

CSCI 310

Experimental Design

Sarah Carruthers
Fall 2014

Look at A3 (retrospective)

- Common problems:
 - focus on backend rather than user experience
 - failure to link to principles
 - in some cases, clear that the concept is understood, but terminology missing
 - no issues identified
 - steps not adequately broken down into actions
 - following instructions!

Some stuff

- Talk by Johnny Blakeborough
 - Gender and Gaming
 - October 28 6-7:30, bldg 355, theatre 203
- You may know that gender issues in the computer game industry are currently a hot topic. Come and learn more about the work (and play) impacts of gender in the computer game industry.
- This event is a guest lecture to a gender studies class on the topic of gender and gaming, and the lecture is being opened to Computing Science and Digital Media students.
- Johnny Blakeborough is a computer technician at VIU, and one of the original Desert Bus team – Desert Bus is a computer game that does not offend anyone, and has raised \$523,520 for charity to date. Johnny is also Director of the Vancouver Island Short Film Festival.

Last class

- Methods for gathering qualitative info from users:
 - Think aloud
 - co-discovery
 - co-operative
 - observation
- Different evaluation approaches are useful for different purposes
- Qualitative data is important: it helps us understand why users act the way they do
- There are many ways of collecting qualitative data, and they all have pros and cons

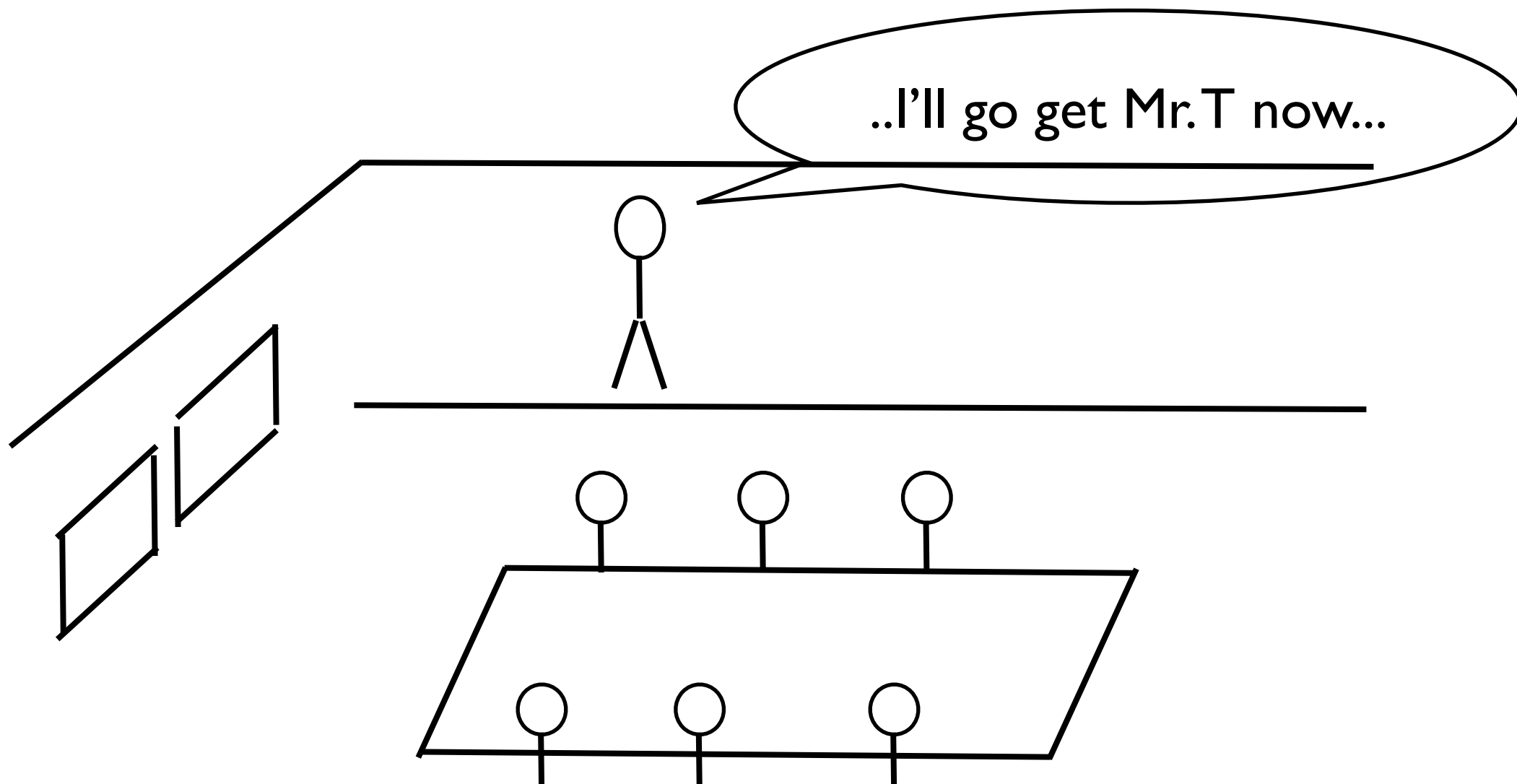
Objectives

- Explain the process of running a user testing, or experiment session
- Describe evaluation scripts and pilot tests and explain why they are important
- Explain the difference between user testing and experiments
- Describe the major building blocks of an HCI experiment
- Identify flaws in experimental designs and suggest better alternatives

