

# 可视化研究中的评估方法

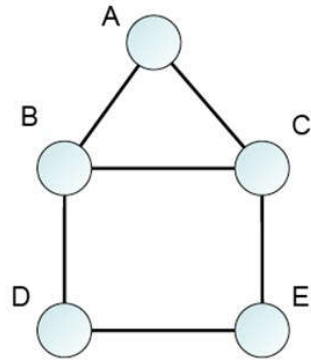
张小龙

# 评估什么呢?

- 可视化、可视分析系统中的某个设计
- 可视分析系统整体

ChinaVis 2016 长沙

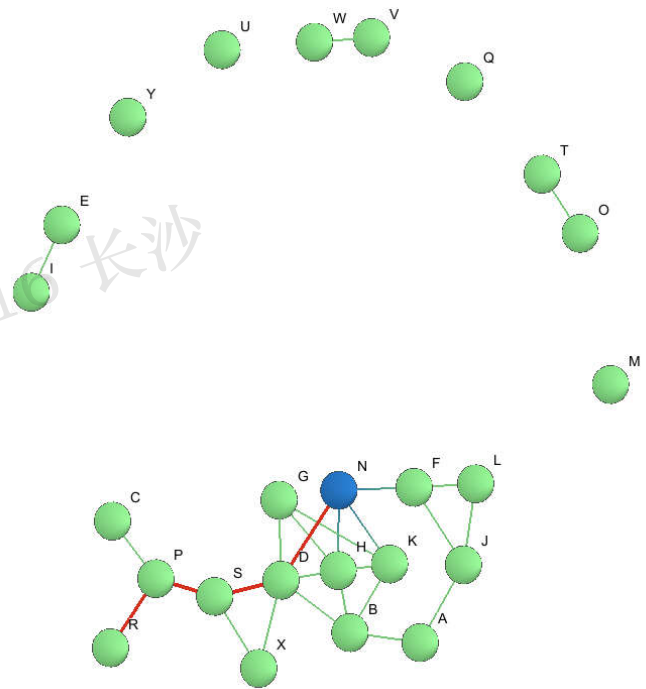
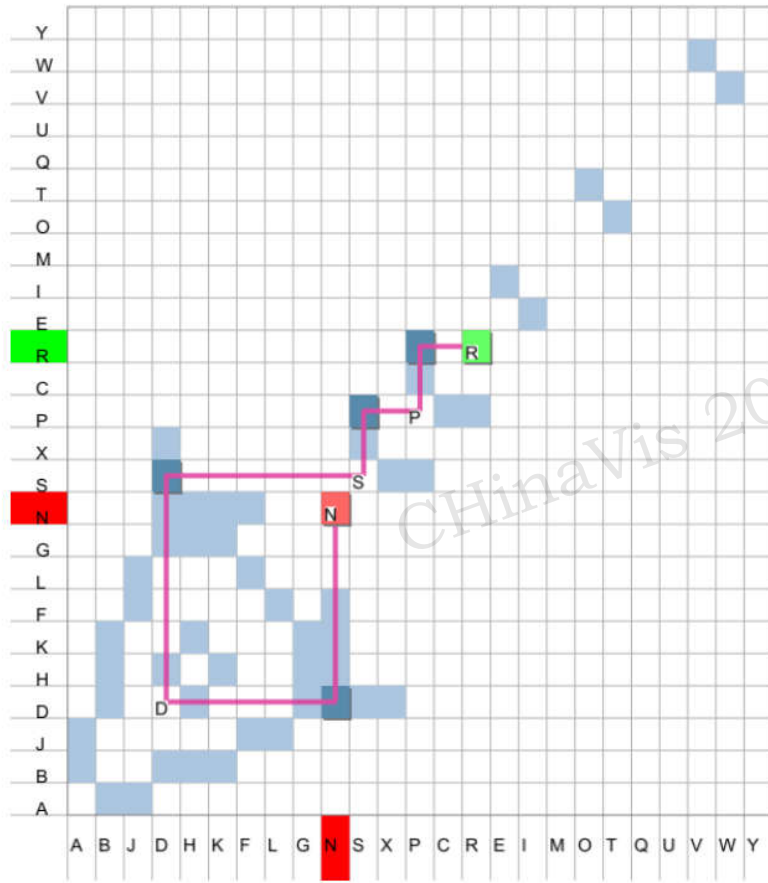
# 对某个设计的评估



vs.

