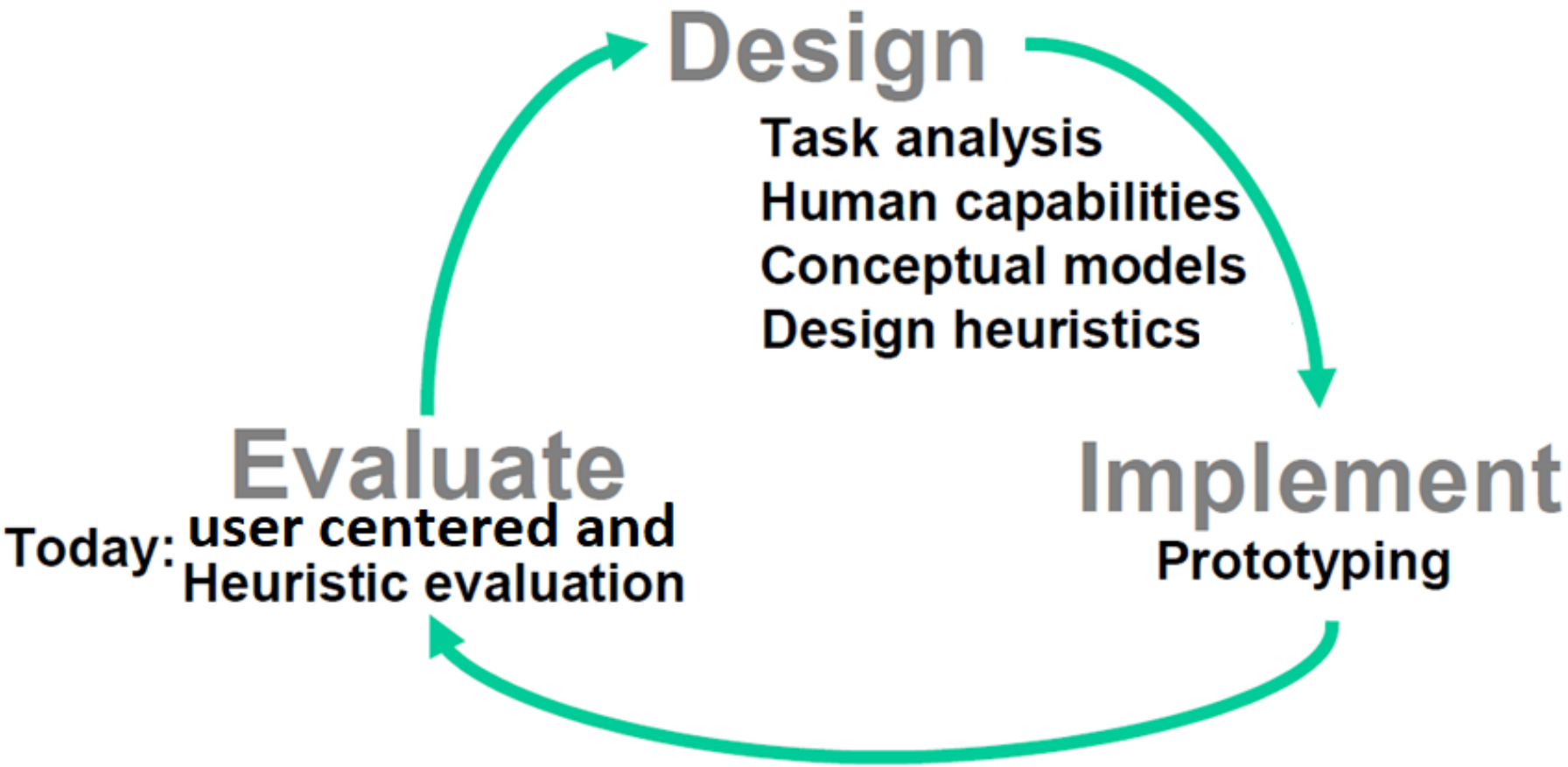
# Part 7 – Evaluation

HCI 2019


Design
Task analysis
Human capabilities
Conceptual models
Design heuristics
Implement
Prototyping
Evaluate
Today: user centered and
Heuristic evaluation

# Introduction

Evaluations assess design and implementation

**Formative** evaluation during development (cook tastes the soup)

**Summative** evaluation at completion of project (guests taste the soup)

Ideally, evaluation should be considered at all stages in the design life cycle

- That's practically difficult but formal methods (such as expert-based methods – coming on later slides) can and should be used
- Later, we'll talk about how to plan and conduct actual user evaluations

# Introduction

Since evaluation should be considered at all life cycle stages, there is a link between:

- evaluation & design techniques
- evaluation & prototyping techniques
- evaluation & implementation techniques
- Etc.

