
Human-Centric Computing

Usability Evaluation

Heuristic Evaluation

Qualitative and Quantitative Methods

Lingyun Yu

Overview

- usability and usability evaluation
- heuristic evaluation techniques
- qualitative evaluation techniques
- quantitative evaluation techniques

Usability

- ***usability***: ease with which ***people*** (users) can use a particular tool or object to achieve a specific ***goal*** (task)
- aspects of usability:
 - ***learnability***: how easy to accomplish tasks the first time?
 - ***efficiency***: once learned, how quickly to complete tasks?
 - ***memorability***: how easy to reestablish proficiency after not having used a design for a period of time?
 - ***errors***: how many, how serious, how easy to recover?
 - ***satisfaction***: how pleasant to use the design?

Heuristic Evaluation

Heuristic Evaluation

- heuristics by Jakob Nielsen (1994) and others
- use of design principles/heuristics to inspect an interface for usability problems
- general approach:
take the interface and check for the interface guidelines/heuristics
- number of evaluations
 - single inspector
 - multiple inspectors

Single-Inspector Heuristic Evaluation

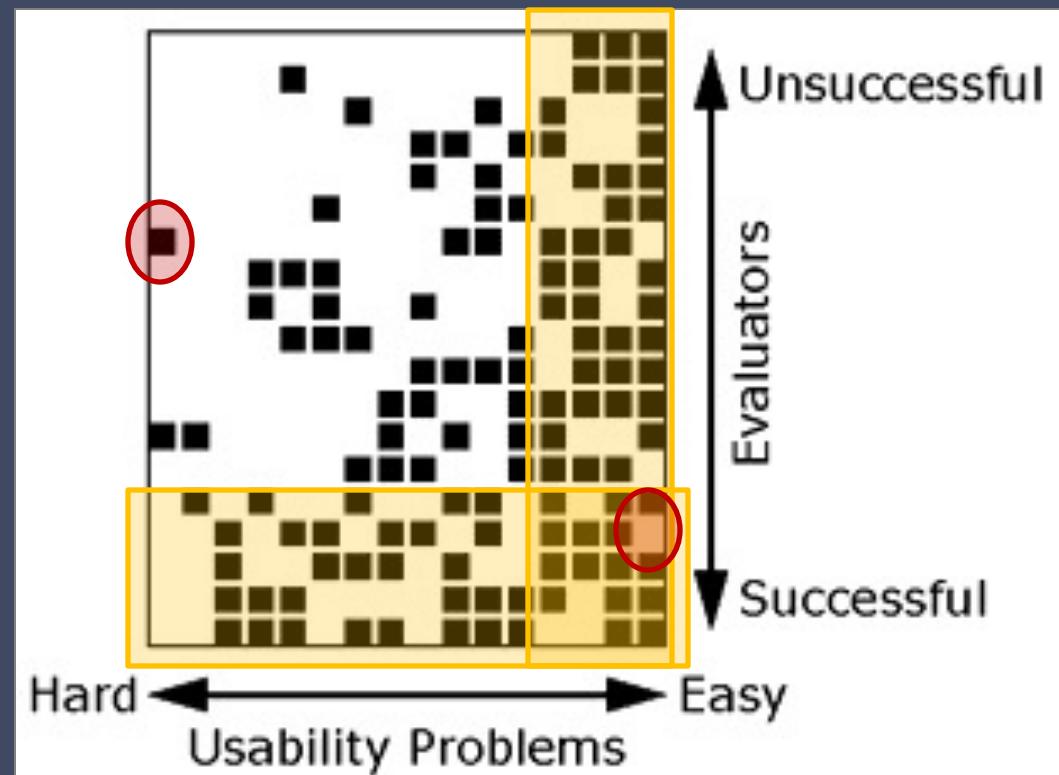
- example:
average over six case studies of heuristic evaluation
 - 35% of all usability problems found
- score not great, but finding some problems with one inspector is better than finding no problems with no evaluators ...

Single-Inspector Heuristic Evaluation

- results vary according to:
 - difficulty of the interface being evaluated
 - expertise of the inspectors
- average percentage of problems found:
 - 22% – novice evaluators (no usability experience)
 - 41% – regular specialists (expertise in usability)
 - 60% – double specialists (expertise in both usability and the particular type of interface being evaluated; also find domain-related problems)
- tradeoff:
novices yield poorer results, but are cheaper

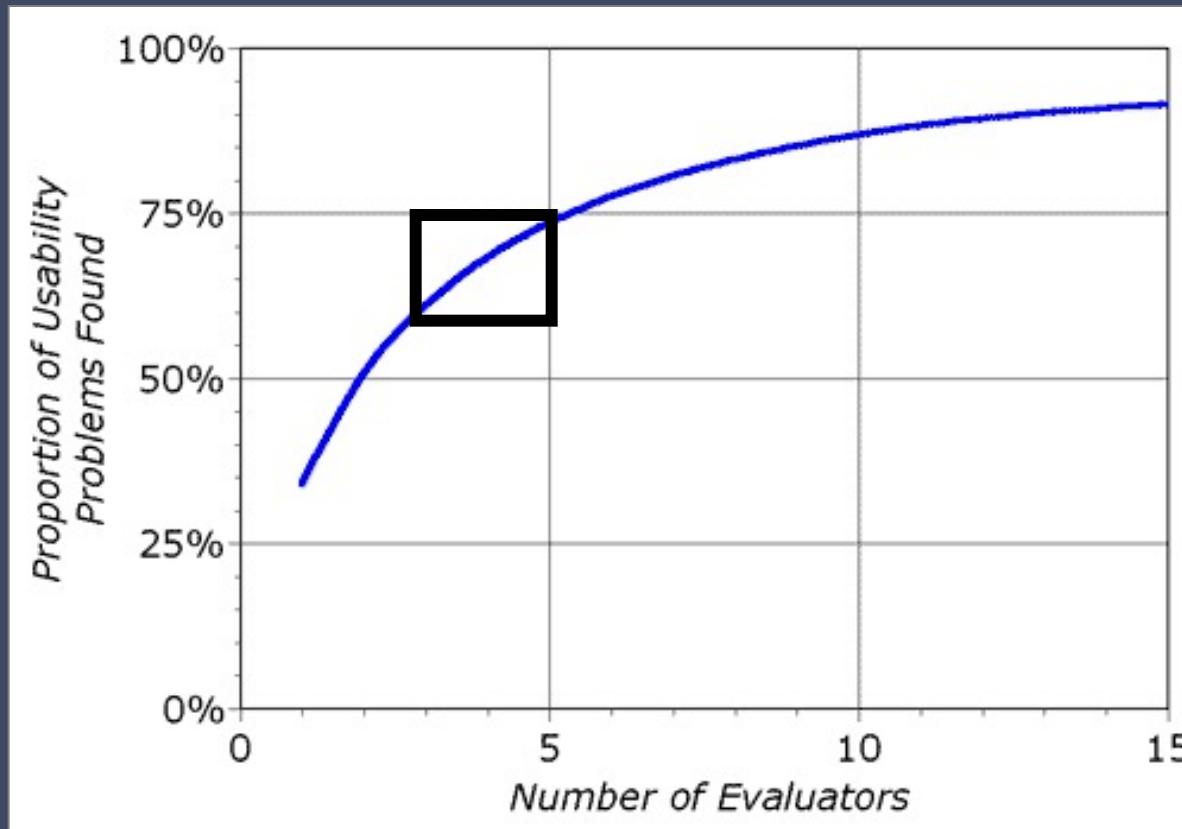
Single-Inspector Heuristic Evaluation

- different people will find different problems
- evaluators may miss both easy and hard problems
 - ‘best’ evaluators can miss easy problems
 - ‘worse’ evaluators can discover hard problems



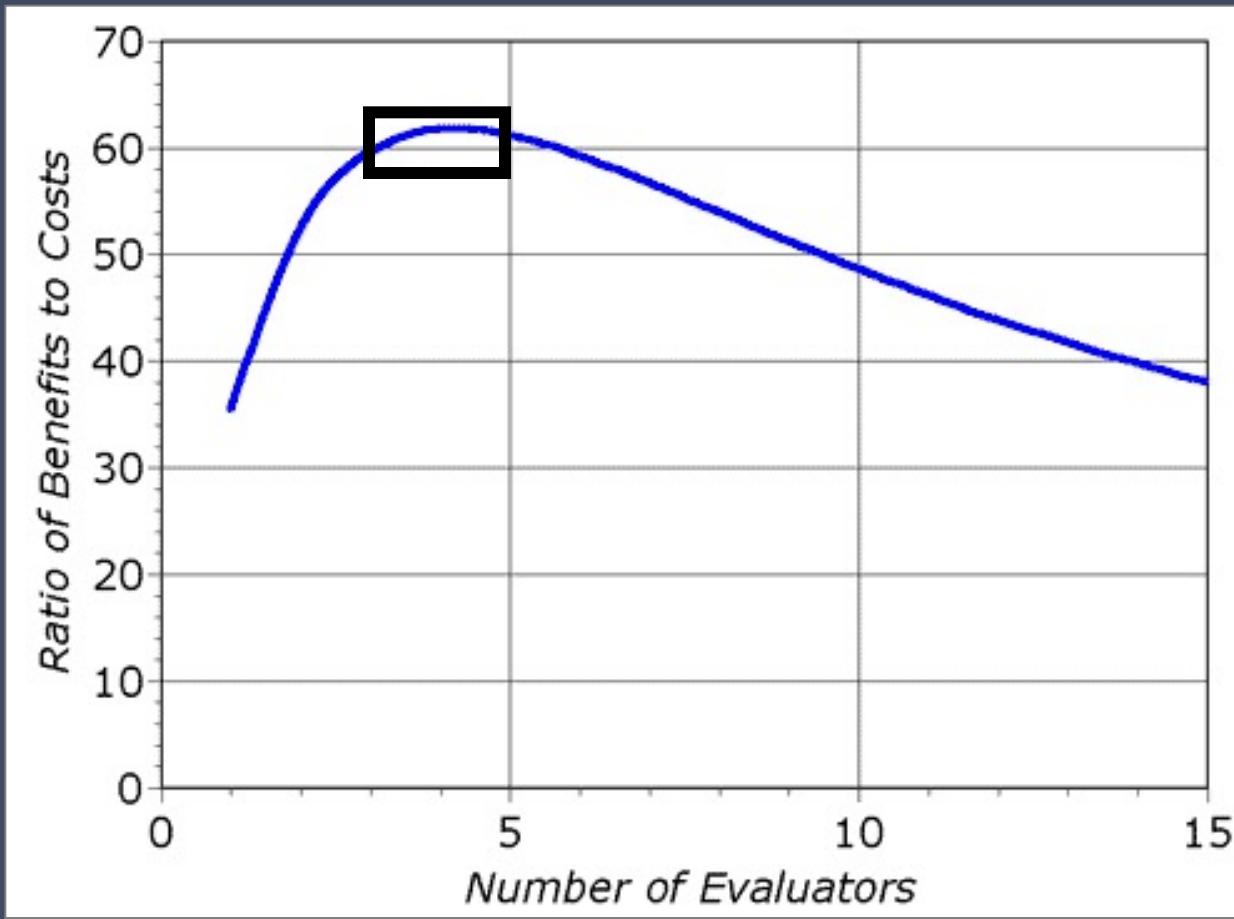
Multiple-Inspector Heuristic Evaluation

- 3–5 evaluators find 66%–75% of usability problems
 - different people find different usability problems
 - only modest overlap between the sets of problems found