User Testing Procedure

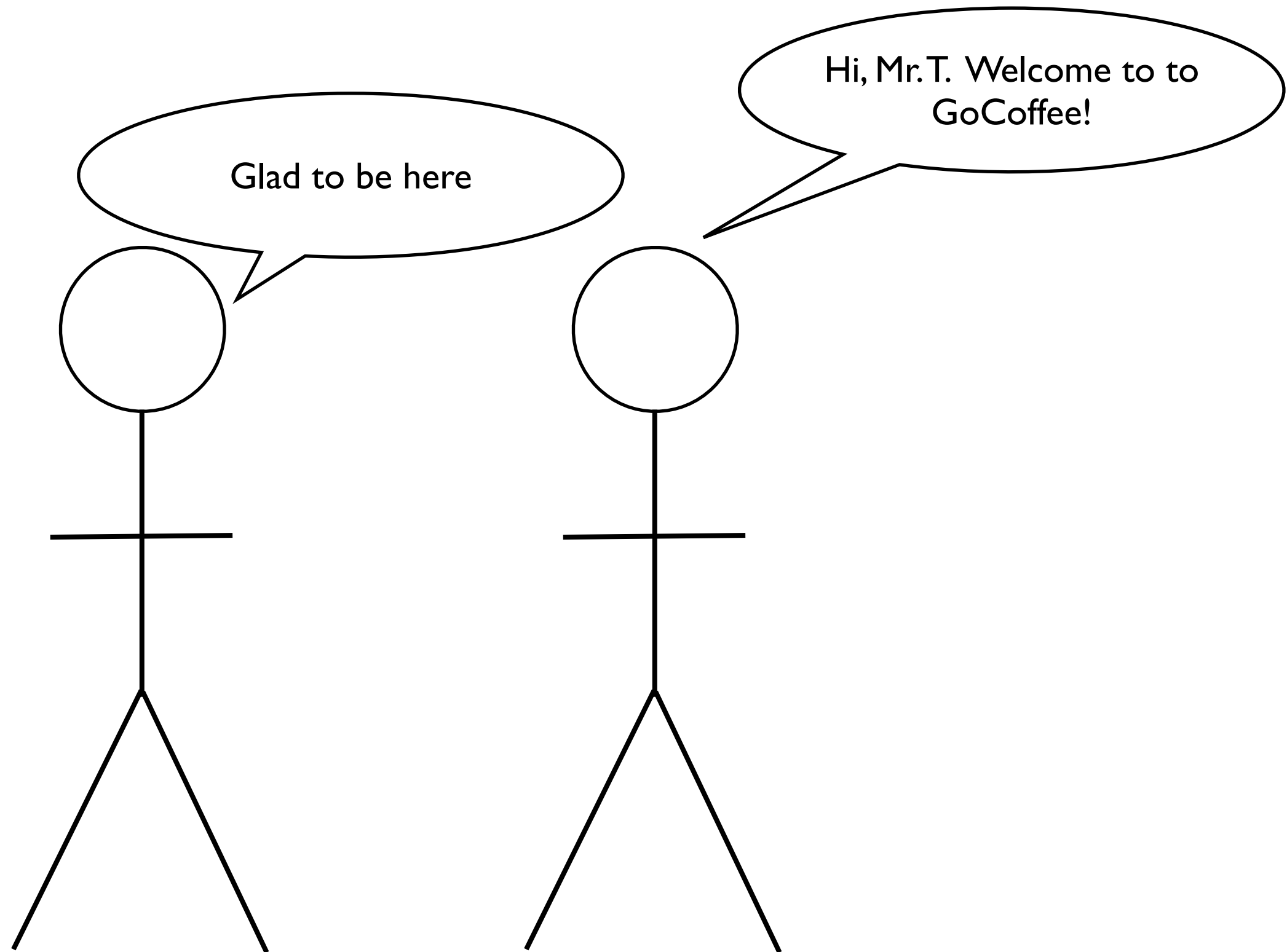
- The user arrives
- Greeting the user
- Brief into to the observation team
- Briefing in the Lab Room
- Test in progress
- Test conclusion
- Dismiss user
- Team records its notes

