Human Computer Interaction

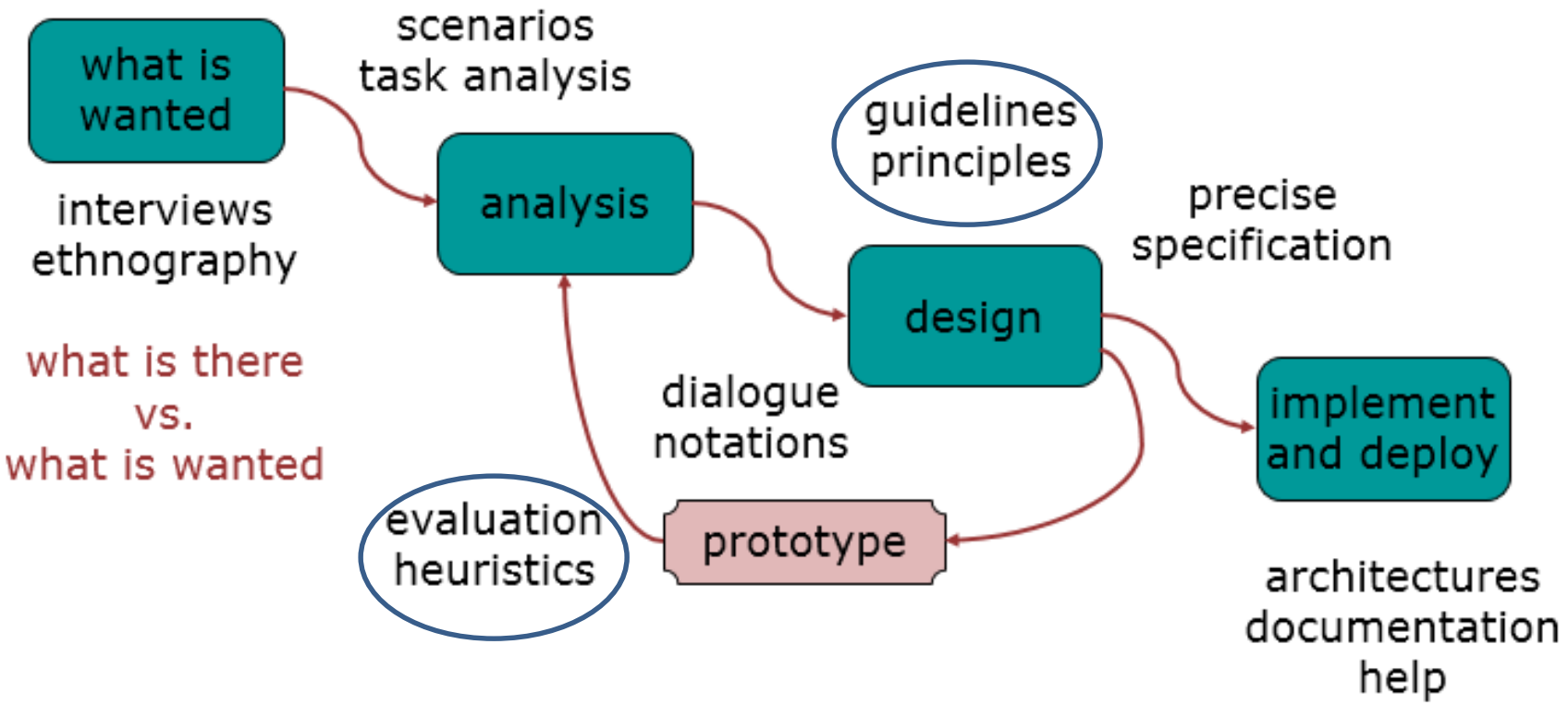
Unit-6 – Evaluating Interactive Systems

Pallavi Rao

Agenda

- Overview of Evaluation
- Evaluation through expert analysis (Analytic)
  - Heuristic evaluation
  - Cognitive walkthrough
  - Model based evaluation (Unit-7)
- Evaluation through user participation (Empirical)
  - Think Aloud
  - Log file analysis
  - Experiments
  - Observation
  - Focus groups
  - Query / Structured Interviews
  - Surveys
  - Eye-tracking
  - Physiological responses

Recall – ISDLC



Overview of Evaluation

- Evaluation
  - Tests usability, functionality and acceptability of an interactive system
  - Occurs in laboratory, field and/or in collaboration with users
  - Evaluates both design and implementation
  - Should be considered at all stages in the design life cycle

Overview of Evaluation

- Usability is a measurable property
- There are methods available to measure usability
- So we have Usability Evaluation

Goals of Evaluation

- Assess extent of system functionality
- Assess effect of interface on user
- Identify specific problems

Evaluation

- Expert analysis
  - Assessing early designs
  - Various methods of expert evaluation (Mostly done in usability lab)
- User participation
  - Later stage of design
  - Requires a working prototype
  - Various methods of user participation (Usability lab vs. Field)

Evaluation through expert analysis

Heuristic Evaluation

Cognitive Walkthrough

Pallavi Rao

Heuristic Evaluation

- Proposed by Nielsen and Molich.
- Usability criteria (heuristics) are identified
- Design examined by experts to see if these are violated
- Heuristic evaluation `debugs' design.
- Used to critique a decision that has already been made.

Heuristic Evaluation

- These heuristics have been reflected in many of the products designed by some of the most successful companies in the world such as Apple, Google and Adobe

- Note that there is considerable overlap between Nielsen and Molich's heuristics and Ben Shneiderman’s 'eight golden rules'.

Heuristic Evaluation

- 10 Heuristics
  - Visibility of system status
    - Always keep users informed about what is going on, through appropriate feedback within reasonable time.
    - For example, if a system operation will take some time, give an indication of how long and how much is complete.
  - Match between system and the real world
    - The system should speak the user's language, with words, phrases and concepts familiar to the user, rather than system-oriented terms.
    - Follow real-world conventions, making information appear in natural and logical order.
  - User control and freedom
    - Users often choose system functions by mistake and need a clearly marked 'emergency exit' to leave the unwanted state without having to go through an extended dialog.
    - Support undo and redo.

Heuristic Evaluation

- 10 Heuristics
  - Consistency and standards
    - Users should not have to wonder whether words, situations or actions mean the same thing in different contexts.
    - Follow platform conventions and accepted standards.
  - Error prevention
    - Make it difficult to make errors.
    - Even better than good error messages is a careful design that prevents a problem from occurring in the first place.
  - Recognition rather than recall
    - Make objects, actions and options visible.
    - The user should not have to remember information from one part of the dialog to another.
    - Instructions for use of the system should be visible or easily retrievable whenever appropriate.

Heuristic Evaluation

- 10 Heuristics
  - Flexibility and efficiency of use
    - Allow users to tailor frequent actions.
  - Aesthetic and minimalist design
    - Dialogs should not contain information that is irrelevant or rarely needed.
  - Help users recognize, diagnose and recover from errors
    - Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.
  - Help and documentation
    - Few systems can be used with no instructions so it may be necessary to provide help and documentation.
    - Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.

Heuristic Evaluation

- To aid the evaluators in discovering usability problems, a set of 10 heuristics are provided.
  - Related to guidelines (Unit-4)
  - Can be supplemented by additional heuristics depending on a particular domain

- Generally conducted with 3 to 5 evaluators (highly recommended)
  - Conducted independently by each evaluator
    - Evaluators are not end users, they are domain experts and have some design ideas
  - Each evaluator assesses the system and notes violations
  - Also assesses the severity of the problems
    - Cosmetic problem
    - Minor usability problem
    - Major usability problem
    - Usability catastrophe

Heuristic Evaluation

- Nowadays, designers are encouraged to establish their own design-specific heuristics to evaluate their products, systems, websites, etc.

- Since Nielsen and Molich developed these heuristics in the 1990s, technology has advanced and they are less attuned to many of the products available in the market today.
  - For instance, Nielsen and Molich's heuristics would be too general to evaluate the usability of designs intended for online communities or mobile devices where the working environment is constantly changing.

- However, the original heuristics are still largely applicable in spite of the specific capabilities and constraints of modern designs.

- Therefore, as a designer it’s crucial that you learn to incorporate Nielsen and Molich’s heuristics into your designs as the first step.

