

## Lecture 25

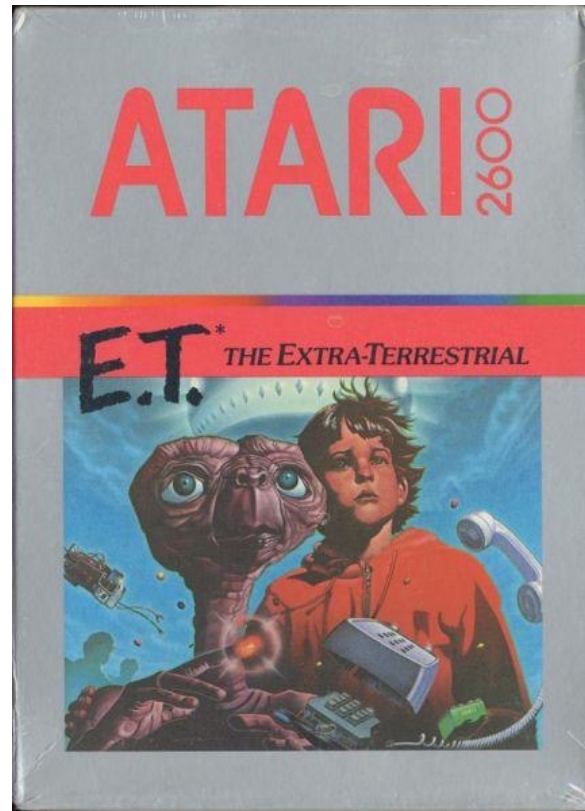
# Playtesting

# Why Player Testing?

---

A good game no one can learn to play...

...is a bad game



# Take Away for the Lecture

---

- Questions to start with
- Different metrics of usability
- Conducting a user study
- Data collection/analysis

# Questions to Start With

---

- Why are you conducting the test?
- What are you going to **learn**?
  - Sources of player difficulty?
  - Typical player strategies?
- How will you use the **results**?
  - Sometimes, to persuade or justify
  - Often, to iterate the design

# Brainstorming Exercise

---

- What matters in a game?
  
  
  
  
  
  
  
  
  
  
- And how do you measure it?

# Accessibility

---

- Players with disabilities is a neglected market
  - Many players have some colorblindness
  - Many deaf people are gamers
  - Blind gamers are not unheard of
- **Example:** *NanoEmpire*
  - Text based game made by James Senter at Cornell
  - 30k plays on Kongregate, but iOS version a flop
  - ... until blind community found they could play it

# Colorblindness Fail: The Witness



# Colorblindness Win: *Witcher 3*

## Colorblind Mode Off



Highlights are traditional hues

## Colorblind Mode On



Highlights are a much brighter palette



# Quantitative Metrics

---

- Time to learn to use a game verb
- Time (ability) to complete a specific task/quest
- Usage (or lack of usage) of gameplay features
- Errors (how many, where)
- Player satisfaction (Likert scale)
- **Problem:** need many users for good stats

# Qualitative Metrics

---

- What does the user say?
- Where/how do they run into trouble?
- What's the first reaction/impression?
- How would they describe the gameplay?
- Would they play it again? Recommend it?
- **Advantage:** More amenable to small groups

# The Key Idea

---

- Put the challenge **where you want it**
  - Some things are meant to be difficult
  - If not explicitly a challenge, should be easy
- **Example:** Fast Travel
  - Allow the user to explore a vast world
  - But keep constant travel from being boring
- Even bigger problem in educational games
  - Are they encountering the “right” challenges?