Usability Test Storyboard



The facilitator informs the team that the user, Mr.T, has arrived

Usability Test Storyboard



Greeting the user

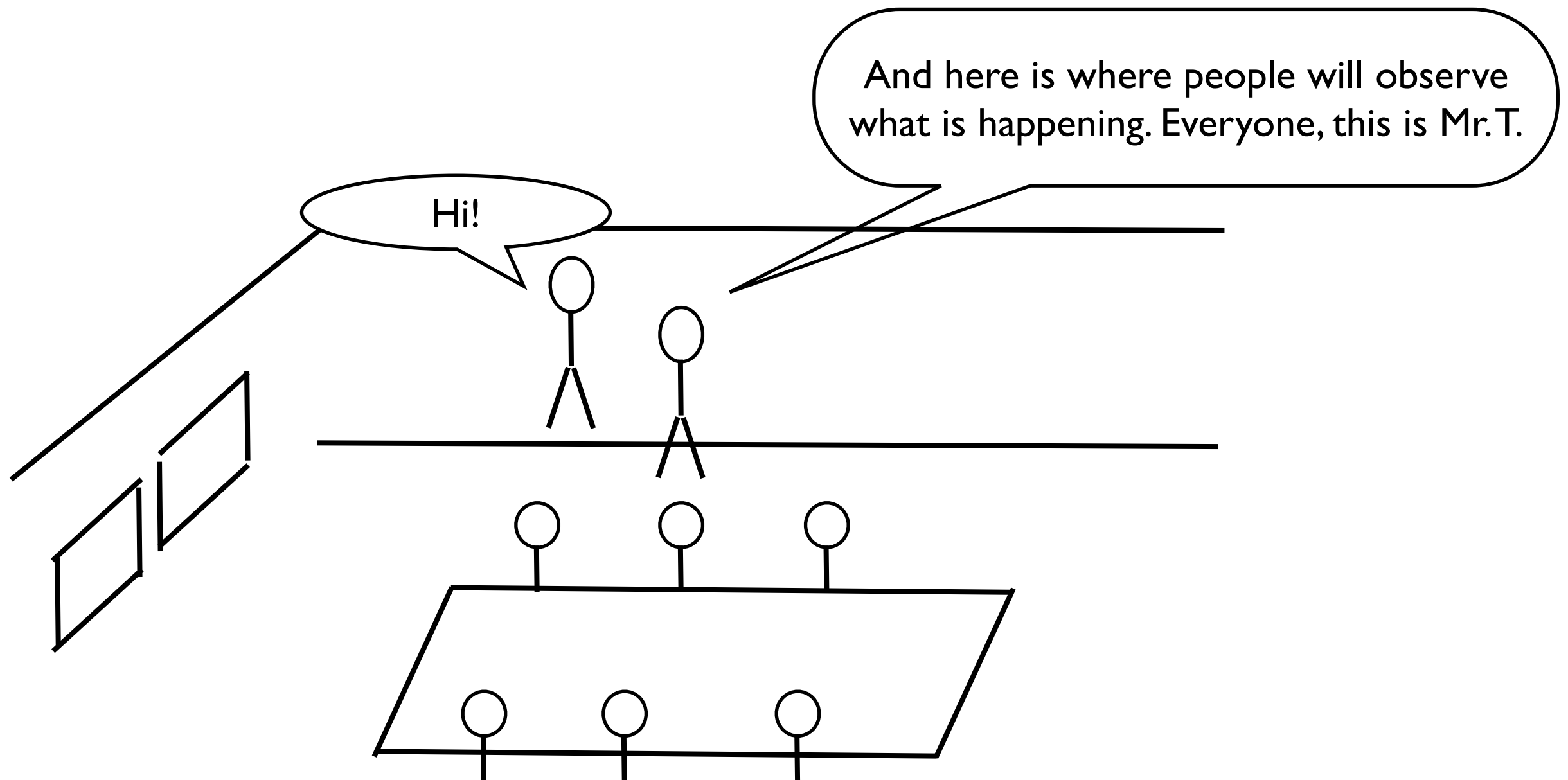
Usability Test Storyboard

10-4, good buddy

Please keep in mind that we are focusing more on that app, than you

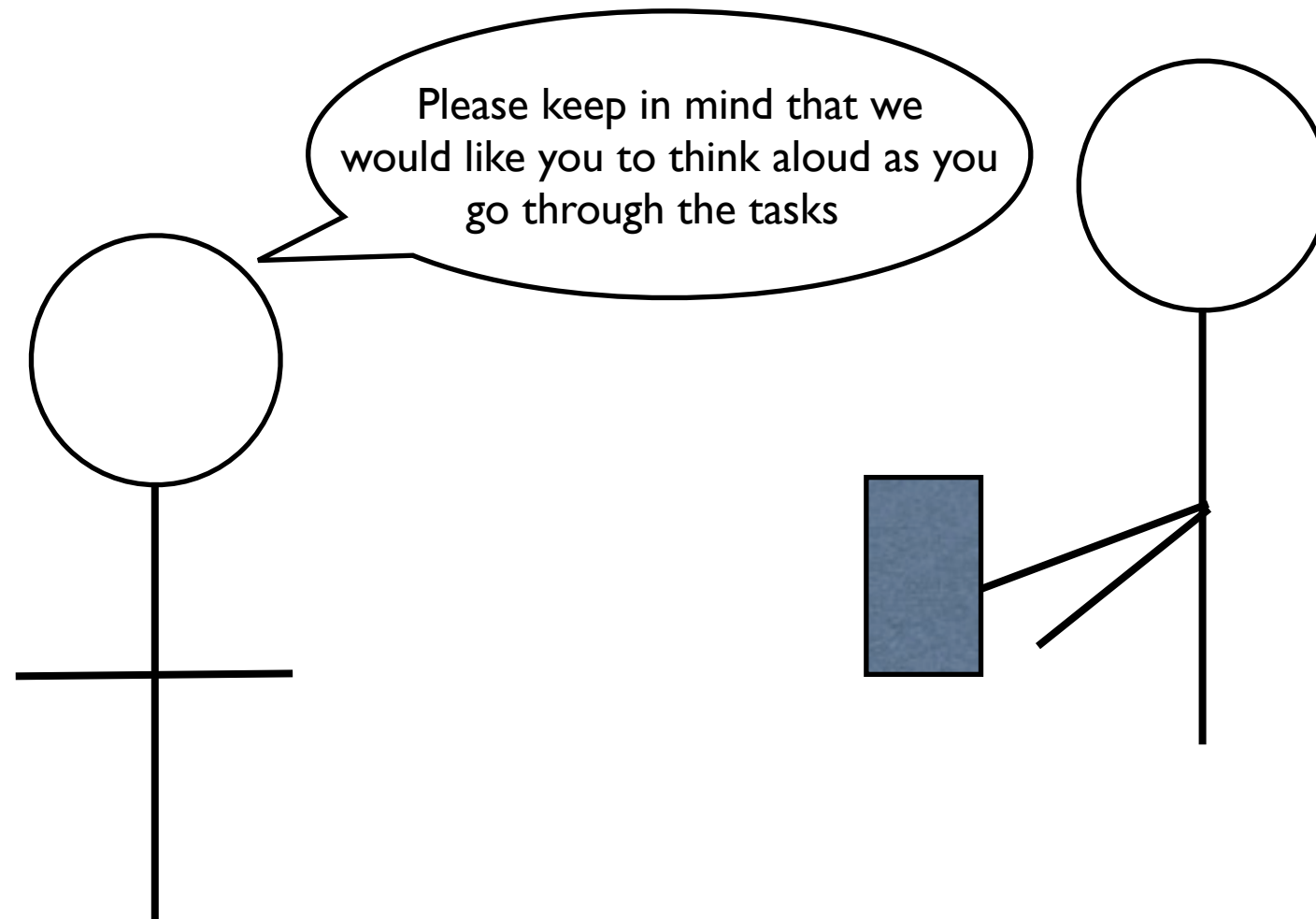