Multiple-Inspector Heuristic Evaluation

- where is the best cost/benefit?
- depends on the costs, but:



Heuristic Evaluation: Individuals vs. Teams

- individual inspectors who look at an interface alone are recommended (according to Nielsen)
- reasons:
 - evaluation not influenced by others
 - independent and unbiased
 - greater variability in the kinds of errors found
 - no overhead required to organize group meetings
- problem: some interfaces require groups, then use several independent groups

Self-Guided vs. Scenario Exploration

- self-guided exploration
 - open-ended exploration
 - not necessarily task-directed
 - good for exploring diverse aspects of the interface
- scenario exploration
 - step through interface using a number of representative end user tasks (remember task-centered design)
 - ensures problems found in relevant parts of the interface
 - ensures that specific features of interest are evaluated
 - limits scope of evaluation – problems may be missed

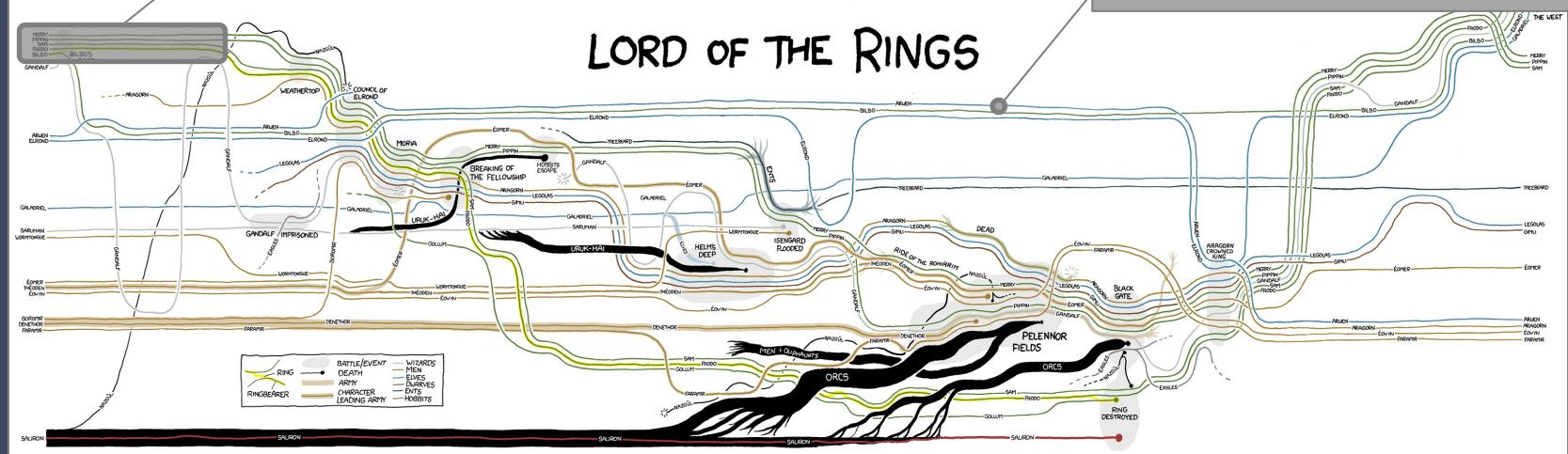
Procedure and results

- Go through the interface at least twice
 - Get a feel for the flow of the interaction and the general scope of the system
 - Focus on specific interface elements
- Output
 - A list of usability problems in the interface, annotated with references to those usability principles that were violated by the design.

iStoryline

The vertical group of lines indicate who are together.

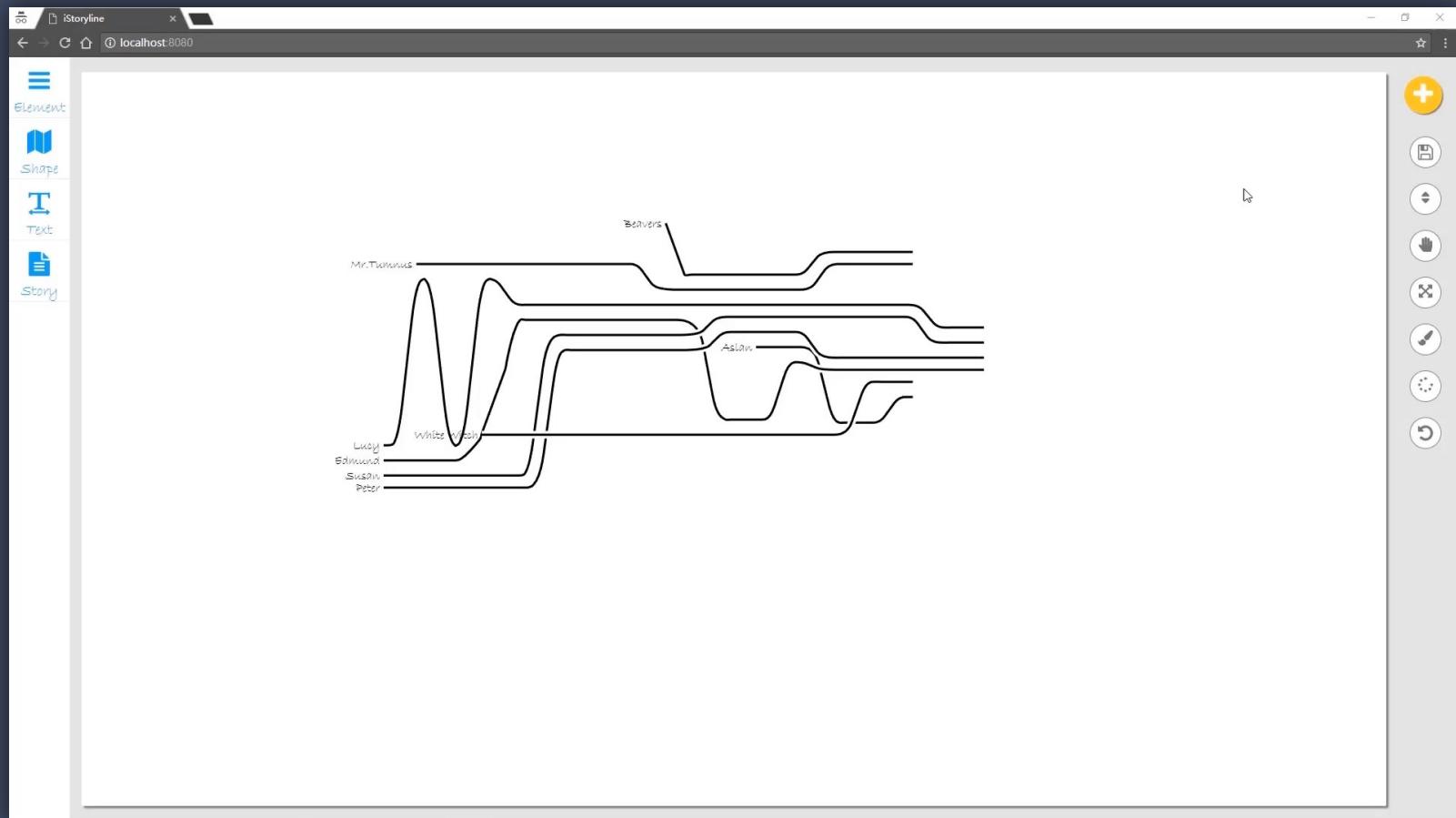
Each line represents one character.



The horizontal axis is time.

R. Munroe 2009

iStoryline: an authoring tool



Storyline Visualization

- Design Principles
 - D1 Lines in the same group should appear next to each other.
 - D2 Otherwise, lines must be far away from each other.
 - D3 A line must remain straight unless its group changes.
- General Design Principles
 - For the authoring tool