# Goals of Evaluation

Assess extent of system performance and functionality (the tasks that users are interested in)

Assess effect of the user interface on the user

- (the user's experience of the interaction e.g., easy to learn, easy to use, satisfaction etc., and the usability attributes e.g., learnability, speed of operation, robustness, recoverability, adaptability)

Identify specific problems (e.g., errors, confusion, unexpected results)

[the above goals are of course interrelated…]

# Evaluation Techniques

Evaluation techniques can be categorized as follows:

- Expert-based
- Model-based
- User-based (user-centered evaluation )

# Expert-based Evaluation Techniques

Expert-based evaluation techniques are also referred to as expert analysis techniques

Evaluation through expert analysis is done because:

- It can be expensive to regularly carry out user tests at all life cycle stages
- Moreover, it can be difficult to get an accurate assessment based on incomplete designs and prototypes

# Expert-based Evaluation Techniques

Expert analysis characteristics:

- Designer or HCI expert assesses a design based on known/standard cognitive principles, design principles or empirical results
- Expert analysis methods can be used at any stage in the life cycle
- Expert analysis methods are relatively cheap
- Expert analysis methods, however, do not assess the actual use of the system

Examples of expert analysis methods:

- Heuristic Evaluation (HE)
- Cognitive Walkthrough (CW)

# Heuristic Evaluation

- Heuristic Evaluation (HE) was proposed by Nielsen and Molich
- In HE, experts scrutinize the interface and its elements against established usability heuristics [another previous lesson]
- The experts should have some background knowledge or experience in HCI design and usability evaluation

# Heuristic Evaluation – the process

- 3 to 5 experts are considered to be sufficient to detect most of the usability problems
- The enlisted experts are provided with the proper roles (and sometimes scenarios to use) to support them when interacting with the system/prototype under evaluation
- They then evaluate the system/prototype individually
  - This is to ensure an independent and unbiased evaluation by each expert
- They assess the user interface as a whole and also the individual user interface elements
- The assessment is performed with reference to a set of established usability principles
- When all the experts are through with the assessment, they come together and compare and appropriately aggregate their findings

# Cognitive Walkthrough

- Cognitive Walkthrough (CW) was proposed by Polson et al.
- CW evaluates design on how well the design supports user in learning the task to be performed [primarily through exploration i.e. hands on]
- CW is usually performed by expert in cognitive psychology
- The expert ‘walks through’ the design [i.e. steps through each step of some known/representative task] to identify potential problems

# Cognitive Walkthrough - requirements

Four requirements in order to perform the CW:

- specification or prototype of the system
- description of the task the user is to perform
- complete, written list of actions constituting the task
- description of the user (including the level of experience and knowledge)

# Cognitive Walkthrough – the process

With the foregoing information, the evaluator steps through each of the actions trying to answer the following 4 questions:

- is the effect of the action the same as the user's goal at that point? [what the action will do/action's effect should be what the user intends/user's goal.]
- will users see that the action is available [when they want it] - visibility at that time?
- once users have found the correct action [as in the foregoing], will they know/recognize it is the one they need? [effective representation of the action, clear representation.]
- after the action is taken, will users understand the feedback they get? [effective confirmation that the action has been taken.]