Heuristic Evaluation

- Pros
  - help the evaluators focus their attention on certain issues
  - does not carry the ethical and practical issues/problems associated with inspection methods involving real users
  - Evaluating designs using a set of heuristics can help identify usability problems with individual elements and how they impact the overall user experience

Heuristic Evaluation

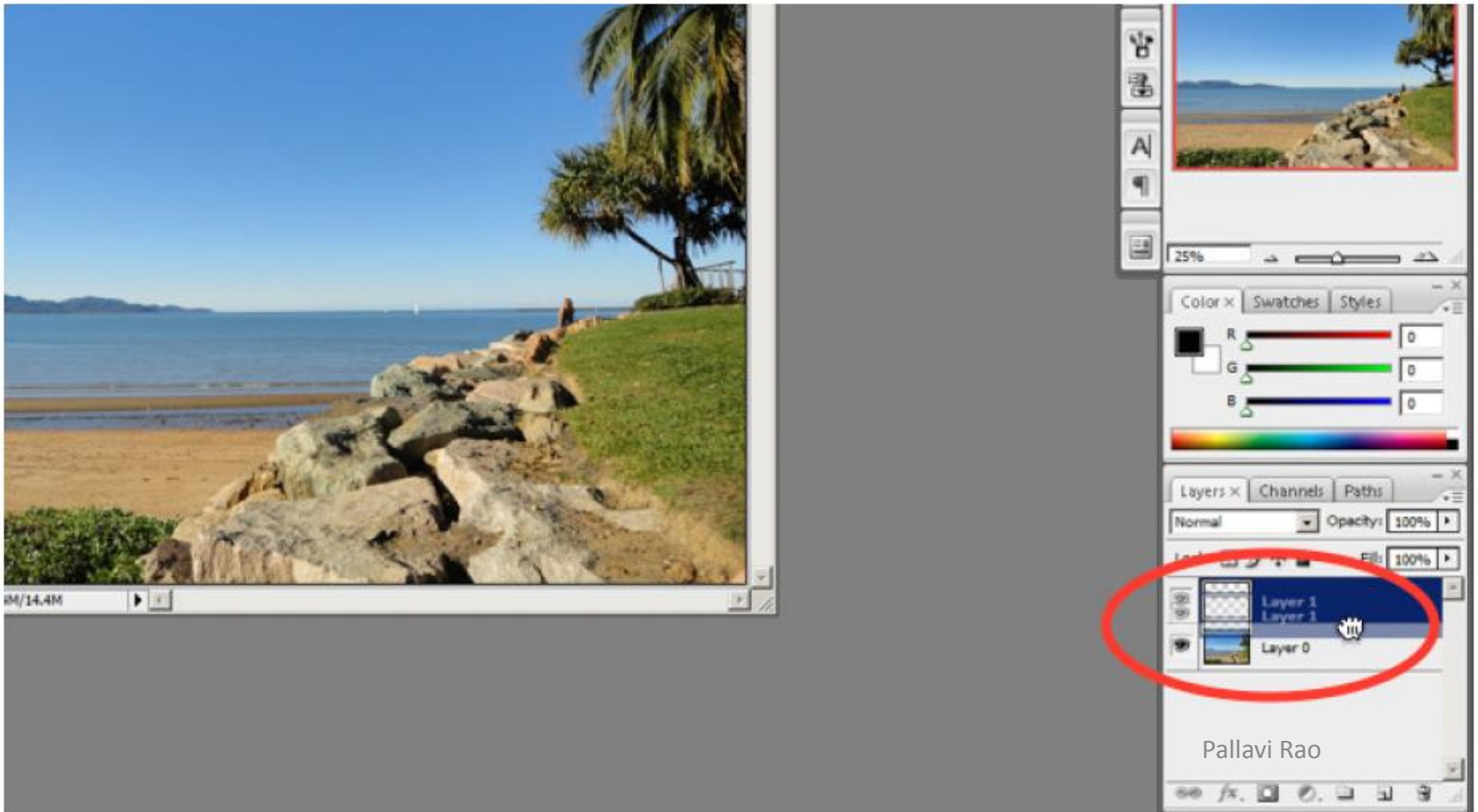
- Cons
  - Choosing appropriate heuristics is extremely important; if the wrong set of heuristics is employed, certain usability problems may be overlooked
  - Time consuming including training evaluators
  - Can lead to false alarms

Heuristic Evaluation

- In practice, in most of the IT companies, Heuristic evaluation (with some additional heuristic – product specific) is conducted followed by a user study (will be explained later in this unit)

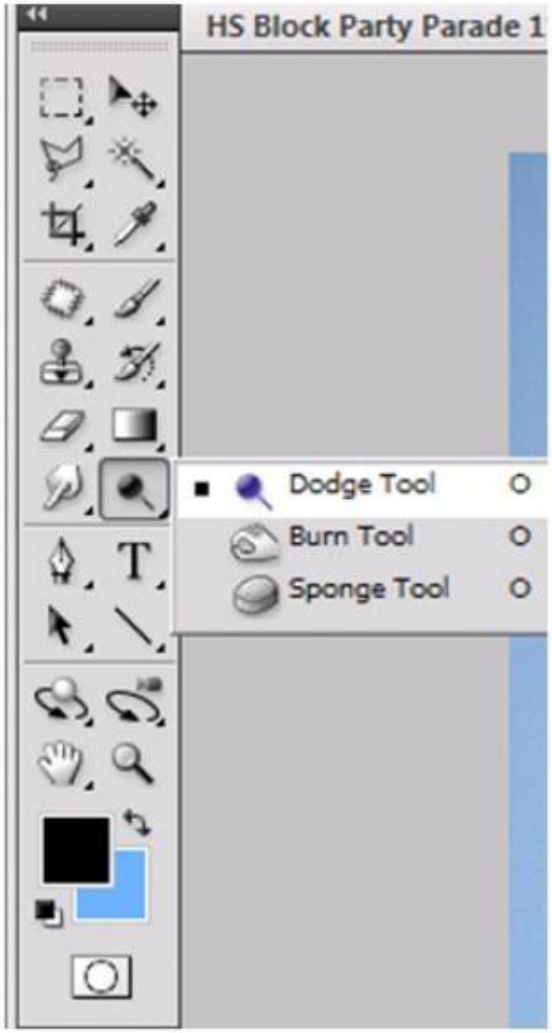
Heuristic Evaluation – Adobe Photoshop Example

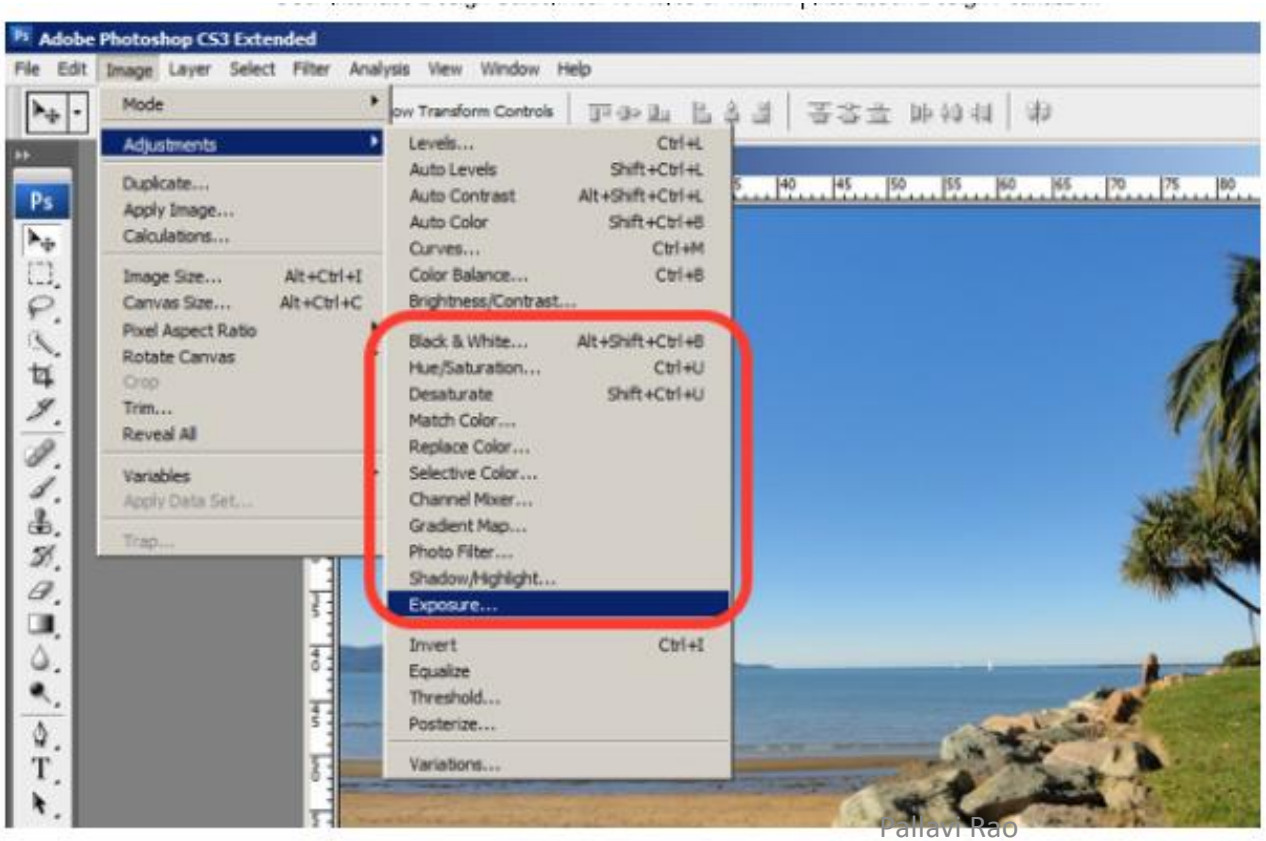
- Visibility of system status
  - Visually showing the user what their actions have led to
  - Example – When users mover layers around in the layers palette, they can visually see the layer being represented as physically dragged within the space