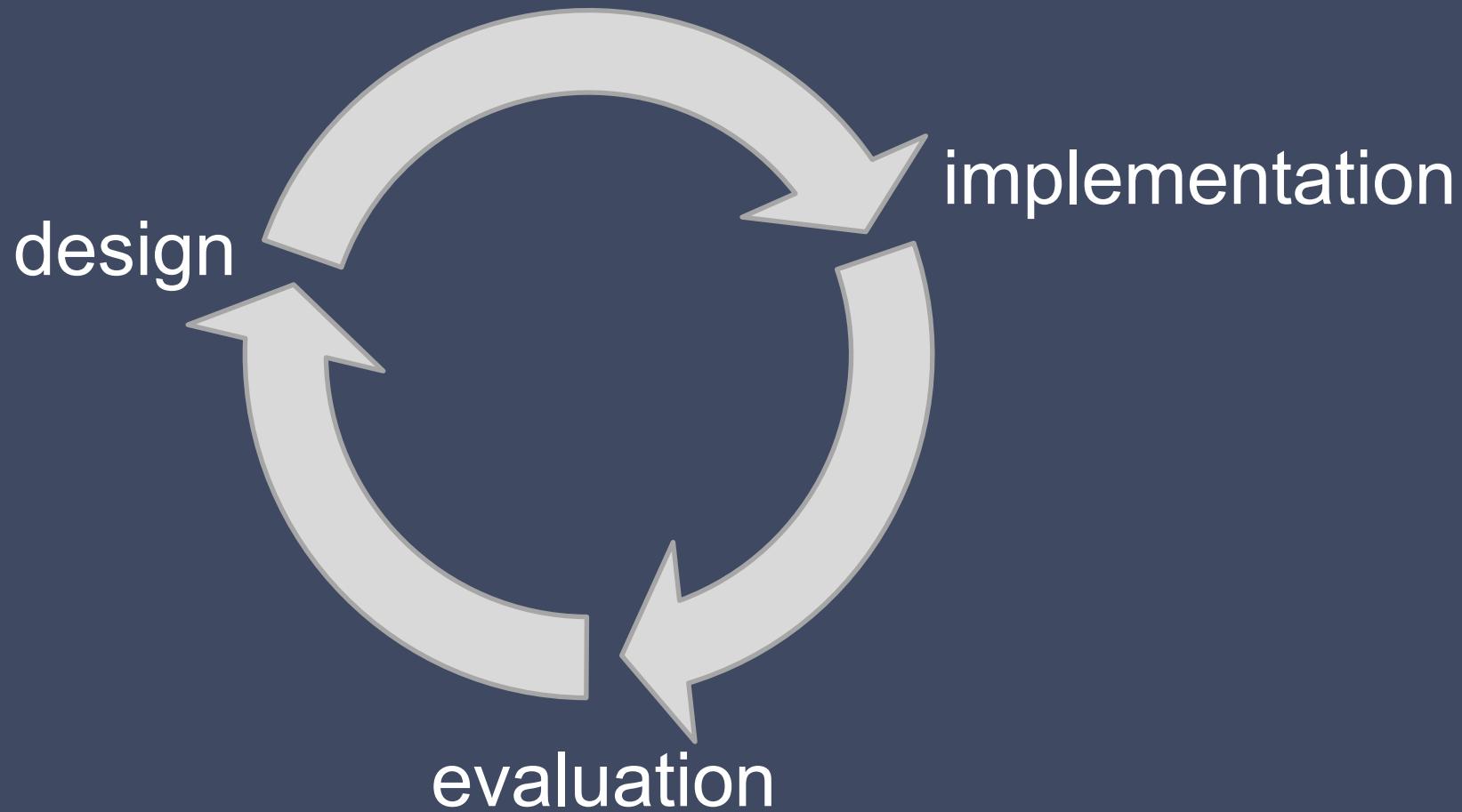
Qualitative and Quantitative Evaluation

Usability Evaluation with ‘Real’ People

- so far: evaluation only by developers or experts
- evaluation with ‘real’ people (users)
- why important
 - pre-design: investing in new expensive system requires proof of viability
 - initial design stages: develop and evaluate initial design ideas with end users
 - iterative design: requirements met? specific design problems? what solutions work?
 - acceptance testing: verify that system meets expected user performance criteria

Usability Evaluation with ‘Real’ People

- essential part of the design life cycle



Ethics

- Testing can be a distressing experience
 - pressure to perform, errors inevitable
 - feelings of inadequacy
 - competition with other subjects

Golden rule

- subjects should always be treated with respect

Ethics – before the test

- Don't waste the user's time
 - use pilot tests to debug experiments, questionnaires etc
 - have everything ready before the user shows up
- Make users feel comfortable
 - emphasize that it is the system that is being tested, not the user
 - acknowledge that the software may have problems
 - let users know they can stop at any time
- Maintain privacy
 - tell user that individual test results will be completely confidential
- Inform the user
 - explain any monitoring that is being used
 - answer all user's questions (but avoid bias)
- Only use volunteers
 - user must sign an informed consent form

Ethics form



Ethical Assessment of Undergraduate Final Year Projects Involving Human Research

FORM FYP-LRR For Use in Low Risk Research Projects

Department:	
Supervisor:	
Project Title:	

**LOW RISK RESEARCH (LRR) PROJECTS MUST TICK 'YES'
TO ALL THESE QUESTIONS.**



(Tick 'Yes' or 'No as appropriate)

Yes No

All research participants are from non-vulnerable groups such as children, people with learning or communication disabilities, people in custody or people engaged in illegal activities.

All research participants are able to provide consent.

All participants will be recruited in China.

All participants take part in the study with their knowledge.

The study does not involve deliberately misleading the participants.

The study does not require discussion of sensitive topics that may cause distress or embarrassment to the participant. Sensitive topics may include death, significant illness, sexuality or religion.

The study does not require disclosure of criminal activity or child protection issues.

The study does not require the administration of drugs, or other substances to participants.

Supervisor's Declaration:

- The information in this form is accurate to the best of my knowledge and belief, and I take full responsibility for it.
- I have read and understand XJTLU's 'General Guidelines on Research' and 'Procedures for the Ethical Assessment of Undergraduate Final Year Projects Involving Human Research'.
- I understand that I am responsible for monitoring the research at all times.
- In the event of serious adverse events or breaches of protocol, I understand that I am responsible for immediately stopping the research and alerting the Research Ethics Committee within 24 hours of the occurrence via ethics@xjtlu.edu.cn.
- I am aware of my responsibility to comply with the requirements of the law applicable to the research and relevant guidelines relating to security and confidentiality of personal data.
- I understand that research records and data may be subject to inspection for audit purposes.
- I understand that all conditions apply to students and personnel involved in the research, and that it is my responsibility to ensure that they abide by them.

Signature of Supervisor:

Date:

SUBMIT TO DEPARTMENT FYP COORDINATOR.

Example of consent form

INFORMED CONSENT FORM
知情同意书

/

*Title of Research Project:
项目名称:

*Researcher(s):
研究人员 :

Please initial box

1. I confirm that I have read and have understood the information sheet dated [DATE] for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.
本人确认已于[日期]阅读并了解了该项目相关研究信息，并已从项目负责人处得到考虑、提问的机会，且得到满意答复。

2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without my rights being affected.
本人知晓对该项目的参与为自愿，且可以随时退出，无需任何理由，同时权利不会受任何影响。

3. I understand that I can at any time ask for access to the information I provide and I can also request the destruction of that information if I wish.
本人知晓可随时要求获取或销毁所提供的个人信息。

4. I agree to take part in the above study.
本人同意参加此项研究。

Participant Name _____ Date _____ Signature _____
参与者 日期 签名

Name of Person taking consent _____ Date _____ Signature _____
知情同意书提供者 日期 签名

Researcher _____ Date _____ Signature _____
研究人员 日期 签名

*The contact details of lead Researcher (Principal Investigator) are: Lingyun Yu
项目负责人的联系方式如下：

SD463, Dept. of Computing, Xi'an Jiaotong-Liverpool University 111, Ben'ai Road, SIP,
Suzhou, Jiangsu, China, 215123
中国江苏省苏州市工业园区独墅湖科教创新区仁爱路 111 号;
西交利物浦大学;理科楼 D 区 (SD) 计算机科学与软件工程系;邮编: 215123
Email/电子邮件: Lingyun.Yu@xjtu.edu.cn Phone/联系电话: +86 (0) 512 8816 1508

V1 for AY 2020/2021 1 for subject, 1 for researcher 1

Ethics – during the test

- Don't waste the user's time
 - never have the user perform unnecessary tasks
- Make users comfortable
 - try to give user an early success experience
 - keep a relaxed atmosphere in the room
 - coffee, breaks, etc
 - hand out test tasks one at a time
 - never indicate displeasure with the user's performance
 - avoid disruptions
 - stop the test if it becomes too unpleasant
- Maintain privacy
 - do not allow the user's management to observe the test

Ethics – after the test

- Make the users feel comfortable
 - state that the user has helped you find areas of improvement
- Inform the user
 - answer particular questions about the experiment that could have biased the results before
- Maintain privacy
 - never report results in a way that individual users can be identified
 - only show videotapes outside the research group with the user's permission

Qualitative vs. Quantitative Evaluation

- can test numeric or non-numeric criteria
- non-numeric (qualitative) criteria
 - e.g., what techniques do people employ to reach a goal?
 - e.g., what do people like or dislike about an interface?
 - subjective opinions of people about interfaces
- numeric (quantitative) criteria
 - e.g., how fast can someone achieve a goal?
 - e.g., how precise are people in their interactions?
 - e.g., how many mistakes do people make?
 - quantitative comparison of techniques
 - statistical evaluation: significance

Qualitative Evaluation Techniques

Qualitative Evaluation Approaches

- naturalistic approaches:
observe in realistic settings
 - pro: real-life situations
 - con: hard to arrange and do, time-consuming, may not generalize easily
- usability engineering approaches:
observe in simulated settings
 - pro: situation can be better controlled, easier and cheaper to arrange and do, less time-consuming
 - con: changed context for interface use

Qualitative Evaluation Techniques

- *direct observations*: observe people while they are using a system
- *interviews*: investigate specific issues
- *continuous evaluations*: monitor system in use

Direct Observations

- evaluator observes people interacting with a system
 - in the lab: completing a set of pre-determined tasks
 - in the field: going through normal tasks/duties
- value:
 - excellent at identifying major design/interface problems
 - validity depends on how controlled/arranged a situation is
- simple technique:
 - just watch the user
 - problem:
no insight into a
person's decision
processes or attitude



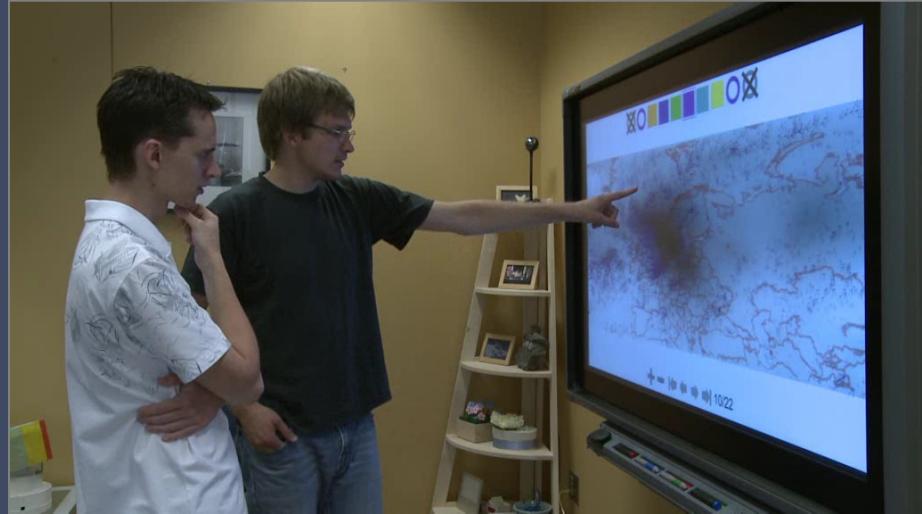
Direct Observations: Think Aloud Method

- person is asked to speak out their thoughts while doing a task, e.g.:
 - what they are trying to do, why they took an action
 - how they interpret what the system did
- gives insight into what the person is thinking
- most widely used technique, but issues:
 - may alter how people perform tasks
 - unnatural (awkward & uncomfortable)
 - hard to talk while concentrating



