Introduction to Usability

Franca Garzotto

Usability: definition

- **Usability = easiness of use**
- A crucial aspect of the quality of an application.

- For an “official” definition:

  Usability is a measure of “the effectiveness, efficiency and satisfaction with which specified users can achieve specified goals in particular environments” (ISO 9241-11).

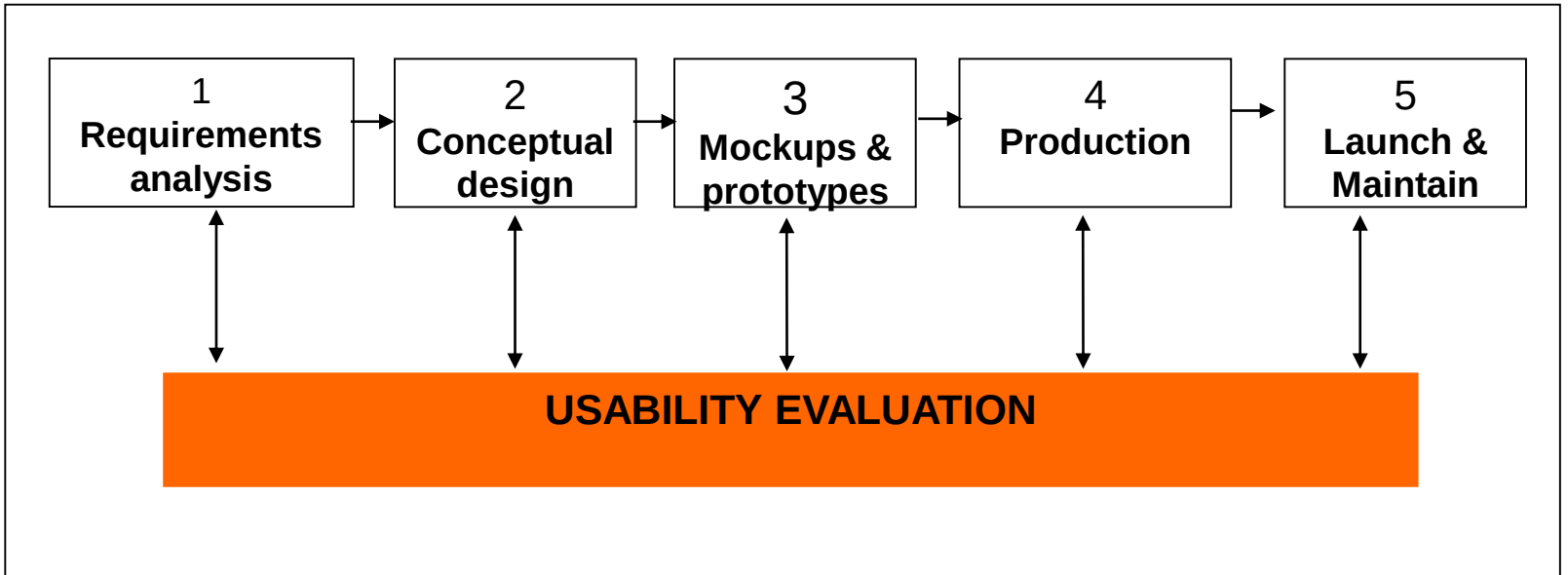
What determines Usability?

- **What** the product is
  - **product characteristics**
- **Who** is using it
  - user characteristics
- What they want to achieve
  - **goals** and tasks
- The **usage context**
  - circumstances and scenarios of use

Usability evaluation in the product lifecycle

- Usability evaluation should be done as early as possible in the development cycle:
  - The later errors/problems/flaws are discovered the more is expensive to fix them

Pervasive Usability (Brink et al.2002)



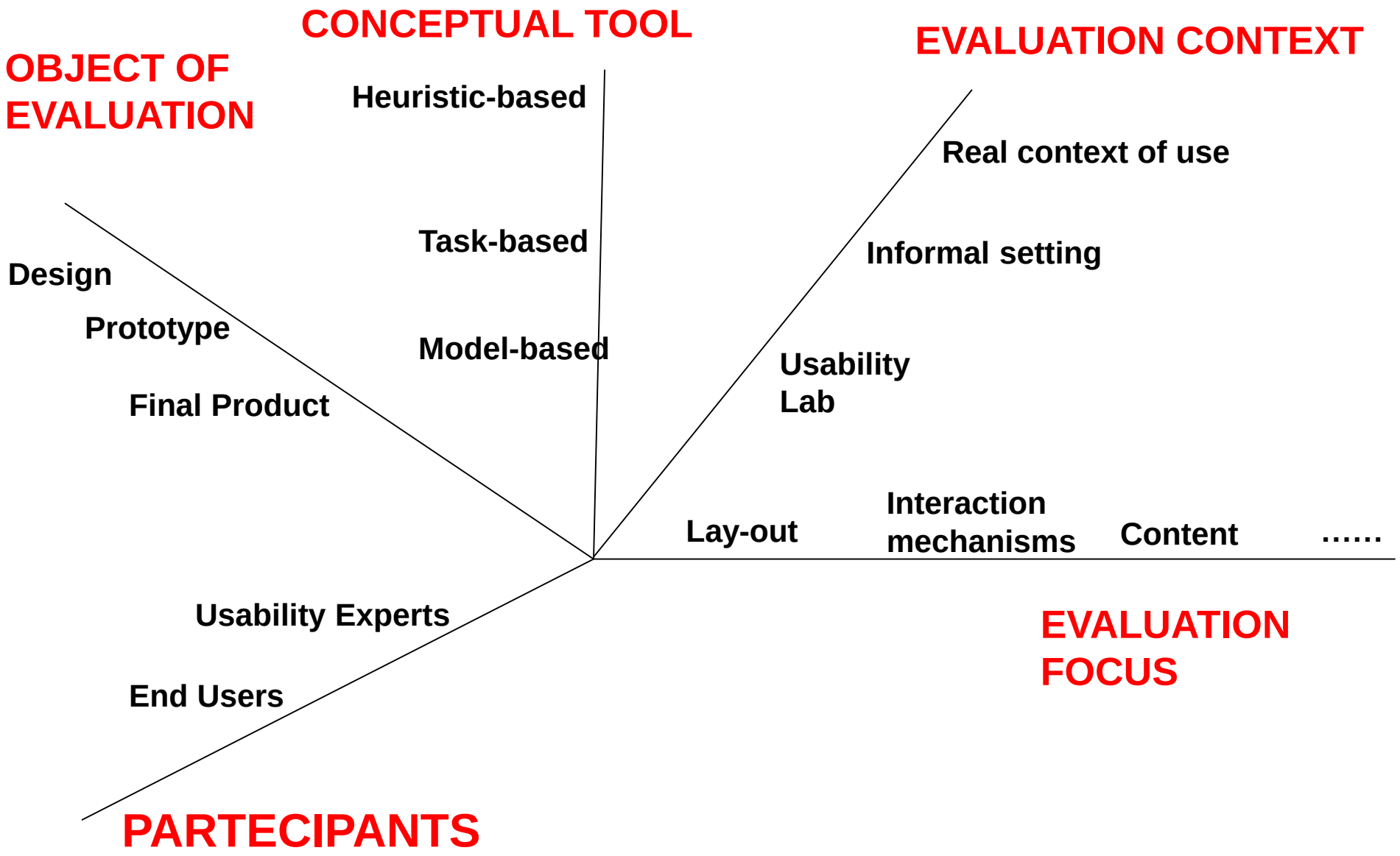
Usability evaluation in the lifecycle (cont.)

- What can be evaluated:
  - How **requirements** meet usability needs
  - The quality of the conceptual **design artifacts**
  - The quality of the design **prototypes**
    - Low fidelity
      - Paper mock-ups, layout sketches
    - High-fidelity
      - Screen shots, running prototypes, working navigation pieces, implemented transactions,…
  - **Implemented** applications

Usability Evaluation

- Goal: to anticipate, identify and characterize **usability problems**
- …and elaborate actionable **recommendations** for design improvement

Usability evaluation :
"variables" that characterize the different methods



Usability Evaluation: Who and How

- **Who** measures usability?
  - **Inspection**
    - **Expert** review
  - **Usability Testing (in the lab or in the field)**
    - **Users** are the measure of quality
- **How** to measure it?
  - **Task-oriented** (scenario-based)
    - "Doing things" with the application
  - **Heuristics-based**
    - Verifying compliance with usability principles
  - ....

Usability Evaluation: Where and What

- **Where** measures usability?
  - In the final context of use (eg., work, home, park, ...)
  - In an artificial setting (a usability lab – see user testing)
  - In a semi-formal settin (a lab)
- **What** to measure for usability
  - Design specifications (e.g., design schemas, mockups)
  - Prototype
  - Final product
- **Which** aspects to focus on
  - Look and feel elements (layout, buttons, background...)
  - Interaction capabilities (e.g., links)
  - How the system «speaks» to the user
    - Button labels, messages, ….

Focus of evaluation (web sites)

CONTENT

INFORMATION ARCHITECTURE

NAVIGATION/INTERACTION

LABELLING (semiotics)

GRAPHICS/LAYOUT

a b c

OPERATIONS

ERRORS MANAGEMENT

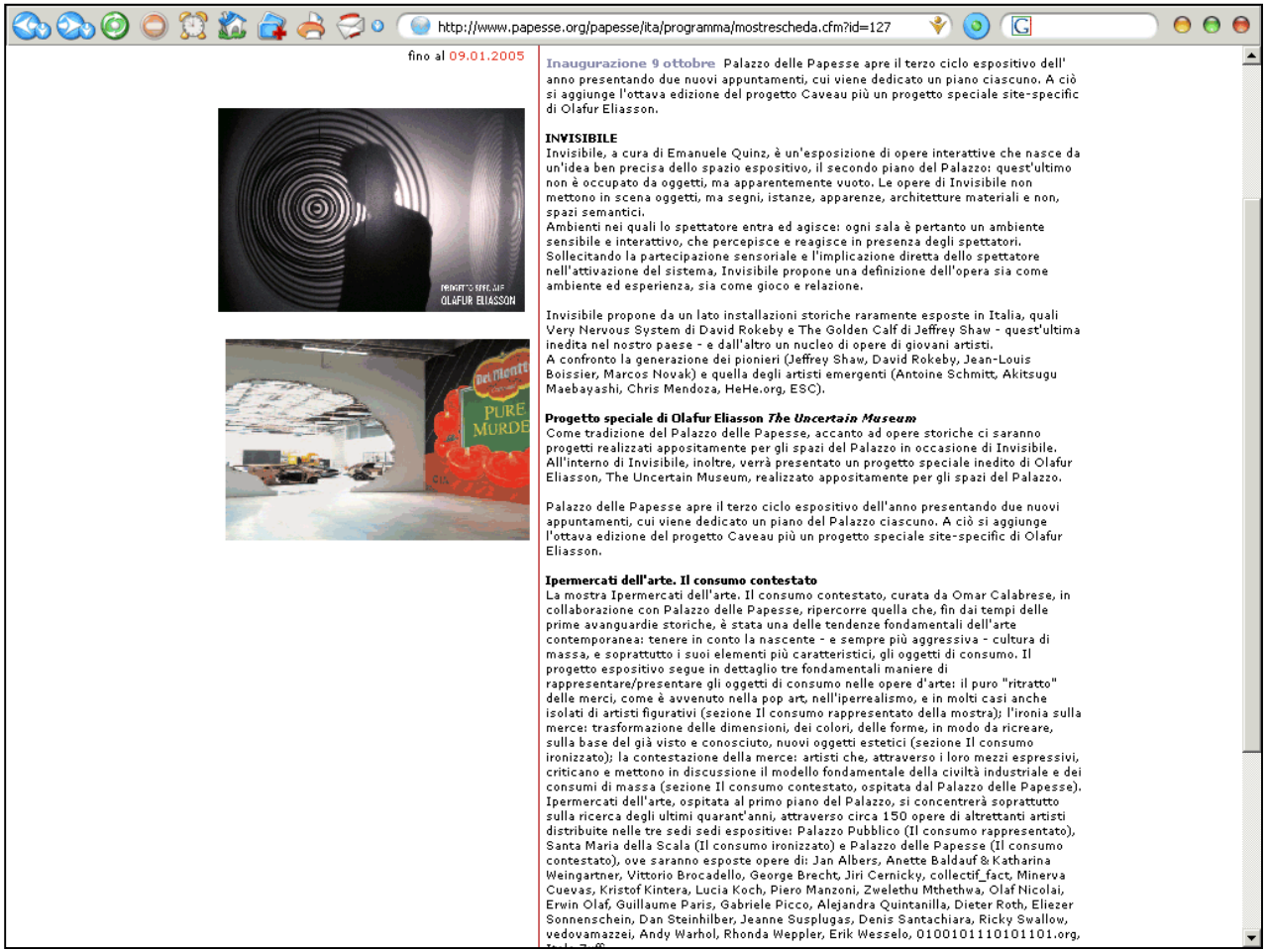
COGNITIVE OVERLOAD

See examples in the next slides

Examples: Content

Currency of Information

Readability (Text Conciseness)



Inaugurazione 9 ottobre Palazzo delle Papesse apre il terzo ciclo espositivo dell' anno presentando due nuovi appuntamenti, cui viene dedicato un piano ciascuno. A ciò si aggiunge l'ottava edizione del progetto Caveau più un progetto speciale site-specific di Olafur Eliasson.

