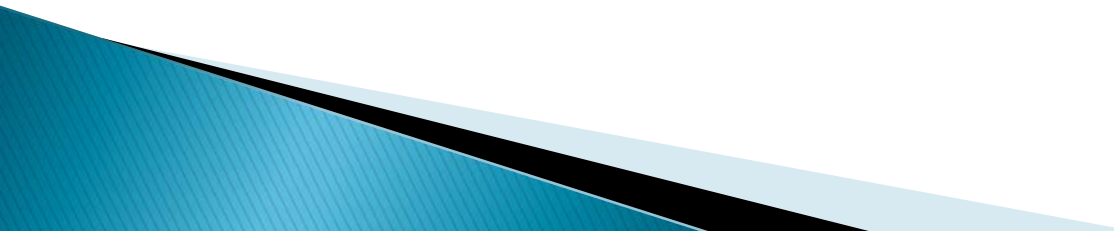Professors d'IDI - UPC

IDI – Usability Testing

Outline

- **Concepts**
- Usability testing
- Formal usability tests
- Simplified usability tests
- Use cases

Concepts

▸ Usability:

◦ **Ease of use** and **acceptability** of a system or product for <u>a particular class of users</u> carrying out <u>specific tasks</u> in a <u>specific environment</u>.

▪ Where "Ease of use" affects user performance (efficacy, efficiency), satisfaction (comfort).

▪ And "Acceptability" affects whether or not the product is used.
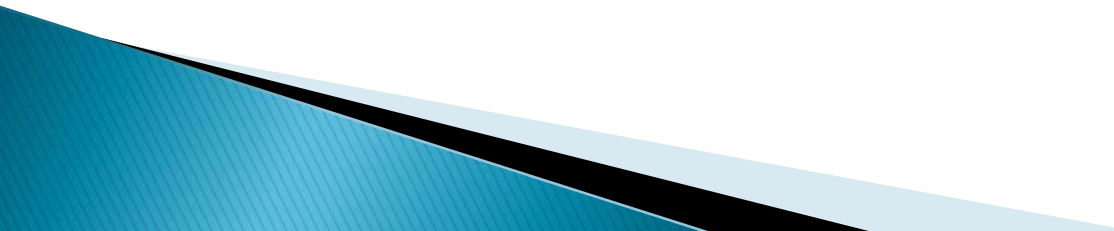
Concepts

▸ Usability:

◦ The extent to which a product can be used by <u>specified users</u> to achieve <u>specified goals</u> with effectiveness, efficiency, and satisfaction in a <u>specified context of use</u>.

◦ To be useful, usability has to be specific. It must refer to <u>particular tasks</u>, <u>particular environments</u> and <u>particular users.</u>

➜ So has to be its testing!

Concepts

▸ How to test?

- Ease of use is inversely proportional to the number and severity of difficulties people have in using software.

- Let's examine the difficulties!!!

Outline

- *Concepts*
- **Usability testing**
- Formal usability tests
- Simplified usability tests
- Use cases

Usability testing

▸ Two major families by goals:

◦ Determine usability problems (*i.e.text editor*):

· Discovery, prioritization, and resolution of usability problems

· <u>Iterative testing</u>

◦ Measure task performance (*i.e. 3D selection*).

Include two fundamental tasks:

· The development of the usability objectives

· <u>Iterative testing</u> to determine if the product under test has met the objectives

Usability testing

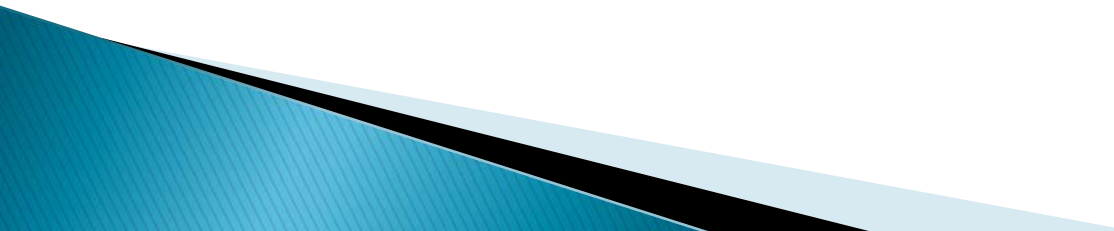- Great variety of usability tests:
  - Can be very informal or very formal
  - Often use think-aloud (TA)
    - more reliable than posterior interviews
    - doesn't affect efficiency
    - better for problem discovery than measurement
  - Remote or local
  - Evaluated software can be varied:
    - Prototypes, under development, competitive products…

Usability testing

- Testing techniques:
  - "Formal" usability tests
  - Remote testing
  - Guerrilla usability testing
  - Steve Krug's "usability testing on 10 cents a day"
  - Heuristic/expert evaluation

Outline

- *Concepts*
- *Usability testing*
- **Formal usability tests**
  - Environment
  - Tasks & roles
  - Development
  - Reporting
- Simplified usability tests
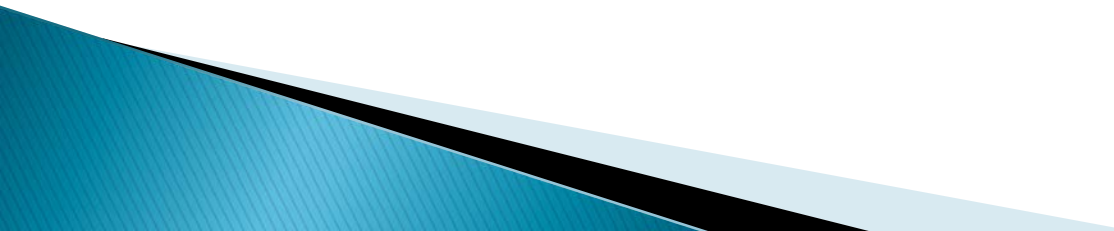- Use cases

Formal usability tests. Environment

- "Formal" usability tests require a controlled environment
  - Inside a room, outside…
    - Illumination conditions (useful for perception studies)
  - Devices used (e.g. computer with Internet connection and a browser, or a mobile…)
  - Other conditions (e.g. connection quality…)
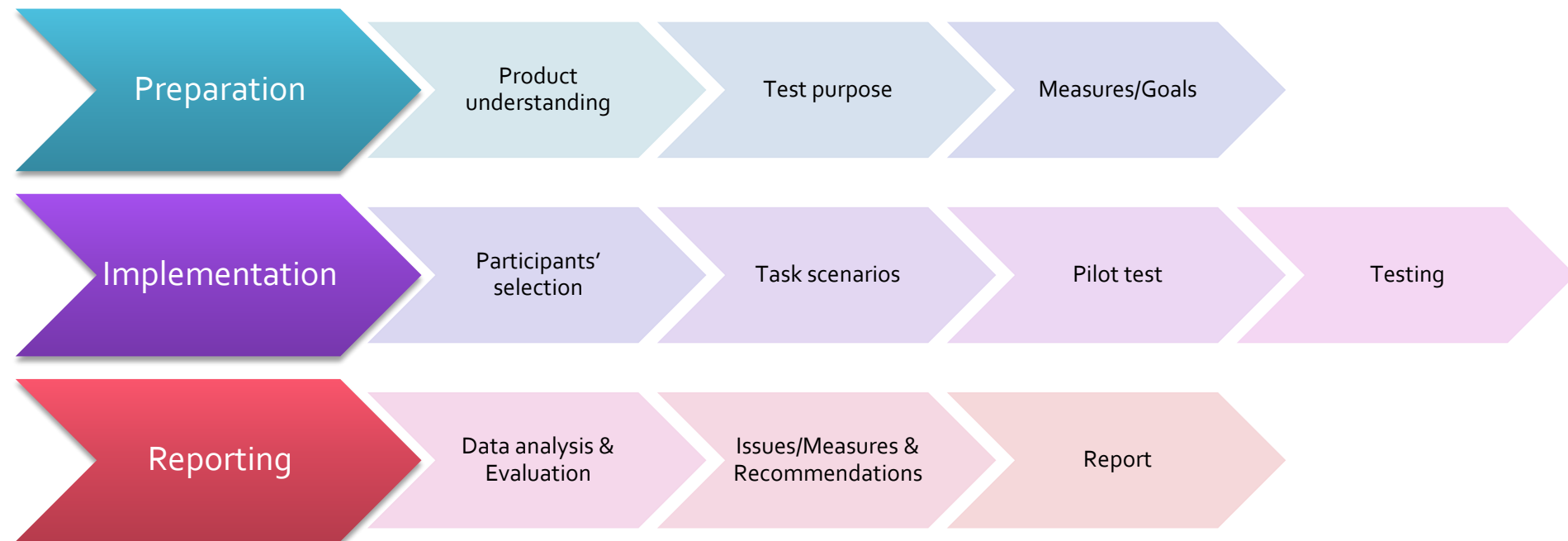 - Usability lab ☺

Formal usability tests. Environment
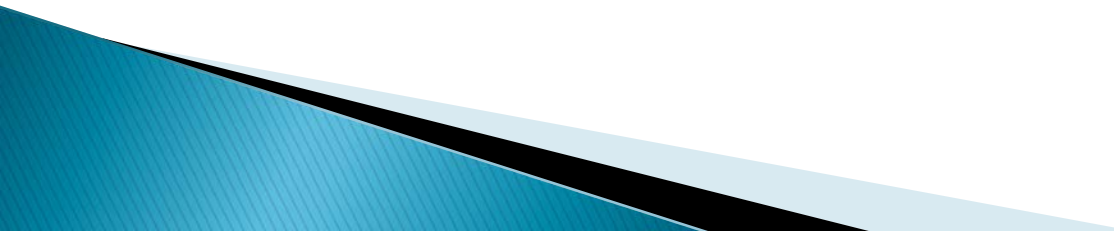
Formal usability tests. Environment

- Set of soundproofed rooms
  - Proper recording and avoiding distractions to participants

- Different areas and equipment
  - Participant area (where the experiment is carried out)
  - Observer area with one-way glass
  - Executive viewing area behind the primary observer area
  - Video cameras and microphones, telephone…

Formal usability tests. Tasks and roles

▸ Usability test workflow:

Preparation	Product understanding	Test purpose	Measures/Goals	
Implementation	Participants' selection	Task scenarios	Pilot test	Testing
Reporting	Data analysis & Evaluation	Issues/Measures & Recommendations	Report	

Formal usability tests. Tasks and roles

▶ Usability test roles:
- **A**: Test administrator
- **B**: Briefer
- **CO**: Camera Operator
- **DR**: Data Recorder
- **HD**: Help Desk Operator
- **PE**: Product Expert
- **S**: Statistician

Formal usability tests. Preparation

▸ **Preparation (A):**

◦ Product understanding: Purpose of the product, parts ready to test, type of users...: A, PE

◦ Test purpose: Product comparison, within/between subjects...: A, S

◦ Measures/Goals: Number of iterations, counting mistakes/errors, timings...: A, S

Preparation ▸ Product understanding ▸ Test purpose ▸ Measures/Goals

Formal usability tests. Preparation (1)

- **Product understanding (A + PE):**
  1. Understand the purpose of the product
  2. Parts of the product are ready for testing
  3. Types of people who will use the product
  4. Determine the use given to the product
  5. Conditions of usage of the product

| Preparation | Product understanding | Test purpose | Measures/Goals |

Formal usability tests. Preparation (2)

▸ Preparation:
  ◦ Product understanding: Purpose of the product, parts ready to test, type of users…: A, PE
  ◦ **Test purpose**: Product comparison, within/between subjects…: **A, S**
  ◦ Measures/Goals: Number of iterations, counting mistakes/errors, timings…: **A, S**

Preparation ▸ Product understanding ▸ **Test purpose** ▸ Measures/Goals

Formal usability tests. Preparation (3)

▸ Preparation:

  ◦ Product understanding: Purpose of the product, parts ready to test, type of users…: A, PE

  ◦ Test purpose: Product comparison, within/between subjects…: A, S

  ◦ **Measures/Goals**: Number of iterations, counting mistakes/errors, timings…: **A, S**

Preparation ▸ Product understanding ▸ Test purpose ▸ **Measures/Goals**

Formal usability tests. Implementation

▸ **Implementation:**

◦ Participants' selection: **A**

◦ Task scenarios: initial conditions, steps: **A**

◦ Pilot test: Members of the team: **A, B, CO, DR, HD**

◦ Testing, **A, B, CO, DR, HD**

Implementation → Participants' selection → Task scenarios → Pilot test → Testing

Formal usability tests. Implementation (1)

- **Participants' selection**
  - It's complicated.
  - Should be representative
    - People that could be real users
      - E.g. no other managers!!!
    - Specialized recruiting agencies are a possibility
  - No-show rates above 10%

Implementation ▸ **Participants' selection** ▸ Task scenarios ▸ Pilot test ▸ Testing

Formal usability tests. Implementation (1)

- ▸ **Participants' selection**
  - ◦ Should be payed. Factors to consider:
    - • Time needed, qualification of the participant, would be the person be payed for her time?
    - • Non-monetary for internal participants that are already being payed for their time
    - • Always offer food and beverages!!!