Direct Observation: Constructive Interaction

- observe two people working together
 - monitor their normal conversations
 - removes awkwardness of think-aloud
- different version: co-discovery learning
 - use semi-knowledgeable ‘coach’ and ‘novice’
 - only novice uses the interface
 - novice to ask questions
 - coach responds



How to Record Observations for Analysis

- paper and pencil
 - primitive but cheap method
 - record events, comments, interpretations
 - hard to get detail (writing is slow)
 - 2nd observer helps
- audio recording
 - good for recording think-aloud
 - hard to tie to on-screen user interactions
- video recording
 - can see and hear what people are doing
 - initially intrusive
 - when several aspects are essential: 2 cameras, mirror



Interviews and Questionnaires

- *interviews* are good for pursuing specific issues:
 - vary question to suit the context
 - probe more deeply on interesting issues as they arise
 - good for exploratory studies via open-ended questioning
 - often leads to specific constructive suggestions
- problems:
 - accounts/answers are subjective
 - interviews are time-consuming
 - evaluator can easily bias the interview

Interviews and Questionnaires

- plan a set of central questions
 - a few good questions to get things started
 - avoid leading/suggestive questions (bias)
 - focus on the interview
 - could be based on results of user observations
- semi-structured interview
 - let user responses lead to follow-up questions
 - follow interesting discussions

Interviews and Questionnaires

- *questionnaires*: ask questions on paper
 - “expensive” to prepare, but cheap to administer
 - can reach wider group (e.g., e-mail, online forms)
 - even generic online questionnaire tools available now
 - some results can be quantified
 - questions have to be good, otherwise of little use
- preparation
 - what information do you want, how to analyze, use of results
 - only ask questions whose answers will be used
 - specify target audience
 - decide on the delivery and data collection (e-mail, online forms, paper forms, etc.)

Interviews and Questionnaires

- gathering numeric values
 - people's degree of agreement to statements
 - people's degree of liking or disliking certain aspects
 - people's amount of previous experience with technique
- approach
 - give ranges, assign numbers to possible answers
 - always give explanations what ranges/numbers mean
- for judgments (agreement/liking) need **Likert** scale
 - *odd number* of possible choices
 - typically 5 or 7 options, e.g. (--) (-) (0) (+) (++)
 - reason: to offer a neutral choice

Questions

- extract opinions/concepts from people without biasing them
- bias already by wording of questions:
 - What do you like about this interface?
 - Do you think this interface lets you perform efficiently?
 - Why do you think this interface is bad?
- question-based interviews will always bias people

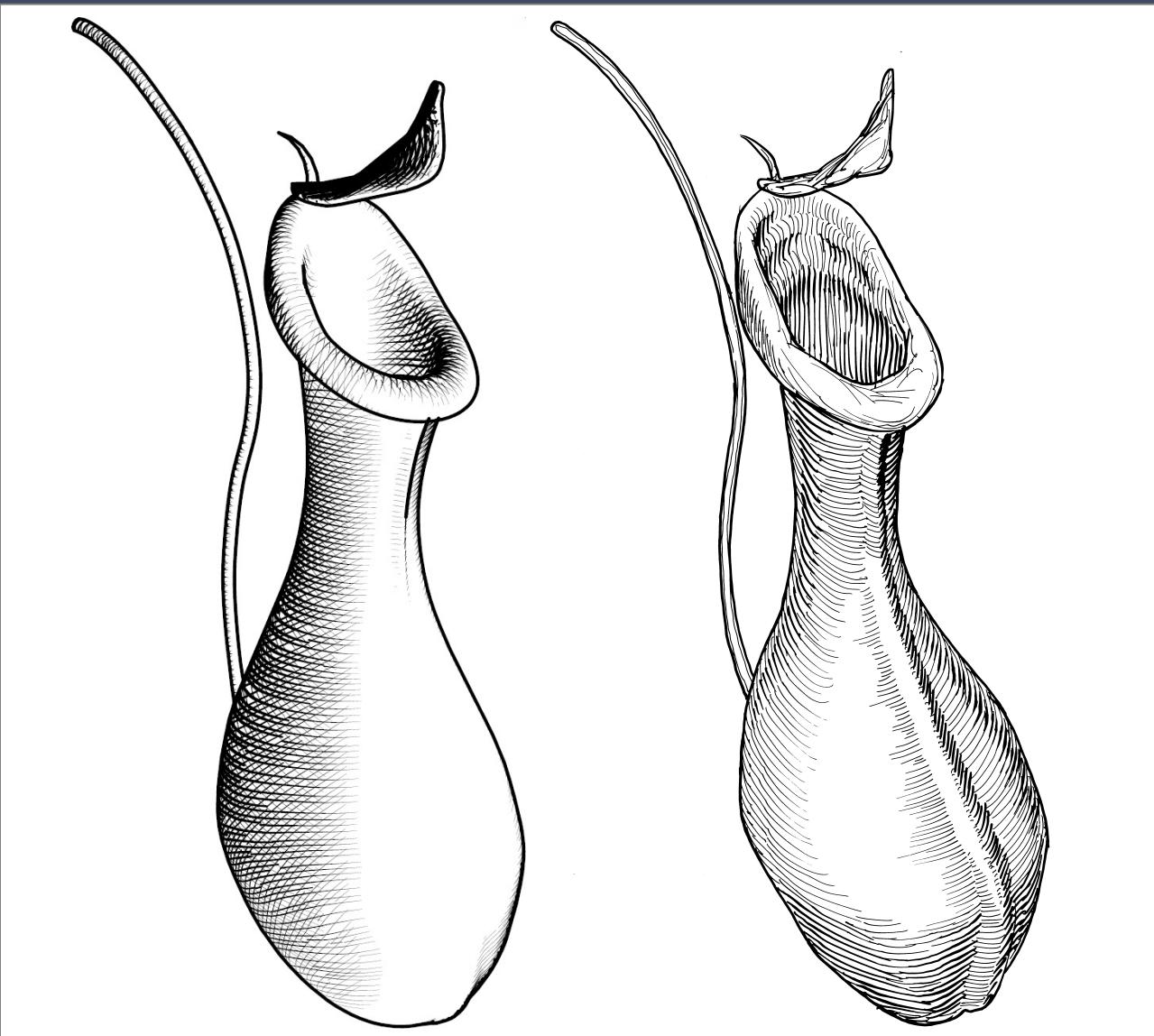
Continuous Evaluation

- monitor systems in actual use
 - usually later stage of development
 - fix problems in next release
- user feedback
 - provide means of reporting feedback: help desks, bulletin boards, e-mail, etc.
 - combine with trouble-shooting facilities
- case/field studies
 - careful study of “system usage” on-site
 - seeing system in “real-life” use
 - monitor through external observers and site visits

Qualitative Evaluation Techniques

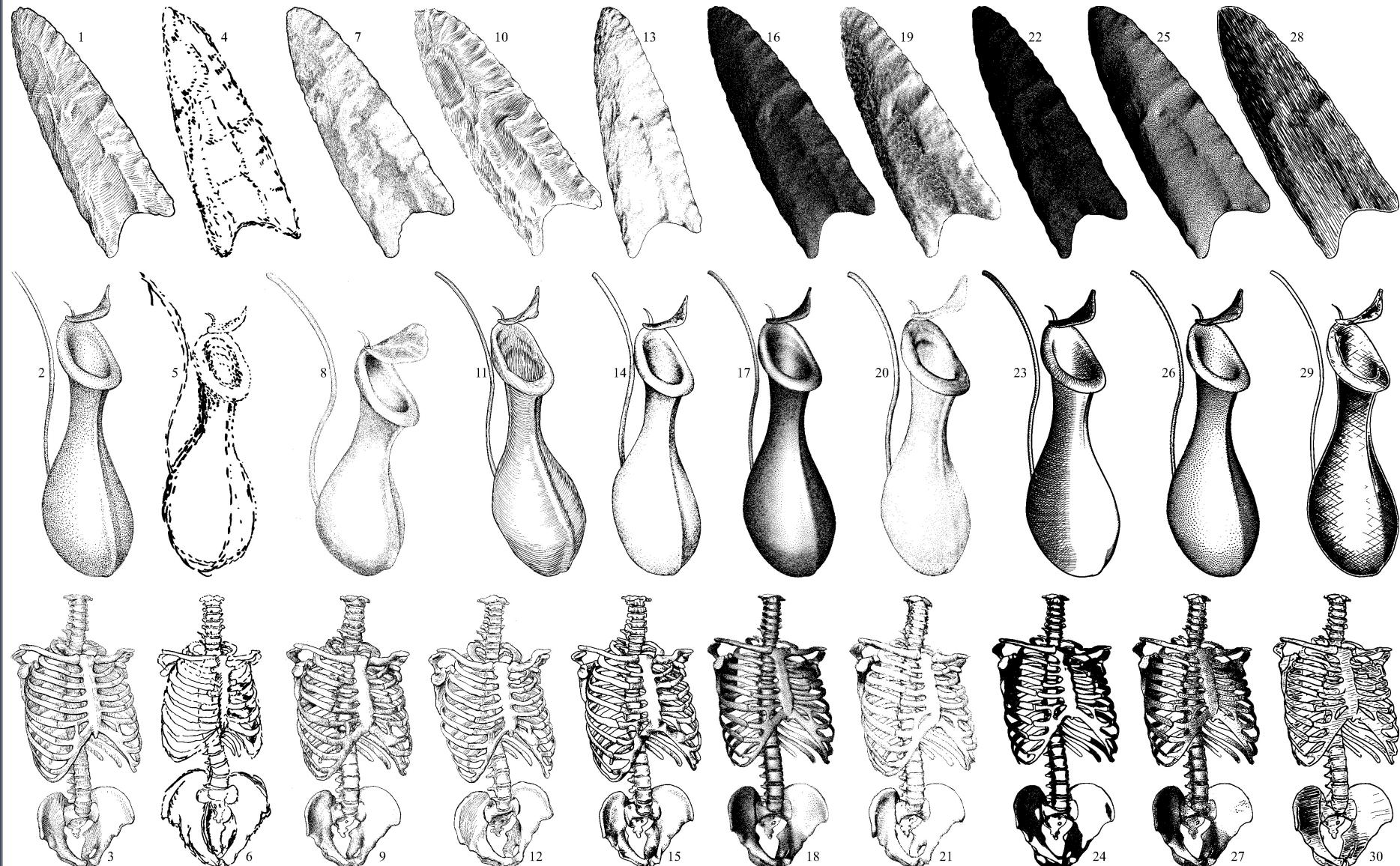
- qualitative techniques to evaluate non-numeric aspects of interfaces
- a number of different techniques for evaluation
- often combinations of techniques employed:
 - e.g., questionnaire with open-ended interviewing
 - e.g., single-person or team direct observation and think-aloud with co-discovery learning
- results often not backed with statistics and may be biased (reduce as much as possible), but yield important insights into what people think
- more general issues discovered