Heuristic Evaluation – Adobe Photoshop Example

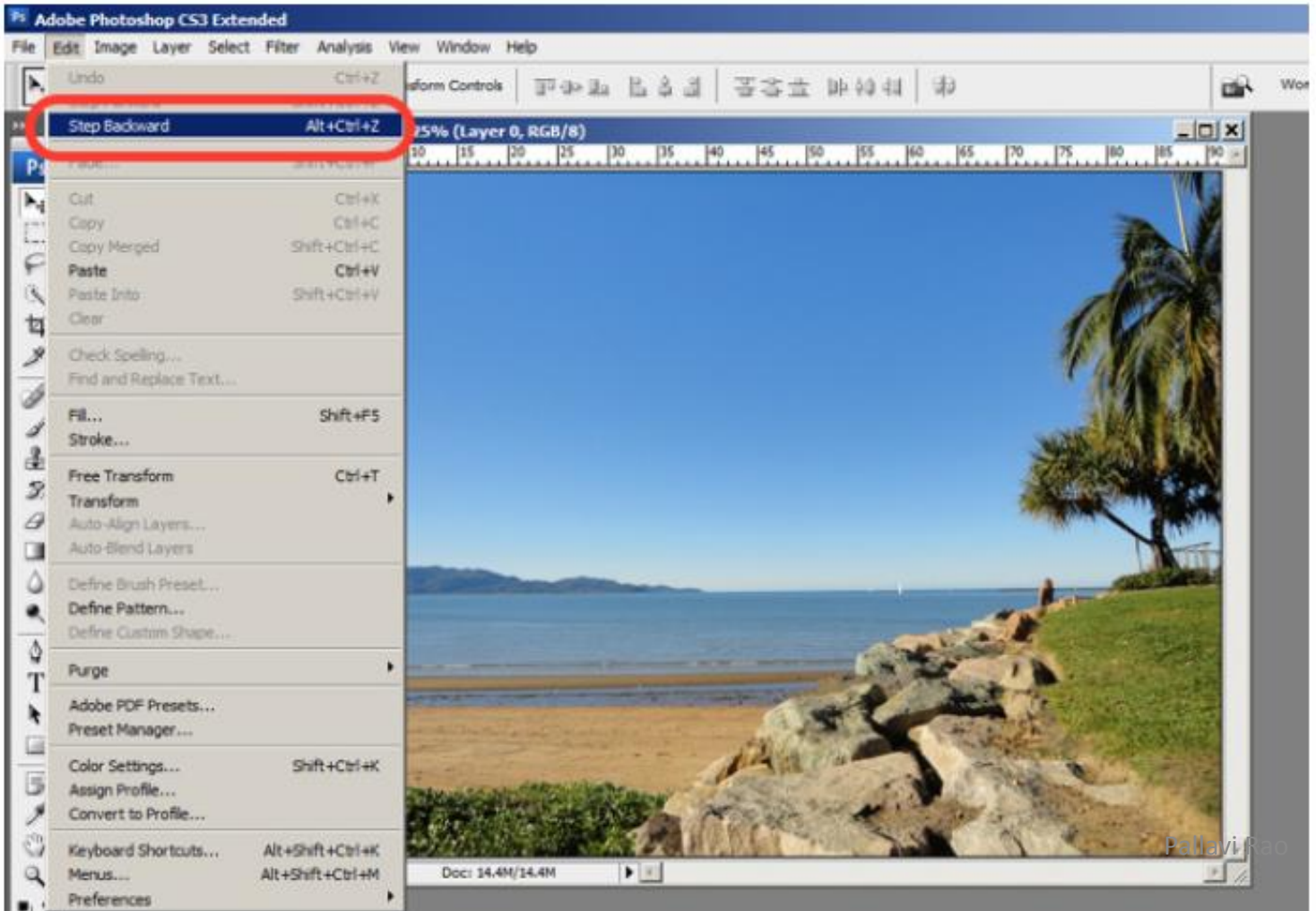
- System match to the real world
  - Terminology used in photography / print media used





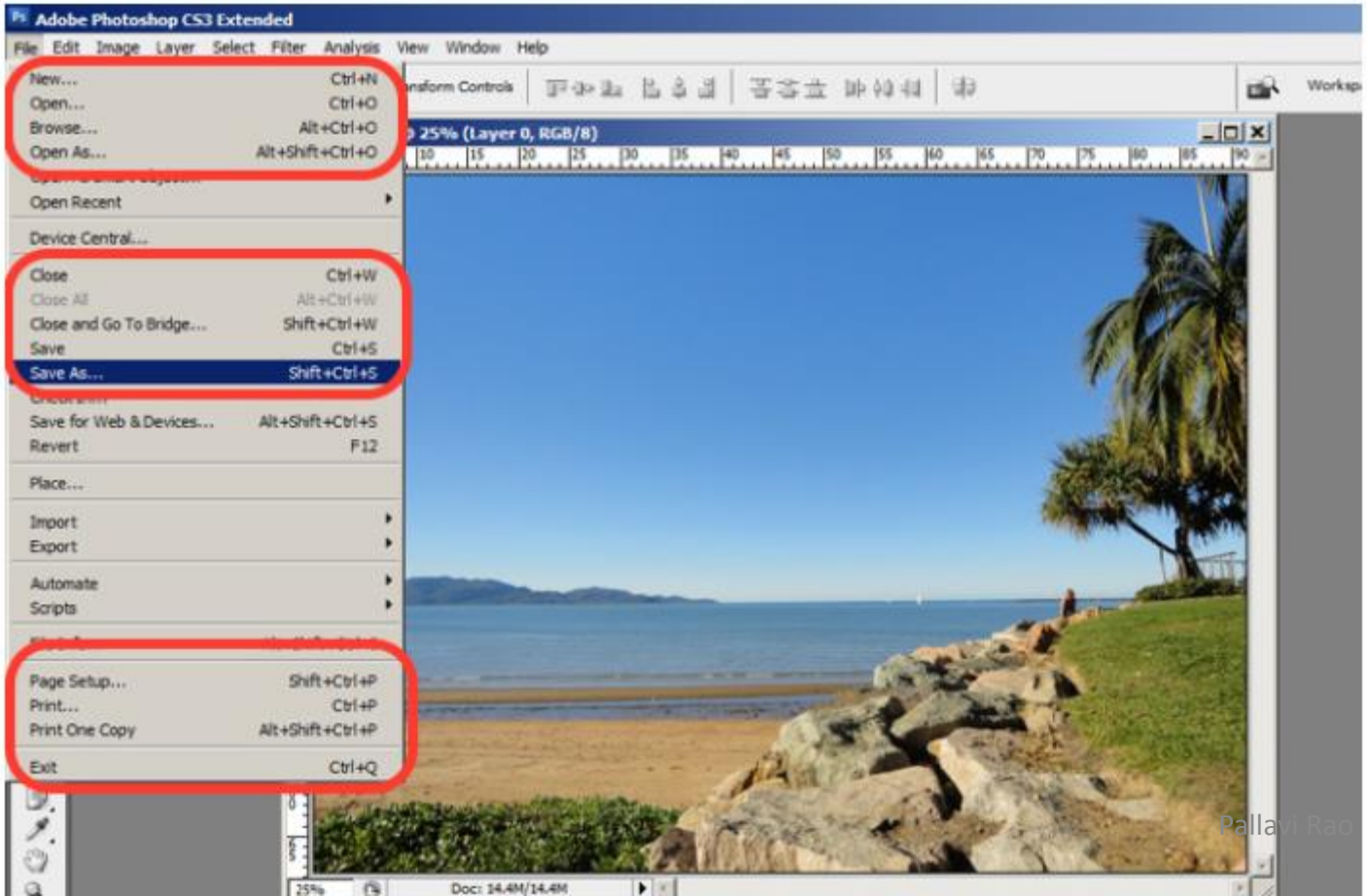
Heuristic Evaluation – Adobe Photoshop Example

- User control and freedom
  - Users are in control as they can take a step backward or step forward under Edit menu