Informal chat: The facilitator helps Mr.T relax. He starts explaining what to expect in the test

Usability Test Storyboard



The facilitator briefly introduces Mr.T to the team, and shows the room in which the team will be observing

Usability Test Storyboard



briefing in the Lab Room

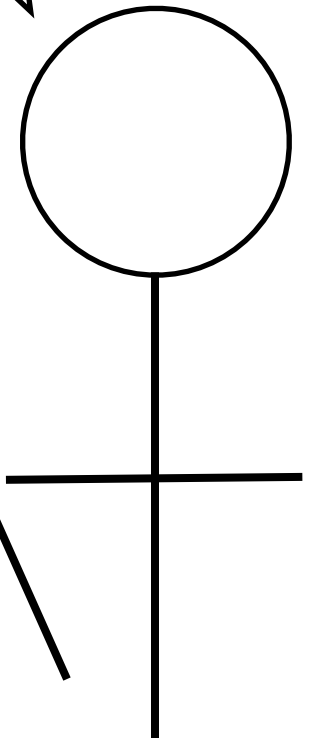
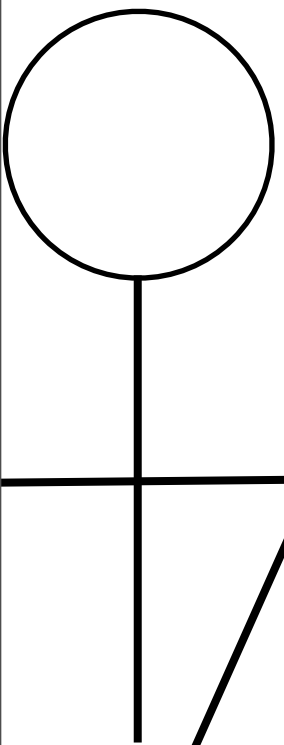
Usability Test Storyboard