Implementation → **Participants' selection** → Task scenarios → Pilot test → Testing

Formal usability tests. Implementation (1)

- **Participants' selection**
  - For statistical significance: 10-12 participants
  - Less formal: 4-5 participants per user group
  - Ensure recruiting criteria reflects user characteristics
    - E.g. for a website, ensure participants have prior experience browsing

| Implementation | **Participants' selection** | Task scenarios | Pilot test | Testing |

Formal usability tests. Implementation (2)

▸ **Test task & scenarios:**

◦ Tasks must be representative

- Core tasks: Features that everybody uses (write a text)
- Peripheral tasks: Features used less often (table insertion)

| Implementation | Participants' selection | **Task scenarios** | Pilot test | Testing |

Formal usability tests. Implementation (2)

- **Test task & scenarios:**
  - Scenarios must be determined
    - Define initial conditions
    - Description of the scenario: what to do and why
      - Should not provide step–by–step instructions but should include details
      - Some action must be taken on finish
      - Not all users must be provided with the same scenarios (may depend on the user profile)

| Implementation | Participants' selection | **Task scenarios** | Pilot test | Testing |

Formal usability tests. Implementation (3)

▶ **Testing** (**A, B, CO, DR, [HD]**):

- Brief participants: **B**
- Initial questionnaire: **B**
- Develop tasks: **B, CO, DR, [HD], A**
- Debrief: **B**
- Final questionnaires: **B**

Implementation | Participants' selection | Task scenarios | **Pilot test** | **Testing**

Formal usability tests. Reporting

- **Reporting** (whole team):
  - Data Analysis & Evaluation: **A, S**
  - Issues/Measures & Recommendations: **A, S**, team
  - Report: **A, S**, team
    - Describe & prioritize the usability problems
    - Present quantitative measurements

| Reporting | Data analysis & Evaluation | Issues/Measures & Recommendations | Report |
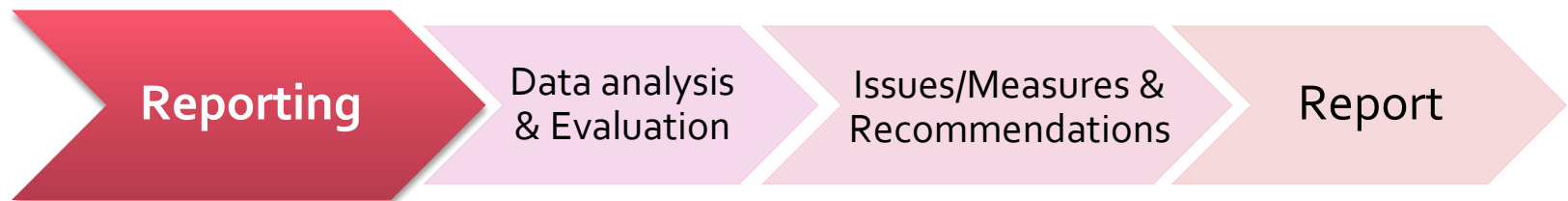
Formal usability tests. Reporting (1)

- **Data analysis & evaluation**:
  - Frequency: Number of users that find a problem divided by the number of users testing the app or web
    - Easy (objective) to evaluate
  - Severity: Importance of the problem
    - Might be completely catastrophic or simply cosmetic
    - Difficult (more subjective) to evaluate

Reporting → **Data analysis & Evaluation** → Issues/Measures & Recommendations → Report
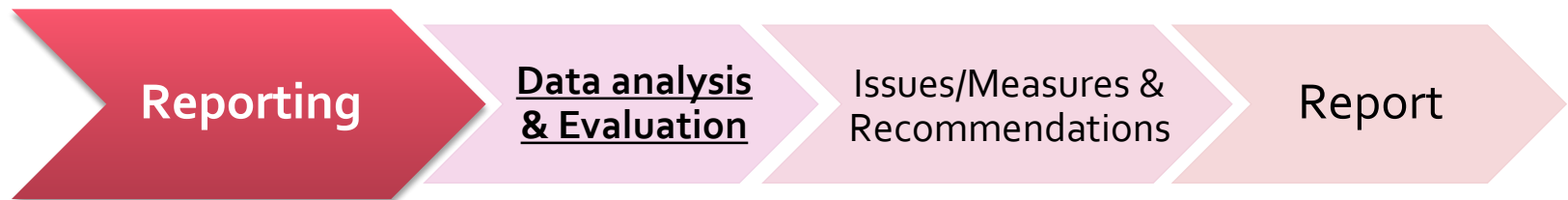
Formal usability tests. Reporting (1)

▶ Usability problems:
  ◦ Should indicate the importance: <u>severity</u>
  ◦ Can be classified:
    • Mistakes: Errors due to incorrect intention
    • Slips: Errors due to appropriate intention but incorrect action
  ◦ Expertise does not affect on the number of errors
    • But affects how fast they are handled

Reporting ▶ **Data analysis & Evaluation** ▶ Issues/Measures & Recommendations ▶ Report
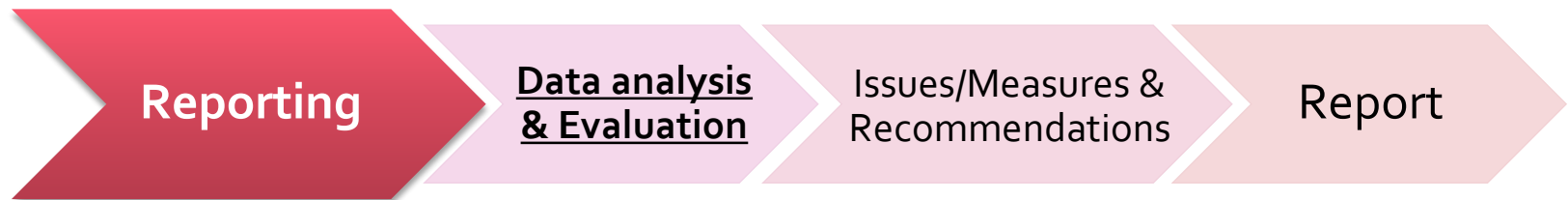
Formal usability tests. Reporting (1)

▸ Problem evaluation. **Dumas and Redish:**
  ◦ **Level 1**: Prevents Task Completion
  ◦ **Level 2**: Creates significant delay and frustration
  ◦ **Level 3**: Problems have a minor effect on usability
  ◦ **Level 4**: Subtle and possible enhancements/suggestions

Reporting ➤ **Data analysis & Evaluation** ➤ Issues/Measures & Recommendations ➤ Report
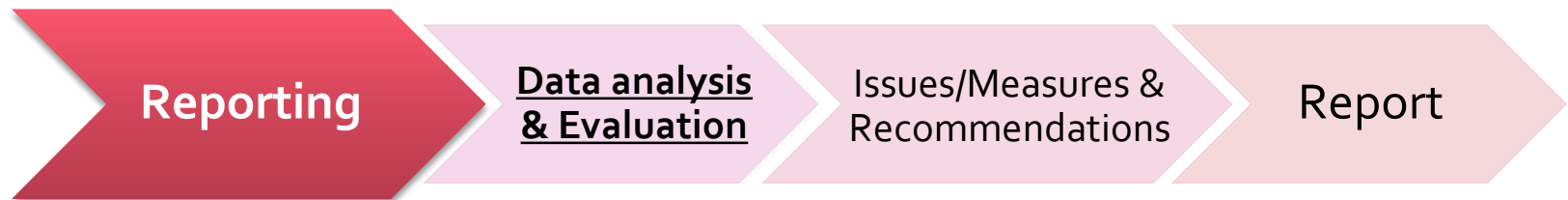
Formal usability tests. Reporting (2)

▸ **Recommendations:**
  ◦ Create a problem grid: frequency/impact
  ◦ Global changes (prevent task completion) first
    • A *missing help* may be a global problem or something related with a concrete UI
  ◦ Try to give at least one recommendation for each problem
    • Present the different trade-offs clearly

| Reporting | Data analysis & Evaluation | **Issues/Measures & Recommendations** | Report |
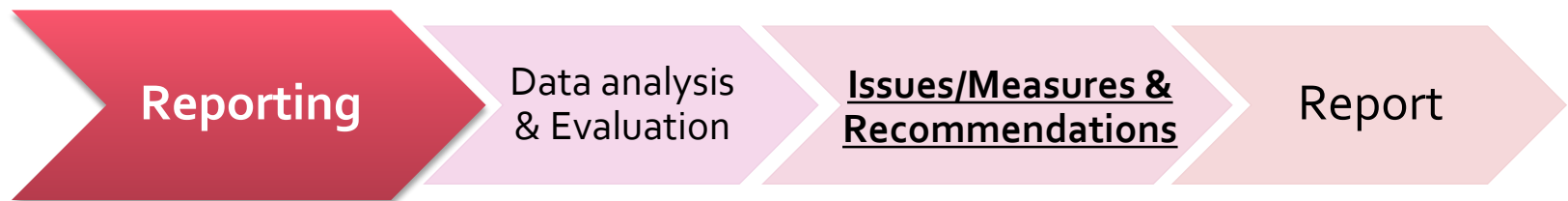
Formal usability tests. Reporting (3)

▸ **Problem evaluation. Conclusions**
  ◦ Do not use a large number of categories
    • Do not get obsessed by the number of categories either
  ◦ Different evaluators may disagree on some problems' severity
  ◦ Treat frequency separately from severity
  ◦ Do not forget to point out positive findings
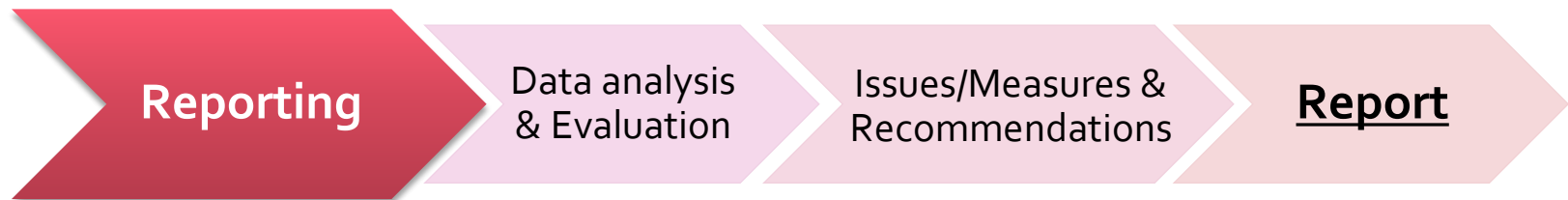
Reporting ▸ Data analysis & Evaluation ▸ Issues/Measures & Recommendations ▸ **Report**

Outline

- *Concepts*
- *Usability testing*
- *Formal usability tests*
- **Simplified usability tests**
  - Guerrilla usability testing
  - Steve Krug's "usability testing on 10 cents a day"
  - Remote testing
  - Heuristic/expert evaluation
- Use cases