INVISIBILE
Invisibile, a cura di Emanuele Quinz, è un'esposizione di opere interattive che nasce da un'idea ben precisa dello spazio espositivo, il secondo piano del Palazzo: quest'ultimo non è occupato da oggetti, ma apparentemente vuoto. Le opere di Invisibile non mettono in scena oggetti, ma segni, istanze, apparenze, architetture materiali e non, spazi semantici.
Ambienti nei quali lo spettatore entra ed agisce: ogni sala è pertanto un ambiente sensibile e interattivo, che percepisce e reagisce in presenza degli spettatori. Sollecitando la partecipazione sensoriale e l'implicazione diretta dello spettatore nell'attivazione del sistema, Invisibile propone una definizione dell'opera sia come ambiente ed esperienza, sia come gioco e relazione.

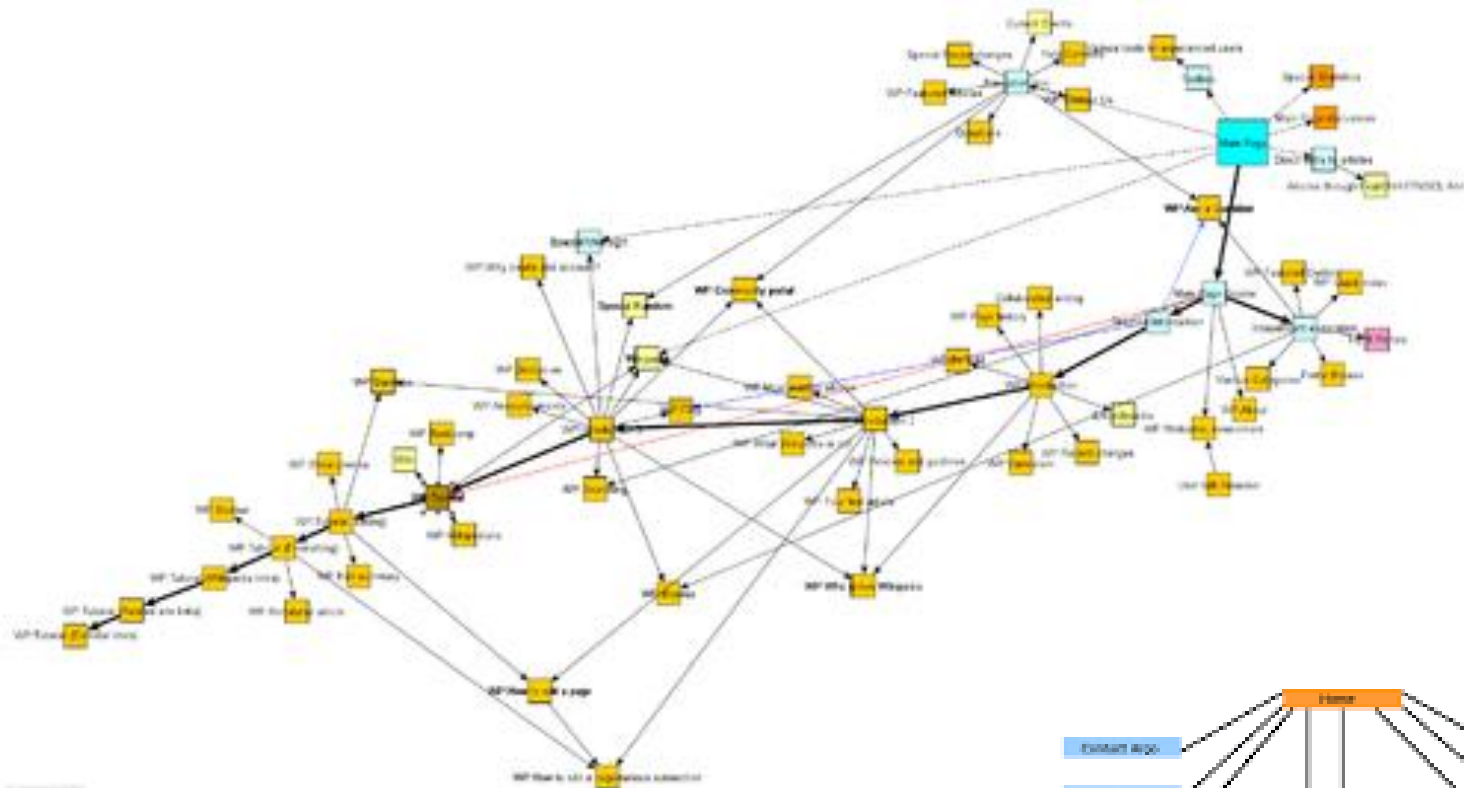
Invisibile propone da un lato installazioni storiche raramente esposte in Italia, quali Very Nervous System di David Rokeby e The Golden Calf di Jeffrey Shaw - quest'ultima inedita nel nostro paese - e dall'altro un nucleo di opere di giovani artisti.
A confronto la generazione dei pionieri (Jeffrey Shaw, David Rokeby, Jean-Louis Boissier, Marcos Novak) e quella degli artisti emergenti (Antoine Schmitt, Akitsugu Maebayashi, Chris Mendoza, HeHe.org, ESC).

Progetto speciale di Olafur Eliasson *The Uncertain Museum*
Come tradizione del Palazzo delle Papesse, accanto ad opere storiche ci saranno progetti realizzati appositamente per gli spazi del Palazzo in occasione di Invisibile. All'interno di Invisibile, inoltre, verrà presentato un progetto speciale inedito di Olafur Eliasson, The Uncertain Museum, realizzato appositamente per gli spazi del Palazzo.

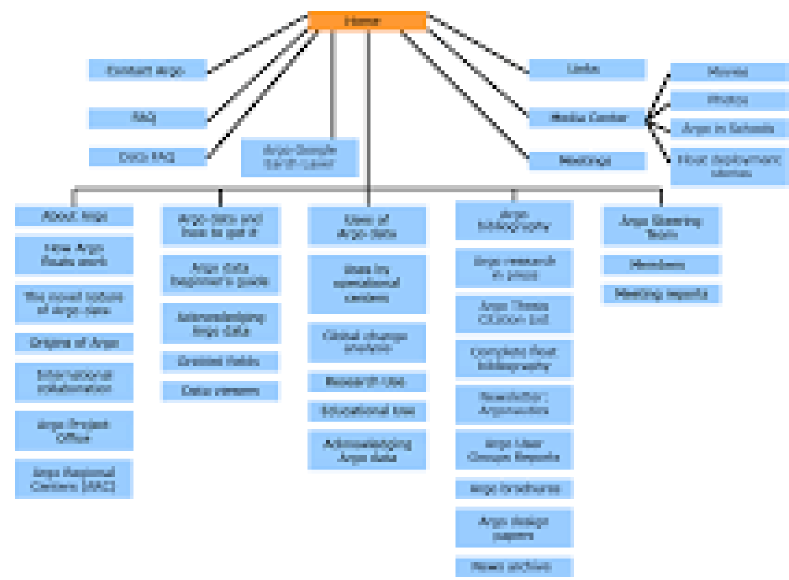
Palazzo delle Papesse apre il terzo ciclo espositivo dell'anno presentando due nuovi appuntamenti, cui viene dedicato un piano del Palazzo ciascuno. A ciò si aggiunge l'ottava edizione del progetto Caveau più un progetto speciale site-specific di Olafur Eliasson.

Ipermercati dell'arte. Il consumo contestato
La mostra Ipermercati dell'arte. Il consumo contestato, curata da Omar Calabrese, in collaborazione con Palazzo delle Papesse, ripercorre quella che, fin dai tempi delle prime avanguardie storiche, è stata una delle tendenze fondamentali dell'arte contemporanea: tenere in conto la nascente - e sempre più aggressiva - cultura di massa, e soprattutto i suoi elementi più caratteristici, gli oggetti di consumo. Il progetto espositivo segue in dettaglio tre fondamentali maniere di rappresentare/presentare gli oggetti di consumo nelle opere d'arte: il puro "ritratto" delle merci, come è avvenuto nella pop art, nell'iperrealismo, e in molti casi anche isolati di artisti figurativi (sezione Il consumo rappresentato della mostra); l'ironia sulla merce: trasformazione delle dimensioni, dei colori, delle forme, in modo da ricreare, sulla base del già visto e conosciuto, nuovi oggetti estetici (sezione Il consumo ironizzato); la contestazione della merce: artisti che, attraverso i loro mezzi espressivi, criticano e mettono in discussione il modello fondamentale della civiltà industriale e dei consumi di massa (sezione Il consumo contestato, ospitata dal Palazzo delle Papesse). Ipermercati dell'arte, ospitata al primo piano del Palazzo, si concentrerà soprattutto sulla ricerca degli ultimi quarant'anni, attraverso circa 150 opere di altrettanti artisti distribuite nelle tre sedi sedi espositive: Palazzo Pubblico (Il consumo rappresentato), Santa Maria della Scala (Il consumo ironizzato) e Palazzo delle Papesse (Il consumo contestato), ove saranno esposte opere di: Jan Albers, Anette Baldauf & Katharina Weingartner, Vittorio Brocadello, George Brecht, Jiri Cernicky, collectif_fact, Minerva Cuevas, Kristof Kintera, Lucia Koch, Piero Manzoni, Zwelethu Mthethwa, Olaf Nicolai, Erwin Olaf, Guillaume Paris, Gabriele Picco, Alejandra Quintanilla, Dieter Roth, Eliezer Sonnenschein, Dan Steinhilber, Jeanne Susplugas, Denis Santachiara, Ricky Swallow, vedovamazzei, Andy Warhol, Rhonda Weppler, Erik Wesselo, 0100101110101101.org,

Examples: Information architecture

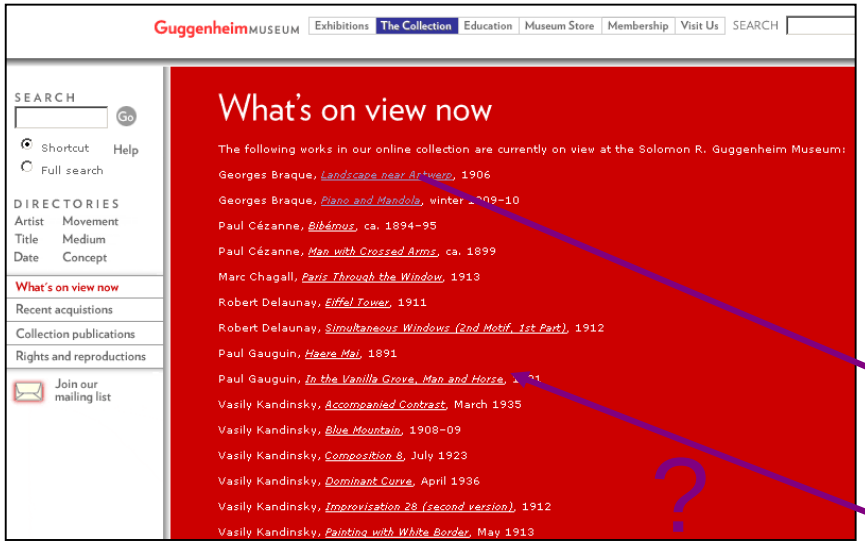



HOME
Download
Software
Video
Product
Where to buy
Support
Press Room
System Integrator
About us
PCI
PCI-E
Express Card
CardBus
USB
TV Box
Capture
Portable Media
FAQ
Document Download
Software Download
Contact Supporter
RMA
Knowledge Center
Press
Award
Event
Mini PCI & Mini Card
USB Pen drive
PCI
PCI-E
Express Card
CardBus
Overview
Executive
Milestone
Technology
site map