Example (1)



with P. Neumann, S. Carpendale, M. Sousa & J. Jorge

Example (1)



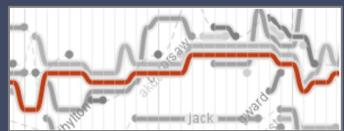
Example (1)



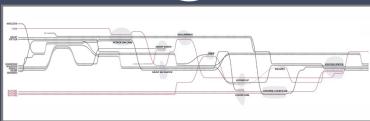
with P. Neumann, S. Carpendale, M. Sousa & J. Jorge

Example (2) - iStoryline

Ogawa et al.



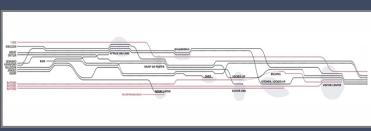
2009



Tanahashi et al.

2012

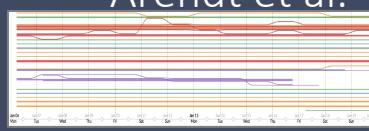
Liu et al.



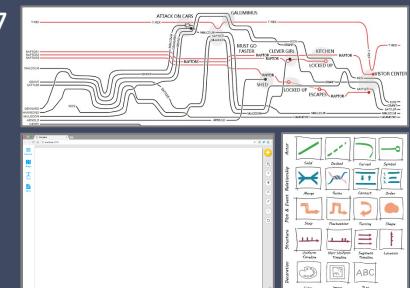
2013



Arendt et al.



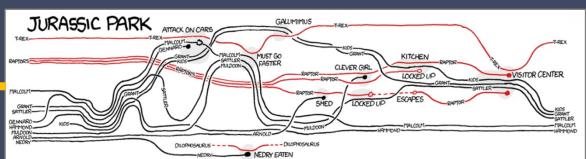
2017



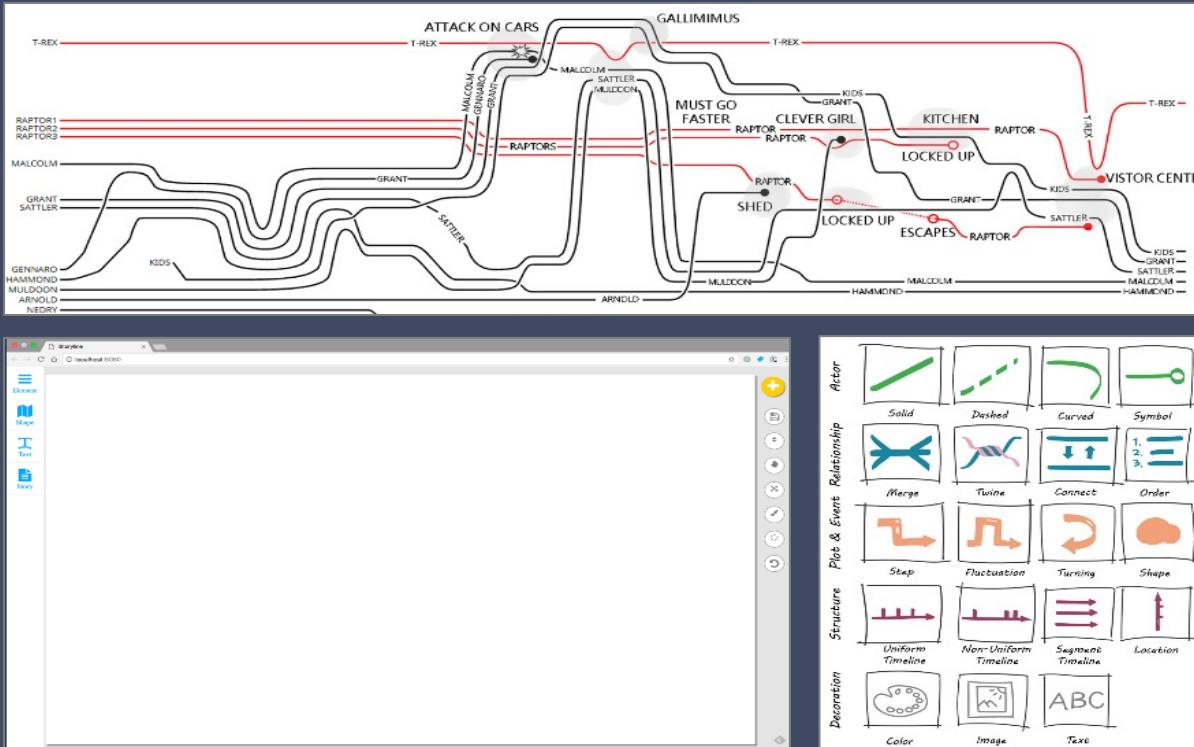
2018

Storyline

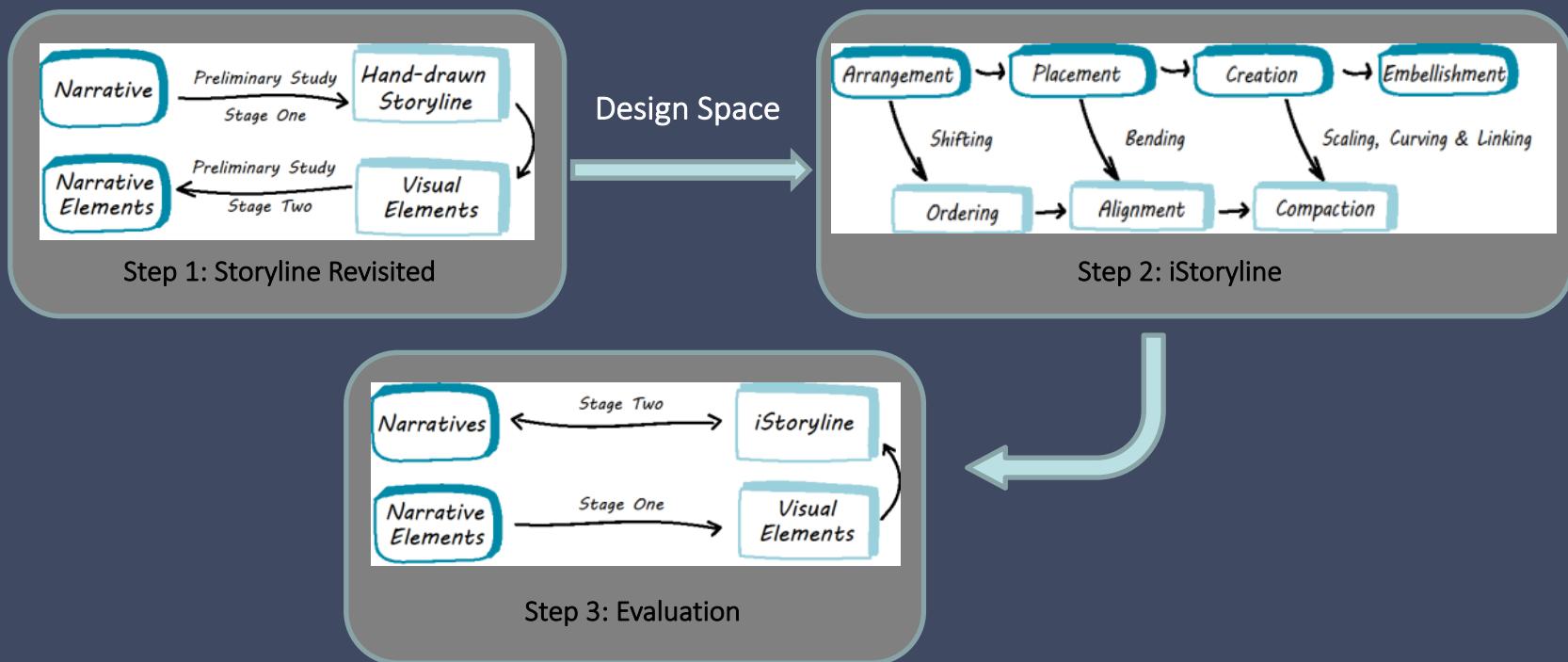
R. Munroe
<https://xkcd.com/657/>



Contribution



Workflow



Stage one: Narrative elements -> Visual elements

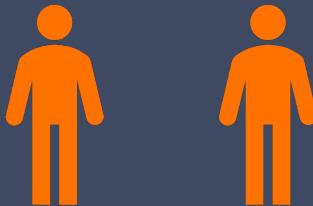
- Participants manually draw storylines for their own stories and follow the basic design principles (D1 and D2) of storyline visualizations.