Hmm, I'm confused

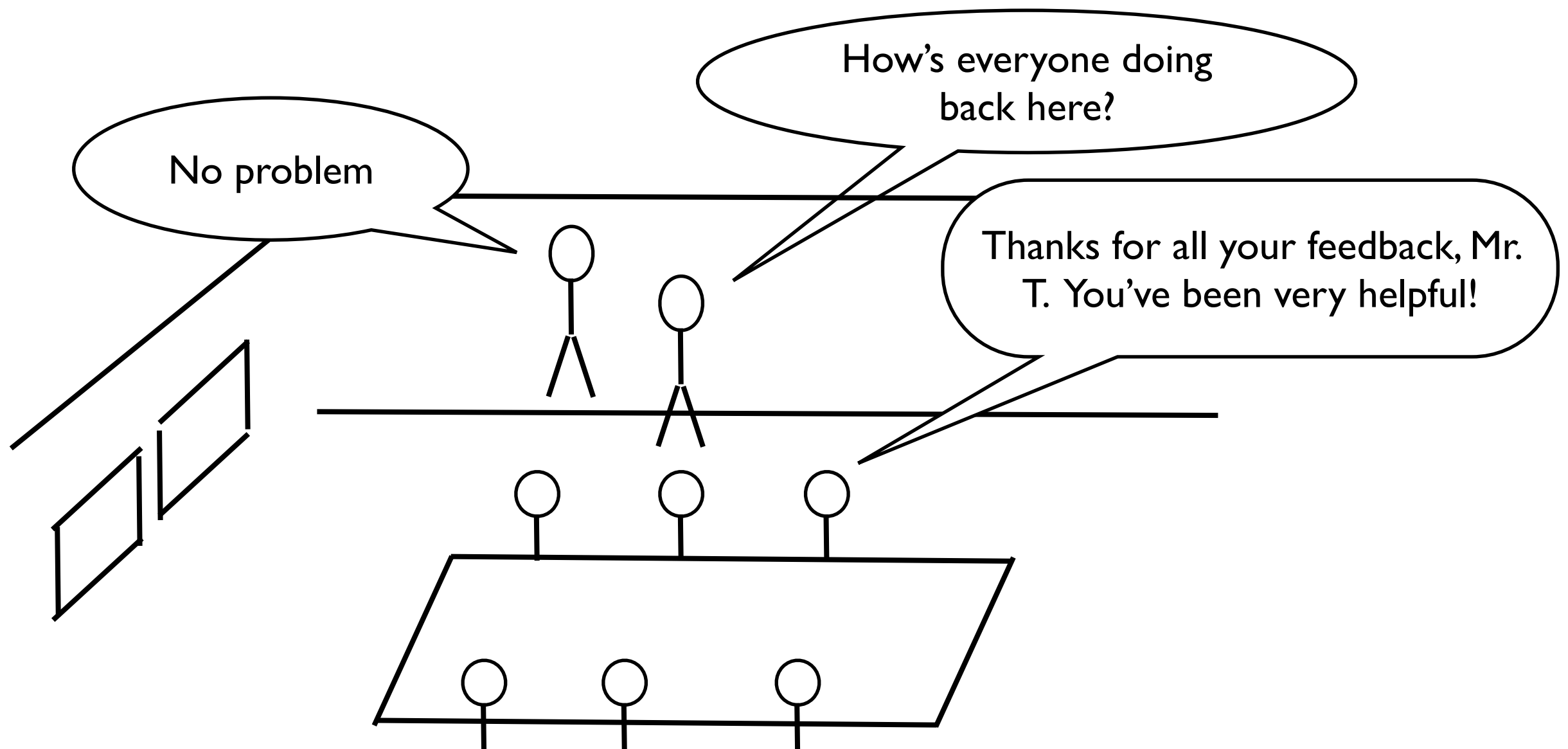
user view

screen view

you know,
Frank, no one has
figured out how to use
our coolness
calculator

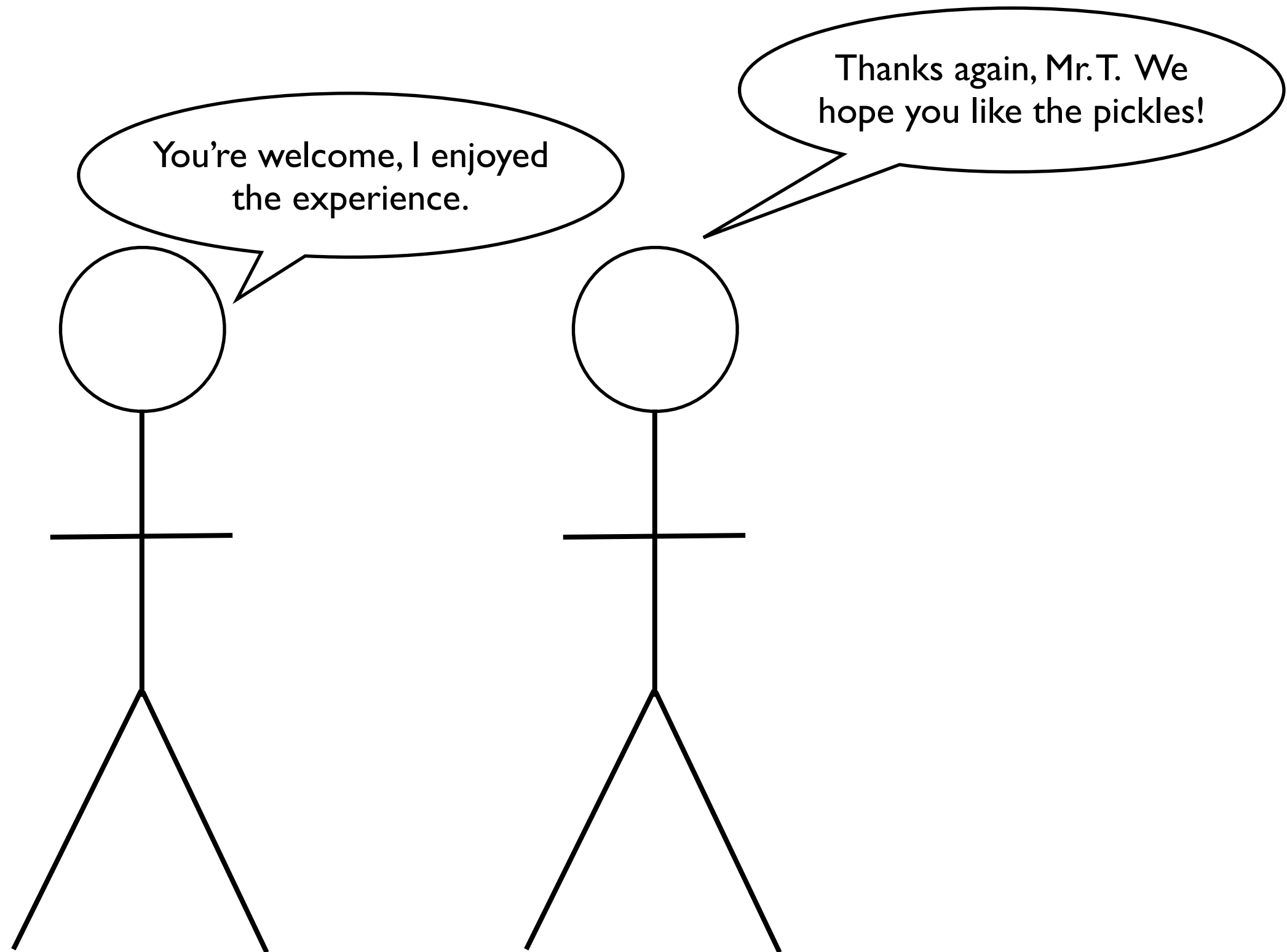


Usability Test Storyboard



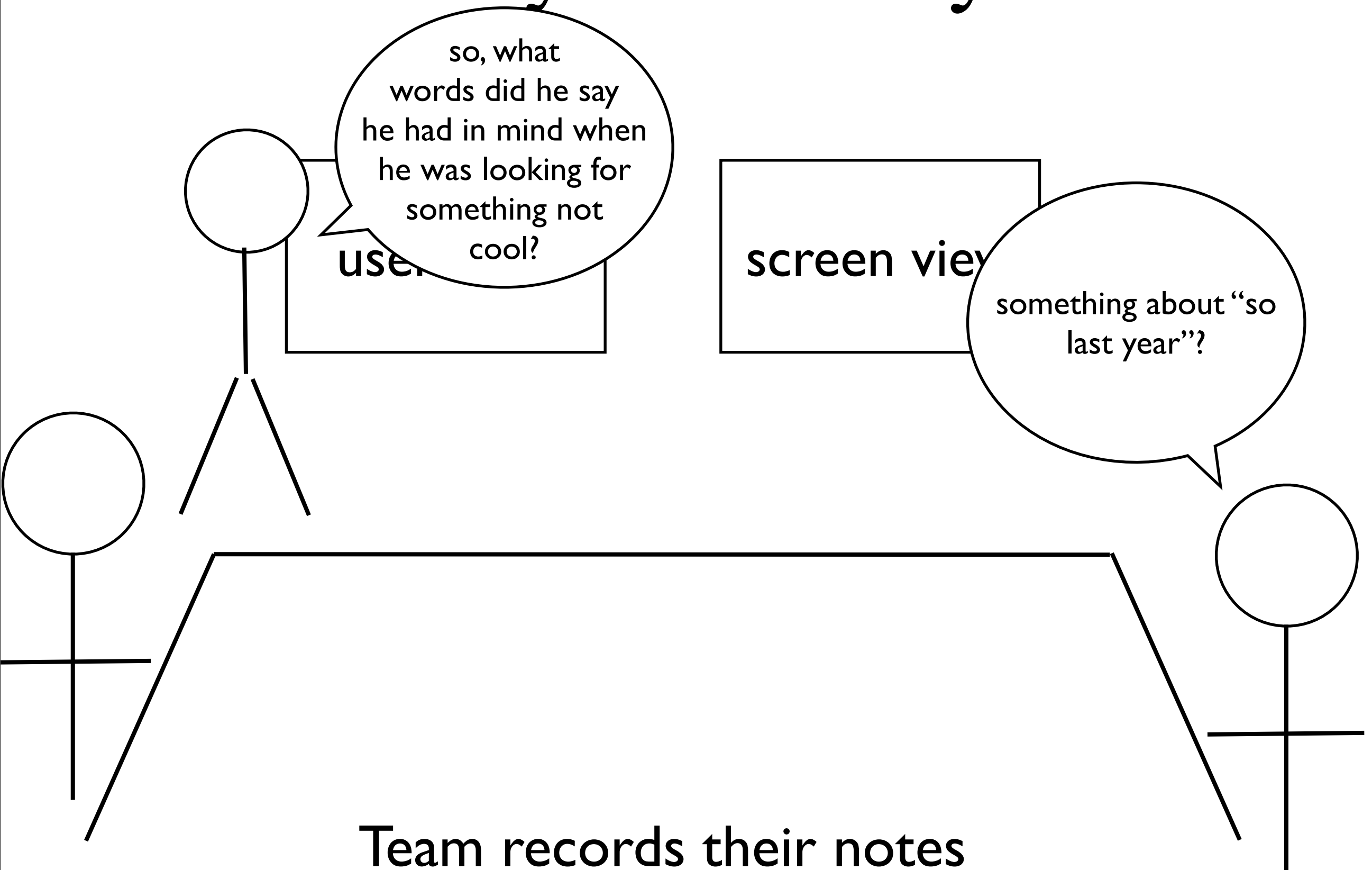
Test concludes

Usability Test Storyboard



Goodbye to the user

Usability Test Storyboard



Welcoming the user

- Why?
 - If they are at ease, they will be:
 - more comfortable sharing honest opinions
 - less stressed during tasks
- How?
 - Thank them for taking the time to help
 - Explain that you are testing the interface, not them

Controlling bias

- Avoid influencing the user
- How else might the experimenter introduce bias?