Home
Contact Argo
FAQ
Data FAQ
Links
Media Center
Meetings
Movies
Photos
Argo in Schools
Float deployment stories
About Argo
How Argo floats work
Origins of Argo
International collaboration
Argo Project Office
Argo Regional Centers (ARC)
Argo data and how to get it
Argo data beginner's guide
Acknowledging Argo data
Gridded fields
Data viewers
Uses of Argo data
Global change analysis
Research Use
Educational Use
Acknowledging Argo data
Argo bibliography
Argo research in press
Argo Thesis Citation List
Complete float bibliography
Argo brochures
Argo design papers
News archive
Argo Steering Team
Members
Meeting reports

Examples: Navigation/interaction

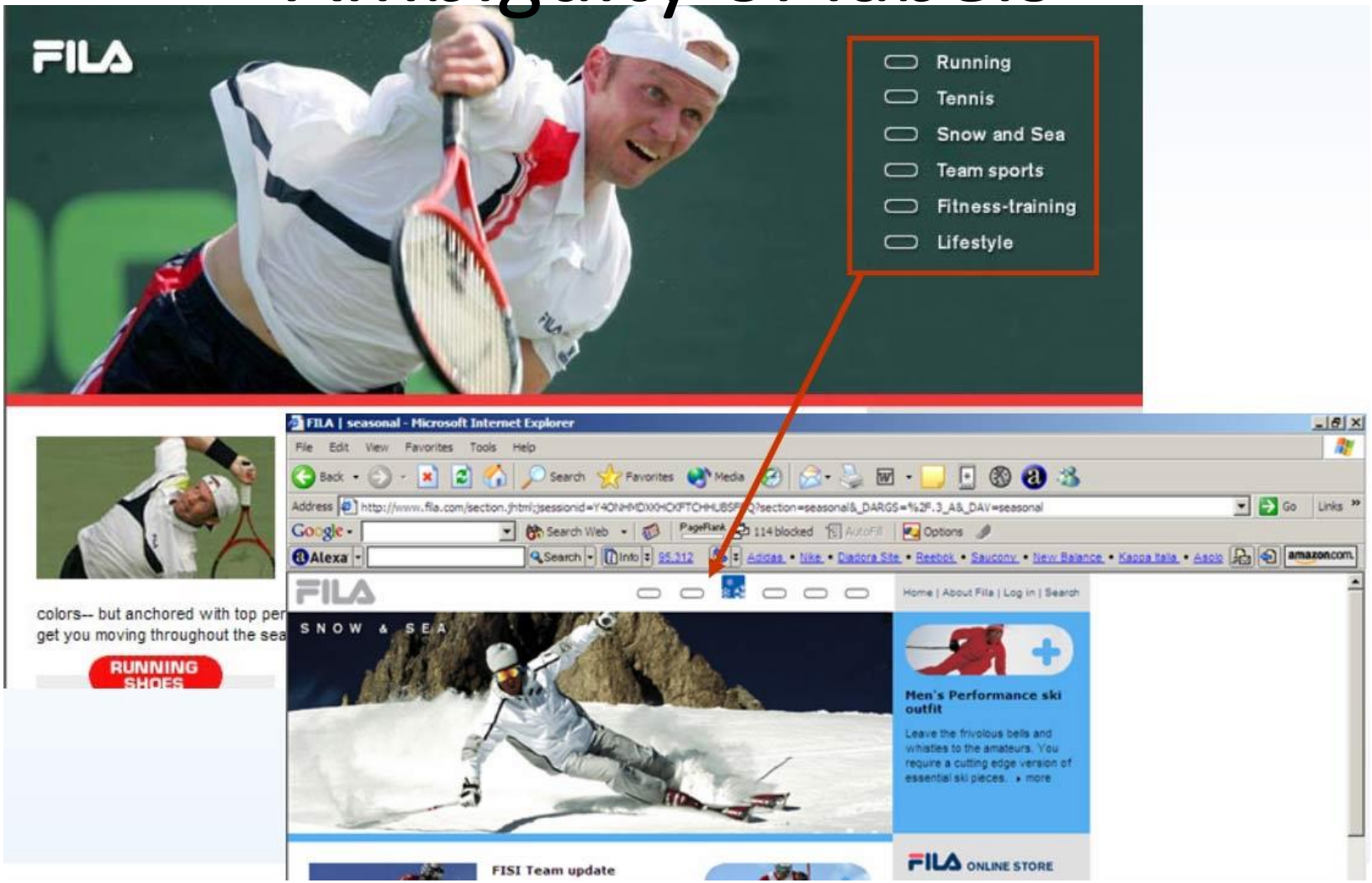
Backward Navigation



www.guggenheimcollection.org/site/on_view_now.html

Examples: Labelling/Semiotics

Ambiguity of labels

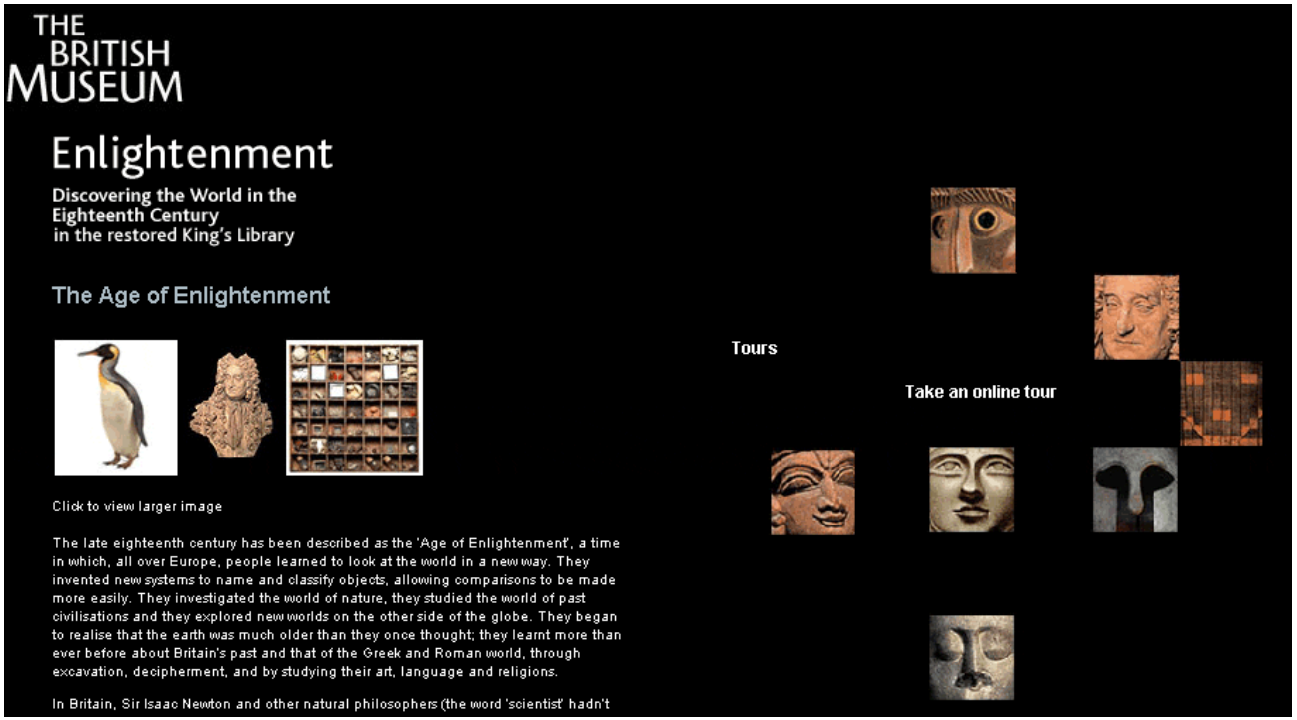


Ambiguity of labels

What happens when I click on the button "Tours"?

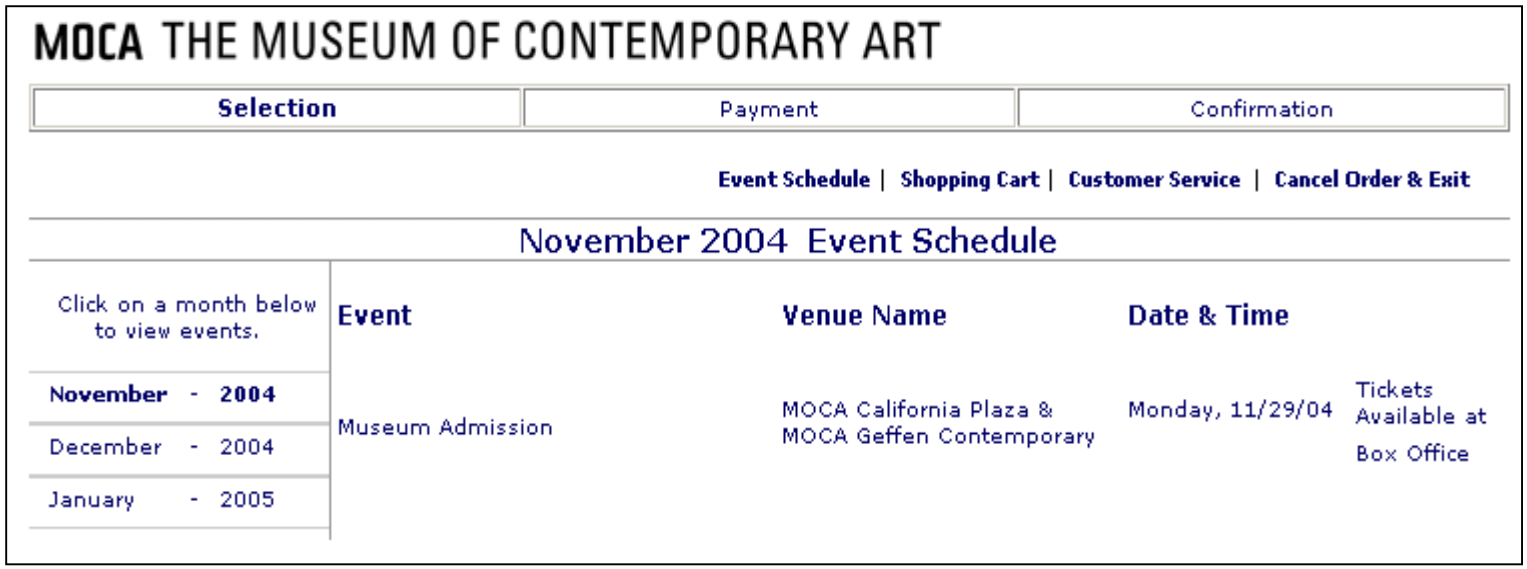
And on the button "Take a online tour?"

Which is the difference?



Examples: Graphics/layout

Anchor identity & Use of Cromatic Codes



MOCA THE MUSEUM OF CONTEMPORARY ART

Selection	Payment	Confirmation

Event Schedule | Shopping Cart | Customer Service | Cancel Order & Exit

November 2004 Event Schedule

Click on a month below to view events.

- **November - 2004**
- December - 2004
- January - 2005

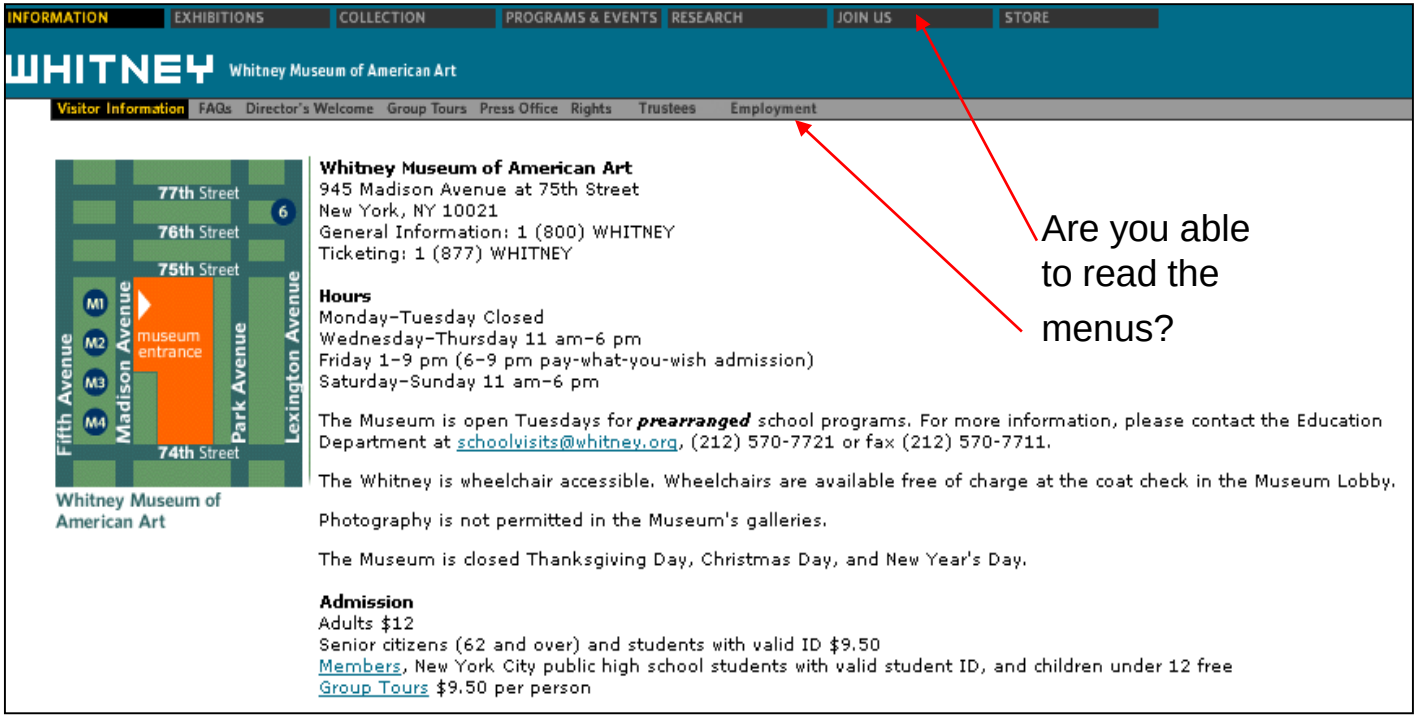
Event	Venue Name	Date & Time	
Museum Admission	MOCA California Plaza & MOCA Geffen Contemporary	Monday, 11/29/04	Tickets Available at Box Office

Which are links?