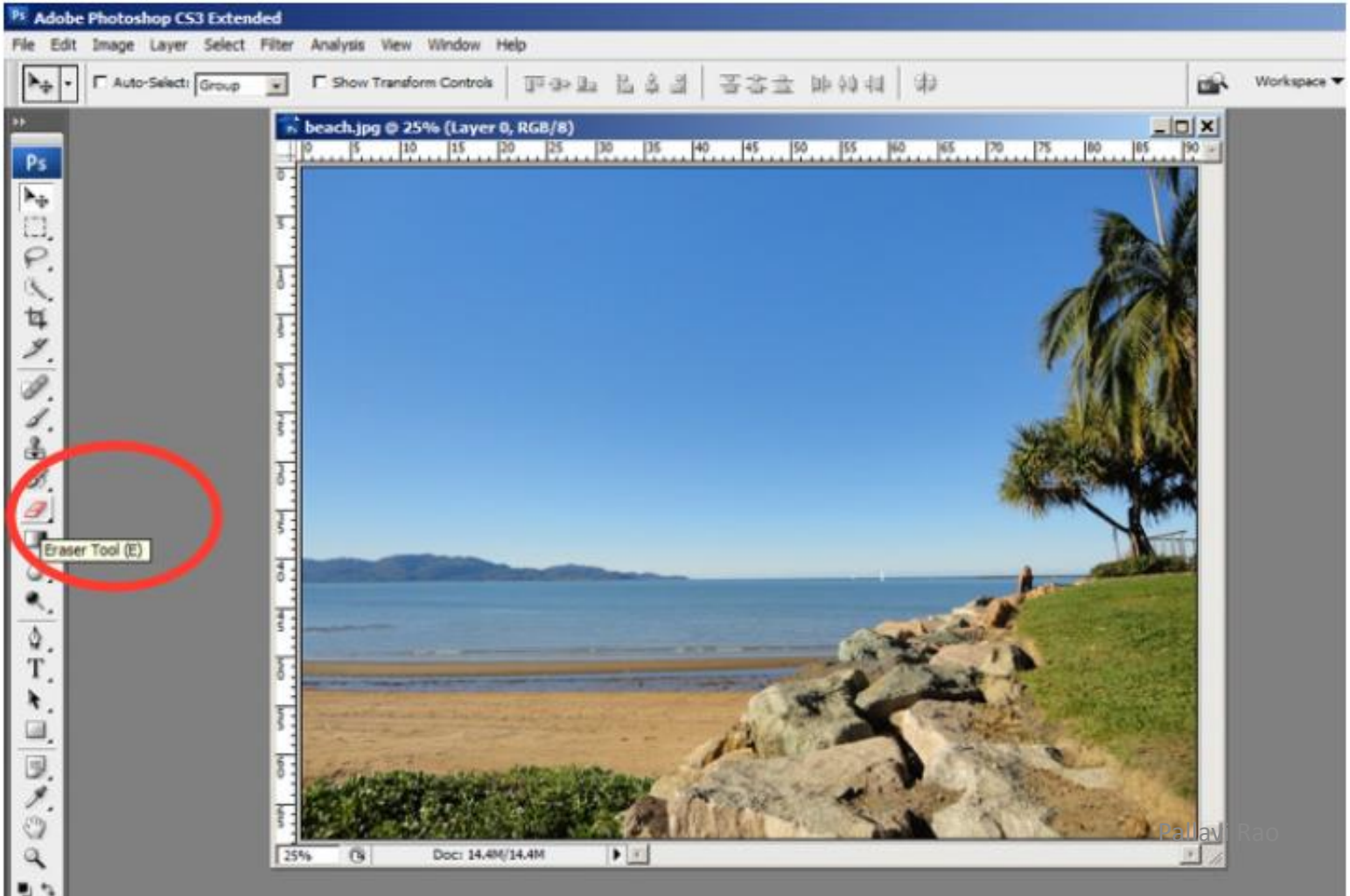
Heuristic Evaluation – Adobe Photoshop Example

- Consistency and standards
  - Standard layout and look and feel
  - Use common terminology
  - File menu displays variety of highly familiar options



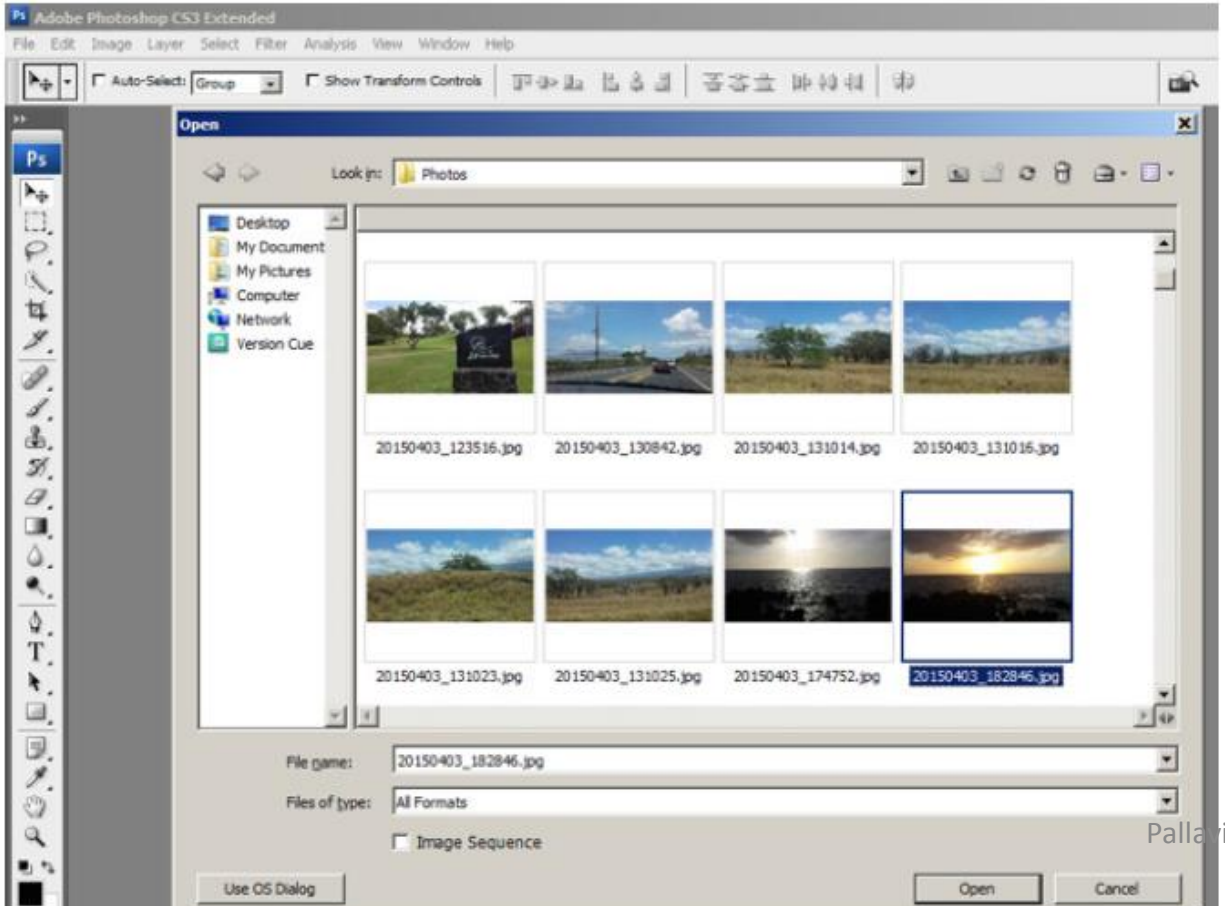
Heuristic Evaluation – Adobe Photoshop Example

- Error prevention
  - Brief description or label of the tools when a user hovers over it to help make sure users are using the proper tool for the task at hand



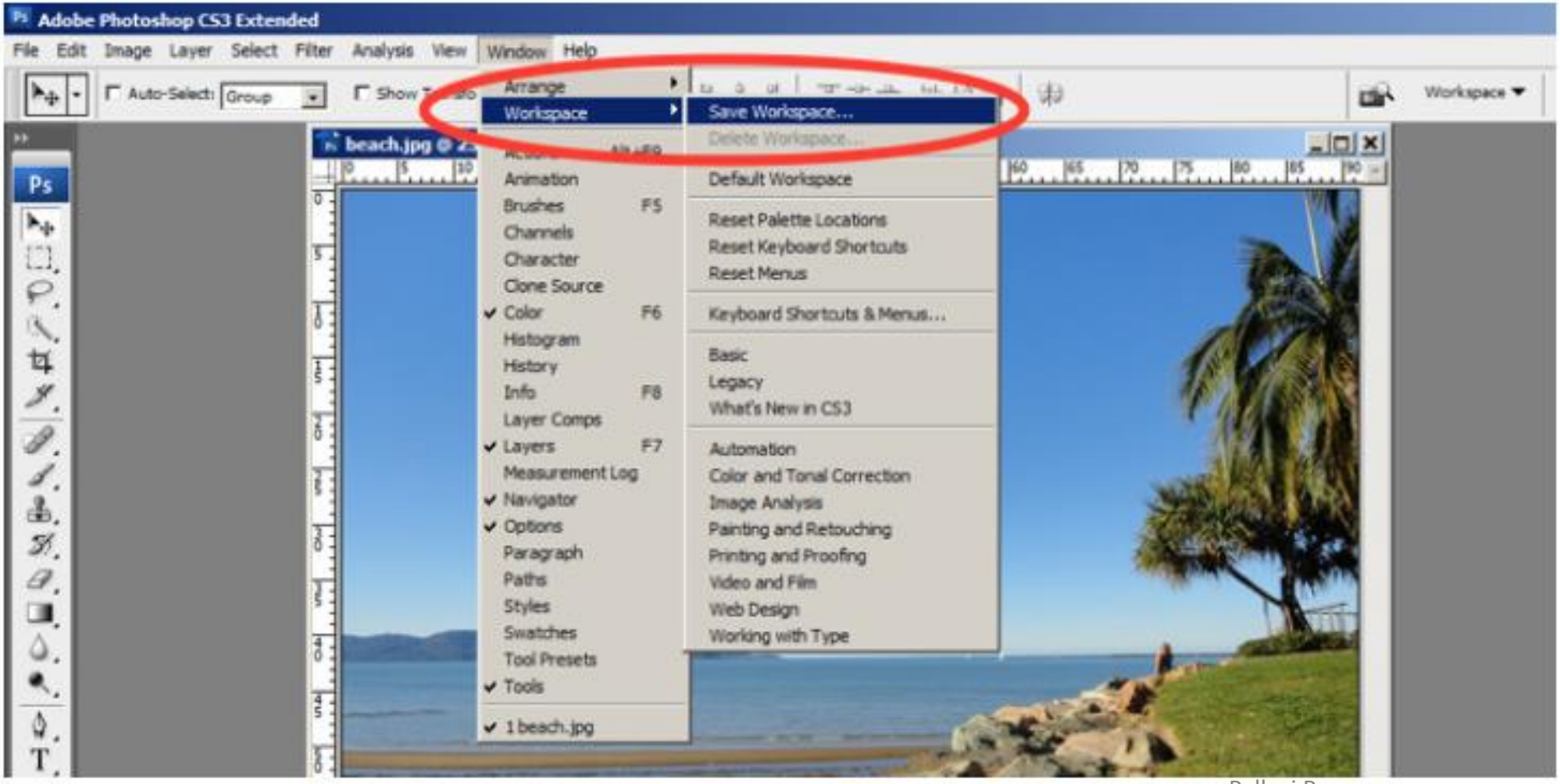
Heuristic Evaluation – Adobe Photoshop Example