Simplified usability tests

- Testing just a single person early is much better than 50 near the end
- The point of testing is to inform your judgment

Simplified usability tests

▸ **Guerrilla usability testing**
  ◦ Take someone in a coffee or public space and ask her to use a website for a couple of minutes
  ◦ Observe users
    • Ask open-ended questions such as "What would you do here?"
  ◦ Get to know them a bit
    • Offer coffee or bagels
  ◦ Analyse captured data
    • Considering your audience

Simplified usability tests

- **"Usability testing on 10 cents a day"**
  - Prepare some tasks to evaluate
  - Grab somebody from the company as user
  - Gather stakeholders in an observing room
  - Let the user do a set of tasks
  - Capture gestures, mouse, record…
  - Discuss over lunch (order pizza for everybody)
  - Report

Simplified usability tests

- **Remote testing**
  - Like traditional tests but participant and facilitator are in different physical locations
    - Participants can do the test at home
    - Facilitator watches remotely
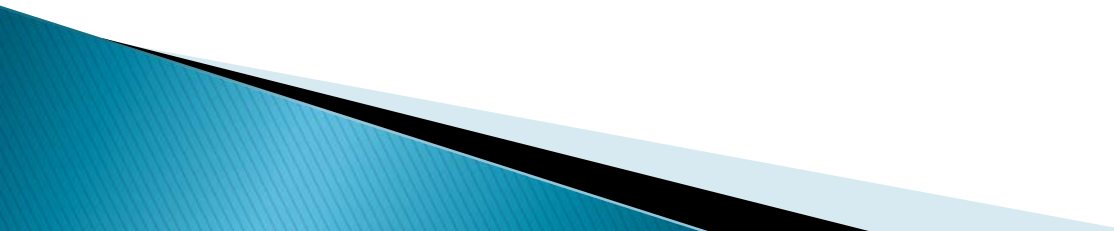
Simplified usability tests

- **Remote testing**
  - Advantages
    - Cheaper and easier test setup
    - Usually faster (in terms of allocating/securing facilities travel…)
    - Can get geographically dispersed users
  - Disadvantages
    - Cannot read body language.
    - Difficult to decide when to talk/interact
    - Variability in participants' motivation
    - No-show rates higher than in-person studies

Simplified usability tests

▸ **Remote testing.** Two types:
  ◦ Unmoderated:
    • Users do the task completely alone
  ◦ Moderated:
    • Users have access to a facilitator

Simplified usability tests
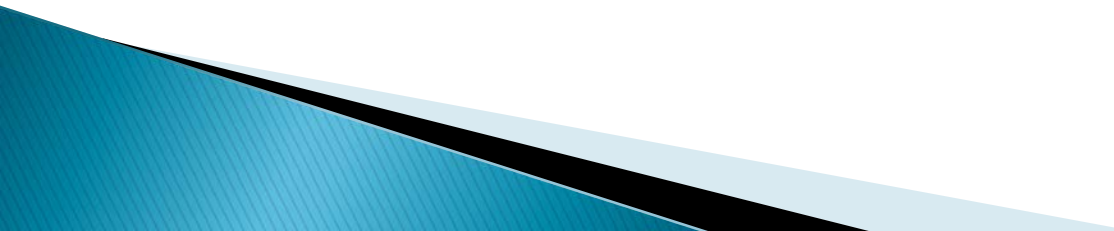
▸ Unmoderated **Remote testing**
- Users don't have real-time support
- Don't get any clue on how the session went
- No opportunity to ask detailed questions
  - Sometimes the software allows to have some of them predefined
- Preferable to work only on a few specific elements than a broad view of a product
- Good for tight timeframes

Simplified usability tests

- Moderated **Remote testing**
  - Facilitator can change or reorder tasks as needed
  - Facilitator can ask follow-up questions or clarifications
  - Participant is less likely to spend time on tasks not related to the test
  - Test sessions can be longer (usually about an hour)
  - Can perform more in-depth tests
  - The team can watch the test and discuss afterwards

Simplified usability tests

▶ **Heuristic evaluation:**

- ◦ 3-5 usability experts evaluate an app or UI
- ◦ Use pre-defined principles (heuristics)
- ◦ Can highlight usability issues before user testing

Simplified usability tests

▶ **Heuristic evaluation.** Advantages

  ◦ Can be quick and cost effective

    • If we have internal resources

  ◦ Can be used early in the design process

  ◦ Can give a comprehensive usability status of a product's UI

  ◦ Is compatible with other usability testing methods

Simplified usability tests

- **Heuristic evaluation.** Process:
  - Collect the UI
  - Understand the business and users' needs
  - Understand user motivations and tasks to accomplish
  - Define the heuristics
  - Use a minimum of 3 experts
  - Set up a consistent evaluation system
  - Highlight problem(s) and its rating
  - Compare and analyse the results of multiple experts

Outline

- *Concepts*
- *Usability testing*
- *Formal usability tests*
- *Simplified usability tests*
- **Use cases** *(Read & Study the document uploaded to the "racó")*
  - Guerrilla testing: WhatsApp web app
  - Measure test: Depth perception in VR

Use case. WhatsApp web app

- Web application

Use case. WhatsApp web app

- Type of usability test: Guerrilla
- Objective
  - Identify common problems on WhatApp web
- Testing parameters
  - What is tested: Just two common tasks
  - Participants: 3 users, 2 never used it previously
  - Test procedure: Observation + interview

Use case. WhatsApp web app

▸ Test process
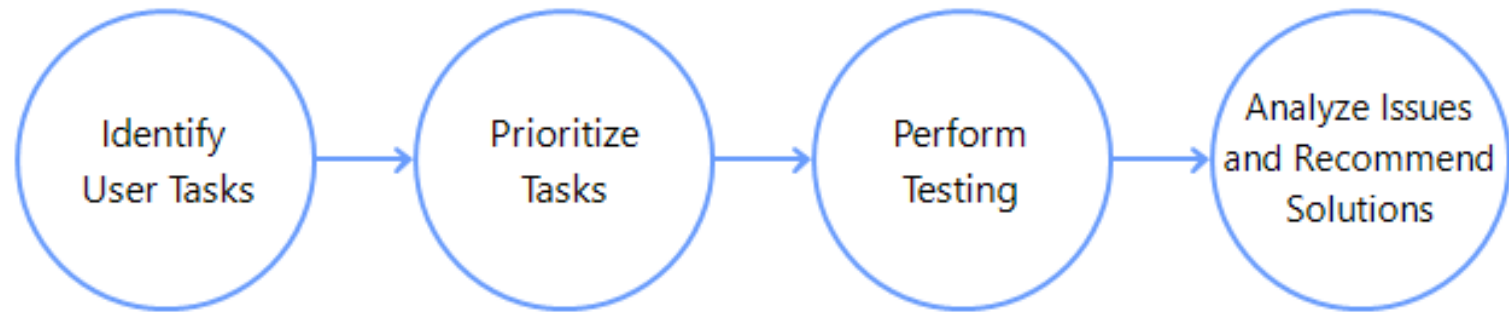
Use case. WhatsApp web app

▶ User tasks
  ◦ Send a message to a friend
  ◦ Share photos with a friend

Identify User Tasks → Prioritize Tasks → Perform Testing → Analyze Issues and Recommend Solutions
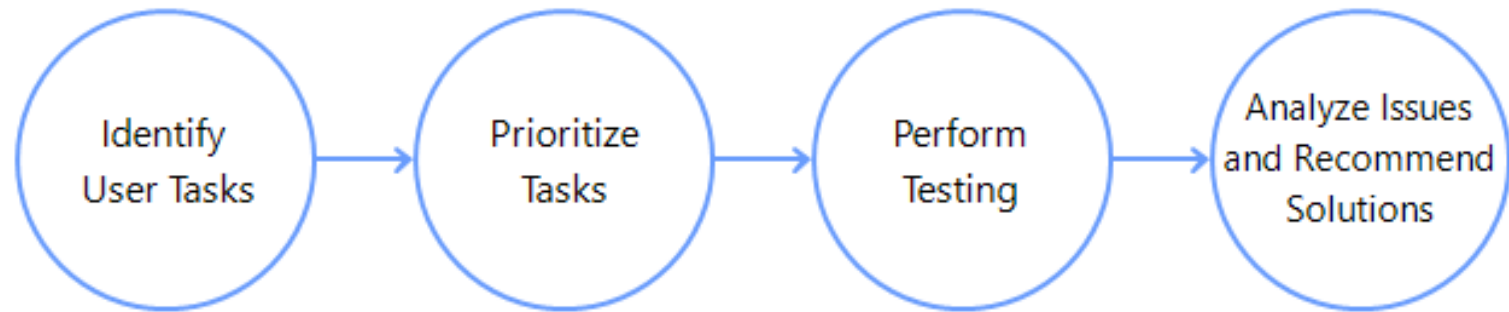
Use case. WhatsApp web app

- Development (perform testing)
  - Give the instructions to the users
    - Users are observed with performing actions
    - Asked about the experience on certain subtasks

Use case. WhatsApp web app

▸ Development of task 1
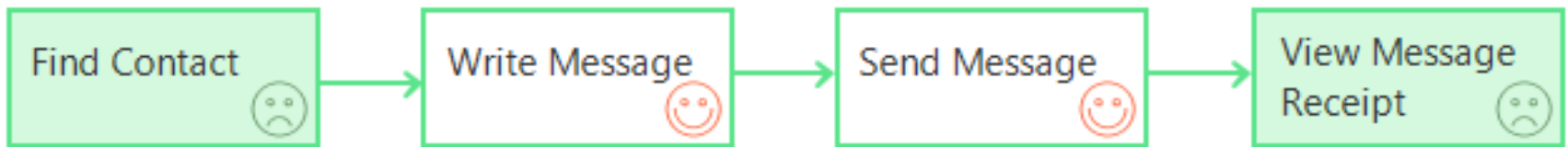
Use case. WhatsApp web app

▸ Development of task 2
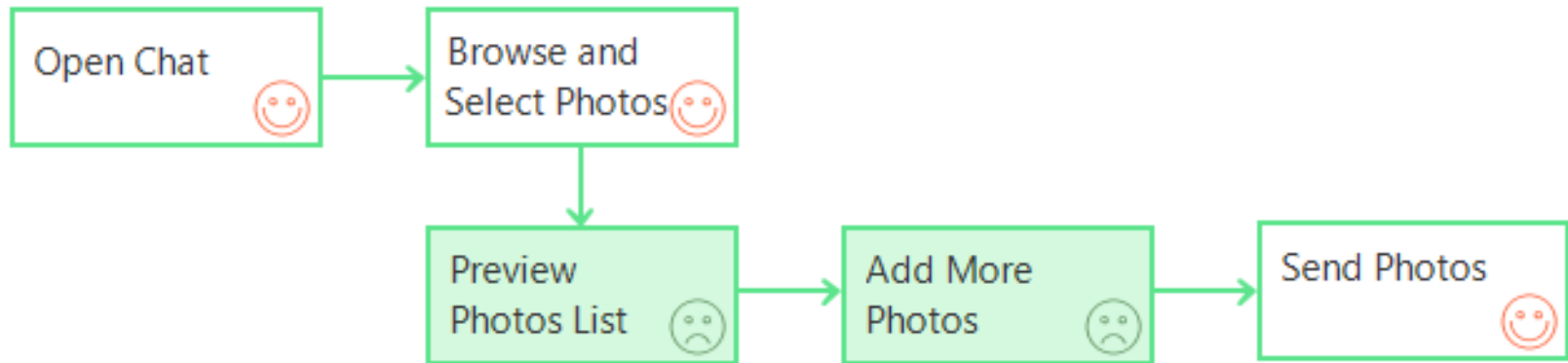
Use case. WhatsApp web app