	A	B	C	D	E
A		■	■		
B	■		■	■	
C	■	■			■
D		■			■
E			■	■	

- 针对某类特定任务的通用性设计
  - 一个新的设计导致更好的任务执行
- 实验室方法：确定因果关系
- 挑战
  - 对实验设计要求很高：内在有效性、外在有效性等
  - 对任务的依赖性很强



(图来自浙江大学数据可视化暑期研讨班课件)

# 对一个系统的整体评估



- 实验的方法评估系统比较难
  - 一个系统涵盖很多耦合度很高的用户任务
  - 设计实验的典型任务很难
- 买点往往集中在某一个特点上
- 已有的方法往往很宏观、很粗略
  - [CSU方法](#) (the Computer System Usability test, Lewis, J.R., 1995)

# 那该怎么办？

- 这个问题不是可视化、可视分析领域独有的
- 人机交互领域也有这个问题
  - UIST会议的常见话题

“如何评价新的系统、新的交互方式呢？”

# 借鉴人机交互领域的一个观点

- 不要拘泥于用实验室方法衡量细小的改进
  - Incremental contribution

## Evaluating User Interface Systems Research

*Dan R. Olsen Jr.*

Brigham Young University

Computer Science Department, Provo, Utah, USA

olsen@cs.byu.edu,

### **ABSTRACT**

The development of user interface systems has languished with the stability of desktop computing. Future systems, however, that are off-the-desktop, nomadic or physical in nature will involve new devices and new software systems for creating interactive applications. Simple usability testing is not adequate for evaluating complex systems. The problems with evaluating systems work are explored and a set of criteria for evaluating new UI systems work is presented.

### **ACM Classification Keywords**

H.5.2 User Interfaces

### **WHY UI SYSTEMS RESEARCH?**

Before addressing the evaluation question we must first consider the value of user interface systems research. The systems we have are stable. Applications are being written. Work is progressing. The users are happy (sort of). Why then does the world need yet another windowing system?

### **Forces for change**

A very important reason for new UI systems architectures is that many of the hardware and operating system assumptions that drove the designs of early systems no longer hold. Saving a byte of memory, the time criticality of dispatching an input event to the right window or lack of

# 情景、任务和用户（STU）

- 重要性（importance）
  - "establishing the importance of the problem and its proposed solution"
  - "consider[ing] the difference that our new technology will make"
- 未解决的问题（problem not previously solved）
  - "The larger and more diverse the STU context is, the stronger the claim to the importance of a solution"
- 普遍性（generality）
  - "The greater the diversity and the larger the number of demonstrated solutions, the stronger the generality claim"



# 情景、任务和用户（STU）

- 降低问题求解的难度（reducing solution viscosity）
  - "reducing the effort required to iterate on many possible solutions"
- 整合的力量（power in combination）
  - "effectiveness by supporting combinations of more basic building blocks"
- 拓展性（can it scale up?）
  - "scale up to large problems"
- 扩大参与（empowering new participants）
  - "introducing new populations"

## 这些目标的一个共同要求

- 所设计的系统需要确实帮助用户解决实际的问题
- **STU**
  - 深入了解用户
  - 准确获取需求和任务
  - 全面掌握应用背景
- 仅仅拿到数据是不够的！

# 我们真的知道用户要什么吗？

- 我们很多时候觉得我们知道
- 但一些认知特性决定了设计研究人员和用户想的的确不一样
  - 心智模型（mental model）
    - 实际问题的理解（如技术系统的作用）
  - 思维定式（einstellung）
    - 解决问题的策略（如使用技术系统的方式）
  - 归因习惯（attribution）
    - 解释问题的方式（如使用系统时的障碍）

