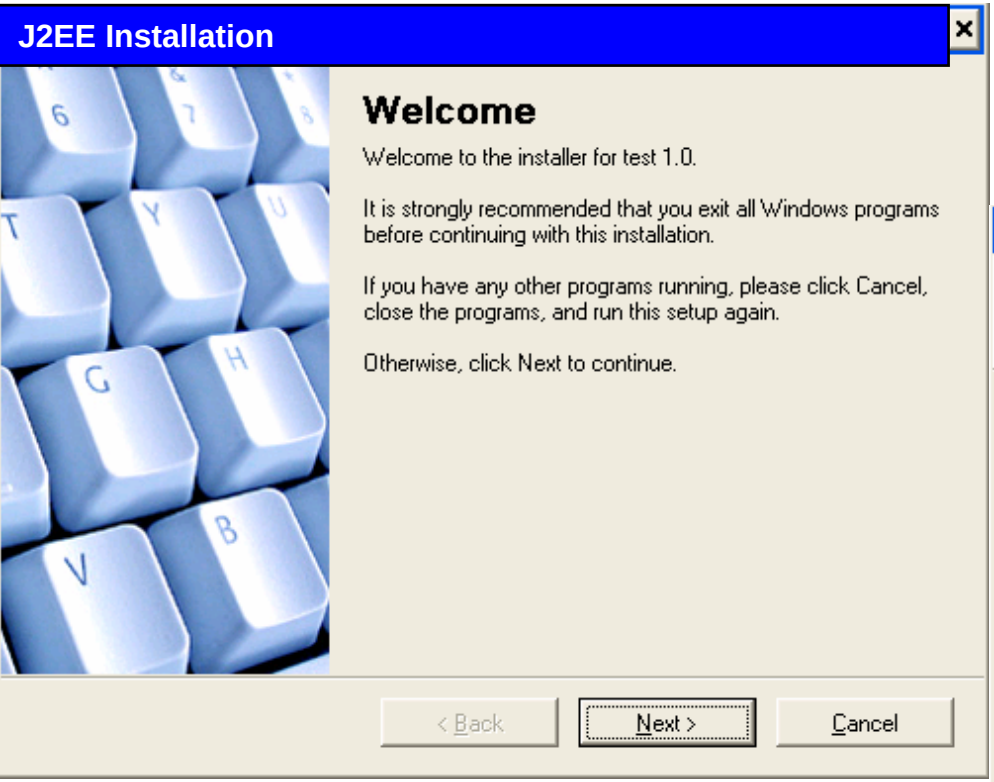
- **Omitting positional feedback (**i.e., by not showing users effects of their actions**) may, counter-intuitively,**
  - reduce error rate,
  - significantly increase perceived complexity & task time.