# Cognitive Walkthrough – the process

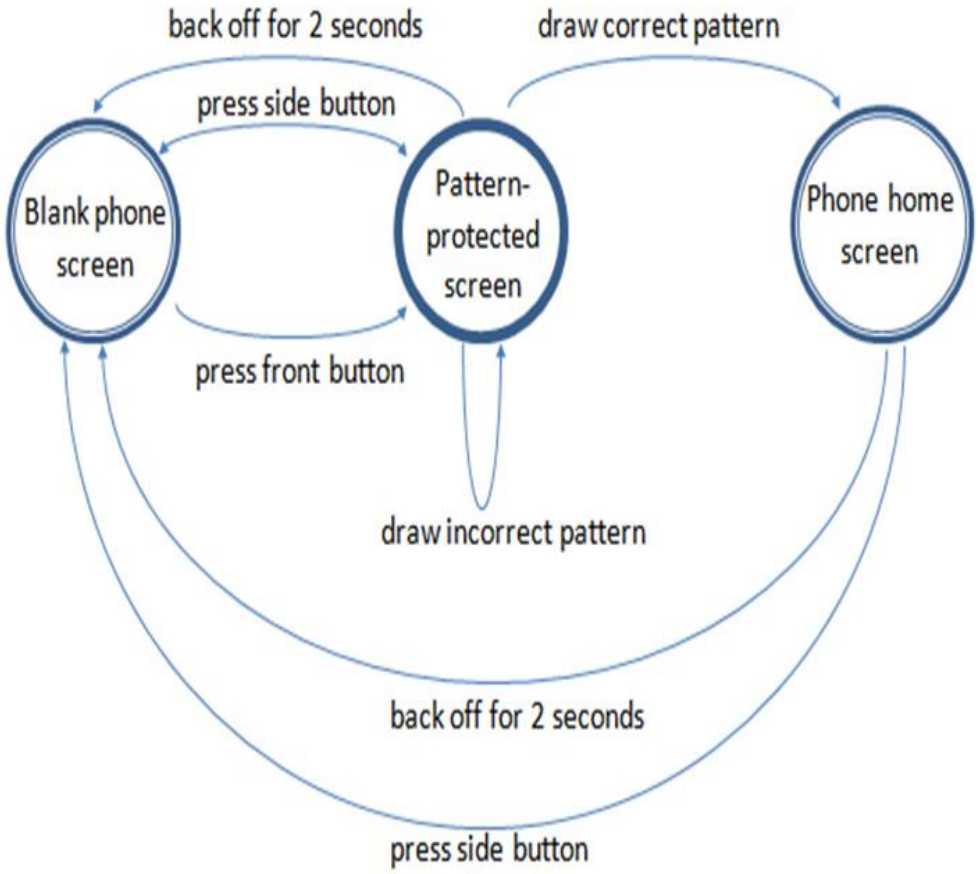
Forms are used to guide analysis e.g.

- cover form [for the four requirements above, date, time, evaluators of the CW],
- answer form [for answering the four questions above],
- usability problem report [for describing any negative answers/problems, severity of the problem e.g. frequency of occurrence and seriousness of the problem, date, time, evaluators]

# Reading Assignment: Model-Based Evaluation

For instance: Dialog models (such as State Transition Networks) can be used to evaluate dialog problems in a user interface e.g. unreachable states, circular dialogs, etc.

**Note:** Model-based evaluation is sometimes classified under expert-based evaluation techniques



# User-Based Evaluation

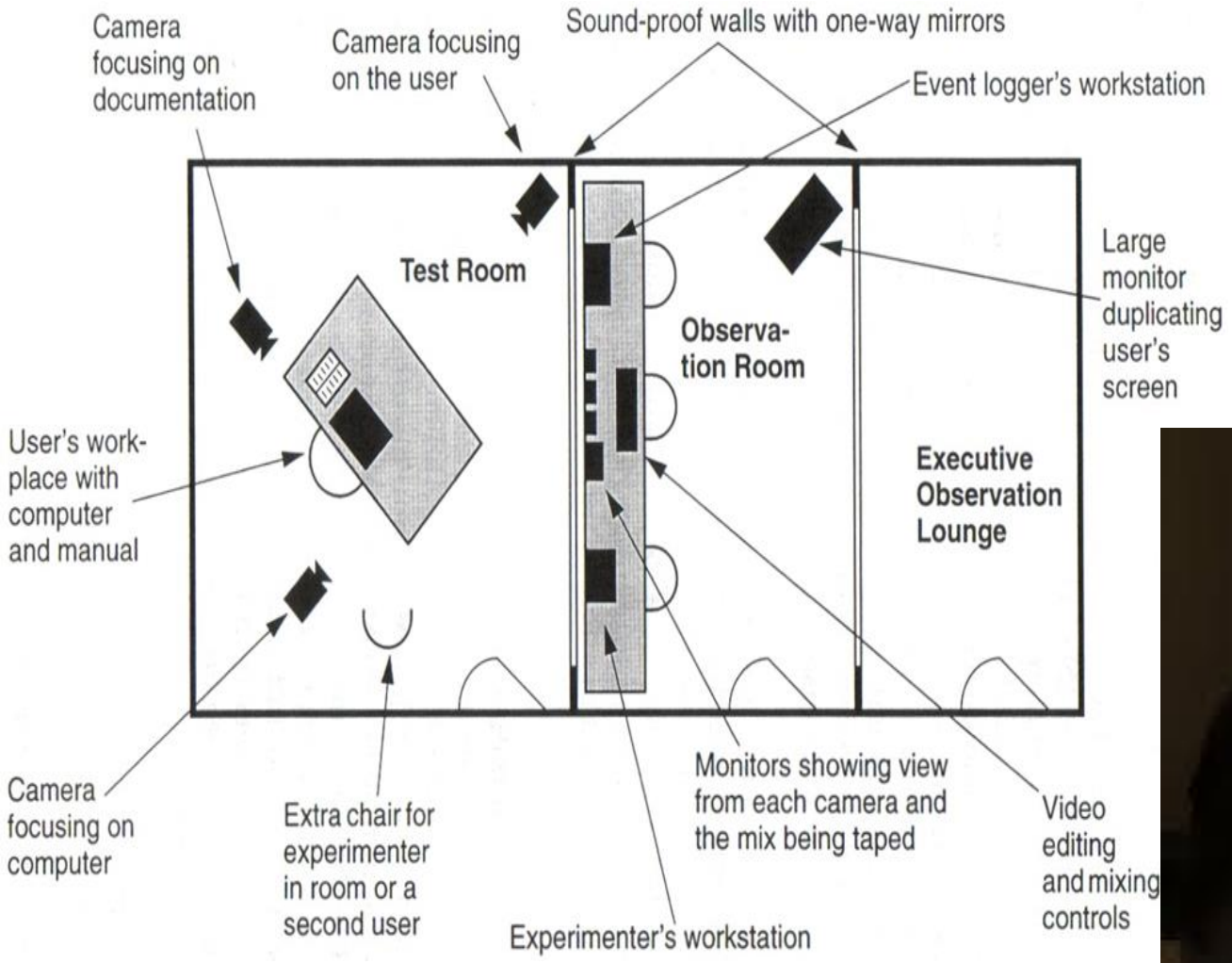
User-based evaluation basically is evaluation through user participation i.e. evaluation that involves the people for whom the system is intended; the users