Background Contrast & Font size

Are you able to read the different information on the screen?

Menu Font Size



Examples: Errors management

System Reaction to User's Error(s)

Billing Information

Title: Mr.

*First name: pippo *Last name: pippo

*Street address: poiu

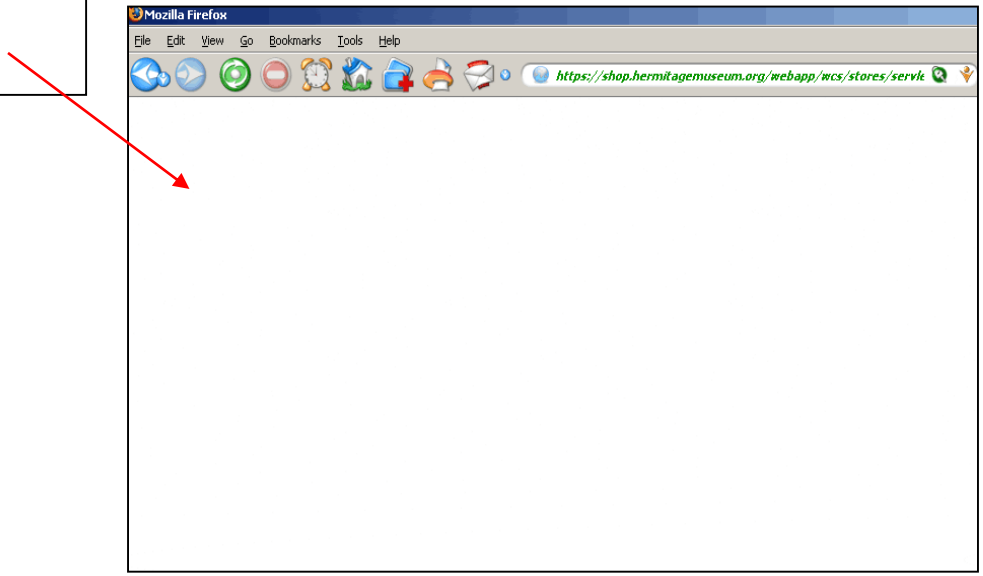
Address 2:

*City: pretews State/Prov: California

*ZIP/Postal code: 38992934 *Country: Belgium

*Daytime number: 1 *Evening number: 1

Fax: *E-mail address: pippo@pippo.com

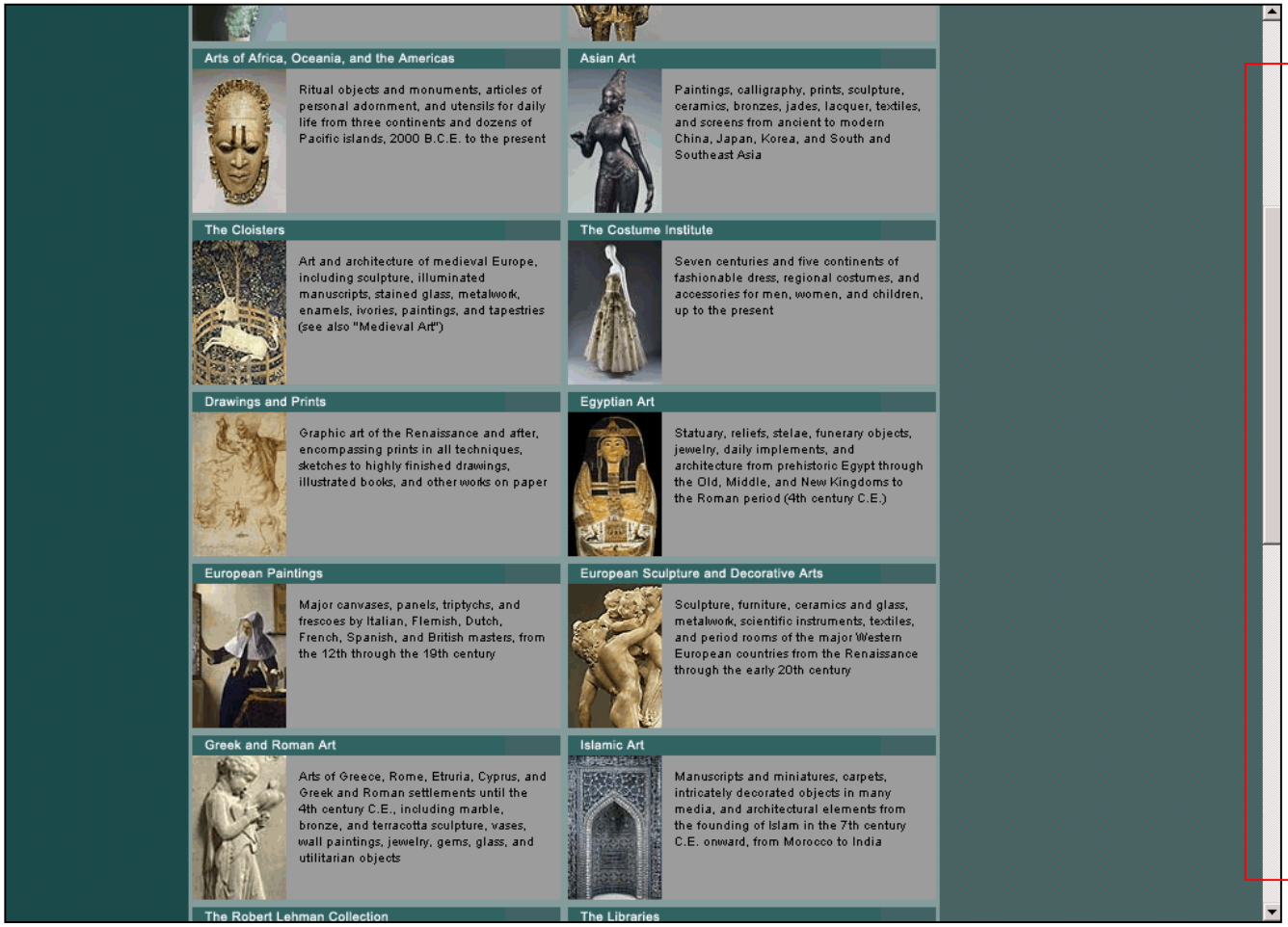


What is the error?

What did I do wrong?

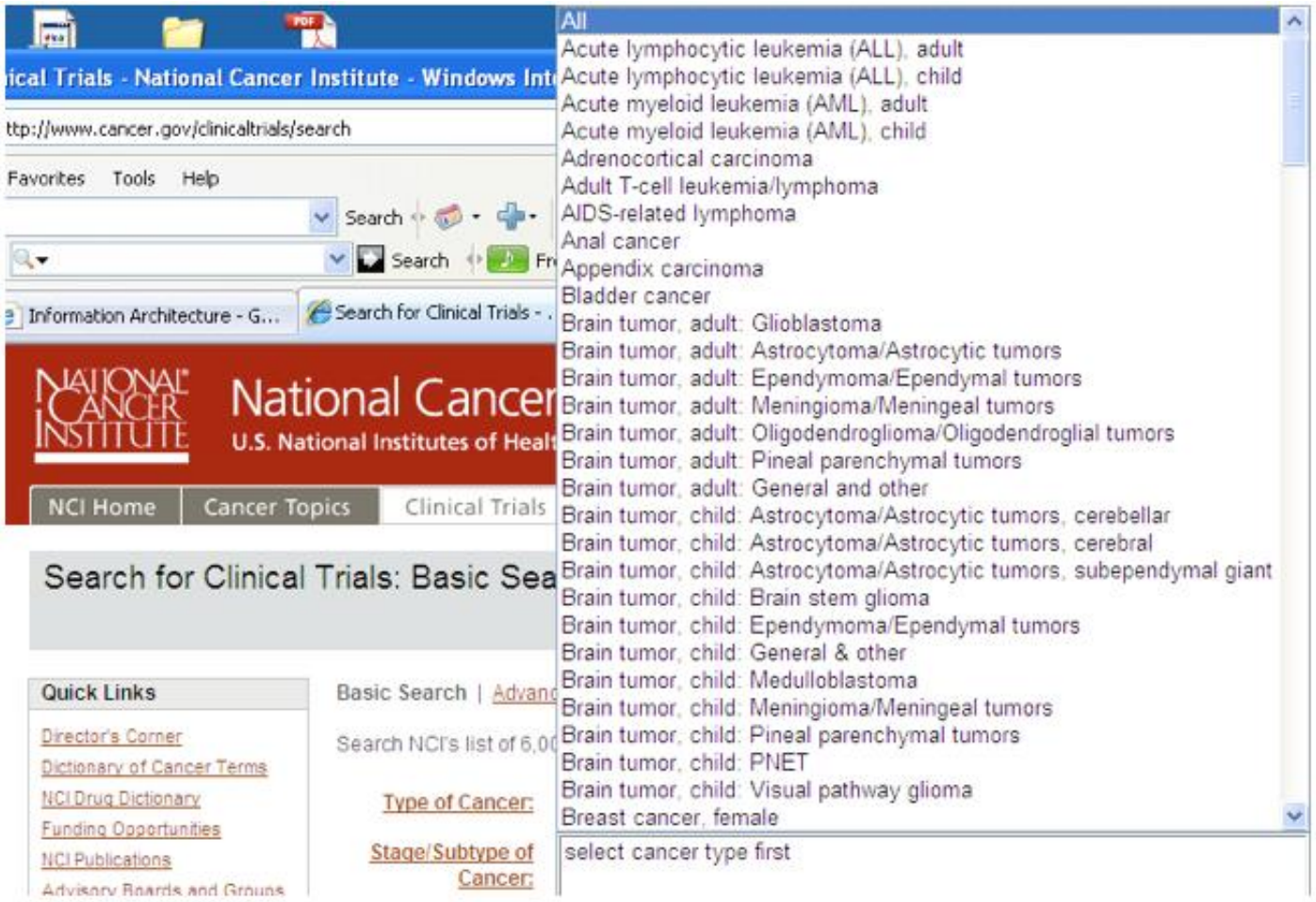
Examples: Cognitive overload

Information Overload (content)



Information Overload (menu)

Information Overload (menu)



How long is this?

Usability Evaluation methods

Who is involved?