▸ Analyse problems. **Finding a contact**:
  ◦ There are two ways to start a new chat:
    • (i) Search within Chats list
    • (ii) Go to New Chat icon on top and search contact
  ◦ The user was not clear to differentiate between these two options
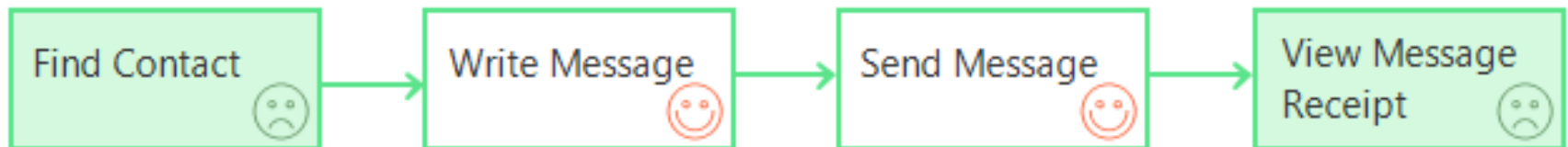
Use case. WhatsApp web app
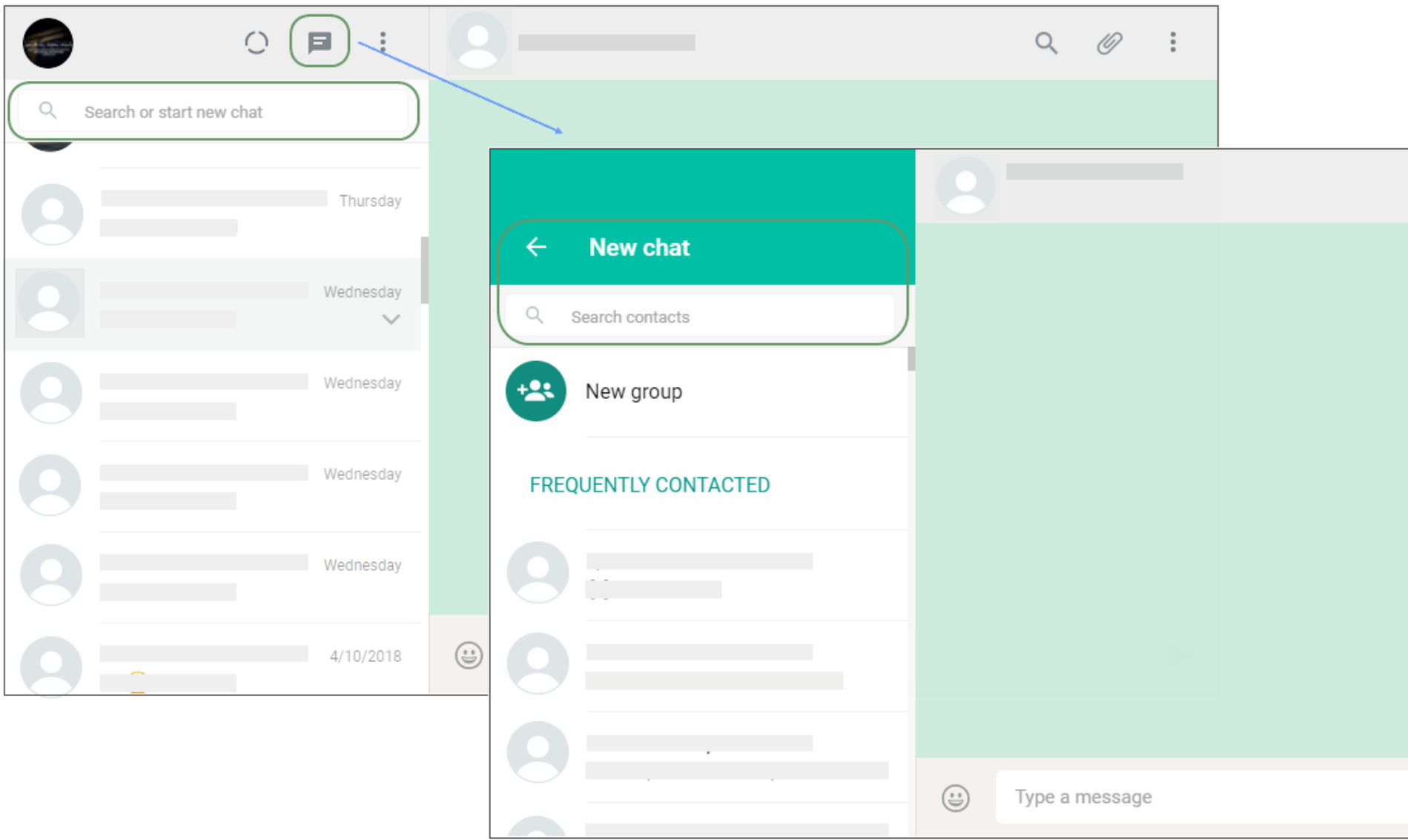
Use case. WhatsApp web app

▸ **Recommendation:**

◦ A clear separation between Chats and Contacts is needed

• Can be done by giving a filter option in Contact list, or a single list can be sorted based on Recent Chats or Contact names.

Use case. WhatsApp web app
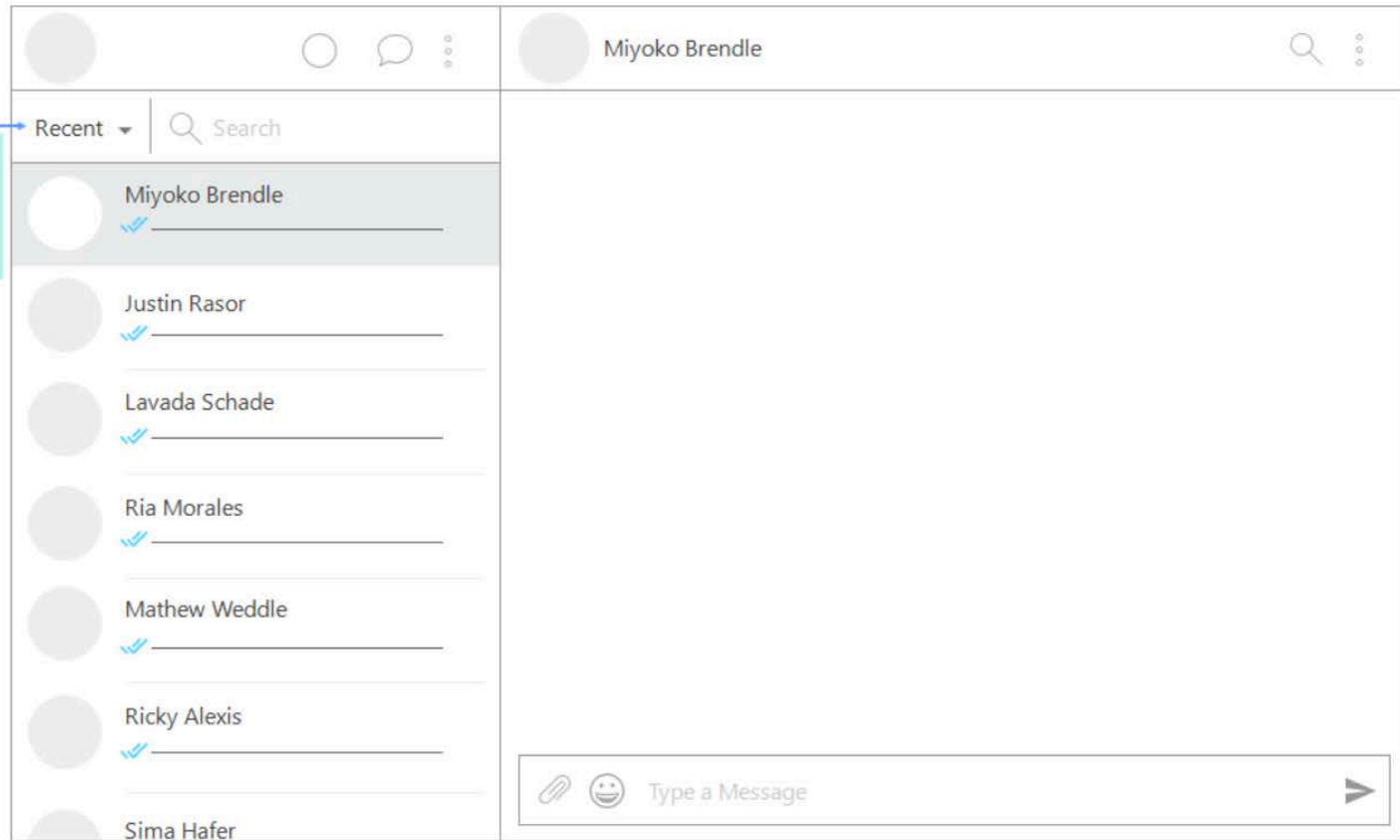
Use case. WhatsApp web app

▸ Analyse problems. **Viewing Message Receipt**:

◦ In Message Info pane, the area showing message status is merged with the Message pane

◦ Also, it is not clear that user is viewing status of which message

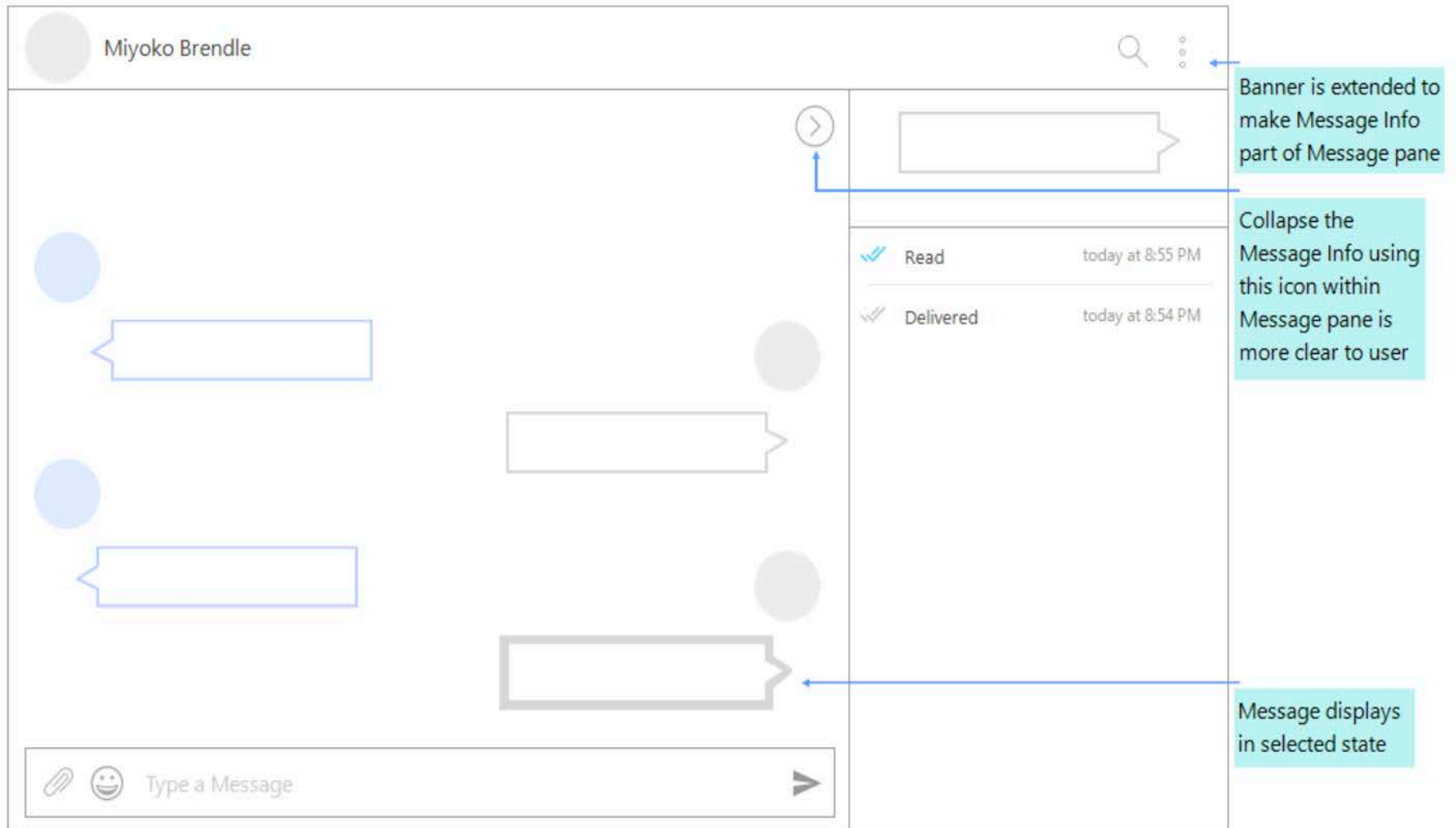◦ Also, it took time for the user to find Close icon on top on Message Info pane