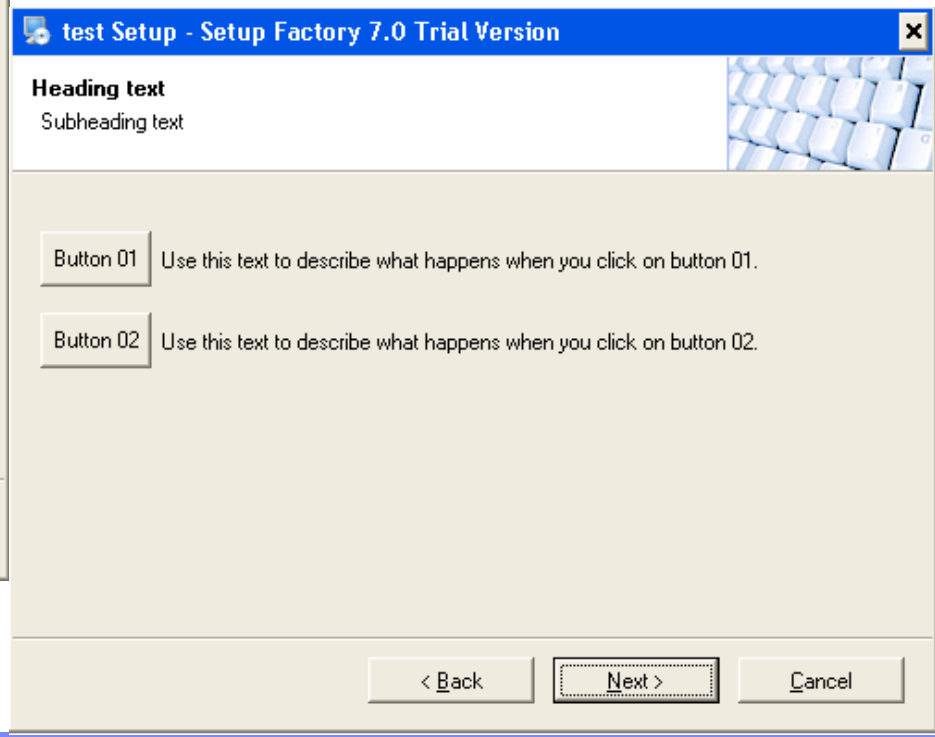
Outline

- Context and previous work
- Motivation
- Model
- User study
- Conclusion
- **Future work**