74 participants

Hand-drawn Storylines

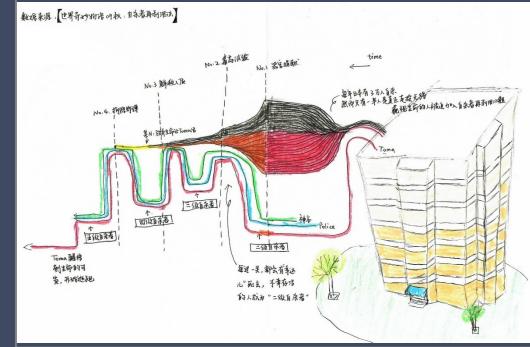
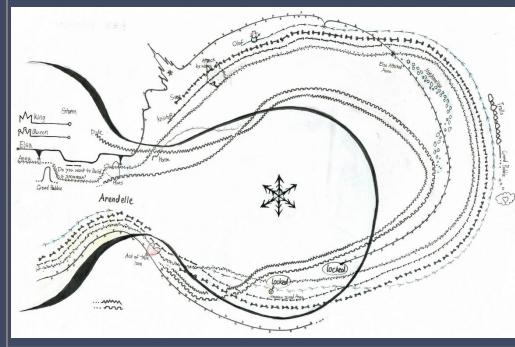
- We select hand-drawn illustrations which have novel visual elements.



74 participants
34 illustrations

Hand-drawn Storylines

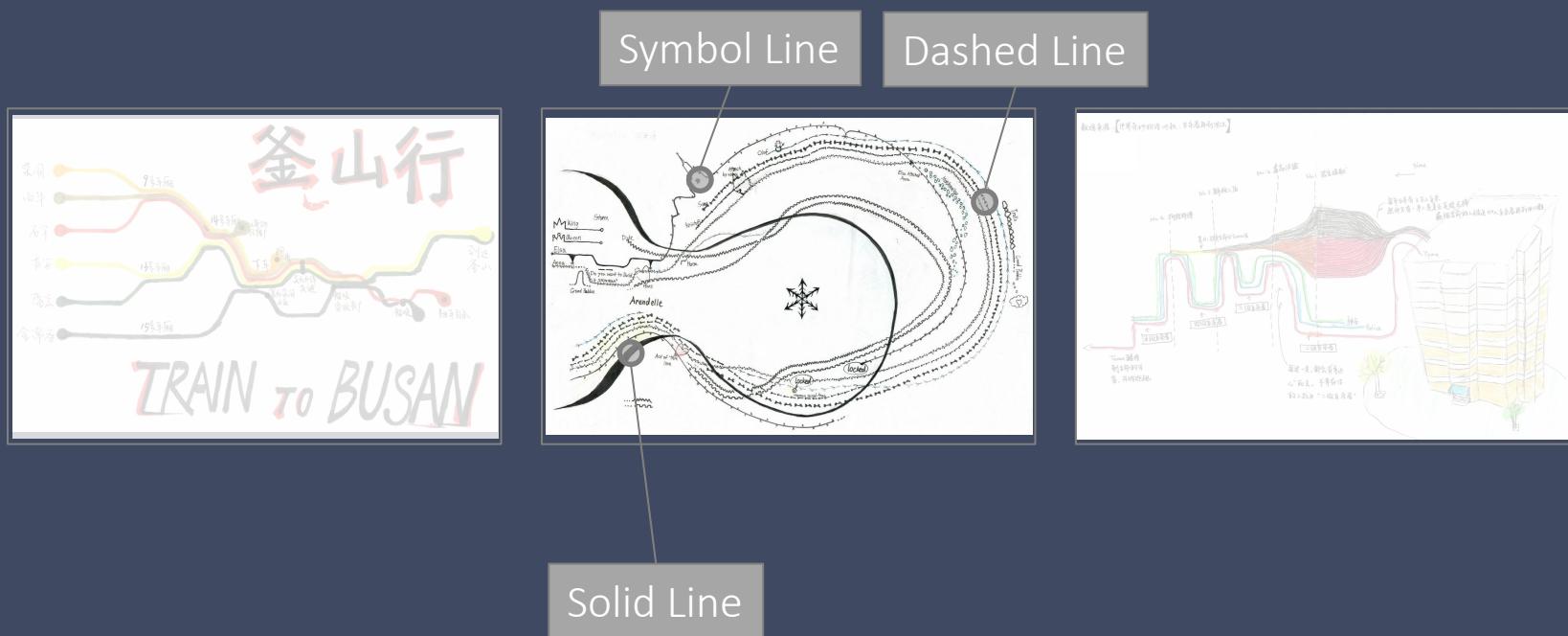
- We identified visual elements for hand-drawn illustrations.



GitHub <https://istoryline.github.io/>

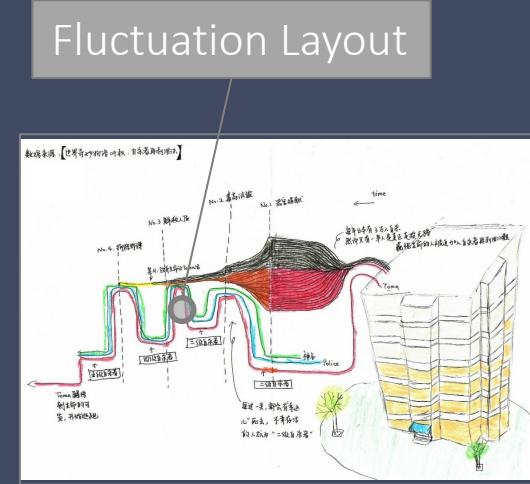
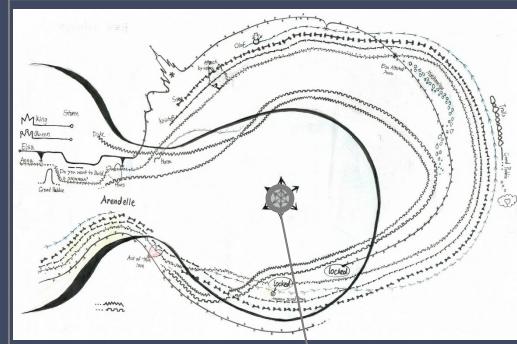
Visual Elements

Character



Visual Elements

Plot



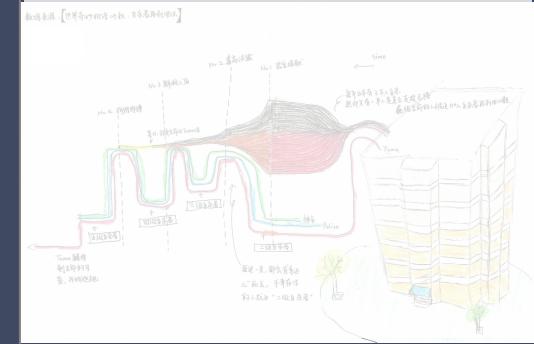
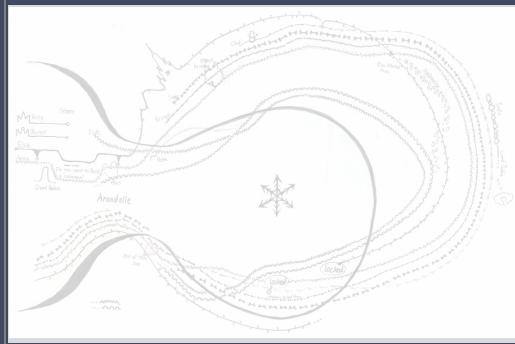
Turning Layout

Visual Elements

Relationship

Merge/Split Group

Twine Group



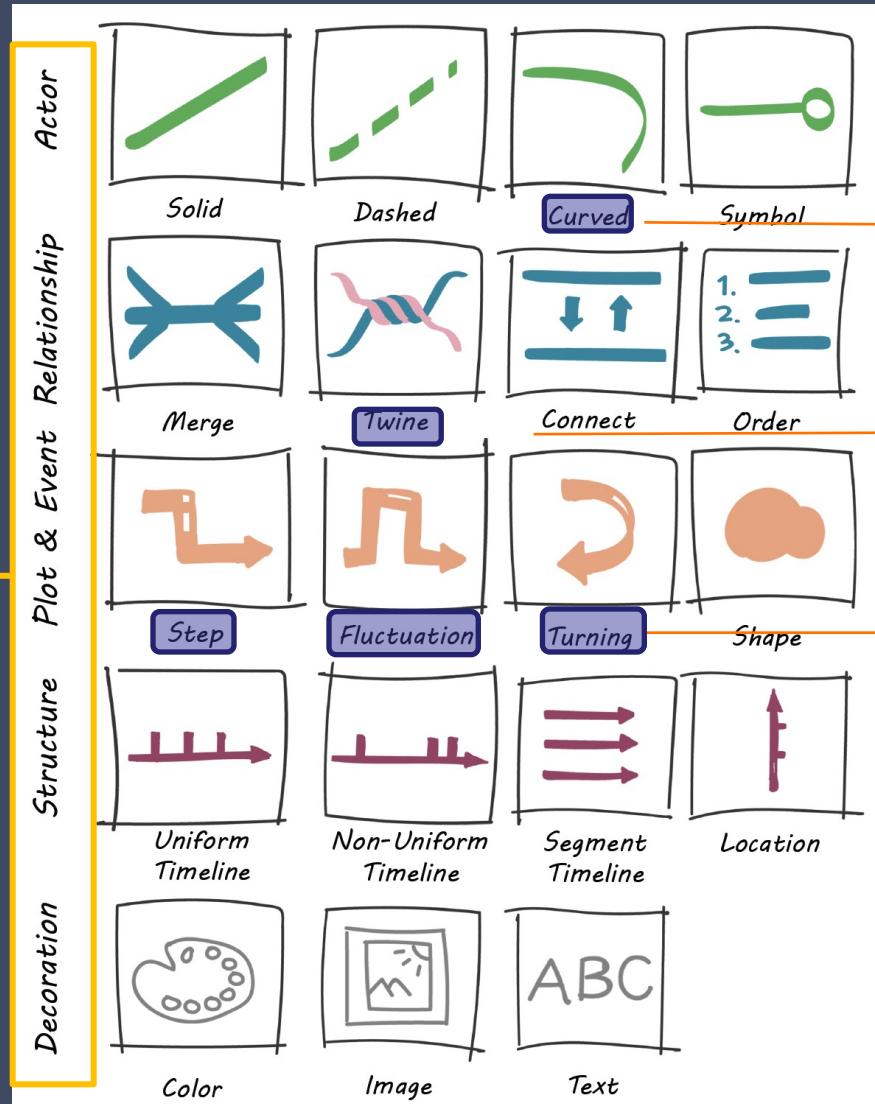
Connect Group

Interviews

- 14 participants who adopted novel visual elements.
- Questions:
 - What elements did you consider as *the most essential* in a movie/story and what *visual elements* were used to convey them?
 - How did you emphasize *minor but affective* elements?
 - What *sequence of actions* did you follow to deal with the innovation in the storylines?