- Recognition than recall
  - Whether making a selection from the menu, or opening a new file, photoshop provides a sample view for users to make right choice – allowing the users to visually recognize what they would be looking for instead of having to recall the name or typing it into search for it



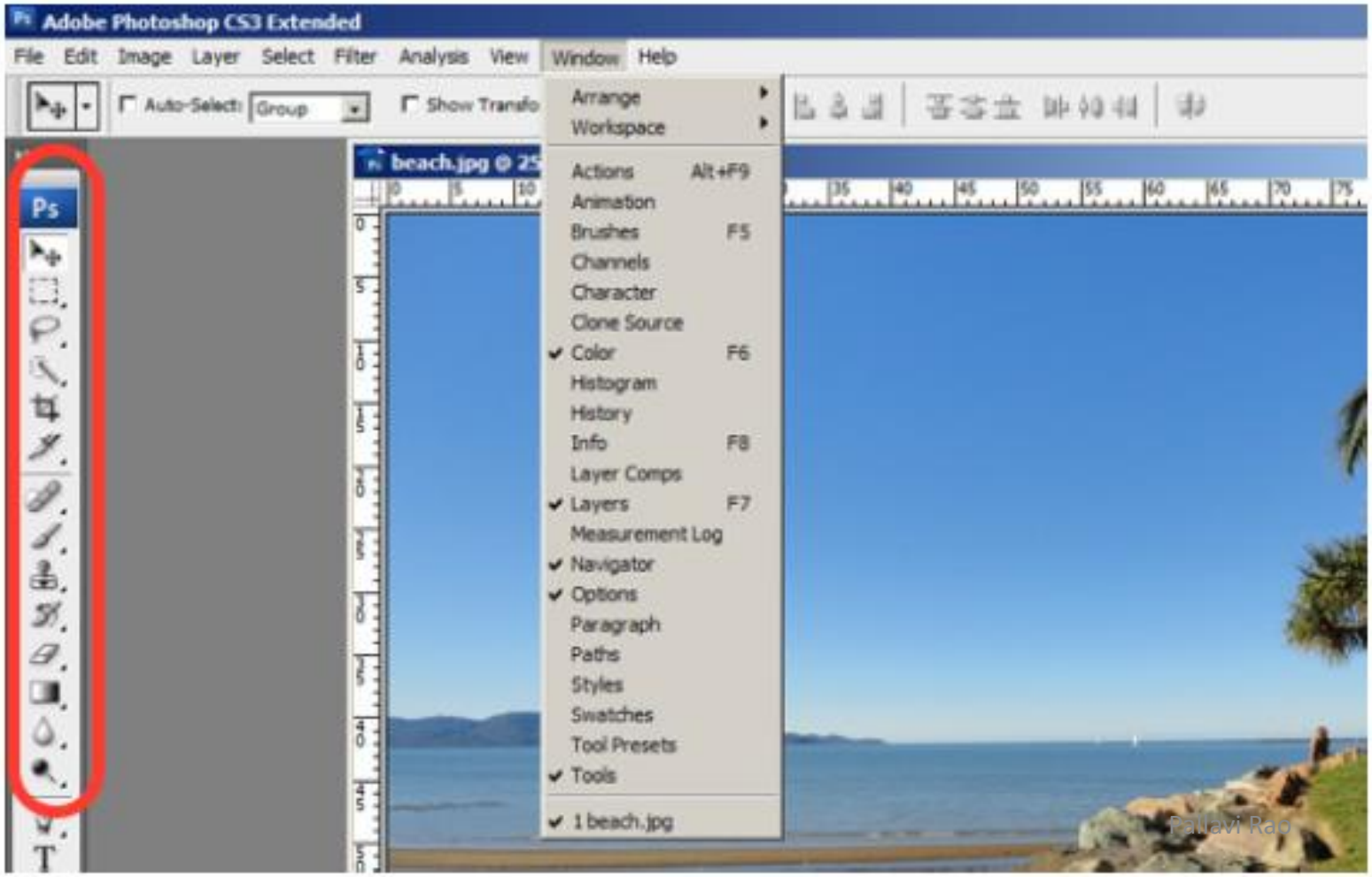
Heuristic Evaluation – Adobe Photoshop Example

- Flexibility and efficiency of use
  - Frequent users are get to save their preferred workspace



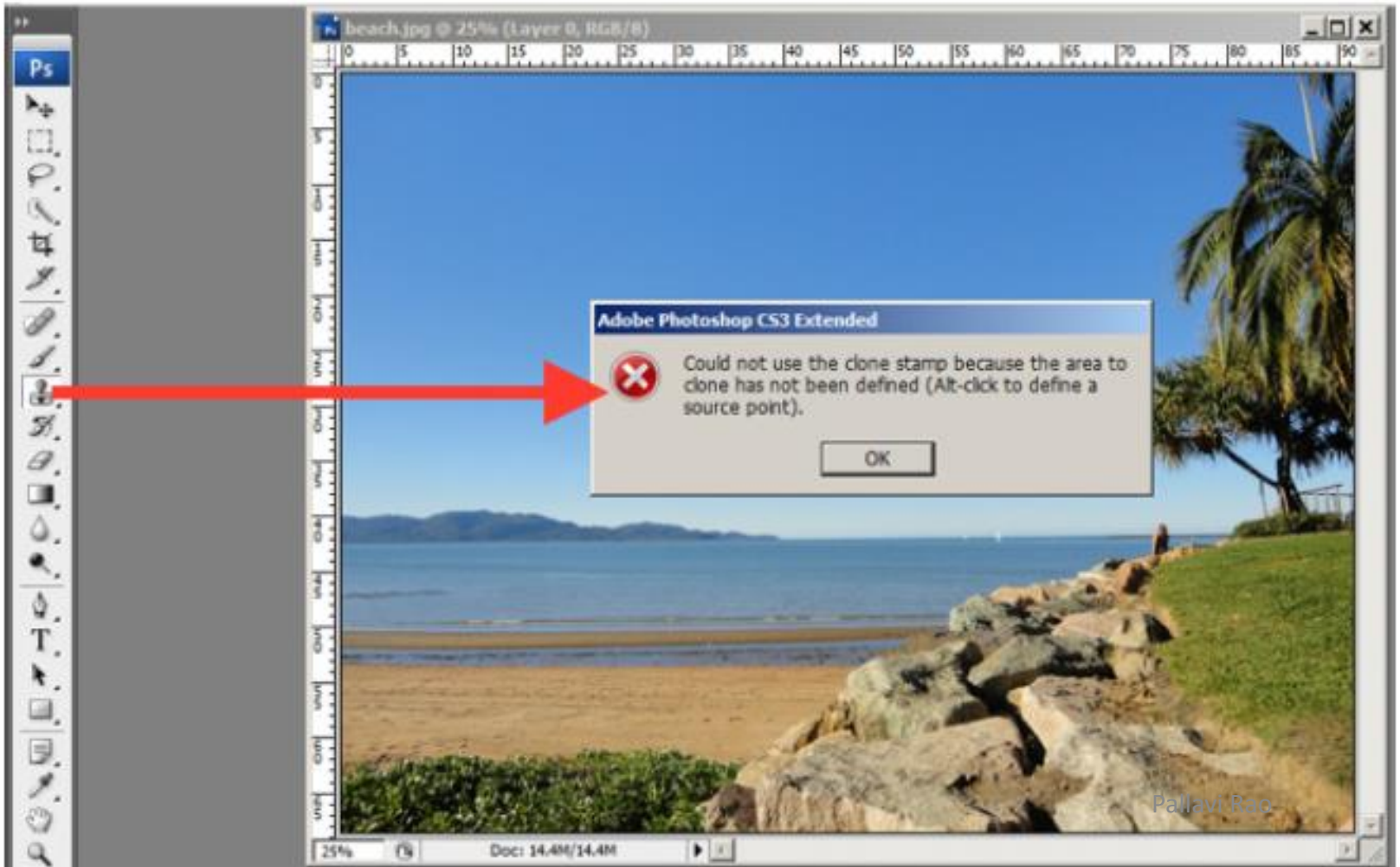
Heuristic Evaluation – Adobe Photoshop Example

- Aesthetic and Minimalist design
  - The toolbar only displays the icons and is neatly tucked to the side to help keep clutter to a minimum, and maintain a minimalist aesthetic.