| Find Contact ☹ | → | Write Message ☺ | → | Send Message ☺ | → | View Message Receipt ☹ |
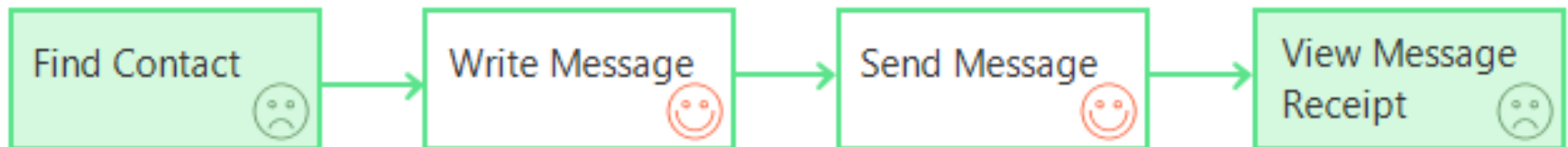
Use case. WhatsApp web app
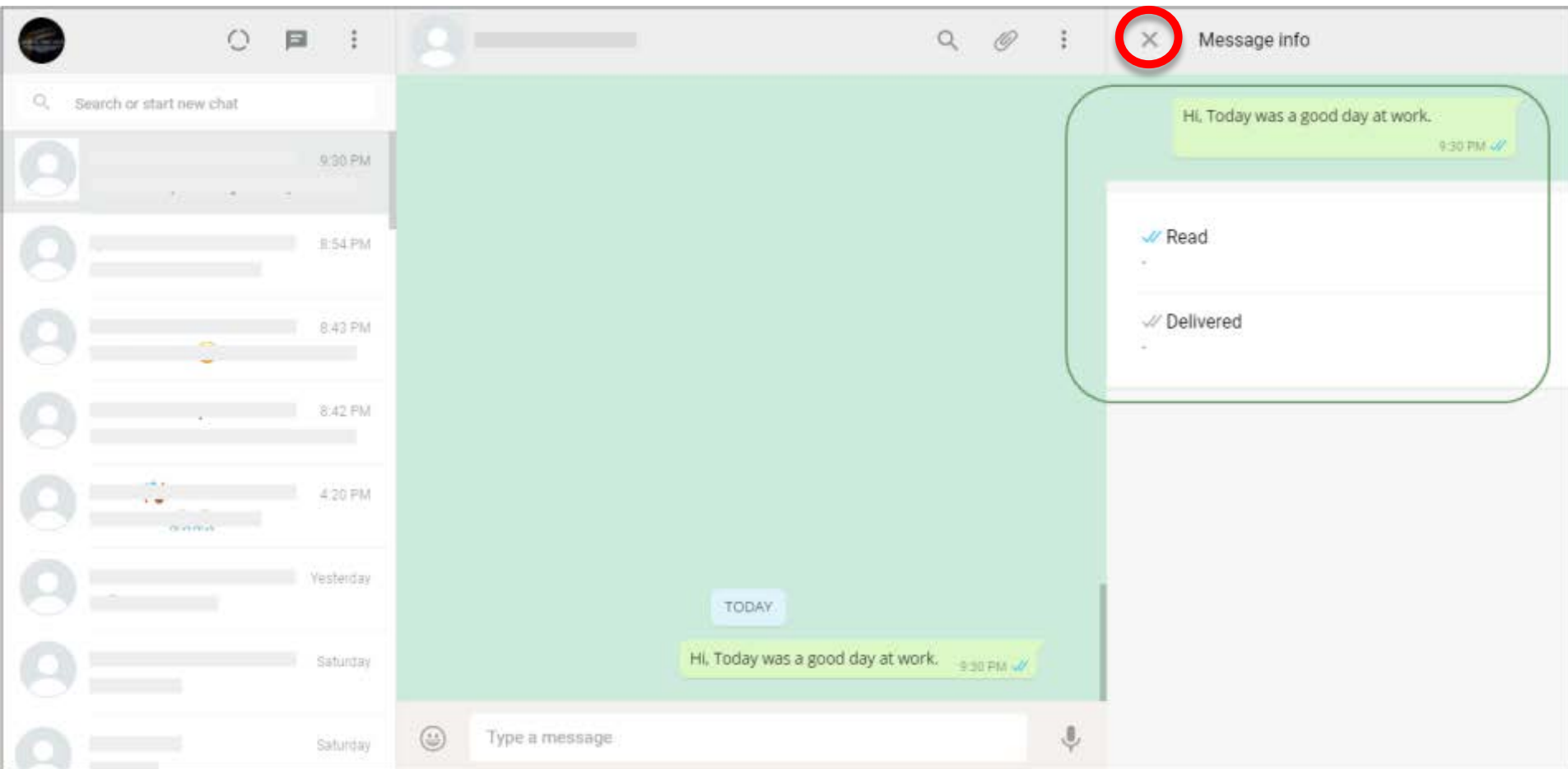
Use case. WhatsApp web app

- **Recommendation:** The area of Message info pane and Message pane needs to differentiate clearly
  - Since this is desktop version and Message area is still visible when Info pane is opened, the link between message and its info could be made more prominent
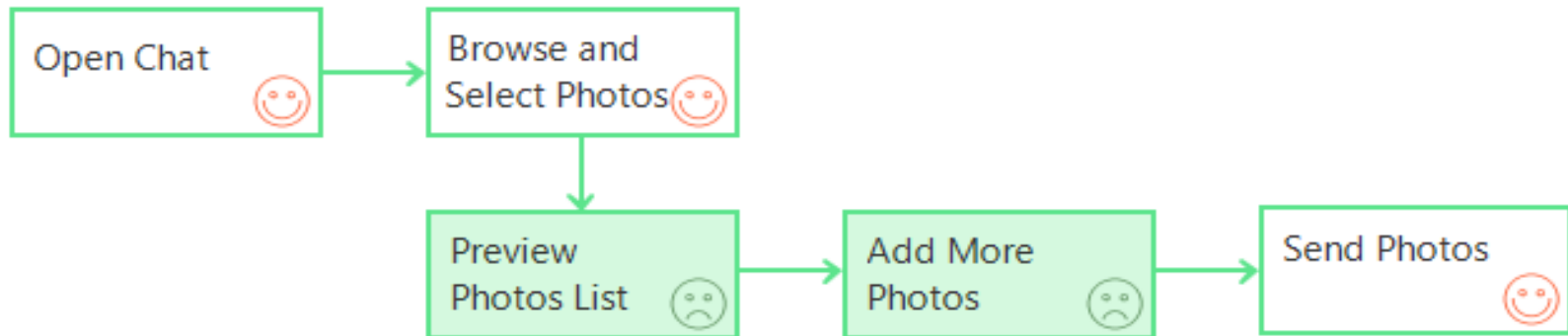
Use case. WhatsApp web app

Miyoko Brendle

Banner is extended to make Message Info part of Message pane

Read — today at 8:55 PM

Delivered — today at 8:54 PM

Collapse the Message Info using this icon within Message pane is more clear to user

Message displays in selected state
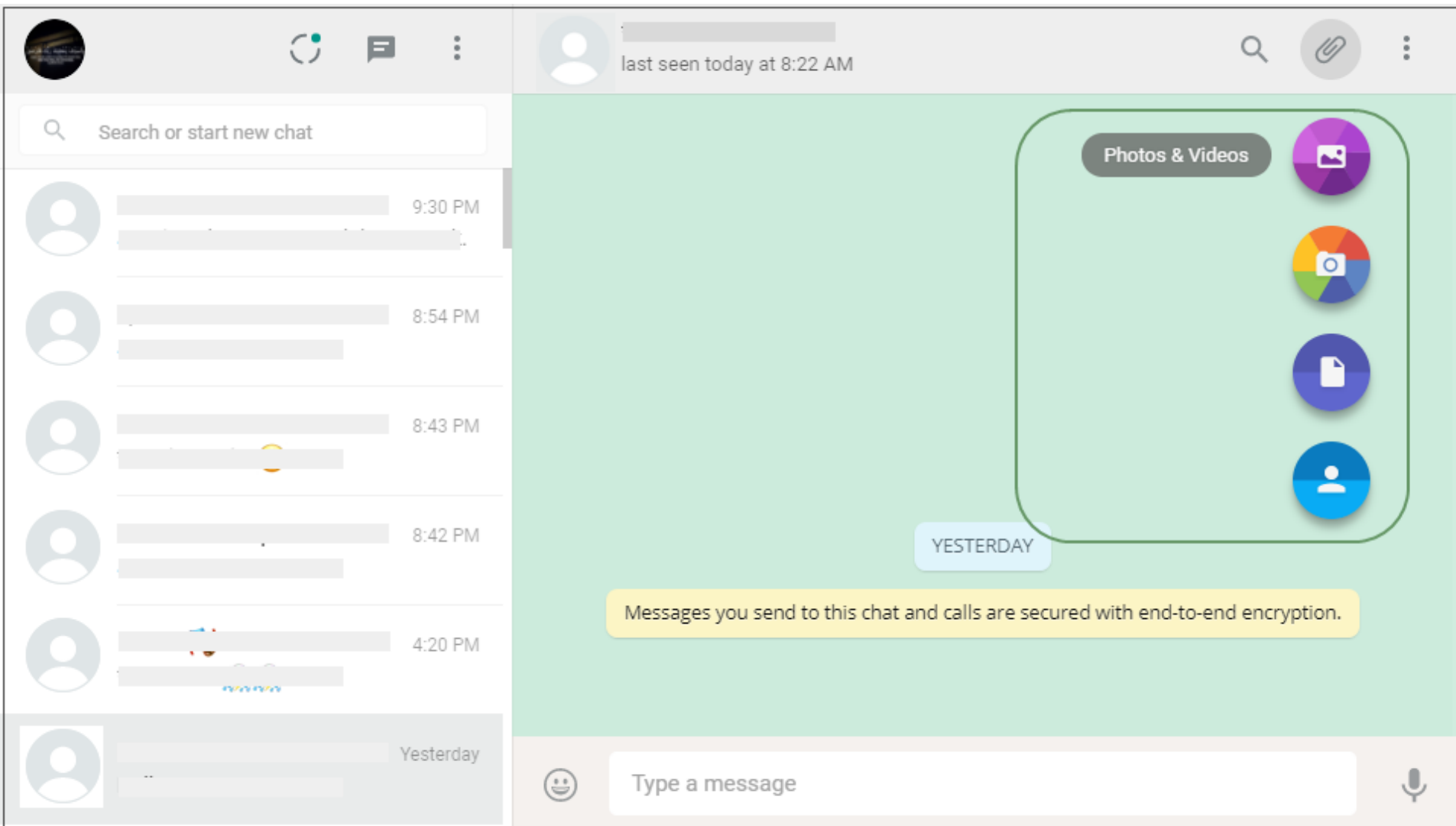
Type a Message