Design Space

Narrative Dimensions



Reducing line wiggles

Reducing line crossings

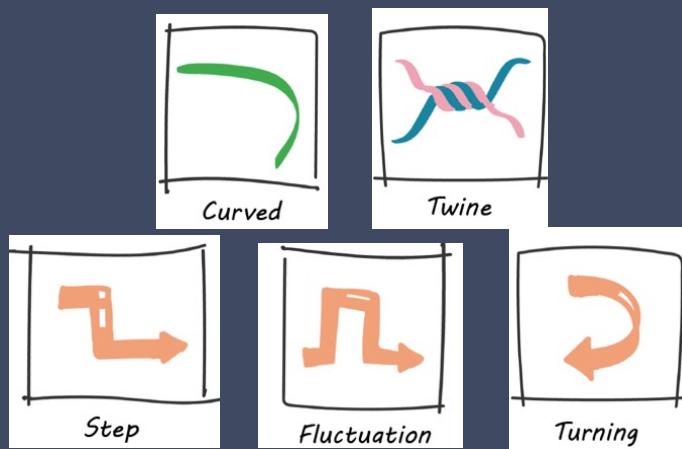
Reducing white space

Stage two: Visual elements -> Narrative elements

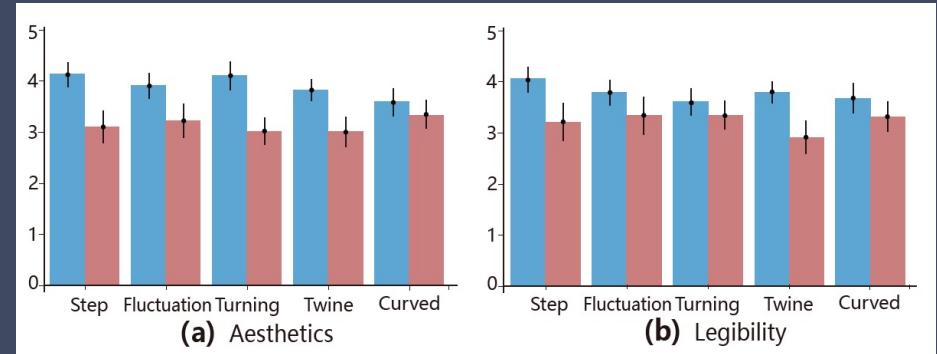
- We conduct another user study to validate the effectiveness of the five novel visual elements.
- 25 participants (8 females)
- Task:
 - Describe *a possible story* for the visual element or associate it with *a narrative element*: an actor, a relationship or a plot/event.
 - Rate the visual element in both the hand-drawn and *automatic storyline* conditions from different aspects.
 - Discuss the *implication* of the visual element in the storylines.

results

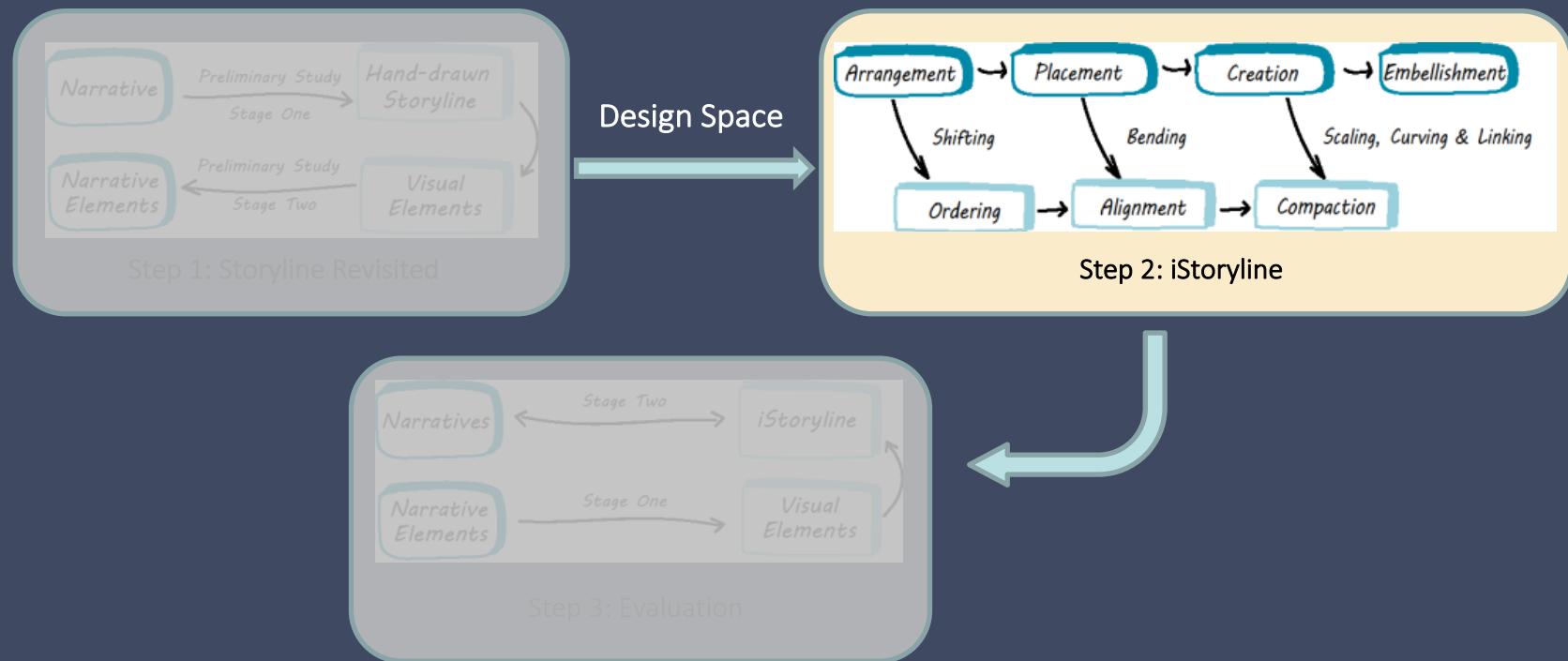
- We collect quantitative data from the participants through comparison experiments.



■ Manually-drawn storylines
■ Automatically-generated storylines



iStoryline



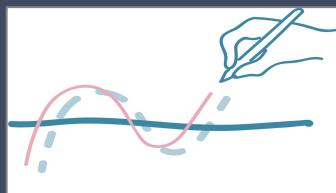
iStoryline

- iStoryline is established on the state-of-the-art automatic method - StoryFlow.