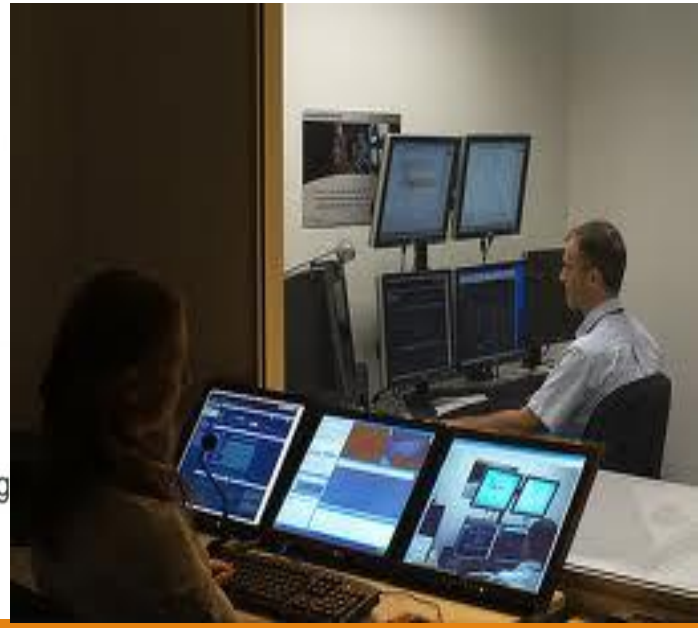
User-based evaluation techniques include:

- experimental methods
- observational methods
- query techniques (e.g., questionnaires and interviews)
- physiological monitoring methods (e.g., eye tracking, measuring skin conductance, measuring heart rate)

User-based methods can be conducted in the laboratory and/or in the field

# Usability Laboratory



# User-Based Evaluation

**Laboratory**

Advantages:

- Specialist equipment available
- Uninterrupted environment

Disadvantages:

- Lack of context
- Difficult to observe several users cooperating

Appropriate:

- If system usage location is dangerous, remote or impractical
- For very constrained single-user tasks [to allow controlled manipulation of use]
  - Ag military systems, airplane systems

# User-Based Evaluation

**Field or Working Environment**

Advantages:

- Natural environment
- Context retained (though observation may alter it)
- Longitudinal studies possible

Disadvantages:

- Field challenges e.g., distractions, interruptions, movements, danger, noise

Appropriate:

- Where context is crucial [especially for longitudinal studies]
  - Customer management system to manage customer preferences

# User-Based Evaluation Techniques

User-based evaluation techniques include:

- experimental methods
- observational methods
- query techniques
- physiological monitoring methods

# Experimental Methods

Experimental methods are also called controlled experiments

Controlled experiments are:

- considered to be the most rigorous of empirical methods
- capable of providing empirical evidence to support a particular claim or hypothesis

**Note:**

- Empirical research, is a way of gaining knowledge by means of direct and indirect observation or experience.
- Empirical evidence, also known as sensory experience, is the knowledge received by means of the senses, particularly by observation and experimentation

# Experimental Methods

What makes up the Experiment?