Use case. WhatsApp web app

▸ Development of task 2
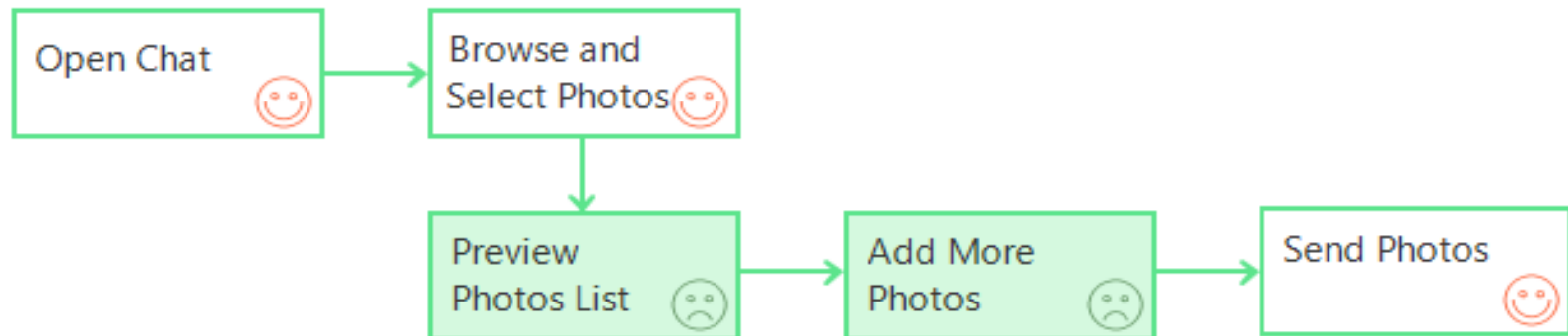
Use case. WhatsApp web app

‣ Analyse problems. **Using attach**

  ◦ The Attach menu and tooltips do not match with the UI

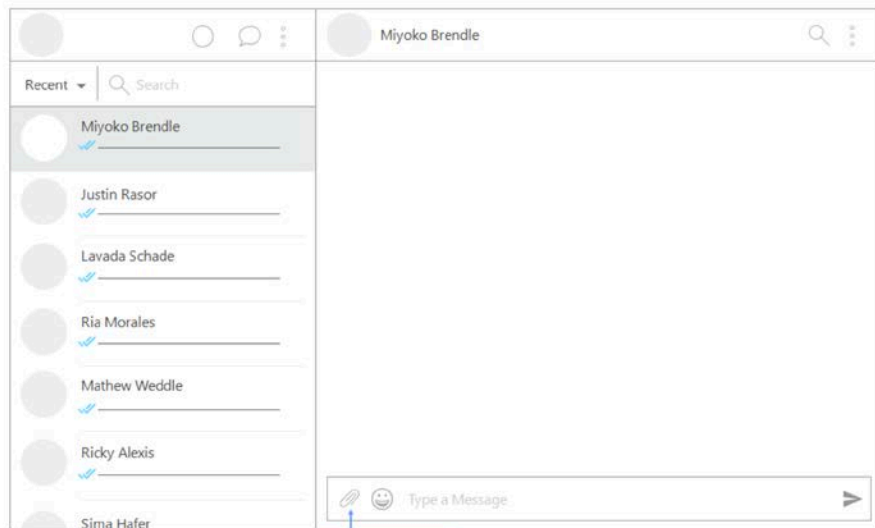   • Shows a totally different theme and experience in current screen
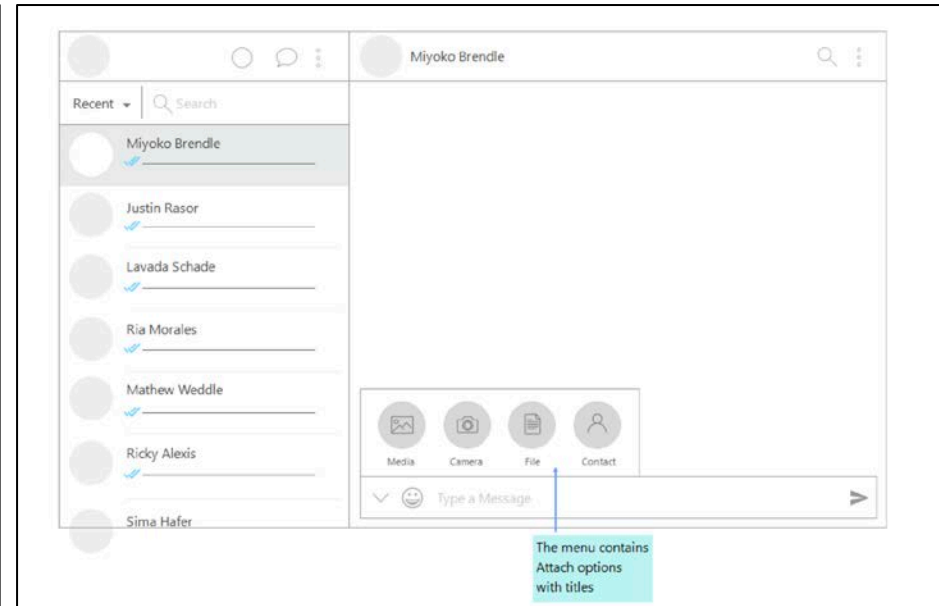
Use case. WhatsApp web app

Use case. WhatsApp web app

▸ **Recommendation**: Make menu placement and theme consistent with UI.

  ◦ Instead of Tooltips, the option names along with icons seems more helpful.

Use case. WhatsApp web app

▸ Analyse problems. **Attaching photos:**

◦ Close icon with Preview title is confusing.

- The user clicked it just to close the preview of selected photos, but it discards all the selected photos.

◦ Adding more files option is not clear.

- The Attach icon still displays on top, but it is not functional. The user clicked on that icon first.

◦ It is difficult to navigate large number of selected files.
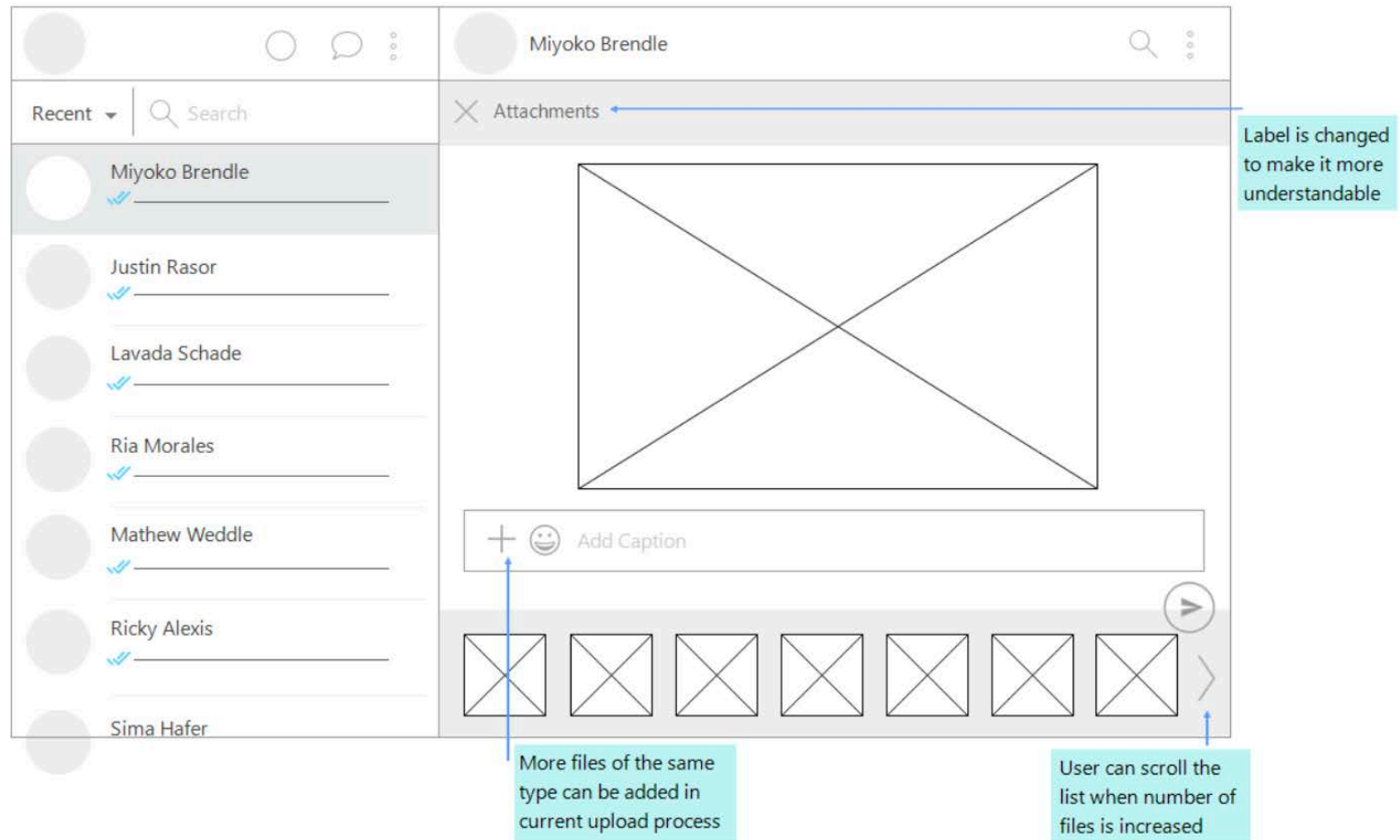
Use case. WhatsApp web app
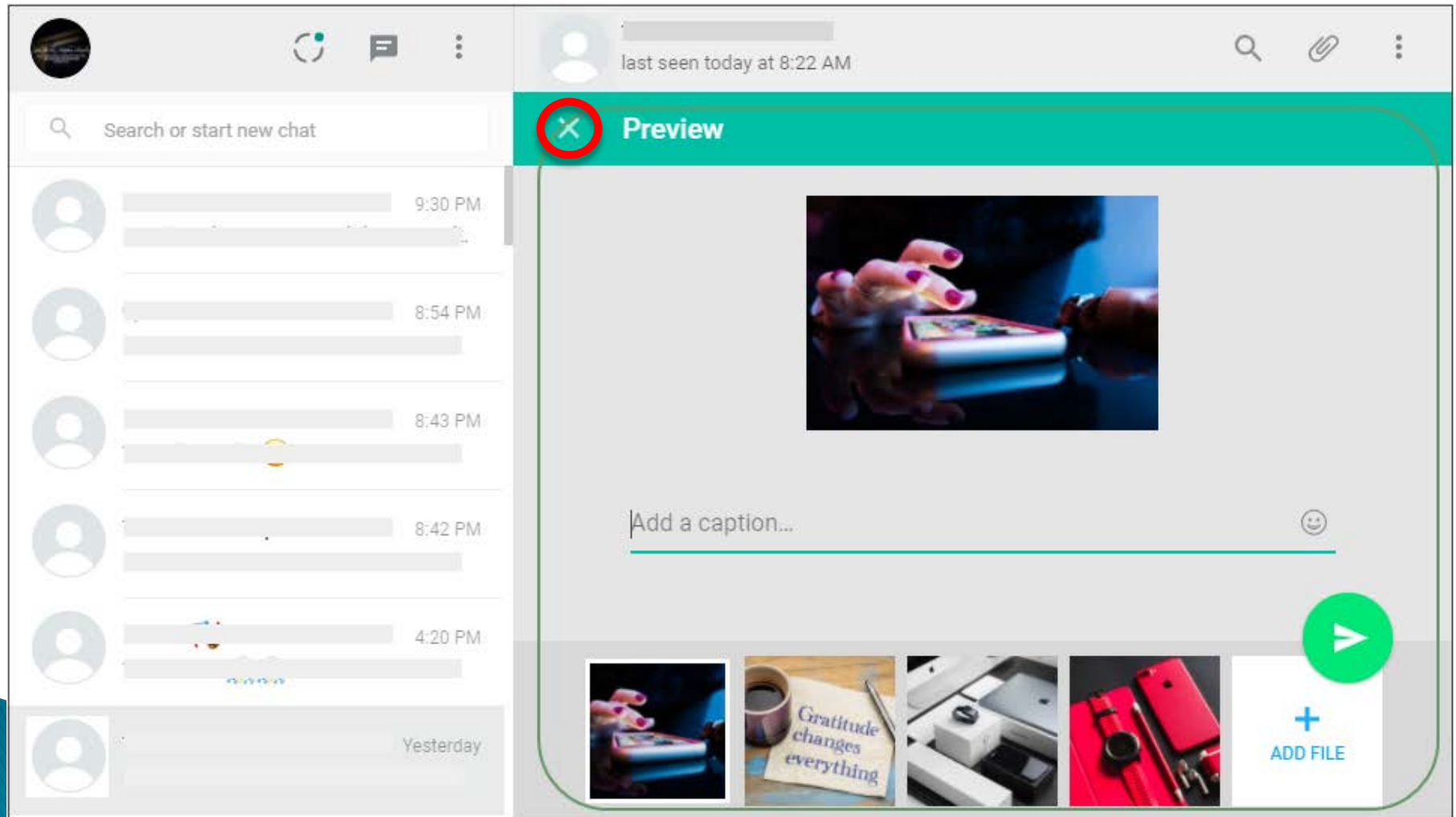
Use case. WhatsApp web app

▶ **Recommendation:**

  ◦ Rename preview area to Attachments to avoid any confusion for the user.