Now you get to work with circular menus. I think you will really like them... I designed them myself!



Evaluation Script

- Read instructions from a script
 - Ensures all participants receive exactly the same instructions
 - Makes sure you don't forget anything
 - Allows multiple researchers to conduct experiments
- Can be awkward to read
 - But, this can be minimized by explaining to participants *why* you're doing it

Pilot test

- A “practice run” of your usability test
 - Gives you practise
 - Evaluates procedures, tasks, instructions, scripts, questionnaires, timing
 - Gives a quick and easy probe to see if you’re answering the questions you care about
- Can use users that don’t exactly match target users
- Don’t include data in final analysis

Experiments vs. User testing

- Experiments:
 - test hypotheses
 - investigate relationship between two or more variables
- User testing:
 - checks that a system is usable by intended user population for their tasks

User Testing

- Aim: improve products
- few participants
- results inform design
- not perfectly replicable
- controlled conditions
- results reported to developers

Research experiment

- Aim: discover knowledge
- many participants
- results validated statistically
- replicable
- strongly controlled conditions
- results reported to community via scientific paper

When to do an experiment...

- Comparison
- Important consequences of decision
- Prediction uncertain
- For research
 - important to show causality

Exercise

- Experiment, or user testing?
 - Will comparison of images be less accurate if done successively on a single screen as compared to simultaneously on two screens?
 - Can radiologists learn to use a trackball, and do they find it to be a satisfactory input device?
 - Do radiologists prefer linearly or hierarchically organized lists of x-rays?
 - Are the number of consultations between radiology and emergency increased by increased availability of workstations in both sites?

What should we measure?

- How often does a particular event occur?
 - requires measurement of that variable
- Is there a relationship or dependency between one variable and another?
 - requires measurement of both variables
- Will the result mean that one of the variables is actually impacting the other?
 - requires manipulation of the variable
 - need to account for the other variables that may be impacting the situation

Experimental Terminology

- **Hypothesis:**
 - A tentative prediction that is testable
- **Independent variable:**
 - systematically manipulated by the experimenter
- **Dependent variable:**
 - what the user does, what you measure
- **Confounding factor:**
 - varies with the independent variable, making it impossible to attribute causality to the independent variable

Validity

- External validity
 - confidence that your results are generalizable to real situations
 - usually good, in natural settings
- Internal validity
 - confidence in our explanation of experimental results
 - usually good in experimental settings
- Trade-off: Natural vs. Experimental
 - precision and direct control over experimental design
 - VS.
 - desire for maximum generalizability in real life situations

User Centric, Inc. tested the iPhone's touch keyboard to conventional QWERTY and numeric phone keyboards...

The goal: to determine how easy it was for a mobile phone users to text using the iPhone.

Participants Texted using Both Conventional Phones and iPhones

A total of 20 participants were brought in for one-on-one usability sessions with a moderator. All sent text messages at least 15 times per week. Ten of the participants owned a phone with a QWERTY keypad, and ten of the participants owned a phone with a numeric keypad. Those who owned a numeric keypad used the "multi-tap" method of entering text messages rather than predictive text. To multi-tap, a user must press a particular key on the numeric keypad multiple times to get the desired character to appear.

During each session, participants were required to use their own phones. In addition, they were provided with an iPhone for use during the study. None of the participants were iPhone owners, and all participants were compensated for their time.

Standardized Messages Were Provided for Texting

Every participant entered six messages using their own phone and six messages on the iPhone. In their sessions, participants were asked to copy 12 standard messages created for this study. Each message was between 104-106 characters in length (including spaces). Six of the messages each contained 8-10 instances of proper capitalization and punctuation, while the remaining six messages contained no capitalization or punctuation but had some abbreviations. Message configuration and phone order were counterbalanced across participants.