## Participants

- Should *match* the expected users as closely as possible e.g., age, education, general computing experience, domain knowledge, etc.
- Sample *size* of the participants should be large enough to be considered representative of the population

## Variables (two types: independent and dependent)

## Hypothesis

# Experimental Methods

## Example:

We want to find out whether users perform faster when using a graphical user interface than when using a command-line interface

Independent variable (IV): interaction mode (with two levels: graphical vs. command-line)

Dependent variable (DV): task completion time

Hypothesis: Users perform faster (DV) when using a graphical user interface (IV level one) than when using a command-line interface (IV level two)

Null hypothesis: There is no difference in performance (DV) when using a graphical user interface (IV level one) or when using a command-line interface (IV level two)

There are two conditions: Graphical (G) and Command-line (C)

# Observational Methods

Observational methods include:

- think aloud
- cooperative evaluation
- protocol analysis,
- post-task walkthroughs

# Observational Methods

**Think aloud**

User is observed performing task

User is asked to describe what s/he is doing and why, what s/he thinks is happening, etc.

Advantages:

- Simplicity - requires little expertise
- Can provide useful insight
- Can show how system is actually used

Disadvantages:

- Subjective [really depends on the user]
- Selective [out of many things, the user may choose what to describe]
- Act of describing may alter task performance

# Observational Methods

**Cooperative Evaluation**

Reading Assignment:

- Describe this evaluation method

# Observational Methods

## Protocol Analysis

Paper and pencil: cheap, limited to writing speed

Audio: good for think aloud, difficult to record sufficient information to identify exact actions in later analysis, difficult to match with other protocols ('synchronization')

Video: accurate and realistic, needs special equipment, obtrusive

Computer logging: automatic and unobtrusive, large amounts of data difficult to analyze

User notebooks: coarse and subjective, useful insights, good for longitudinal studies

**Note:**

Mixed use in practice

Audio/video transcription difficult and requires skill

Some automatic support tools available e.g., EVA (Experimental Video Annotator), Observer Pro (from Noldus), Workplace project (Xerox PARC), etc.

# Observational Methods

**Post-task Walkthrough**

Transcript played back to participant for comment i.e. user reacts on action after the event

Used to fill in intention i.e. reasons for actions performed and alternatives considered

It also is necessary where think aloud is not possible

Advantages:

- Analyst has time to focus on relevant incidents
- Avoids excessive interruption of task

Disadvantages:

- Lack of freshness
- May be post-hoc interpretation of events

# Query Techniques

## Query techniques

- Questionnaires
- interviews

# Physiological Monitoring Methods

Physiological Monitoring Methods such as eye tracking, measuring skin conductance, measuring heart rate etc.

**Example: Eye-tracking**

Head or desk mounted equipment tracks the position of the eye

Eye movement reflects the amount of cognitive processing a display requires

Measurements include: fixations, scan paths, etc. For instance:

- number of fixations
- duration of fixation
- scan paths: moving straight to a target with a short fixation at the target is optimal

# Physiological Monitoring Methods

**Physiological Measurements**

*Emotional* response linked to *physical* changes

These may help determine a user's reaction to a user interface

Measurements include: heart, sweat, muscle, brain. For instance:

- heart activity: e.g. blood pressure, volume and pulse.
- activity of sweat glands: Galvanic Skin Response (GSR)
- electrical activity in muscle: electromyogram (EMG)
- electrical activity in brain: electroencephalogram (EEG)

There is some difficulty in interpreting these physiological responses; more research is needed

# Choosing an Evaluation Method

Factors that can influence the choice

- when in process : design vs. implementation
- style of evaluation : laboratory vs. field
- how objective : subjective vs. objective
- type of measures : qualitative vs. quantitative
- level of information: high level vs. low level
- level of interference: obtrusive vs. unobtrusive
- resources available: time, subjects, equipment, expertise

# Reading Assignment

In user-centered design, what are mental models?

- What is a designer mental model and a user mental model?

In HCI, what is a Loop of Interaction?

- What are the main aspects in loop of interaction?

Why is it crucial to include error messages during the design of user interfaces?

Why is the correct use of color important in user interface design?

What are the characteristics of a poorly designed website?