  ◦ Scrolling in thumbnails area

  ◦ User should be able to add more files by clicking an Add icon with caption

Use case. WhatsApp web app

Use case. WhatsApp web app

▸ **More Observations**
  ◦ Using a scrollbar requires high accuracy to hold the bar and scroll it
    • Cursor is changed to resize when user tries to scroll Message pane
    • No keyboard scroll allowed in Contacts & Contact/Group Info
  ◦ Little visibility of actions' visual feedback (bottom left)
    • Were skipped multiple times
  ◦ Status cannot be updated on desktop version
    • Users cannot see others' status

Professors d'IDI - UPC

IDI – Usability Testing

Use case. Depth perception in VR

- Goal:
  - Evaluate performance of shading technique in VR environments
- Context:
  - Perception of complex, volume datasets is difficult in VR
  - Shading techniques may enhance shape and depth perception

Use case. Depth perception in VR

- Purpose of the test:
  - Analyze whether shading techniques influence the perception of shapes and depth in VR
- Methodology:
  - Provide images under different shading conditions
  - Ask the users to classify two points of the scene placed at different depths
  - Analyze the results obtained

Use case. Depth perception in VR
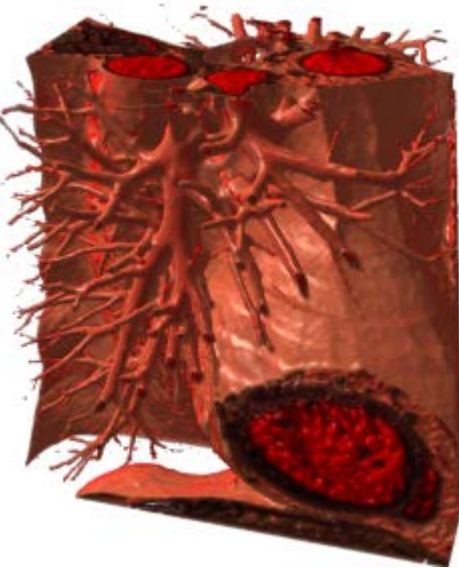
- Sample images:

Use case. Depth perception in VR

▸ Test preparation:
- Select shading models (4)
- Select models (likely unknown to users)
- Determine number of participants, iterations
  - Low level perception problem -> should be > 10
  - Latin squares balance results -> 16 per experiment
- Two tasks
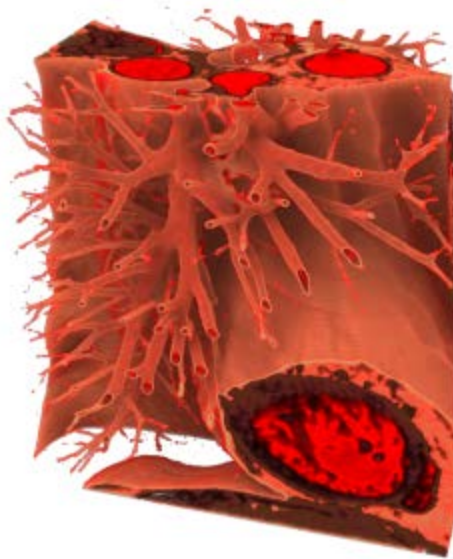
Use case. Depth perception in VR

▸ Shading techniques:

Phong

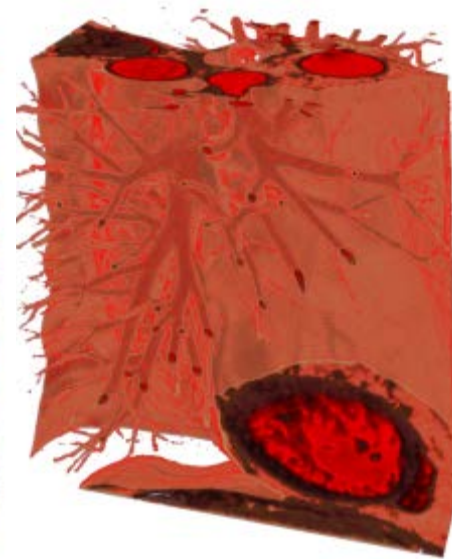Soft shadows

Hard shadows

No shading

Use case. Depth perception in VR

- Images selection:
  - Select models likely unknown
    - Avoid previous knowledge
  - Random shading sorting
    - Avoid learning (shading)
  - Random model sorting
    - Avoid learning (model)
  - Latin squares
    - Avoid fatigue and learning (within users)

Use case. Depth perception in VR

▸ Task: Select the closer point. 2-alternative forced choice (2AFC)
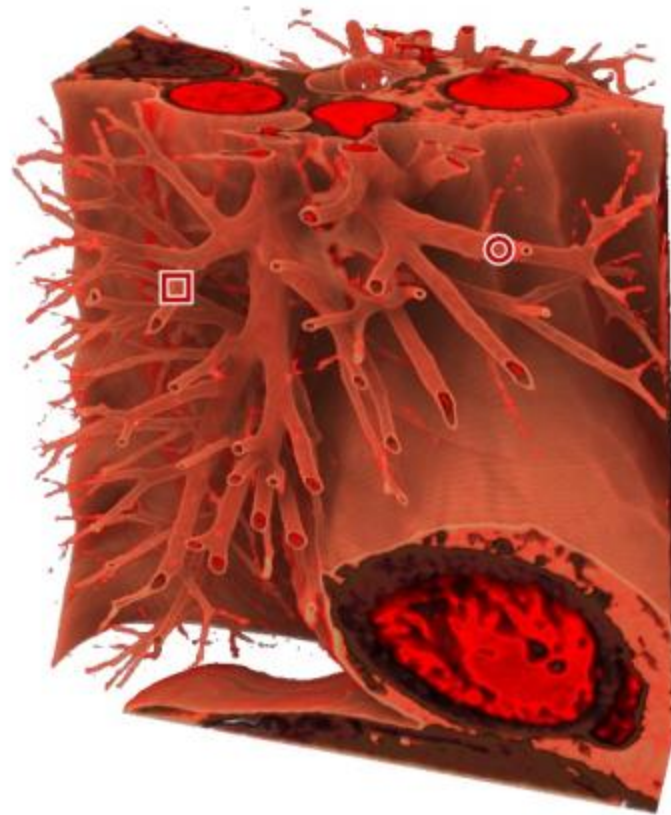
Use case. Depth perception in VR

▶ Measures (what we measure in the test):
  ◦ Time to answer
  ◦ Correctness

Use case. Depth perception in VR

- Variables to include in the analysis (to discard confounding or correlating variables)
  - Shading technique
  - Depth values
    - May analyze if absolute difference correlates with correctness
  - Previous VR background
  - Information of images for left and right eye
  - Luminance of the points' environment
  - Correlation between depth and shading maps

Use case. Depth perception in VR

- Experiment setup:
  - 3D TV
  - Users placed at fixed distance
    - Chair to reduce movements
      - Avoid parallax as confounding variable
  - Dark room
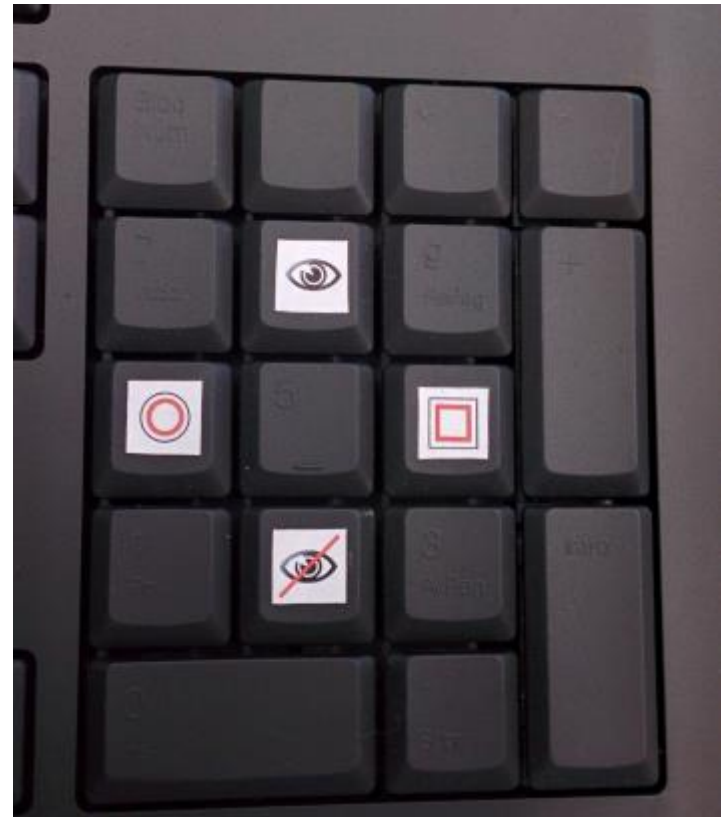 - External light (for virtual light source consistency analysis)

Use case. Depth perception in VR

- Experiment setup:

Use case. Depth perception in VR

▸ Experiment setup:
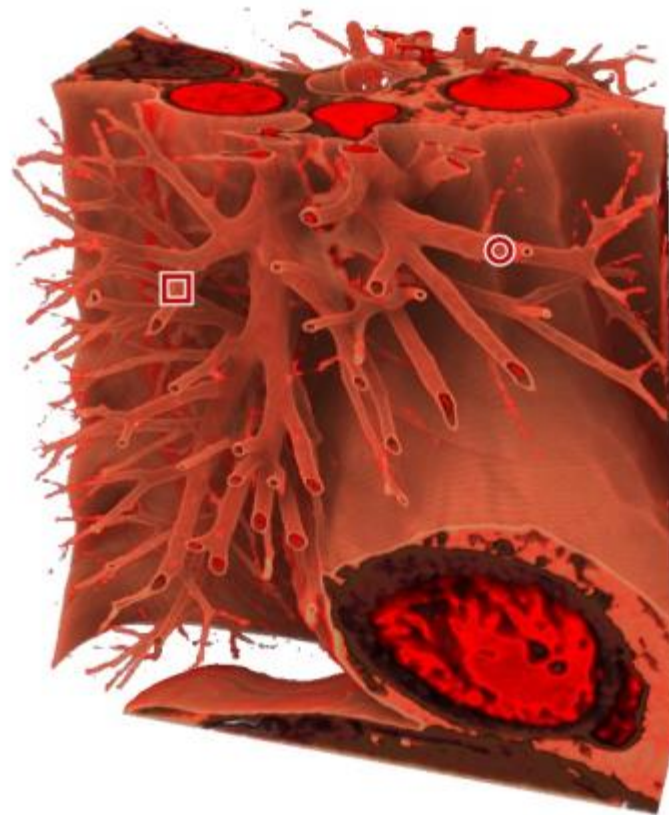 ◦ Modified keyboard to facilitate entry
    • Will compute timings