Usability Experts →Inspection (Expert review)

End users → Usability Testing (UT)

Inspection methods - 1

- A category of methods based on **having expert evaluators** analytically examine usability-related aspects of an application.
- Systematic and in-depth analysis of the application
  - Inspectors (or expert reviewers) are often usability experts, preferable having some experience in the domain
- Inspection methods have achieved widespread use in industrial environments
- Nielsen (1994)

Inspection methods - 2

- Benefits
  - Costs/benefits ratio
    - Inspection can “save users”
    - No special equipments required
  - Quality of results
    - Inspectors can perform in-depth analysis
    - The inspector alone can detect most part of the usability problems of a complex system in a limited amount of time
  - Reliability of results
    - Method-based inspection can be repeated getting similar results

An inspection method: Heuristic evaluation - 1

- Heuristic-driven evaluation is based on checklists and usability principles
  - Heuristics = a set of principles that guide in the discovery (of the usability flaws)

- The quality of the web site is assessed against these principles
- Examples: The famous Nielsen's 10 heuristics (see 4.1)
  - *Visibility of system status, Match between system and the real world, User control and freedom, Consistency and standards, Error prevention, Recognition rather than recall, Flexibility and efficiency of use, Aesthetic and minimalist design*, …

User Testing (UT) Methods

User Testing (UT) Methods

- Usability properties are assessed by observing how the system is actually used by some representatives of **real users**
- User sessions are simulated in a **usability lab** to detect where users find obstacles, get disoriented, get anger, don't find information while trying to achieving their goals.
- On-the-field usability testing are performed **"in-situ"**
- "User experiences" are recorded and analyzed

Various Forms of User Testing

- Usability Testing has been a **fundamental practice** in HCI work for more than 30 years.
- There are various ***forms*** *(*or styles) of Usability Testing, whose selection depends on a number of factors:
  - Resources/infrastructure available
  - Meta-objective of the evaluation
  - Phase in the project/product lifecycle
  - Expertise of the persons involved
- Dozens of books on the topic

User Testing: Goal

- To **uncover actual difficulties** that users have when interacting with the application

- To obtain a **systematic feedback** on the effectiveness and efficiency of use of your product

Typical questions:

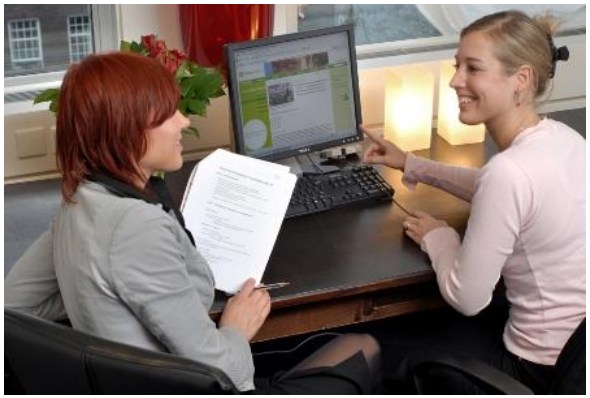
- Can users *complete* the expected tasks?
- Can they do it in *acceptable* times?
- What *difficulties* do they encounter?
- What is their *perception* of these difficulties?
- …

User Testing (UT): General approach

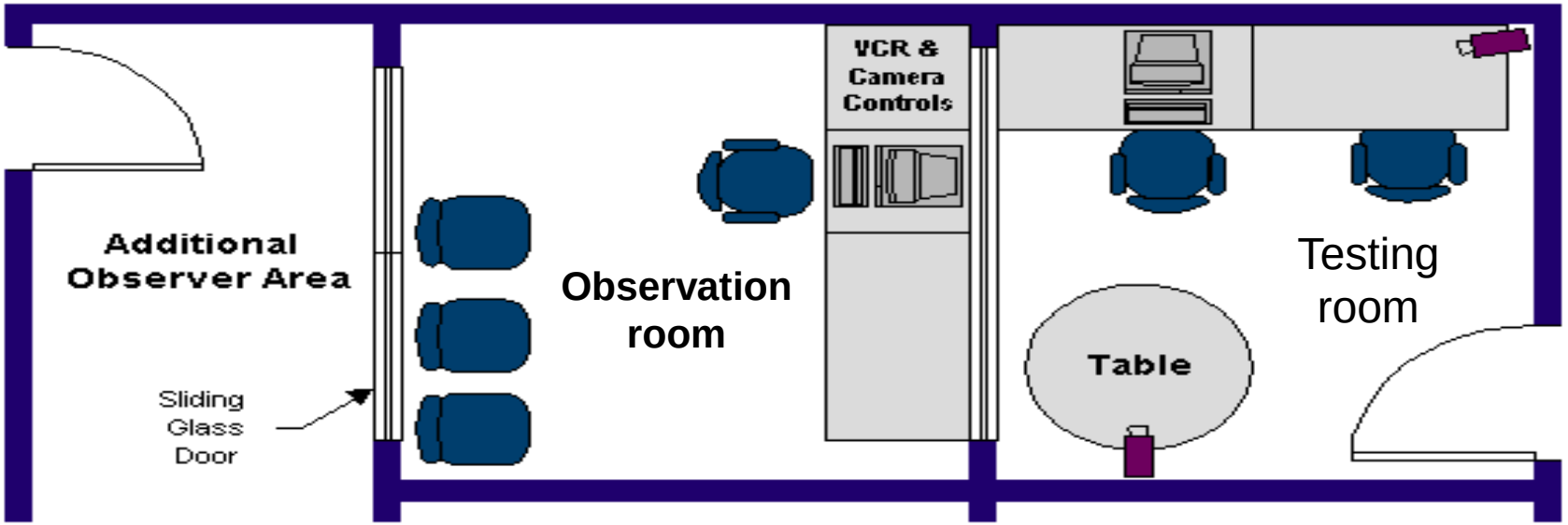
- Involving a **sample** of the "potential" or "actual" **target users** of the application
- Location: **Controlled set-up (e.g., lab) or "ecological" context** (the real environment of use for the studies system)
- "**Observing**" users in action, i.e., **gathering information** about users behaviors, judgements, and feeling resulting from the execution of a set of predefined activities
- Users are assigned some pre-defined **tasks** to perform on the application
- Gathering data DURING or AFTER the experience
  - **Quantitative data on performance** (e.g., time on task, number of errors, task success rate, …)
  - **Qualitative feedback**, e.g., responses to micro-interactions, what is liked/disliked, satisfaction, frustration, "conceptual fit", perceived benefits..

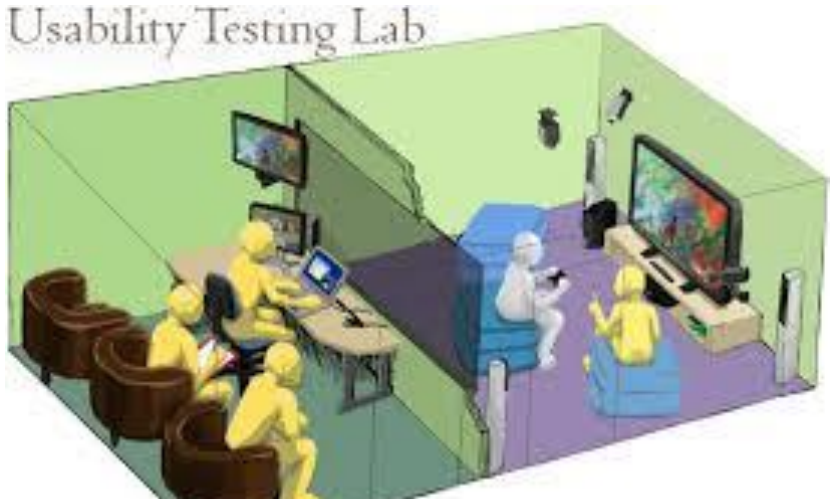
UT – Roles & Physical Setting

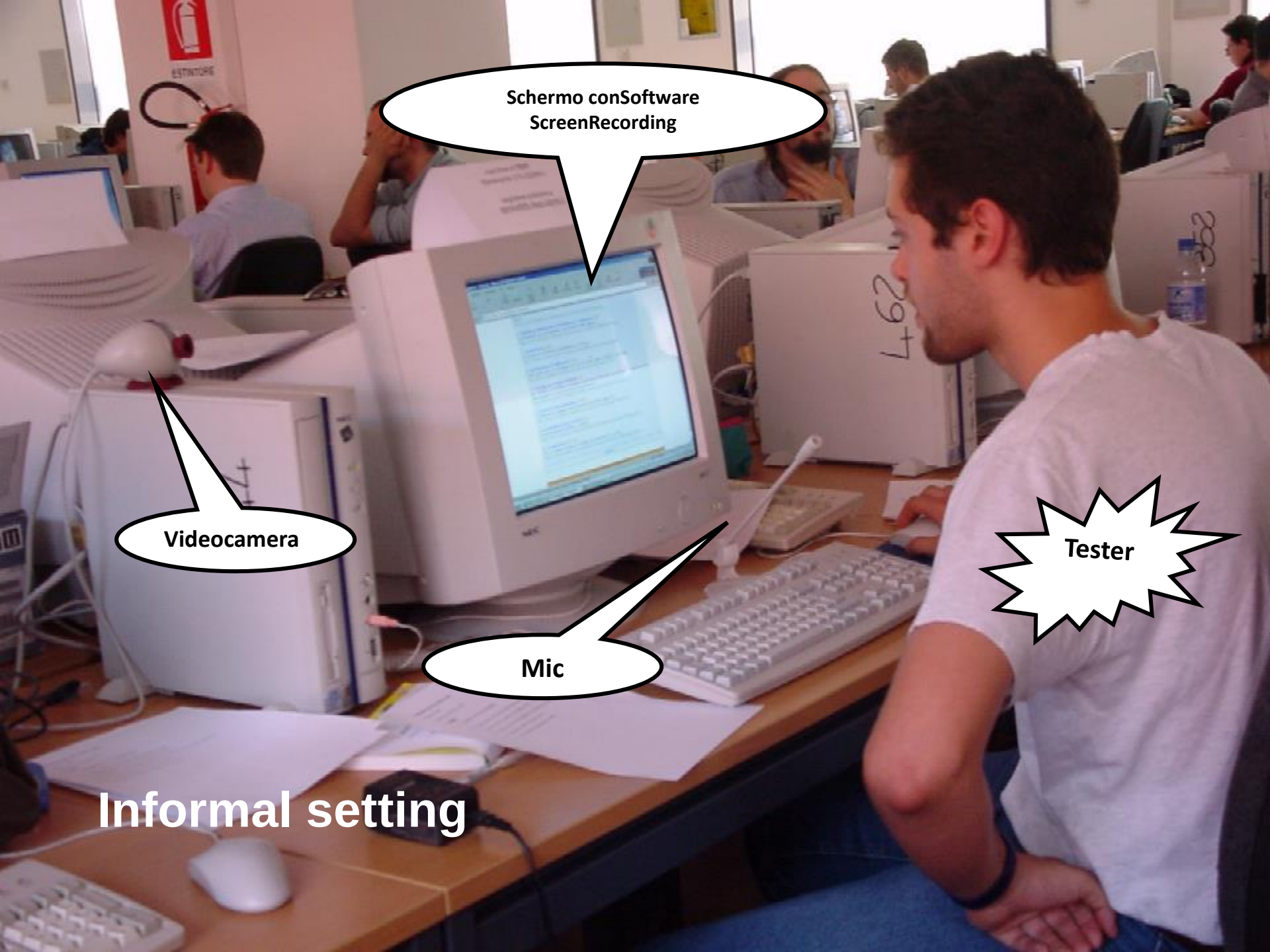
- 1-n Users
- 1 Moderator
- 1-n “Observers”



Usability lab







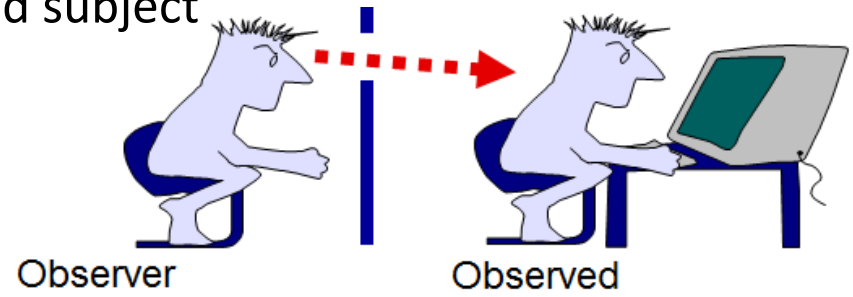
Informal setting

Roles / Attitude

- Users
  - They are not the object of the evaluation
  - They should not feel in any way evaluated or judged
  - They need to feel to be partners (valuable resources) to help you evaluate the application
  - *They are* (not the facilitator) the protagonists of the experience

Roles / Attitude

- 1 Moderator
  - The contact (accompanying) person for the user throughout the test
  - Also called “facilitator”
  - Manage the test (logistics, dialogue, pace) and coordinate its successful completion
  - The general attitude of the moderator is to ***elicit*** **and** ***observe*** **the experience**, **not to lead to consensus or get agreement**.
  - **Refrain from making your point (it does not matter) on the site, on the design, on the test or on the user**. The user should not understand *in any way* whether you agree/disagree with his/her opinion, or whether you like or criticize the design
  - Avoid to interfere with the observed subject