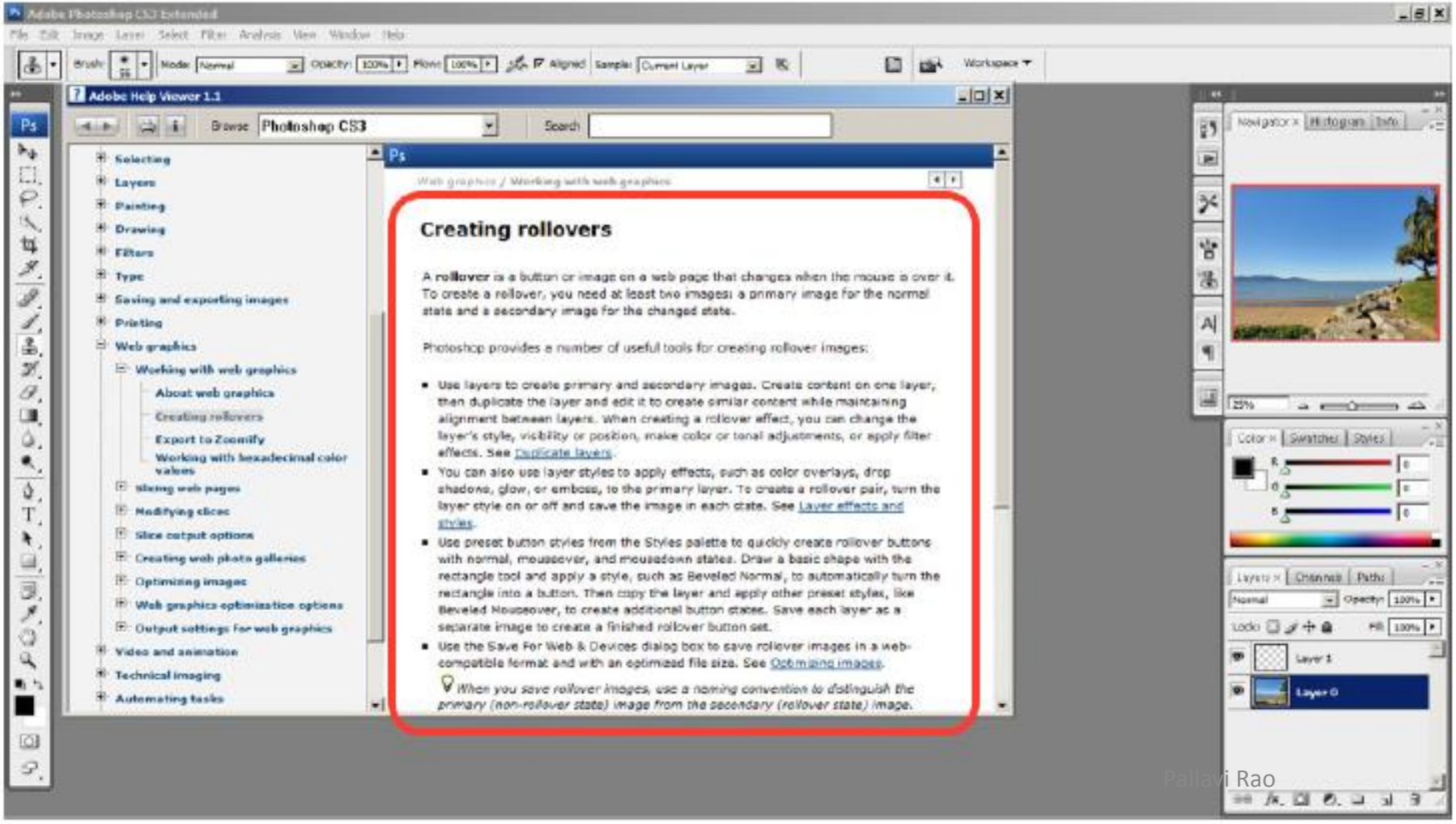
Heuristic Evaluation – Adobe Photoshop Example

- Help users recognize, diagnose and recover from errors
  - Whenever there's an error, it provides dialogue that lets users know what went wrong and how to fix it



Heuristic Evaluation – Adobe Photoshop Example

- Help and documentation
  - Can be accessed easily via the main menu bar.



Cognitive Walkthrough

Proposed by Polson *et al.*

- evaluates design on how well it supports user in learning task
- evaluator ‘walks though’ design to identify potential problems using psychological principles
- forms used to guide analysis

Cognitive Walkthrough

- Cognitive walkthroughs are designed to see whether or not a new user can easily carry out tasks within a given system
  - Interaction centered
- It is a task-specific approach to usability (in contrast to heuristic evaluation which is a more holistic usability inspection).
- The idea is that if given a choice, most users prefer to do things to learn a product rather than to read a manual or follow a set of instructions.
  - The main focus of CW is to establish how easy a system is to learn
  - Focus is on learning through exploration

Cognitive Walkthrough

- 4 things required:
  1. Prototype of the system
  2. Task
     - It is these tasks that the cognitive walkthrough will examine for usability
  3. List of actions needed to complete the task
  4. Users, their experience and knowledge

Cognitive Walkthrough

- Evaluator tries to answer the following questions for each step in the action sequence
  1. Is the effect of action same as users' goal at that point?
  2. Will users see that action is available?
     - About the visibility of the action
  3. Once the users have found the correct action, will they know it is the one they need?
     - Whether its meaning & effect is clear
  4. After the action is taken, will the users understand the feedback they get?

Cognitive Walkthrough

- 1. Is the effect of action same as users' goal at that point?
  - This question is trying to examine whether the interface is making assumptions about users' level of experience / knowledge that are not accurate
  - It also helps to identify when a user's expectations of an action don't align with the actual action taken
  - Example - if the effect of the action is to save a document, is 'saving a document' what the user wants to do?

Cognitive Walkthrough

- 2. Will users see that action is available?
  - About visibility of the action
  - Hidden or obscure controls are a problem for users
  - The more choices present on the interface; it is less likely that user will know what to do
  - Most common example is TV remote control. They provide an overwhelming number of options to users and are often confusing to use
    - In many cases manufacturers try to tackle this by placing rarely used functions underneath a sliding cover, by reducing options the control is easier to use

Cognitive Walkthrough

- 3. Once the users have found the correct action, will they know it is the one they need?
  - Whether its meaning & effect is clear
  - If your use of language is poor, for example you use overly complex words or industry jargon, it can be hard for a user to work out what is needed to achieve their outcome.
  - Steve Krug, says in his book “Don’t Make me Think” – “It doesn’t matter how many times I have to click, as long as each click is a mindless, unambiguous choice.”

Cognitive Walkthrough

- After the action is taken, will the users understand the feedback they get?
  - This will help you investigate when feedback in the system is missing /badly worded /easy to miss / ambiguous.
  - You need to let your users know about the progress they are making in the task.
  - That's why it's always a good idea to have a "thank you" note when a customer finishes making an order.

Cognitive Walkthrough

- Proper documentation required
  - 4 things identified above
  - Separate form consisting of 4 questions for each action
  - Any negative answer for any question should be documented in a separate usability problem sheet

Cognitive Walkthrough

- For each task walkthrough considers
  - what impact will interaction have on user?
  - what cognitive processes are required?
  - what learning problems may occur?
- Analysis focuses on goals and knowledge:
  - does the design lead the user to generate the correct goals?
- “Walk a mile in the users’ shoes”

Cognitive Walkthrough

- Just like HE, CW also is conducted with 3-5 assessors / evaluators
  - After all the “walkthroughs”, all the usability problem sheets are summarized into single report and prioritize issues for fixing

Cognitive Walkthrough

- Benefits
  - Cost effective
  - Fast to carry out
  - Easy to carry out
  - System to be examined from users' perspective and provide quick feedback to enable decision making in the design process
  - Can be implemented during the design phase which can give rapid insight before budget is spent developing an unusable product

Evaluation through user participation

Evaluation through user participation

- Expert evaluation (such as Heuristic Evaluation, Cognitive Walkthrough) – evaluation at the early design phases
- Interactive system design is not complete unless it is evaluated with end users
- Users (actual users or sample users) are involved in evaluation
- Users are involved - not only for evaluating the interfaces but also for investigating the effect of technology on users
  - i.e., the positive or negative effects of attributes of software, hardware, user capabilities and usage environments.
  - By empirical evidence

Evaluation through user participation

- Empirical research
  - It is a way of gaining knowledge by means of direct and indirect observation or experience.
  - Broadly defined as the observation-based investigation seeking to discover and interpret facts, theories, or laws.
  - Testable research questions are asked
  - Can be analyzed quantitatively or qualitatively.