Use case. Depth perception in VR

- Experiment setup:
  - Initial questionnaire (background, VR exposition…)
  - Initial training
  - Tasks
    - May rest between tasks
  - Post questionnaires
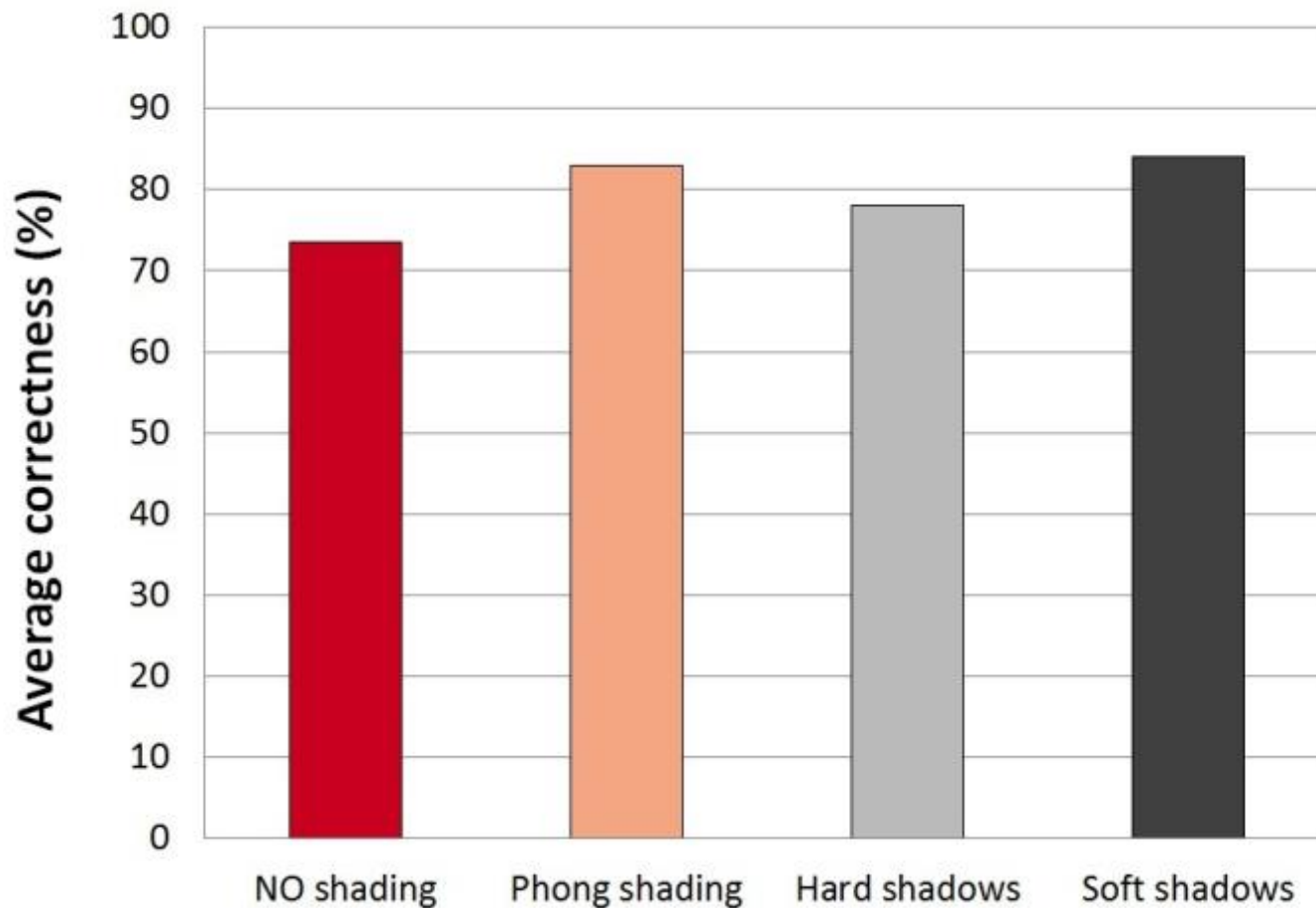
Use case. Depth perception in VR

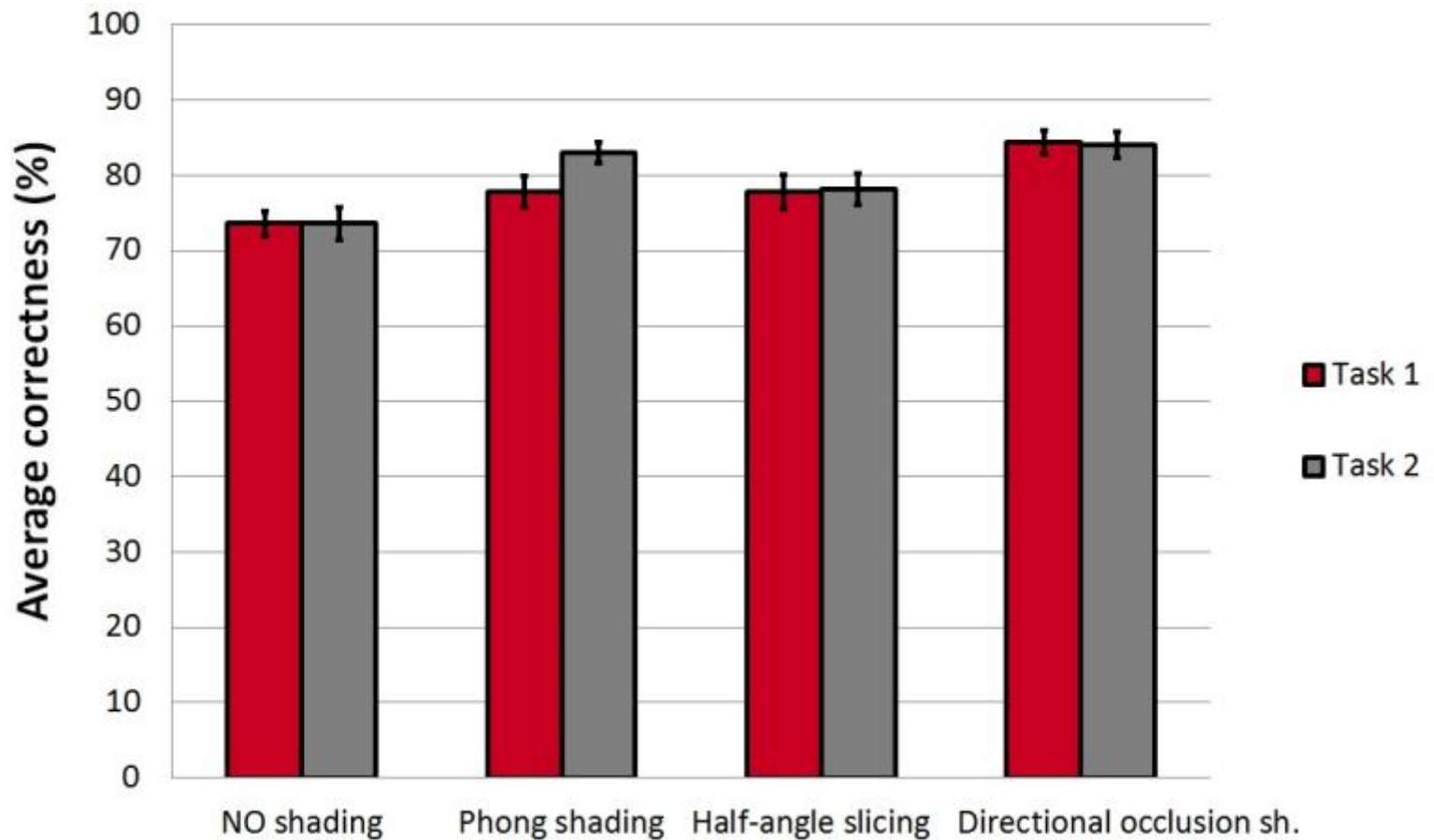- Task: Select the closer point. 2-alternative forced choice (2AFC)

Use case. Depth perception in VR

▸ Analysis of results

Use case. Depth perception in VR

▸ Analysis of results

Use case. Depth perception in VR

▸ Statistical analysis:
  ◦ **ANOVA test:** One-way analysis of variance to reject the null hypothesis that all correctness means are equal between shading techniques.
  ◦ For a significance level of $\alpha = 0.05$, a Bonferroni post-hoc test with the same acceptance level to reveal differences between the individual shading techniques
  ◦ **Result:** reject the null hypothesis when $p < 0.05$

Use case. Depth perception in VR

- Statistical analysis.
  - Chi-square test of association for the categorical variables relative depth and users' answers from tasks 1 and 2

Variables	χ^2	p value	Correct answers for each depth category
T1: relative depth vs. users' answers	5.991	<0.0001	<0.05: 66 % 0.05–0.1: 88 % >0.1: 86 %
T2: relative depth vs. users' answers	5.991	<0.0001	<0.05: 63 % 0.05–0.1: 86 % >0.1: 87 %

Use case. Depth perception in VR

▸ Statistical analysis:

◦ The ANOVA analysis ($\alpha = 0.05$, $p < 0.0001$) of the NMI values shows that there is a significant difference between the images shaded with DOS with respect to the images shaded using HA or PH. A further Bonferroni's test revealed that DOS provides a significantly higher NMI (average NMI = 3.327) than HA (average NMI = 1.84) and PH (average NMI = 1.88).

  • Instead, there is no significant difference between the NMI means of HA and PH.

Use case. Depth perception in VR

- Guidelines and recommendations
  - Using advanced volumetric shading improves depth perception
    - Among the tested shading models the simulation of soft shadows by using directional occlusion shading for desktop-based VR seem to yield better results

Use case. Depth perception in VR

▶ Guidelines and recommendations

◦ Real illumination does not affect depth perception when using advanced volume illumination techniques

◦ External lighting may be carefully controlled to provide a pleasant environment

• Specular highlights on the screen, reflections, or over-illuminated areas will certainly affect the correct perception of the data
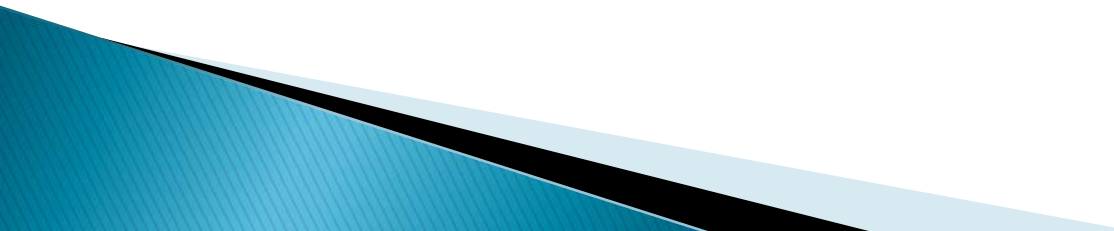
Use case. Depth perception in VR

- Guidelines and recommendations
  - When trying to judge depth in volume models, the X/Y relative position of the markers or the luminance of the points to classify seems to have no importance

Usability. Test Planning: Measures

- For problem discovery:
  - Focus on prioritizing problems
    - Include frequency of occurrence
    - Likelihood of occurrence in normal usage
    - Magnitude of impact on the participants
  - Pre-planned number of iterations
  - Number of participants small, but multiple iterations,…

Usability. Test Planning: Measures

- For measurement tests:
  - Categories
    - Goal achievement indicators (success rate and accuracy)
    - Work rate indicators (speed and efficiency)
    - Operability indicators (error rate and function usage)
    - Knowledge acquisition indicators (learnability and learning rate)

Usability. Test Planning: Measures

▶ For measurement tests:
- Fundamental global Measures
  - Successful task completion rates
  - Mean task completion times
  - Mean participant satisfaction ratings (on a task-by-task basis)
    - There are standardized questionnaires for this
  - Other measurements could be:
    - Number of tasks completed within a specified time limit, number of wrong menu choices, number of user errors, number of repeated errors (same user)

Usability. Test Planning

- After measurements choice, <u>usability objective</u> can be determined
  - It's usually better to set goals that make reference to an average (mean) than to a percentile
    - Sample means drawn from a continuous distribution are less variable than sample medians
    - Unless there is missing data due to participants failing to complete tasks
  - Percentile goals require large sample sizes
    - You can't measure accurately at the 95 percentile unless there are at least twenty measurements