Designing a usability study

Recruiting users

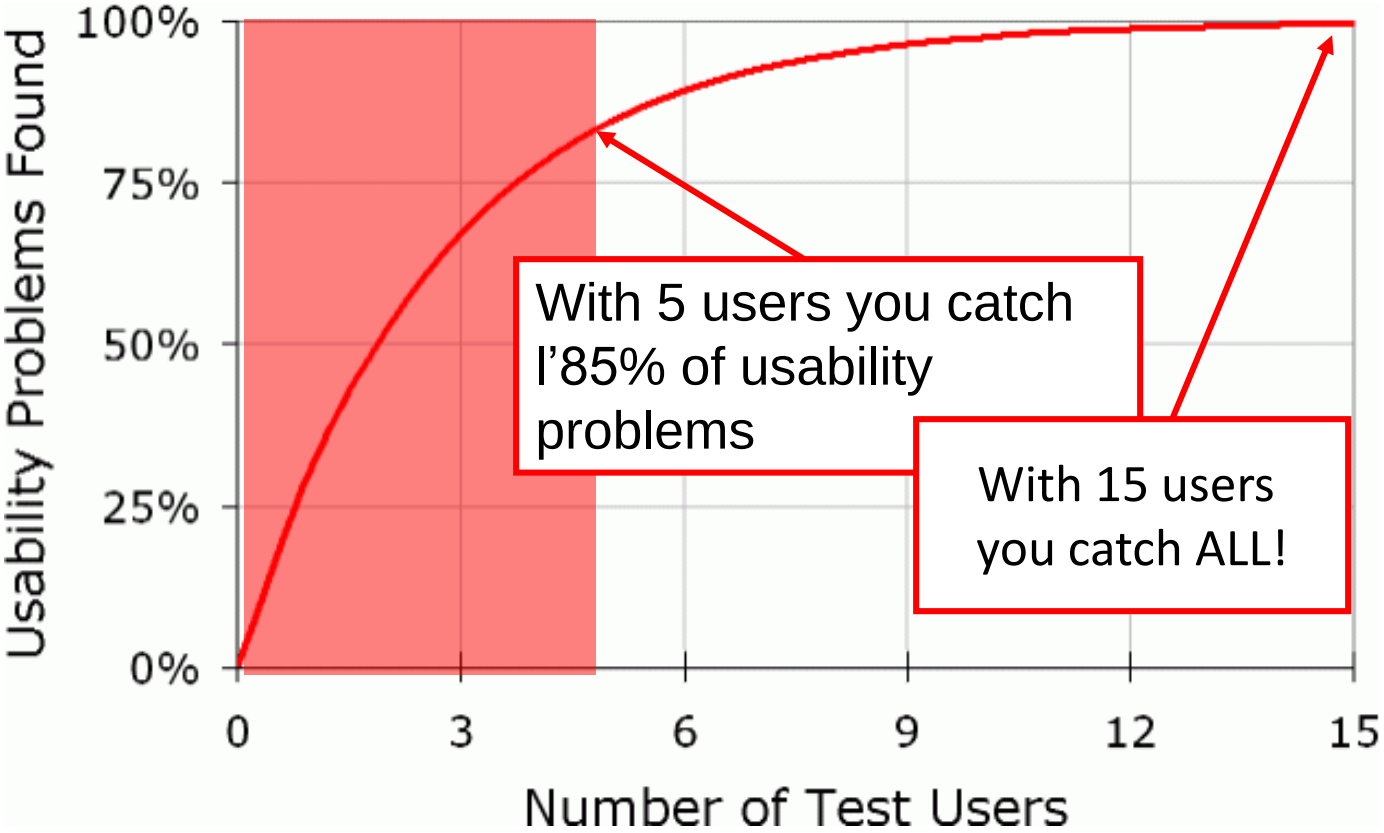
- Define user *profiles* and user *goals* to segment your target audience and recruit users
- Example:
  - Profiles: Young man, married woman, retired, etc..
  - Goals: Willing to pay bill online, to check account, to evaluate the possibility of buying new products
- Recruiting the “right” users is crucial for the usefulness of the results
- Users should correspond to user profiles
- Ideal user:
  - someone who is actually going to need the (type of) application in the near future or is familiar with the *genre* of application

Recruiting users

- How many? Debated issue in the field
- 6-12 is a typical range used
  - Keeping constant tasks / user profile
  - 10 users per condition (*user profile* or *user goal*) is a typical sample size to start seeing consistent results
- "the more the better", BUT
- many "quick and small" tests better than one "large" test

- TIPS:
  - Avoid mixing *first time users* and *frequent users*
  - Do not take users who are not familiar with interacting with computers/web in general
  - You can use more "generic" profiles (e.g. extending the age range, but keep the tasks constant)
  - Set out to get *comparable* results across participants

Nielsen's Rule



J.Nielsen, "Why You Only Need to Test With 5 Users", http://www.useit.com

Main phases

A. Preparation

B. Execution

C. Analysis of results

Preparation

1. Define test GOALS
2. Define qualitative and quantitative data to collect
3. Define scenarios and tasks to assign to the users
4. Define how to recruit users
5. Build the materials (data collection forms)
6. Create sw and hw setting
7. “Test the test” with a mini-sample
8. If needed, go-back and re-design the test

Define scenarios

- Define a set of scenarios at the **proper level of abstraction**

- Prioritize scenarios on the basis of:
  - Relevance to application's mission
  - Relevance to stakeholders
  - Most frequently used
  - Focusing on "new" features
  - Focusing on "highly publicized" features
  - Gravity of usability problems detected during inspection
  - . . .
- Select the most important scenarios (2-3) for each user profile and/or context of use

Selecting Task Scenarios

- Select/elaborate **high-quality task scenarios** on the basis of:

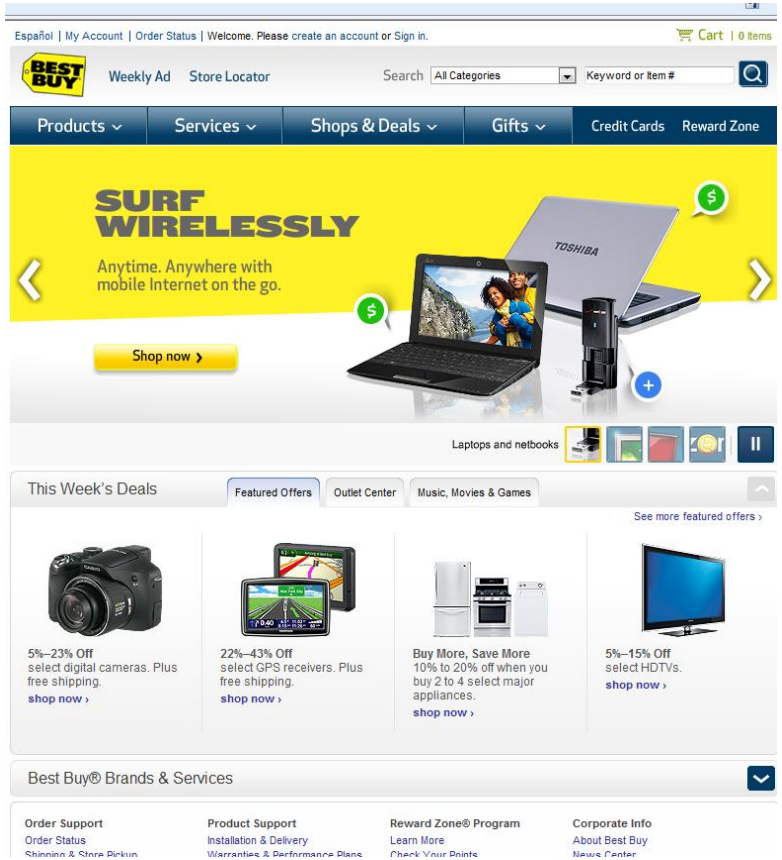
  A. Relevance of the **tasks** to the application's mission / stakeholders
     - *Objective*: test whether and how well users are able to accomplish MUST-DO, critical tasks, crucial to the goals of the application

  B. Severity of the **problems** identified during inspection
     - *Objective*: see how users perform in those critical situations, if they fail completely, have difficulty, or find workaround to succeed in accomplishing critical tasks with sever (anticipated) problems

Elaborating tasks

- Assigned Tasks should be:
  - Meaningful and reasonable
  - Motivating
  - Goal-oriented
  - Not revealing tips
  - Supported
  - In a realistic sequence
- Relating tasks to a narrative scenario helps “sense making” for the participants
- Elaborate a **“warm up”/exploratory task** (not relevant scenario, but to get familiar with the application)
- *Randomize/alternate* the **order** of tasks (to minimize learning effect on the same task)
  - [Task 1 always worst performance of all….?]

Bad Task Formulation



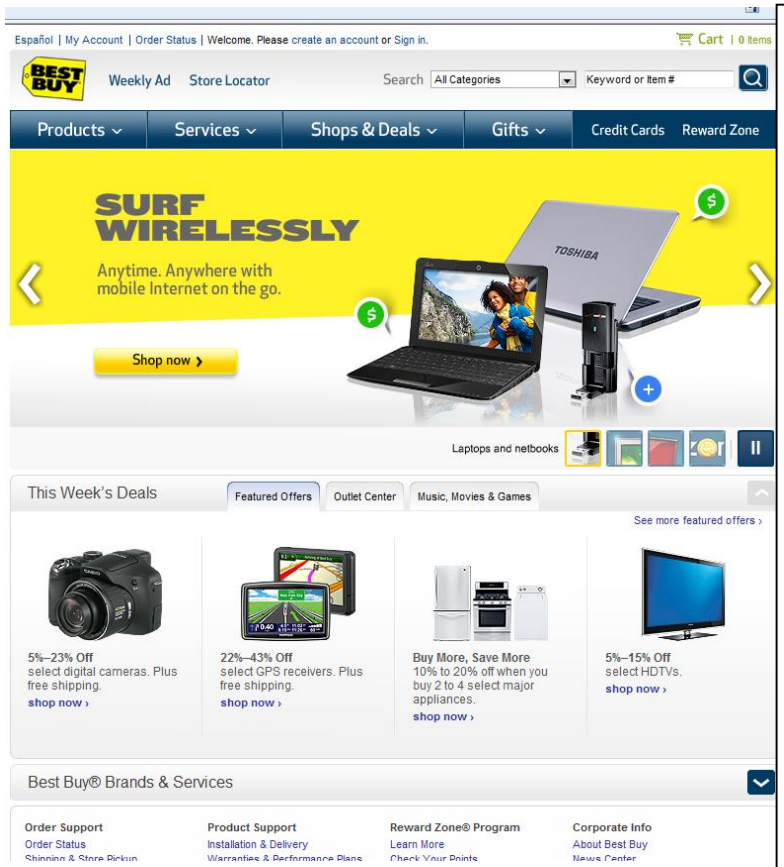
Task 1:
Click on surf wireless
to explore the current offers

Task 2:
Find out the Services offered
by Best Buy

Task 3:
Find if it is possible to reserve a
Digital
Camera, and then pick it up
at the store.

Task 4:
Browse through the home theater
products and select a LCD TV of
your choice.

Task 5:
Is there a gift you would like to give
to
someone?



Bad Task Formulation

Task 1:
Click on surf wireless
to explore the current offers

Task 2:
Find out the Services offered by Best Buy

Task 3:
Find if it is possible to reserve a Digital
Camera, and then pick it up at the store.

Task 4:
Browse through the home theater products and select a LCD TV of your choice.

Task 5:
Is there a gift you would like to give to
someone?

Improved Task Formulation

Task 1:
Find the most expensive digital camera on sale and identify how much you are going to save

Task 2:
Your iPod does not work properly. Find a way to have it fixed.

Task 3:
Find the latest Apple laptop, and reserve it for later pick up at the store.

Task 4:
Find the technical specifications of the cheapest 50' LCD TV

Task 5:
You would like to find something nice to give to your dad for his birthday. Find it and put it in your shopping cart.

Elaborating tasks

- Task length
  - 5-6 tasks per user profile
  - Ca. 45 mins/1 hour per participant is a typical time to spend for a test

- Prepare the instrumentations (papers) for data collection (see following slides “Procedure for data collection and Templates”)

Data to Capture

- Quantitative indicators:
  - Effectiveness (task success rate)
  - Efficiency (time on task)
  - Errors (wrong paths or actions)
  - Perceived tasks difficulty
- Qualitative indicators:
  - Disorientation, stops, frustrations, waiting periods, wandering periods
  - Unexpected behaviors
  - Satisfaction,
  - Engagement
  - …

Equipment

- Tools useful for:
  - Recording user sessions (screen)
    - Interface screen
    - User's voice
    - (Recording expert's voice)
  - Recording user behavior (face and moves)
- Camera and mic
- Testing software - examples:
  - Camtasia - www.techsmith.com
  - CamStudio - http://camstudio.org/

Try it out! Conduct a pilot test

- Pilot test everything before rolling out
- The specific instances of material and procedure you are using needs to be tested
- Little mistakes must not affect the ENTIRE set of testing session – *discover them early*
- Pre-test and practice at least ONCE ALL the procedure, material and equipment with one user (cheap to recruit)
- Revise and correct material, procedure, organization and flow

Execution

- User: executes the task
- Moderator and Observers: observe and gain data

Introduce and Explain Purpose

- Describe the **purpose of the observation** (in general terms). Set the participant at ease by stressing that you're trying to find problems in the product.
- For example, you could say:
  - **You're helping us** by trying out this product in its early stages.
  - **We're looking for** places where the product may be difficult to use.
  - If you have trouble with some of the tasks, it's the **product's fault, not yours.** Don't feel bad; that's exactly what we're looking for.
  - If we can locate the trouble spots, then we can go back and improve the product.
  - Remember, we're testing the product, not you.