Evaluation through user participation

- Empirical Research
  - Some of the non-testable questions
    - What are its performance limits and capabilities?
    - What are its strengths and weaknesses?
    - How much practice is required to become proficient?

Evaluation through user participation

- Empirical Research - Testable questions
  - Suppose you have designed a new text entry technique for mobile phones. You think the design is good. In fact, you feel your method is better than the most widely used current technique, multi-tap. You decide to undertake some empirical research to evaluate your invention and to compare it with multi-tap? What are your research questions?
    - Weak question - Is the new technique better than multi-tap?
    - Better - Is the new technique faster than multi-tap?
    - Better still - Is the new technique faster than multi-tap within one hour of use?
    - Even better - If error rates are kept under 2%, is the new technique faster than multi-tap within one hour of use?

Evaluation through user participation

- Empirical Research - Testable questions
  - The questions are testable (we can actually conduct experiments to test the answer to the questions)

Evaluation through user participation

- User studies explained in this unit are part of empirical research
  - In the context of HCI, it is about evaluating a design with users.
  - They are also used for testing research hypotheses in various studies

Evaluation through user participation

- Empirical research – Data collection
  - Data can be gathered in two ways
    - Manual: In this case, a human observer manually records all the relevant observational data
    - Automatic: The observation can be recorded automatically, through the use of computers, software, sensor, camera and so on

Evaluation through user participation

- Empirical research – Data analysis
  - Empirical evidence can be analyzed quantitatively or qualitatively
  - Quantitative
    - Statistical tests to test the hypotheses
  - Qualitative
    - Various methods – Content analysis, Creating coding themes

Styles of evaluation

- Styles of evaluation
  - Laboratory studies
  - Field studies

Laboratory studies – Usability lab

Laboratory studies

- Advantages:
  - specialist equipment available
  - uninterrupted environment
- Disadvantages:
  - lack of context
  - difficult to observe several users cooperating
- Appropriate
  - if system location is dangerous or impractical for constrained single user systems to allow controlled manipulation of use

Field Studies

- Advantages:
  - natural environment
  - context retained (though observation may alter it)
  - longitudinal studies possible
- Disadvantages:
  - distractions
  - noise
- Appropriate
  - where context is crucial for longitudinal studies

Various Evaluation techniques (with users)

Evaluation Techniques		Data Analysis
Experiments		Quantitative
Observational Techniques	Think aloud	Qualitative
	Cooperative evaluation	
	Protocol analysis	
	Post-task walkthroughs	
Query techniques	Interview	Qualitative/Quantitative
	Questionnaire / Surveys	Qualitative/Quantitative
	Focus groups	Qualitative
Monitoring physiological responses	Eye tracking	Quantitative
	Measures of heart rate, Skin conductance	

Evaluation through user participation

- Challenges
  - Identifying potential participants
  - Cost
  - Institutional (Ethical) review board

Evaluation through user participation

- Generally, before actually conducting the user studies, pilot studies are conducted with 1 or 2 users
- A pilot study is a small trial run of the main study
  - Aim is to make sure that the plan is viable before embarking on the real study
  - For example, checking the equipment and instructions for its use, practice interviewing skills, check the questions in a questionnaire are clear or that an experimental procedure works properly

Experiments

Experiments

- Controlled evaluation of specific aspects of interactive behaviour
- Evaluator chooses hypothesis to be tested
- A number of experimental conditions are considered which differ only in the value of some controlled variable.
- Changes in behavioural measure are attributed to different conditions

Experimental factors

- Subjects
  - who – representative, sufficient sample (for experiments more numbers preferred because of statistical analysis)
- Variables
  - things to modify and measure
- Hypothesis
  - what you'd like to show
- Experimental design
  - how you are going to do it

Variables

- Independent variable (IV)

  characteristic changed to produce different conditions

  e.g. interface style, number of menu items

- Dependent variable (DV)

  Variable that can be measured

  characteristics measured in the experiment

  e.g. time taken, number of errors.

Hypothesis

- prediction of outcome
  - framed in terms of IV and DV

    e.g. “error rate will increase as font size decreases”

- null hypothesis:
  - states no difference between conditions
  - aim is to disprove this

    e.g. null hyp. = “no change with font size”

Experimental design

- Within groups design
  - each subject performs experiment under each condition.
  - transfer of learning possible
  - less costly and less likely to suffer from user variation.

- Between groups design
  - each subject performs under only one condition
  - no transfer of learning
  - more users required
  - variation can bias results.

Experiments - Analysis of data

- Quantitative analysis
  - Various statistical techniques
- Before you start to do any statistics:
  - look at data (look for outliers if any)
  - save original data
  - Descriptive statistics
- Choice of statistical technique depends on
  - type of data
  - Questions we want to answer

Design an experiment-1

- Design an experiment to test whether adding color coding to an interface will improve accuracy.
  - Identify your hypothesis, participant group, dependent and independent variables, and tasks users need to perform.

Design an experiment-1

- Participants: Taken from user population
- Hypotheses: Color coding will make selection more accurate.
- IV – Color coding
- DV – Accuracy (measured as number of errors)
- Experimental design - Between-groups to ensure no transfer of learning (or within-groups with appropriate safeguards if participants are scarce).

Design an experiment-1

- Task:
  - The interfaces are identical in each of the conditions, except that, in the second, color is added to indicate related menu items.
  - Participants are presented with a screen of menu choices (ordered randomly) and verbally told what they have to select. Selection must be done within a strict time limit when the screen clears.
  - Failure to select the correct item is deemed an error.

Design an experiment-2

- Design an experiment to test whether search speed improves as the number of menu items decreases.
  - Identify your hypothesis, participant group, dependent and independent variables, and task.

Design an experiment

- More complex experiments may have more than one independent variable.
- For example, in the previous experiment, we may suspect that the speed of the user's response depends not only on the number of menu items but also on the choice of commands used on the menu.
- In this case there are two independent variables.
- If there were two sets of command names (that is, two levels), we would require six experimental conditions to investigate all the possibilities
  - (three levels of menu size × two levels of command names).

Experiments

- Limitations
  - May not be a good representation of users' typical interaction behavior
  - Its been reported that participants behave differently in lab-based experiments due to stress of being observed

Observational Methods

Think Aloud
Cooperative evaluation
Protocol analysis
Post-task walkthroughs
Ethnography

Pallavi Rao

Observational methods

- Evaluator observing users (either in usability labs or field)
  - Users are asked to complete a set of pre-determined tasks
  - Evaluator watches and records users' actions
- Less number of users required than experiment
  - Generally 5 users from representative sample enough
- Qualitative data analysis
- Great way to find out more about how a product is used and to identify any problems that they encounter
  - Easy to conduct, but difficult to analyze
  - Generally combined with Interviews

Think Aloud

- user observed performing task
- user asked to describe what he is doing and why, what he thinks is happening etc.

- Advantages
  - simplicity - requires little expertise
  - can provide useful insight
  - can show how system is actually use
- Disadvantages
  - subjective
  - selective
  - act of describing may alter task performance

Cooperative evaluation

- variation on think aloud
- user collaborates in evaluation
- both user and evaluator can ask each other questions throughout

- Additional advantages
  - less constrained and easier to use
  - user is encouraged to criticize system
  - clarification possible

Protocol analysis

- paper and pencil – cheap, limited to writing speed
- audio – good for think aloud, difficult to match with other protocols
- 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

- Mixed use in practice.
- audio/video transcription difficult and requires skill.
- Some automatic support tools available

Post-task walkthroughs

- transcript played back to participant for comment
  - immediately → fresh in mind
  - delayed → evaluator has time to identify questions
- useful to identify reasons for actions and alternatives considered
- necessary in cases where think aloud is not possible

Ethnography

- Has its roots in anthropological studies of non-Western cultures
  - In attempting to develop deep understanding of unfamiliar civilizations
  - True understanding of complex human practices and contexts requires in-depth study
- Generally conducted in the field (not in the lab)
- Field can be virtual also (for example, ethnographical observations in online communities)
- Additional information can be captured through interviewing and questionnaires

Ethnography

- How it is done?
  - Observe, ask questions and record what is seen/heard (you need to be aware of people's feelings and sensitive to where you should not go)
  - Collect a variety of data (such as notes, still pictures, audio, video), Interviews and questionnaires can be used

- Data analysis – Mostly qualitative

Ethnography

- Challenges
  - When to stop?
  - How to convert broad observations into design ideas?