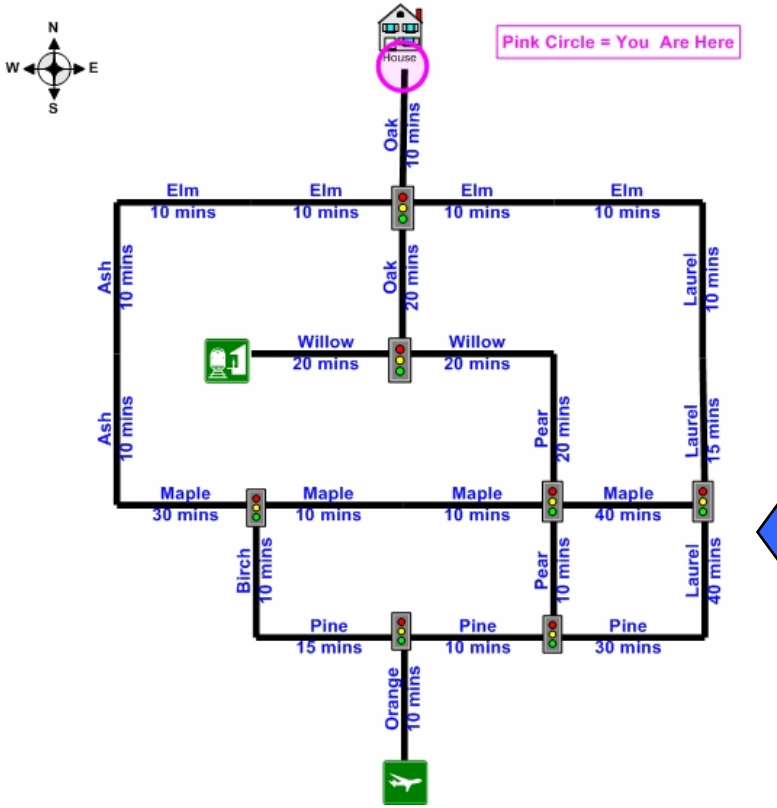
Proposal for a new user study

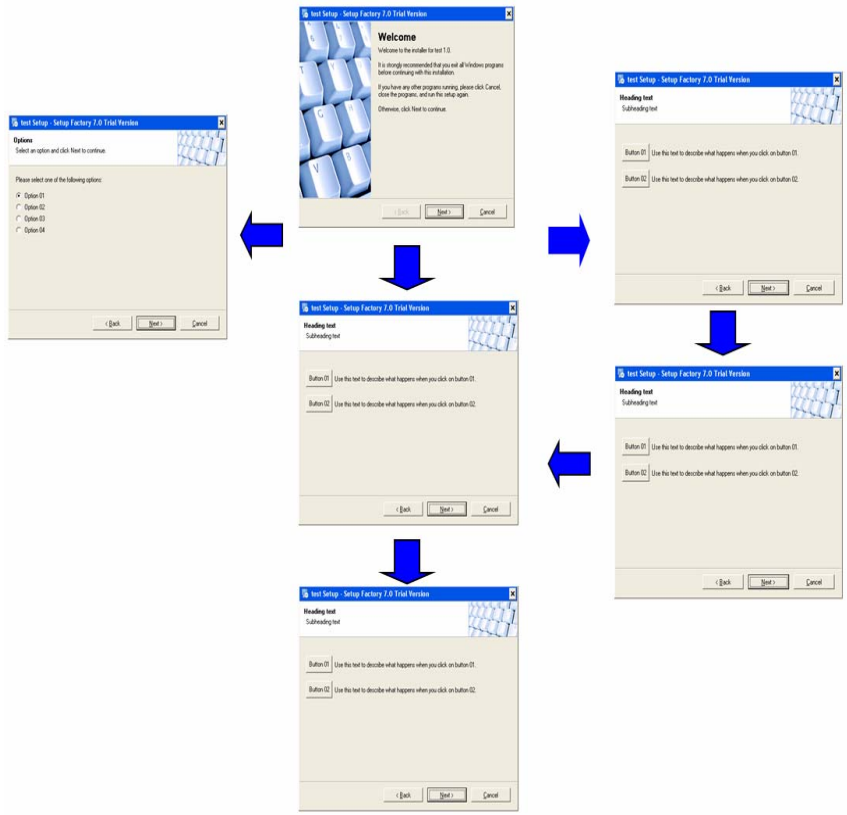
- **Validate the model in the IT configuration domain**





Analogy between two studies



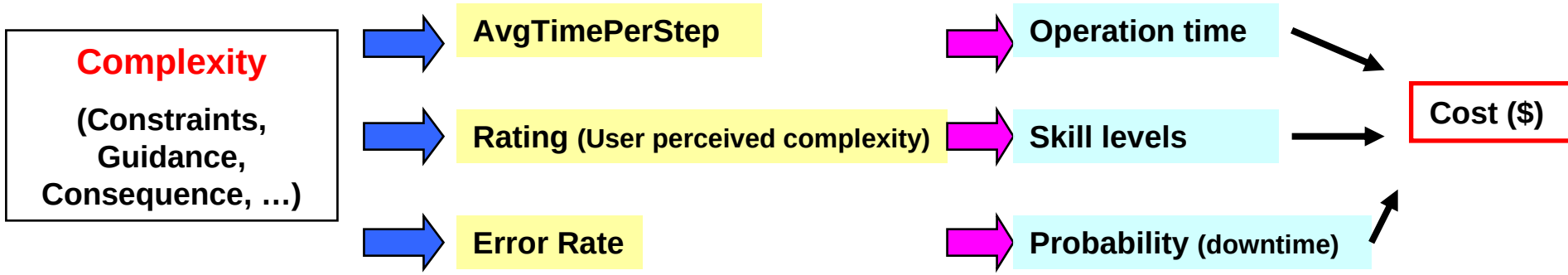


- Driving time per segment
- Global map
- Traffic
- Goal (reach the destination)

- Number of features achieved per step
- Flowchart of the overall process (text)
- Soft compatibility / machine capacity limit
- Achieve the max number of features

Further step

- **Apply the model to assess IT decision complexity**



Conclusions

- **We investigated decision complexity in IT configuration procedures**
  - Developed an initial model of decision complexity
  - Used an carefully-mapped analogous domain to explore complexity space
  - Conduct an extensive user study
  - Quantitative results showing the key factors
  - Next steps are to explore further in simulated IT environment

- **My PhD thesis topic: “Human-directed Adaptation”**
- **Thesis ideas: Using direct human input to solve optimization problems in adaptive and autonomic computing systems.**
- **Website: http://www.cs.northwestern.edu/~blin**

Thank you!