Explain Possibility to Quit

- If applicable, have each participant sign an *informed consent (*IRB-approved*)*
- Make sure you inform participants that **they can quit at any time** if they find themselves becoming uncomfortable.
- Participants shouldn't feel like they're **locked into completing tasks**. Say something like this:
  - Although I don't know of any reason for this to happen, if you should become uncomfortable or find this test objectionable in any way, you are free to quit at any time.

Explain Setting

- Talk about the **equipment** in the room.
- Explain the purpose of each piece of equipment
  - hardware, software, video camera, microphones, etc.
  - and how it is used in the test

Explain “Thinking Aloud”

IF THINKING ALOUD IS ONE OF YOUR DATA COLLECTION TECHNIQUES

- Explain how to “**think aloud**”. Ask participants to think aloud during the observation, saying what comes to mind as they work.
- By listening to participants think and plan, you can see their expectations for your product, as well as their **intentions, explicit reactions and their problem solving strategies**.
- You could say:
  - We have found that we get a great deal of information from these informal tests if we ask people to think aloud as they work through the exercises.
  - All you do is speak your thoughts as you work.
  - If you forget to think aloud, I'll remind you to keep talking.

Note on Thinking aloud

method invented in the 30ies by pyschologist Édouard Claparède (1873-1940)

- Basically: the moderato asks the user to say what he or she is thinking while performing a given task:
  - What is (s)he doing?
  - What does (s)he see on the screen
  - How does (s)he think (s)he should proceed?
  - Which are her/his doubts or difficulties?
  - Which are her/his feelings?
  - …

Assign Tasks

- Introduce the product and describe the tasks.
- Explain what the participant should do and in what order.
- Give the participant written instructions for the tasks
  - Use formal printed papers or (clearly handwritten) “task cards”
  - “write-in” sheet or multiple choice

- Ask if there are any questions before you start

then

- begin the task execution and observation

Observe

- You just *observe* and make non-leading *questions* – everything else is not useful
- During observation:
  - Track time on task
  - Video and audio record
  - Watch carefully and take notes
  - Ask questions **at the end of the task** if you feel necessary to explore important issues or reactions
  - Be neutral and objective

Observe – tips

- Avoid “praising” the user
  - "How are you doing?“
  - *"Good Job! You're doing great!"*
- Do NOT ask:
  - *"Do you like this feature?“*
    - Focus groups are for eliciting opinions, usability tests focus on eliciting **behaviors**
  - *"Is this what you were expecting to be on this page?“*
    - *Typical answer: “Oh, I don’t know, I guess so, ...”*

Track Task Success

- **Task Success** (or task completion) is an indicator of the effectiveness of the interaction
- Various levels of success
  - **Complete success**
    - With assistance
    - Without assistance
  - **Partial success**
    - With assistance
    - Without assistance
  - **Failure**
    - Participant thought it was complete, but it wasn't
    - Participant gave up

Success rate metric: example

	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6
User 1	F	F	S	F	F	S
User 2	F	F	P	F	P	F
User 3	S	F	S	S	P	S
User 4	S	F	S	F	P	S

Note: S = success, F = failure, P = partial success

Success rate: (9 + (4*0.5)) / 24 = **46%**

- successful trials
- partially successful
- total trials

“Giving Assistance”

- Moderator **may** give some assistance – but not intrusively
  - Carefully consider to **WAIT some time** before providing assistance (and do so only when the situation is really problematic)
- RECORD the situations where assistance was needed
- Examples of assistance:
  - Moderator takes the participant **back to a homepage** or resets to an initial state.
  - Moderator **restates** the task (the user forgot)
  - Providing **hints**
    - E.g. “Why not try the ‘plan your visit’ tab?”
  - Moderator **answers a question** or provides information.

Scoring Methods for Task Success

- Method 1
  - Complete success (without assistance) = 1.0
  - Partial success, or if assistance given = 0.5
  - Gives up or wrong answer = 0.0
- Method 2
  - 1 = No problem
  - 2 = Minor problems
  - 3 = Major problems
  - 4 = Failure/gave up

Determining Unsuccessful Tasks (0 points) - “Stopping Rules”

- Rule 1
  - The participant **should continue to work** on each task until they either complete it or reach the point at which they give up/wrong answer or seek assistance.
- Rule 2
  - “**Three strikes** and you’re out.”
    - Three wrong paths, or three attempts from the start.
- Rule 3
  - Define **cut-off time** (threshold) based on precise design requirements (typically not specified)
    - E.g. Find a way to fix your iPod [cut-off: 4 minutes]
  - Then:
    - “Call” the task after cut-off time
    - Better: *Follow Rule 1 but record Failure after threshold*

Track Time-on-Task

- **Time-on-task** is typically used as measure of efficiency for most interactive products (where efficiency is crucial)
- While the absolute time might not be useful per se, it **allows analysis of outliers and patterns** as a diagnostic tool.
- It can be used to compare different versions of the product
- Tools for measuring Time-on-Task
  - Stopwatch
  - Videotaping
- Clarify the rules for turning on and off the clock (e.g. stop when assistance given, and then resume after assistance has been given)

Track Time-on-Task (cont.)

- It is defined as the time that goes from the moment the user:
  - *has finished* reading the task
  - *has understood it* (no more questions asked)
  - *directs* his/her attention to the application
- Until the moment that one of the predefined stopping rules is met
- **IF used with "Thinking Aloud"**, make sure to keep users **focused on the task** (and ***not on off-task commentary***)

Capture Errors

- Errors are incorrect actions that may lead to task failure
- What constitutes an error:
  - Entering incorrect data into a form field
  - Making the wrong choice in a menu or drop-down list
  - Taking an incorrect sequence of actions
  - Failing to take a key action
- Why to measure errors
  - An error will result in a significant inefficiency
  - An error will result in significant costs
  - An error will result in task failure

Capture Errors (cont.)

- For specific applications – where interactions is very constrained – it is useful to define a priori the error opportunities for each task
  - Error opportunity = a chance to make a mistake
- E.g. Task #4 – Insert a picture into a word document
  - Error opportunity 1: click a wrong button
  - Error opportunity 2: insert a wrong picture
  - Error opportunity 3: insert the picture into a wrong place
- Then, you take notes of how many actual errors users do in a given task (e.g. 5) and you calculate an error rate per task (e.g. 5 out of 48 → 10%)
- Defining error opportunities before the UT (in the design phase) may facilitate data collection and reporting

Post-TASK questions: Quantitive

- It may be of value to measure the **perceived difficulty** of a task **after each task has been performed**
- Post-TASK questions are straightforward to answer (fresh memory of the task experience) and provide additional diagnostic information that post-test questionnaires (at the end of the session) do not provide
- EXAMPLES → **use only one**!

Overall, this task was:

Very easy	□	□	□	□	□	□	□	**Very difficult**

I was satisfied with the ease of completing this task:

Strongly Disagree	□ 1	□ 2	□ 3	□ 4	□ 5	□ 6	□ 7	**Strongly Agree**

I was satisfied with the amount of time it took to complete the task:

Strongly Disagree	□ 1	□ 2	□ 3	□ 4	□ 5	□ 6	□ 7	**Strongly Agree**

Post-TASK questions: Qualitative

- Qualitative insights can be gained after each task
  - Ask about **any interesting behaviors** you would like the participant to explain to you
  - Go back and **re-visit pages** if necessary to ask about specific issues you observed

Interview “Post-Test”

- When the test is over (all tasks have been performed):
  - Answer any remaining questions the participant may have
  - Ask additional questions in open-ended or close format to:
    - elicit self-reported reflections on the overall experience
    - elicit response on *specific aspects* of the design
    - See examples in Templates and Kuniavski’s.

Examples of Templates for Data Collection

Hand it out to participant

DO KEEP TRACK of Identifiers for your tests!

Insert Task Formulation

TASK SHEET

For the Subjects

User Number: _____ Track #: _____

Site Address: _____

Evaluator(s) _____ Date _____

Scenario General Introduction: ……

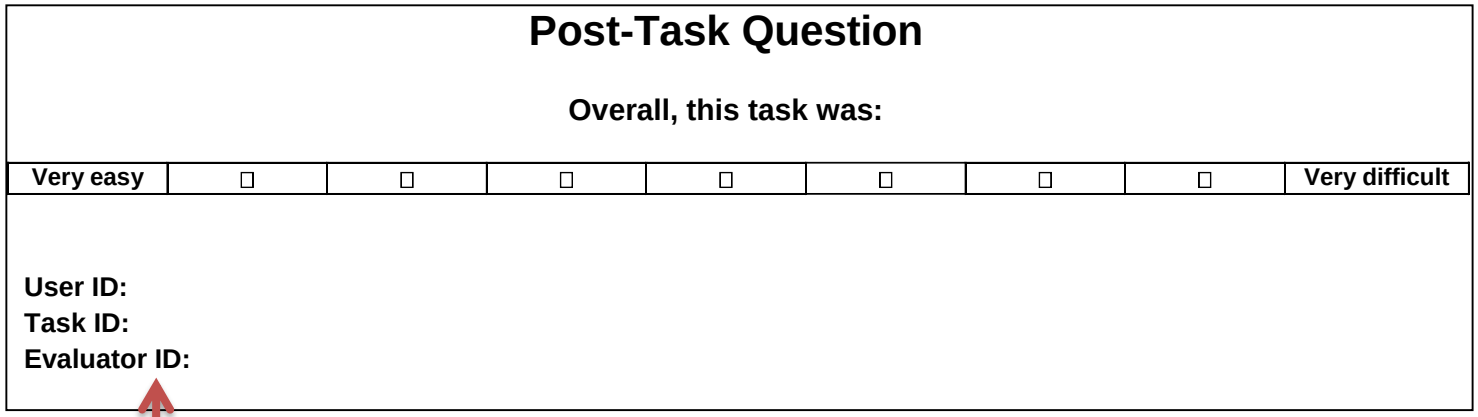
TASK NO.	TASK DESCRIPTION
1	*Task narrative….* *+ (for answer -based tasks: ANSWERS check list).*
2	
3	
4	
5	

Task Answers Formats

- Structured Answer
  - What is the ticket price of the Children's Museum?
    ___$10 ___$11 _X__$12 ___$13 ___$14

- Write-in Answer
  - Find the earliest time you could enter the museum on Friday.
    __*4pm*_____.