# 我们该如何理解STU呢？

- 需要比较深入的实地调查（field studies）
  - 了解用户
    - 知识背景、技术背景
  - 了解行为
    - 任务行为
    - 系统使用行为
  - 了解背景
    - 各种约束条件
- 听其言、观其行

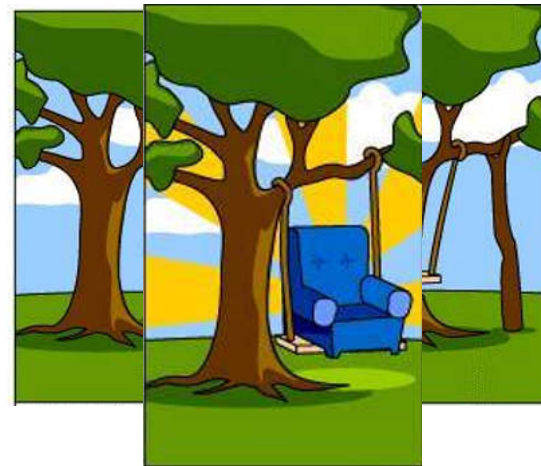
CHinaVis 2016 长沙

# 深入了解人的行为习惯很难！

- 人类行为很微妙
  - The devil is in the details.



- 常用的做法
  - 与用户"厮守"在一起
  - 需要专业的眼睛来观察用户行为
    - 术业有专攻



Examples of participants' words	Properties	Open code
Mentored Trained to specialize in a needed area Assigned to new accounts Included in new business planning Involved in conversations about the long-term outlook of the department Meeting about long-term goals and incentive packages Sent to professional development sessions Paid for graduate school Face time with the client	Being mentored Getting to work on new accounts Getting to have face time with the client Being included in discussions about personal long-term goals and organization's long-term goals Getting funding for graduate school and skills workshops	Being groomed
Intriguing work Constantly learning, training	Having intriguing work Developing professional skills	Constantly learning
Asked if I'm happy Talk about the future Get regular reviews Constant congratulations Get messages about growing the company from within Very little turnover	Feeling appreciated Noticing low turnover and receiving messages about growing the company from within	Receiving verbal encouragement and making observations
Personal development fund Lacking a personal touch (negative evidence) [Senior exec.] like a second mother Long hours, low pay (negative evidence)	Caring about personal well-being by both the organization and senior management Encouraging and enabling a healthy personal life	Being cared for as a whole person

Open codes	Axial codes	Selective code
Wanting experiential learning; constantly learning; working in a good environment;pioneering social media and easily adapting to change; feeling entitled due to unique qualifications, as compared to previous generations; possessing the personal skills and characteristics needed; being groomed	Believing they are ready to be set loose on accounts	Wanting to make a difference
Craving immediate feedback and being motivated by feeling appreciated; detesting getting called out; receiving verbal encouragement and making observations	Seeking external validation	
Mind reading and expectations for a miracle worker;getting called out; not being heard	Silently blaming employers for failures	
Advocating a work-life balance; being cared for as a whole person; accommodating interests and preferences	Wanting a meaningful experience at work and outside of work	

# 评估不是课题的最后一步

- 需要从课题的设计初期就开始考虑
  - STU
- 需要贯穿整个课题的实施

实地调查  
(Field Study)

系统开发

实地调查  
(Field Study)

ChinaVis 2016 长沙



# 所面临的挑战

- 文章发表的紧迫性
  - 需要一年做一个VIS文章，哪有功夫去调研和实地部署、评测
- VAST领域对相关工作的认可度
  - 技术为主导，对相关行为研究缺乏足够的重视
- 跨学科研究的艰巨性
  - 方法的难度
    - 学科往往由方法来界定
  - 合作的动力
    - 我们的课题对他人的价值在哪里？