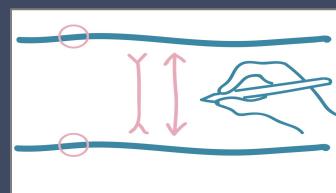
Interactions



Bending



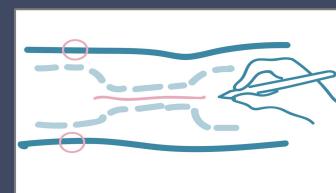
Curving



Scaling

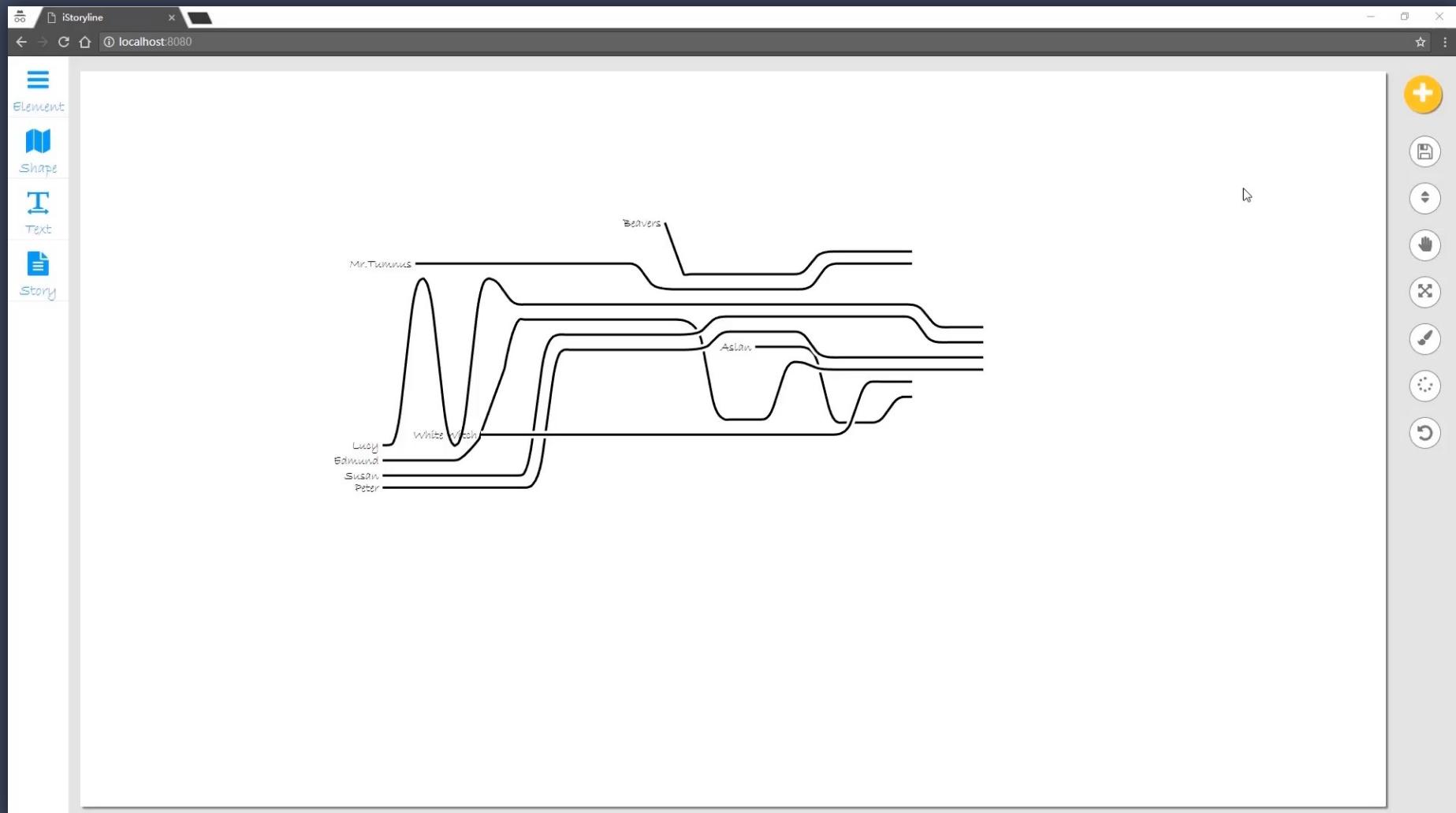


Shifting

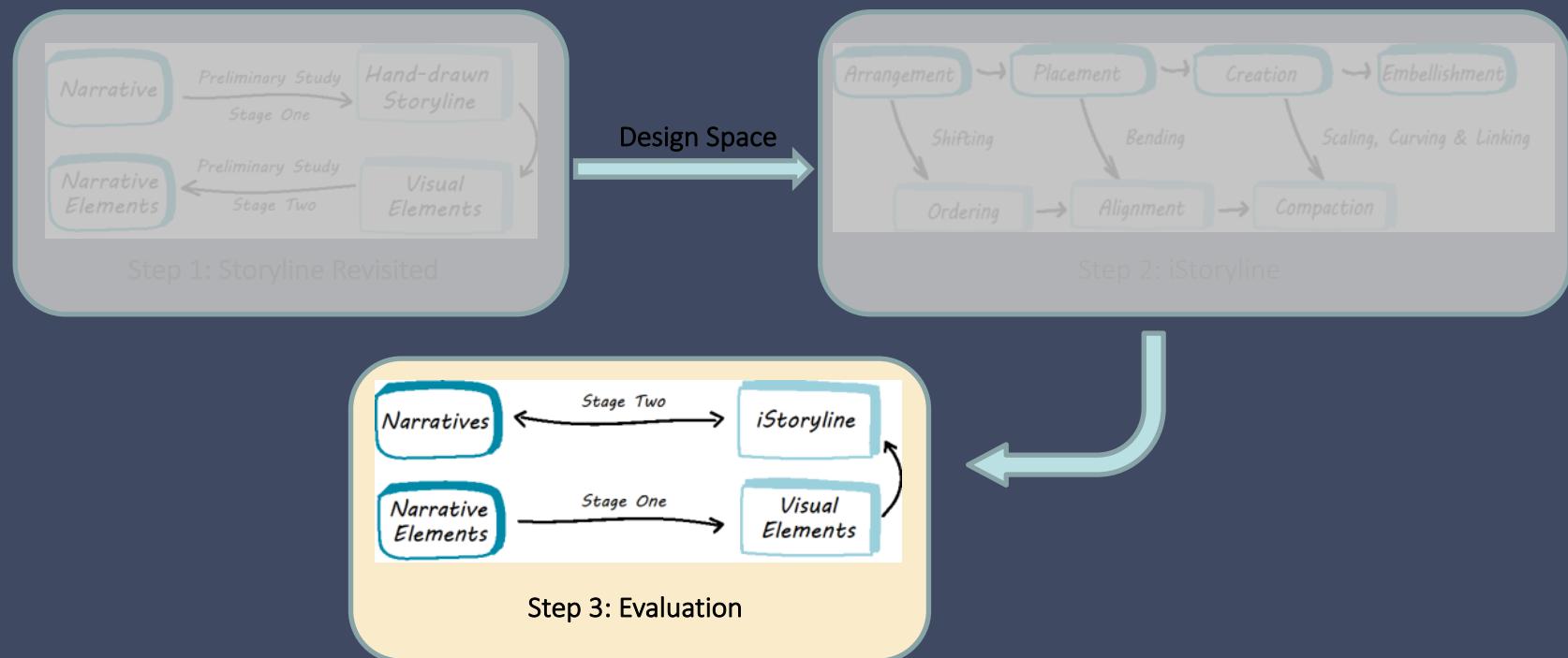


Linking

iStoryline: an authoring tool

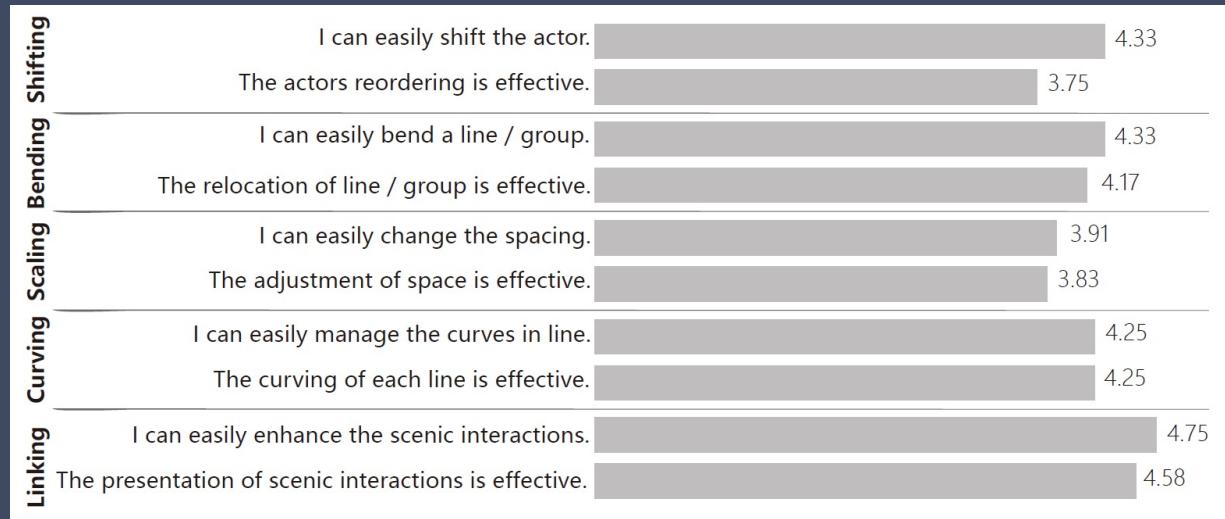


Evaluation



Stage one: Narrative elements -> Visual elements

- We conduct a task-based study (usability test) to assess effectiveness and intuitiveness of iStoryline.
- 12 students (6 females)
- Results:



Stage two: Visual elements <-> Narrative elements

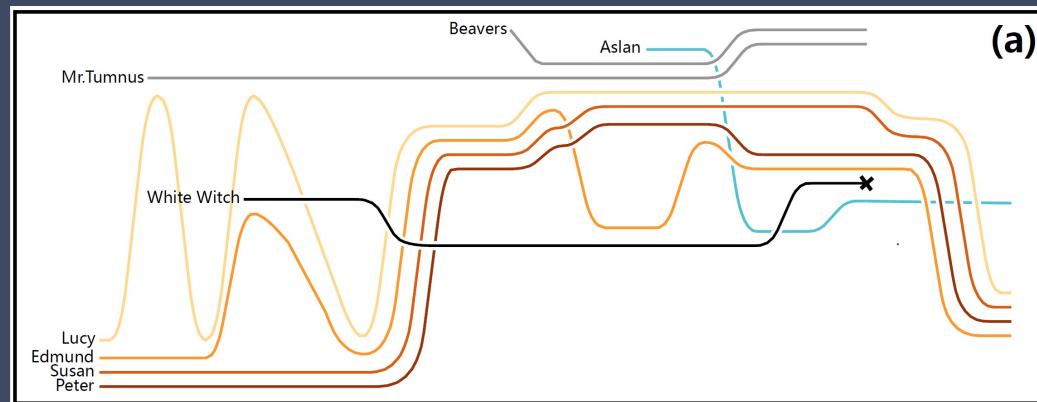
- We conduct an interviews with experts to compare the various interpretations of the iStoryline layout with the original narrative.
- Three experts: a professor (PD) in Industrial Design, a professor (PM) in Digital Media, and an advanced engineer (PE) from a leading e-commercial company.

Tasks

- Tasks:
 - Given *four short narrative events* (by oral storytelling), *mark the corresponding visual elements* in the iStoryline layout.
 - Given *three visual elements* in the context of the whole iStoryline layout, discuss *the possible plots that may occur in the original narrative*.

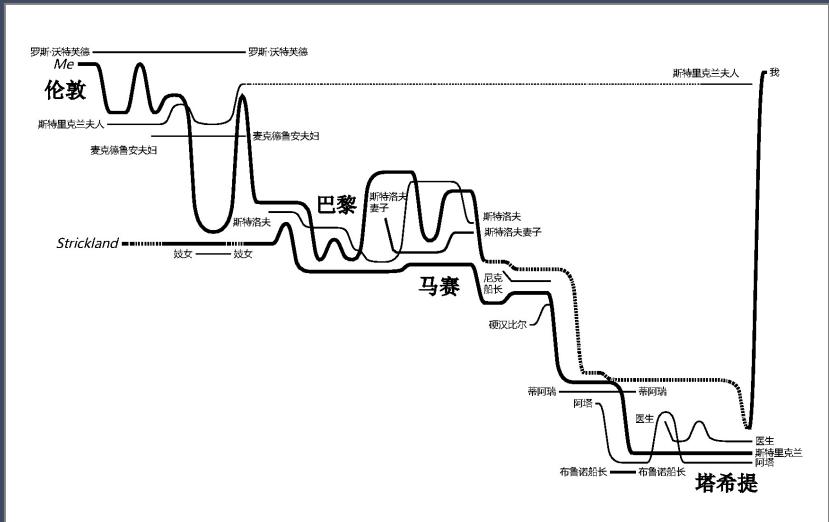
Task one: questions

- When the Pevensie children enter Narnia together and when they return to the real world.
- When the first time Edmund and Lucy enter Narnia together and Edmund meets the Witch.
- When Edmund betrays his siblings and meets the Witch secretly.
- The final battle.



Task two: questions

- Who is the main actor in the story.
- The main location where most subplots take place.
- Discuss Strickland's life before he meets Dirk Stroeve.



Discuss

- What you want to evaluate (goals)?
- Who will be the participants (domain experts or normal users, a single user or a group of users)?
- How do you want to evaluate (procedures)?
- What questions will you ask?
- What results will you collect?

References

- Chapter 14, 15, 16, 16.2
“Interaction Design:
Beyond Human-Computer
Interaction” by Helen
Sharp, Yvonne Rogers,
and Jenny Preece