Post-Task Question



DO KEEP TRACK of Identifiers for your data

Hand it out to participant (separately from task description) AFTER each task

EVALUATOR SHEET Task Record Sheet Example 1 – PER INDIVIDUAL PARTICIPANT

Participant ID: **Evaluator ID:**

Task	Start Task	End Task	Task Completion	Elapsed Time	Time on Task (mins)	Comments/ Observations
1	10:29:06 AM	10:34:09 AM	Complete	0:05:03	5.0	
2	10:35:18 AM	10:43:21 AM	Incomplete	0:08:03	8.1	
3	10:43:28 AM	10:50:30 AM	Complete **with assist**	0:07:02	7.0	
4	10:51:05 AM	10:56:02 AM	Complete	0:04:57	4.9	
5	10:57:15 AM	10:59:50 AM	Complete	0:02:35	2.6	
6	11:01:06 AM	11:09:45 AM	Complete	0:08:39	8.6	

Task Record Sheet Example 2

		Participants				
TASK		1	2	3	4	5
1	Task Time: Min. & Sec.					
	Task completed: Yes or No					
	Comments on observed behavior.					
2	Task Time: Min. & Sec.					
	Task completed: Yes or No					
	Comments on observed behavior.					
3	Task Time: Min. & Sec.					
	Task completed: Yes or No					
	Comments on observed behavior.					
4	Task Time: Min. & Sec.					
	Task completed: Yes or No					
	Comments on observed behavior.					
5	Task Time: Min. & Sec.					
	Task completed: Yes or No					
	Comments on observed behavior.					
	TOTAL					

POST-TEST QUESTIONNAIRE EXAMPLE

DEEP: DEsign-oriented Evaluation of Perceived Web Usability

Name:_____ Task:_____ Date:_____

Perceived Content

The wording of the text was clear

[ ] Strongly Disagree [ ] Disagree [ ] Neutral [ ] Agree [ ] Strongly Agree [ ] Not Applicable

The content (including text, pictures, audios, and videos etc.) was easy to understand

[ ] Strongly Disagree [ ] Disagree [ ] Neutral [ ] Agree [ ] Strongly Agree [ ] Not Applicable

The text was useful

[ ] Strongly Disagree [ ] Disagree [ ] Neutral [ ] Agree [ ] Strongly Agree [ ] Not Applicable

The text was relevant

[ ] Strongly Disagree [ ] Disagree [ ] Neutral [ ] Agree [ ] Strongly Agree [ ] Not Applicable

Perceived Structure and Information Architecture

I could quickly get to know the structure of the website by skimming its homepage

[ ] Strongly Disagree [ ] Disagree [ ] Neutral [ ] Agree [ ] Strongly Agree [ ] Not Applicable

The organization of the website was clear

[ ] Strongly Disagree [ ] Disagree [ ] Neutral [ ] Agree [ ] Strongly Agree [ ] Not Applicable

Under each section of the website, the web pages were well organized

[ ] Strongly Disagree [ ] Disagree [ ] Neutral [ ] Agree [ ] Strongly Agree [ ] Not Applicable

Perceived Navigation

It was easy to find the information I needed on the website

[ ] Strongly Disagree [ ] Disagree [ ] Neutral [ ] Agree [ ] Strongly Agree [ ] Not Applicable

This website helped me find what I was looking for

[ ] Strongly Disagree [ ] Disagree [ ] Neutral [ ] Agree [ ] Strongly Agree [ ] Not Applicable

I got what I expected when I clicked on things on this website

[ ] Strongly Disagree [ ] Disagree [ ] Neutral [ ] Agree [ ] Strongly Agree [ ] Not Applicable

DEEP Questionnaire Cont.

Perceived Cognitive Effort						
Using this website was effortless	[ ] Strongly Disagree	[ ] Disagree	[ ] Neutral	[ ] Agree	[ ] Strongly Agree	[ ] Not Applicable
Using this website made me feel tired	[ ] Strongly Disagree	[ ] Disagree	[ ] Neutral	[ ] Agree	[ ] Strongly Agree	[ ] Not Applicable
I learned to use this website quickly	[ ] Strongly Disagree	[ ] Disagree	[ ] Neutral	[ ] Agree	[ ] Strongly Agree	[ ] Not Applicable
Perceived Layout Consistency						
The layout of the pages throughout the website was consistent	[ ] Strongly Disagree	[ ] Disagree	[ ] Neutral	[ ] Agree	[ ] Strongly Agree	[ ] Not Applicable
I noticed abrupt changes in the layout of the pages	[ ] Strongly Disagree	[ ] Disagree	[ ] Neutral	[ ] Agree	[ ] Strongly Agree	[ ] Not Applicable
The layout under each section of the website was consistent	[ ] Strongly Disagree	[ ] Disagree	[ ] Neutral	[ ] Agree	[ ] Strongly Agree	[ ] Not Applicable
Perceived Visual Guidance						
The colors helped me to distinguish different sections of the website	[ ] Strongly Disagree	[ ] Disagree	[ ] Neutral	[ ] Agree	[ ] Strongly Agree	[ ] Not Applicable
The highlighted areas of a page helped me locate the information I needed	[ ] Strongly Disagree	[ ] Disagree	[ ] Neutral	[ ] Agree	[ ] Strongly Agree	[ ] Not Applicable
I got to know the content of a page by skimming the highlighted areas	[ ] Strongly Disagree	[ ] Disagree	[ ] Neutral	[ ] Agree	[ ] Strongly Agree	[ ] Not Applicable

Reference

Yang, T., Linder, J., Bolchini, D., DEEP: Design-oriented Evaluation of Perceived Usability, International Journal of Human-Computer Interaction (IJHCI), 1, 2011. doi:10.1080/10447318.2011.586320 , First posted on: 16 May 2011 (iFirst).

http://discern.uits.iu.edu:8670/DEEP/deep.html

Participant ID: _____ Site: _____ Date: ___/___/___

System Usability Scale (SUS)

Instructions: For each of the following statements, mark <u>one</u> box that best describes your reactions to the website *today*.

		Strongly Disagree				Strongly Agree
1.	I think that I would like to use this website frequently.	☐	☐	☐	☐	☐
2.	I found this website unnecessarily complex.	☐	☐	☐	☐	☐
3.	I thought this website was easy to use.	☐	☐	☐	☐	☐
4.	I think that I would need assistance to be able to use this website.	☐	☐	☐	☐	☐
5.	I found the various functions in this website were well integrated.	☐	☐	☐	☐	☐
6.	I thought there was too much inconsistency in this website.	☐	☐	☐	☐	☐
7.	I would imagine that most people would learn to use this website very quickly.	☐	☐	☐	☐	☐
8.	I found this website very cumbersome/awkward to use.	☐	☐	☐	☐	☐
9.	I felt very confident using this website.	☐	☐	☐	☐	☐
10.	I needed to learn a lot of things before I could get going with this website.	☐	☐	☐	☐	☐

Please provide any comments about this website:

POST-TEST QUESTIONNAIRE

User Number: _____ ____

Site Address: _____ _____

Evaluator(s) _____ _____ _____ _____ Date _____ _

Please answer the following questions based on your experience using the Web site or software.

LEGEND SA = Strongly Agree, A = Agree, N = Neither, D = Disagree, SD = Strongly Disagree

Write the questions which are RELEVANT for your evaluation

1. The language on the Task List you were given was easily understood?	SA	A	N	D	SD
2. The amount of information on the home page was adequate?	SA	A	N	D	SD
3. The use of color was appropriate?	SA	A	N	D	SD
4. Information was grouped consistently?	SA	A	N	D	SD
5. The navigation was inherently intuitive?	SA	A	N	D	SD
6. Colors and navigation was consistent through out the site?	SA	A	N	D	SD
7. There was too much information on individual pages?	SA	A	N	D	SD
8. There was adequate cross-referencing of topics and information?	SA	A	N	D	SD
9. Topic and page headings were self-explanatory?	SA	A	N	D	SD
10. It was necessary to scroll often to reach desired information?	SA	A	N	D	SD
11. The Table of Contents was helpful?	SA	A	N	D	SD
12. The site "Search" was helpful and reliable?	SA	A	N	D	SD
13. The terminology was understandable throughout the site?	SA	A	N	D	SD
14. The graphics were clear and sharp?	SA	A	N	D	SD
15. The text and graphics were presented in a visually aesthetic manner?	SA	A	N	D	SD
16. Overall, the pages were quick to load?	SA	A	N	D	SD

Please add any comments or suggestions in the space provided that you feel will help us evaluate the usability of the Web site or software.

_____

_____ _____

_____

_____ ____

FINAL STEP: Analysis and interpretation of collected data

- Organize the data
- Translate qualitative data into quantitative data
  - see classification/clustering techniques
- Analyse the data (statistical method)
- Discuss the results

OUTPUT

- Presentation of results
  - Qualitative: the main problems detected
  - Quantitative

- Discussion and final recommendations

Problem reporting

- Group problems by different criteria:
  - users
  - Task/scenario
  - “desig dimensions”
  - Application sections/functionality
  - priority
    - defined on the basis of different factors, e.g., was the problem found by all/many users?

Final recommendations

Organize recommendations by priority, e.g.,

- **Priority 1**: mandatory and urgent interventions
- **Priority 2**: needed but less urgent interventions
- **Priority 3**: hopeful interventions

Final Observations

Usability Testing vs Design

Working with User's Feedback

- Usability Testing captures user's **reactions/response to the design**, user's **preferences**, or **poor performance**
- Usability testing is useful to detect what **went wrong**
- It is typically unproductive to ask the user how to make the **right design**
- It is important to understand the **reason** and the underlying **need** to these reactions - Ask user and your team WHY
- Go deeper than the surface reaction

Working Beyond the Surface

- *"The lesson: while you* do *need to* ***notice*** *what the* ***user's problems*** *are, as designers we are responsible for looking* ***beyond the surface*** *to discover the underlying* ***issues*** *and offering a real* ***solution****".*

Pete Bickford, Usability Testing, in *AppleDirections,* May *1997.*

Usability Testing and Design

"If I had asked my customers what they wanted," Henry Ford said, "they would have said a faster horse." Customers don't envision the future, they inform the present" [1].

Corollary: with Usability Testing a *faster horse* would have never generated the idea of *car*

[1] From book: "The Ten Faces of Innovation", IDEO's Strategies for Beating the Devil's Advocate & Driving Creativity Throughout Your Organization By Tom Kelley with Jonathan Littman

Usability Testing & Multiple Designs

"*Contrary to our expectations, our results also suggest that **usability testing** by itself, even when multiple designs are presented, is not an effective vehicle for soliciting constructive suggestions about how to improve the design from end users. It is a means to **identify problems**, not provide solutions*".

From: Tohidi, M., Buxton, W., Baecker, R., and Sellen, A. 2006. *Getting the right design and the design right*. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Montréal, Québec, Canada, April 22 - 27, 2006). R. Grinter, T. Rodden, P. Aoki, E. Cutrell, R. Jeffries, and G. Olson, Eds. CHI '06. ACM, New York, NY, 1243-1252. DOI= http://doi.acm.org/10.1145/1124772.1124960

Inspection and User Testing

- Inspection and user testing are complementary
- Used in combination
- The importance (and the effort) of doing more inspection vs. user testing (or viceversa) depends on:
  - Application domain
  - Resources available
- Both inspection and user testing may be based on two families of approaches to evaluation:
  - Task-driven evaluation
  - Heuristic evaluation

User Testing Methods - Benefits

- Effective to assess the “look and feel” of the interface, as it is possible to verify at “real-time” the reactions of the users
- It should be used to
  - Validate inspection results
  - Finding unexpected user reactions
  - Detect “false positives”
- Good to justify the results (of an inspection)

User Testing Methods: Drawbacks

- Difficult to properly select correct user samples
- Difficult to adequately train them to manage advanced functions of an application
- Difficult, in a limited amount of time, to reproduce the actual usage situation
  - "Hawthorne effect" (Roethlisberger & Dickson, 1939): **observed users can be affected by observation alone**.
- Failures in creating real-life situations may lead to "artificial" conclusions rather then realistic results
- Considerable in terms of time, effort and cost

Wrap Up

Usability

- Usability is the capability of the application of effectively supporting users in accomplishing their goals

- 2 key Methods for evaluating usability:
  - Inspection methods
  - Usability testing methods

- Usability Evaluation is VALUABLE to make the “design right” rather than the “right design”

User Testing Methods: General Process

- Plan the test
  - Time, scope, samples, incentives
- Recruit users (5-10) according to the user profiles
- Introduce users to the test
- Testing execution
  - Users perform predefined tasks on the applications
  - Collect usability measures
- Debriefing and interviewing
- Analyes and organized results from tests
- Report results

User test Methods: a Task-driven approach

- Usability is assessed by trying to complete actions with the website.

- Tasks are defined which describe potential goals or sequences of actions that users might want accomplish with the application

- Tasks are defined on the basis of the requirements and the user scenarios

User Testing Methods – 6
Data collection

- Some techniques
  - Thinking aloud
  - Contextual inquiry
  - Interview-based testing
    - Tasks are defined by the users together with the inspectors

References and Resources

- Sauro, J. and Dumas, J. S. 2009. Comparison of three one-question, post-task usability questionnaires. In Proceedings of the 27th international Conference on Human Factors in Computing Systems (Boston, MA, USA, April 04 - 09, 2009). CHI '09. ACM, New York, NY, 1599-1608.
- Yang, T., Linder, J., Bolchini, D., DEEP: Design-oriented Evaluation of Perceived Usability, International Journal of Human-Computer Interaction (IJHCI), 1, 2011. doi:10.1080/10447318.2011.586320 , First posted on: 16 May 2011 (iFirst)
- Tohidi, M., Buxton, W., Baecker, R., and Sellen, A. 2006. *Getting the right design and the design right*. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Montréal, Québec, Canada, April 22 - 27, 2006). R. Grinter, T. Rodden, P. Aoki, E. Cutrell, R. Jeffries, and G. Olson, Eds. CHI '06. ACM, New York, NY, 1243-1252. DOI= http://doi.acm.org/10.1145/1124772.1124960
- Brooke, J. (1996). SUS: A Quick and Dirty Usability Scale. In Jordan, P. W., Thomas, B., Weerdmeester, B. A., & McClelland, I. L. (Eds.), Usability Evaluation in Industry. London: Taylor & Francis.