Query Techniques

Interviews

Questionnaires

Pallavi Rao

Query techniques

- Talking to the user directly
- Advantage
  - Get the users view point directly
  - May reveal issues that have not been considered by designers
- Disadvantage
  - Self reporting and hence very subjective
- Useful supplementary materials to other methods

Interviews

- analyst questions user on one-to -one basis usually based on prepared questions
- informal, subjective and relatively cheap

- Advantages
  - can be varied to suit context
  - issues can be explored more fully
  - can elicit user views and identify unanticipated problems
- Disadvantages
  - very subjective
  - time consuming

Interviews

- What interviews can not do
  - Can’t answer specific questions
    - Should the *Buy* button be red or orange?
    - Is it better to use a drop-down menu or a set of radio buttons for a certain set of choices?
- What interviews can do
  - Explore users’ general attitudes
  - Good for exploratory studies

Questionnaires

- Set of fixed questions given to users
- Advantages
  - quick and reaches large user group
  - can be analyzed more rigorously
- Disadvantages
  - less flexible
  - less probing
- Need careful design
  - what information is required?
  - how are answers to be analyzed?

Questionnaires

- Styles of question
  - General
    - Background of users
  - open-ended
    - Subjective information from users
    - But difficult to analyze; Can be used as support for some findings
  - Scalar

    It is easy to recover from mistakes.

    Disagree 1 2 3 4 5 Agree
  - multi-choice
    - Multiple choices of responses
  - Ranked
    - Ordering of items in a list, to get users preferences

Design a questionnaire

- You have been asked to measure users' satisfaction after using an e-commerce portal. Design a questionnaire to find out users satisfaction level.

- Tip: You can measure users' satisfaction by the following variables
  - Ease of use
  - Recover from mistakes
  - Easy to get help when needed
  - Provides feedback

Physiological methods

Eye tracking

Physiological measurement

Pallavi Rao

Evaluation through monitoring physiological responses

- One of the problems with most evaluation techniques is that we are reliant on observation and the users telling us what they are doing and how they are feeling.

- What if we were able to measure these things directly?

- Interest has grown recently in the use of what is sometimes called objective usability testing, ways of monitoring physiological aspects of computer use.

- This will allow us not only to see more clearly exactly what users do when they interact with computers, but also to measure how they feel.

Eye tracking

Pallavi Rao

Eye tracking

- Head or desk mounted equipment tracks the position of the eye
- Eye movement reflects the amount of cognitive processing a display requires
  - Measuring not only where people look, but also their patterns of eye movement, tell us which areas of a screen they are finding easy or difficult to understand

Eye tracking

- Measurements include
  - fixations: eye maintains stable position. Number and duration indicate level of difficulty with display
    - Number of fixation - The more fixation the less efficient search strategy
    - Fixation duration – Longer fixation difficulty with a display
  - saccades: rapid eye movement from one point of interest to another
  - scan paths: moving straight to a target with a short fixation at the target is optimal

Eye tracking

- Eye tracking for usability is still very new and equipment is prohibitively expensive for everyday use.
- However, it is a promising technique for providing insights into what really attracts the eye in website design and where problem areas are in system use.
- More research is needed to interpret accurately the meaning of the various eye movement measurements, as well as to develop more accessible and robust equipment

Physiological measurements

- Emotional response linked to physical changes
- These may help determine a user's reaction to an interface
- Measurements include:
  - heart activity, indicated by 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)
- Some difficulty in interpreting these physiological responses - more research needed

Physiological measurements

- Challenges
  - One of the problems with applying these measurements to interaction events is that it is not clear what the relationship between these events and measurements might be.
  - For example, if increased pulse rate is observed during an interactive task, does that indicate frustration with the interface or stress at being unable to complete the task?

- However, offers a potential means of objectively capturing information about the user's emotional state

Choosing an Evaluation Method

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
Immediacy of response:	immediate vs recall
Level of interference:	obtrusive vs. unobtrusive
Resources available:	time, subjects, equipment, expertise

Classification of evaluation techniques

Table 9.4 Classification of analytic evaluation techniques

	Cognitive walkthrough	Heuristic evaluation	Review based	Model based
Stage	Throughout	Throughout	Design	Design
Style	Laboratory	Laboratory	Laboratory	Laboratory
Objective?	No	No	As source	No
Measure	Qualitative	Qualitative	As source	Qualitative
Information	Low level	High level	As source	Low level
Immediacy	N/A	N/A	As source	N/A
Intrusive?	No	No	No	No
Time	Medium	Low	Low–medium	Medium
Equipment	Low	Low	Low	Low
Expertise	High	Medium	Low	High

Classification of evaluation techniques

Table 9.5 Classification of experimental and query evaluation techniques

	Experiment	Interviews	Questionnaire
Stage	Throughout	Throughout	Throughout
Style	Laboratory	Lab/field	Lab/field
Objective?	Yes	No	No
Measure	Quantitative	Qualitative/ quantitative	Qualitative/ quantitative
Information	Low/high level	High level	High level
Immediacy	Yes	No	No
Intrusive?	Yes	No	No
Time	High	Low	Low
Equipment	Medium	Low	Low
Expertise	Medium	Low	Low

Classification of evaluation techniques

Table 9.6 Classification of observational evaluation techniques

	Think aloud[1]	Protocol analysis[2]	Post-task walkthrough
Stage	Implementation	Implementation	Implementation
Style	Lab/field	Lab/field	Lab/field
Objective?	No	No	No
Measure	Qualitative	Qualitative	Qualitative
Information	High/low level	High/low level	High/low level
Immediacy	Yes	Yes	No
Intrusive?	Yes	Yes[3]	No
Time	High	High	Medium
Equipment	Low	High	Low
Expertise	Medium	High	Medium

1 Assuming a simple paper and pencil record
2 Including video, audio and system recording
3 Except system logs

Classification of evaluation techniques

Table 9.7 Classification of monitoring evaluation techniques

	Eye tracking	Physiological measurement
Stage	Implementation	Implementation
Style	Lab	Lab
Objective?	Yes	Yes
Measure	Quantitative	Quantitative
Information	Low level	Low level
Immediacy	Yes	Yes
Intrusive?	No[1]	Yes
Time	Medium/high	Medium/high
Equipment	High	High
Expertise	High	High

1 If the equipment is not head mounted

Choosing an Appropriate Evaluation method

- For usability evaluation, in practice, a combination of methods is chosen depending on the context and users
- Generally, one analytic method is chosen and an empirical one with users
- There are many possible evaluation techniques that could be appropriate to the scenarios described.

Choosing an Appropriate Evaluation method

For the following scenario, choose appropriate usability evaluation methods. Choose any ONE expert evaluation and ONE user studies giving proper justification.

Note: For the expert evaluation method, you need to select the method and explain briefly why the selected method is viable and for user studies, you need to mention the subjects, technique selected, tasks to carry out, measurement and an outline plan.

You have designed and implemented a **new game system for teenagers** and want to evaluate it before release.

Choosing an Appropriate Evaluation method

Expert evaluation: Cognitive walkthrough (CW)

Justification for the method selected:

- As gaming is more an interactive application, a task oriented evaluation method such as CW is appropriate than the holistic HE. Using CW, evaluators can walk in users' shoes and evaluate whether the interaction provided by the system matches with users' mental model.

Choosing an Appropriate Evaluation method

User studies:

1. Subjects:
   - Teenagers who are interested in games
   - A few teenagers who are not interested in games can also be selected to see if the developed system has a wider appeal
2. Technique: Observation - Think Aloud followed by a Questionnaire and Interview
3. Tasks:
   - Playing a game and achieving the score (Task can be very specific also)
4. Measurement
   - Qualitative measures of Extent of game mastered, Gamers satisfaction
   - Quantitative measures of Scores achieved , Speed of response etc.
   - Qualitative (think aloud and Interview) and Quantitative data (Questionnaire)
5. Outline Plan
   - Allow subjects to play game and talk as they do so. Record the data
   - Follow up with questionnaire and Interview
   - Analyze qualitative and quantitative data collected

Summary

- Analytical
  - System oriented – HE
  - Interaction oriented – CW
  - Model based (GOMS, KLM) – next unit
- Empirical (based on actual usage data)
  - Scientific way of collecting and analyzing observational data from end users on an interactive system
    - Experiment, Query
    - Observation
    - Physiological

Research Methods in HCI

- As HCI is inter-disciplinary, research methods taken from different disciplines
  - Psychology (Experiments)
  - Social Science (Interviews and Questionnaires)
  - Business research methods (Statistical analysis)

- However, some are from HCI discipline
  - Analytic (HE, CW, model based)
  - Observational methods

Research Methods in HCI

- Change in topics of research over time
  - Now its not much about simple task performance
  - Focus is more towards user-generated content
  - Even that will change
  - Shift in measurement also
    - Example – Many people spend time in creating and editing Wikipedia entries.
    - What causes them to do so? What do they get out of the experience?
    - Task and time performance not appropriate metrics
    - What metrics should be used? Joy? Emotion?
  - Shift in the focus of research, need to adapt/develop new research methods