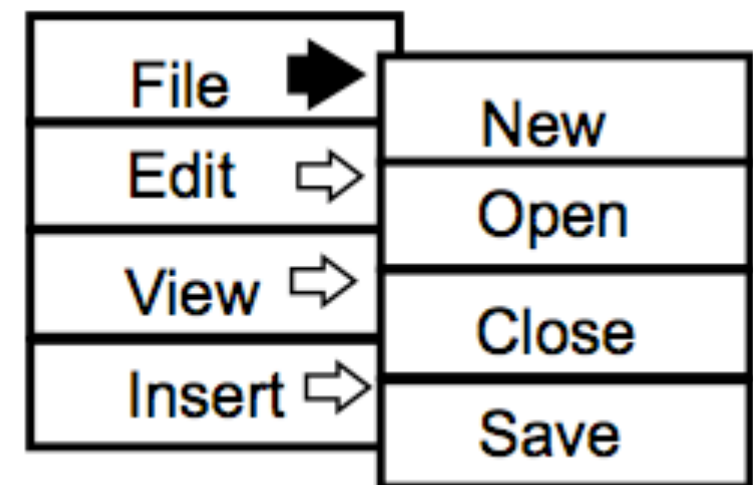
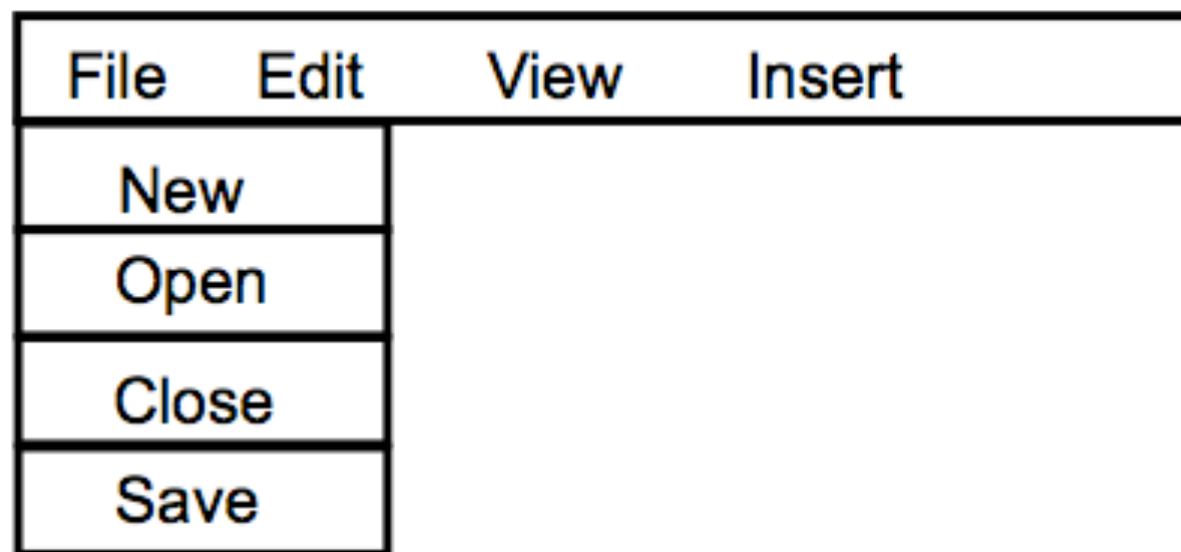
Texting on iPhone Took Twice as Long as Texting on QWERTY Phones

In general, participants took longer to enter text messages on the iPhone than on their own phone. Despite the keyboard similarities, QWERTY phone users took nearly twice as long to enter comparable messages on the iPhone compared to their own phone. On the other hand, multitappers did not experience a significant difference in the time it took them to type messages on the iPhone. (Multitap text entry usually takes about 2 times as long as QWERTY text entry). Participants also made more typing errors on the iPhone. This phenomenon was expected since users had much more experience with their own phones.

<http://www.usercentric.com/news/2007/08/15/qwerty-texters-demonstrated-drop-efficiency-when-texting-iphone>

Exercise: Hypothesis

- Improve this hypothesis:
 - Popup menus are fast and accurate to use



Possible solution

- User performance is faster and more accurate when selecting a single item from a pop up vs. a pull down menu of 4 items, regardless of the subject's previous expertise in using a mouse or using the different menu types

Exercise

- Name the *independent* and *dependent* variables:
 - Is reading speed faster from a high resolution or low resolution screen?
 - Will radiologists prefer a mouse, lightpen, or trackball as a pointing device?
 - Can air traffic controllers enter data faster with a QWERTY or ABC keyboard?

Exercise: Confounds

- Name the intended independent variable and the confounding variable. Suggest a better design.
- In a study to compare legibility of 2 visual displays, one display is located near a window, and the other is in a windowless room
- To determine whether computer training leads to more positive opinions of computers, attitudes of students in a 1st year computer course are compared to attitudes of 1st year students who are not taking a computing course

Exercise: Confounds

- Volunteers compared 2 command selection techniques: function keys and mouse. All volunteers used the function key approach followed by the mouse approach.
- An experiment compares three colour scales: a blue-yellow scale, a greyscale, and a rainbow scale. All participants complete the task first with the blue-yellow scale, then with the greyscale, and then with the rainbow scale.

Assigning participants to conditions

- Between subjects:
 - each participant sees only one condition
- Within subjects:
 - each participant sees all conditions
 - * need to counter-balance order of conditions
- Matched:
 - Between-subjects design where the groups are matched on some factor
 - e.g. IQ, or computer experience

Pros/cons

	Advantages	Disadvantages
Between-subjects	<ul style="list-style-type: none">• no order effects	<ul style="list-style-type: none">• need more participants• individual differences
Within-subjects	<ul style="list-style-type: none">• Few participants• No group differences	<ul style="list-style-type: none">• Order effects
Matched	<ul style="list-style-type: none">• Individual differences less than between-subjects	<ul style="list-style-type: none">• Matching may be imperfect• may not know best matching factor

Key points

- Run pilot tests
- Set participants at ease
- Use experiments when you need to show causality, otherwise user testing may suffice
- Experiments must be defined very carefully to avoid confounds