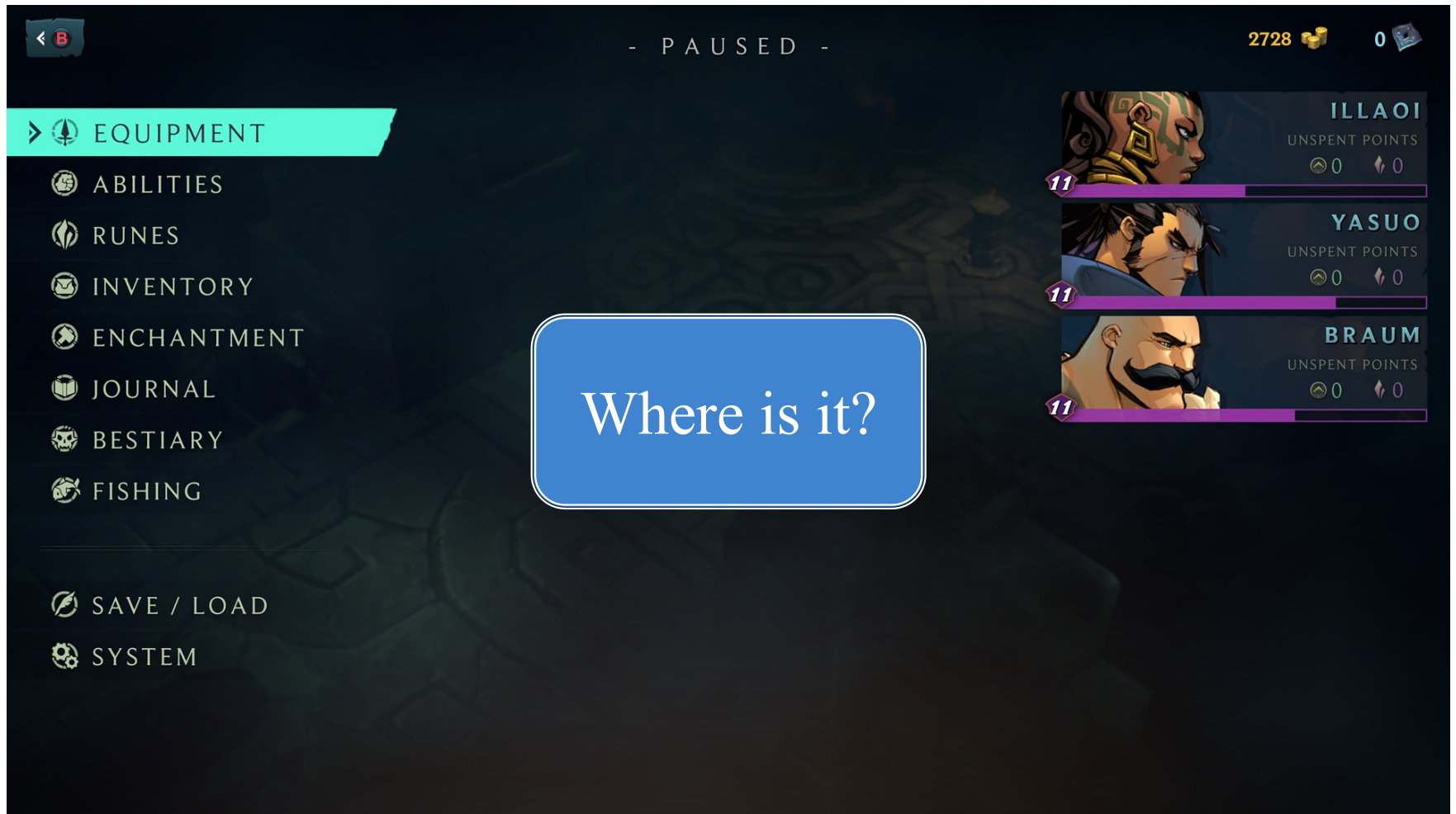
# Save Game Fail: *Ruined King*



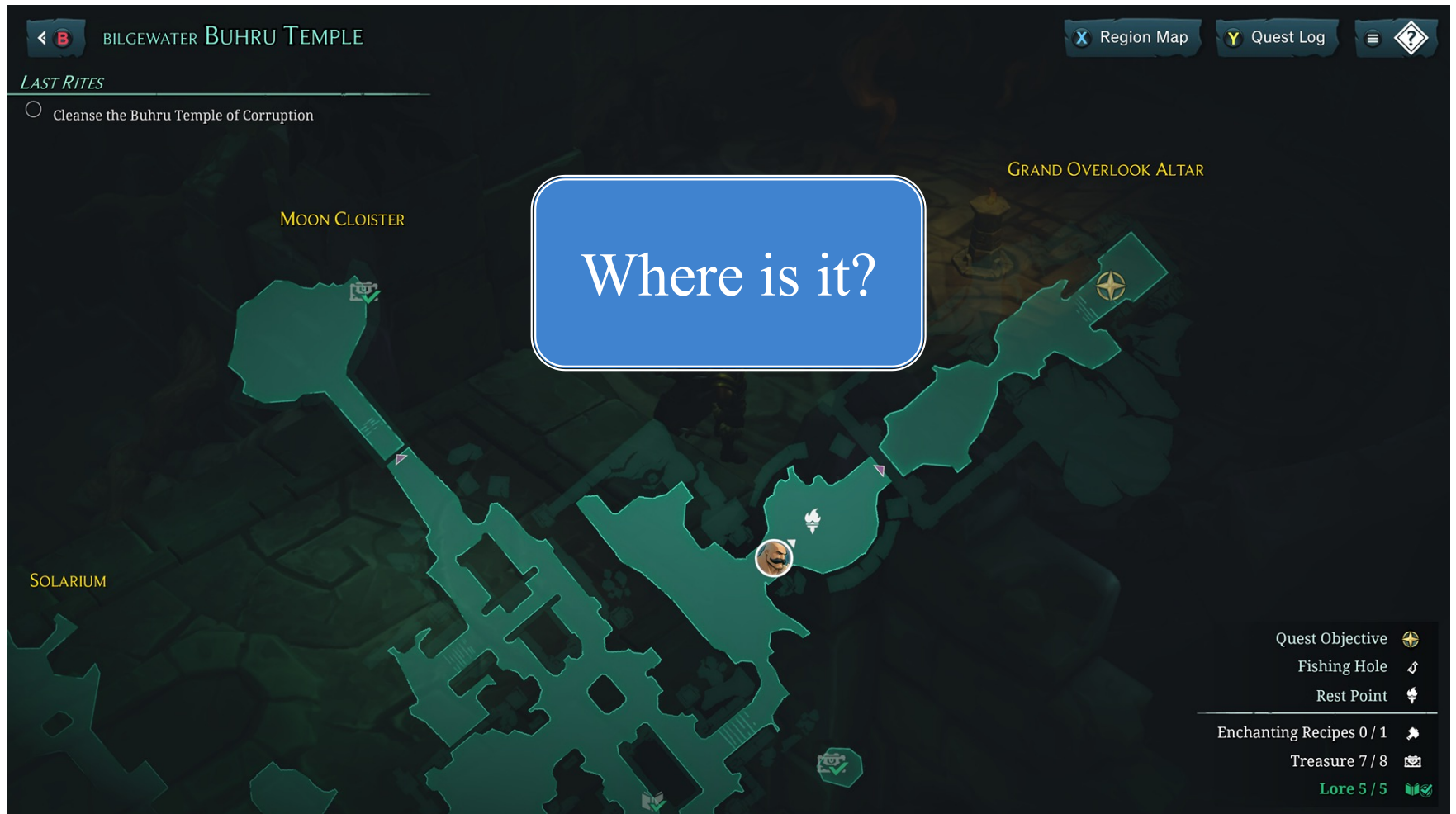
# Save Game Fail: *Ruined King*



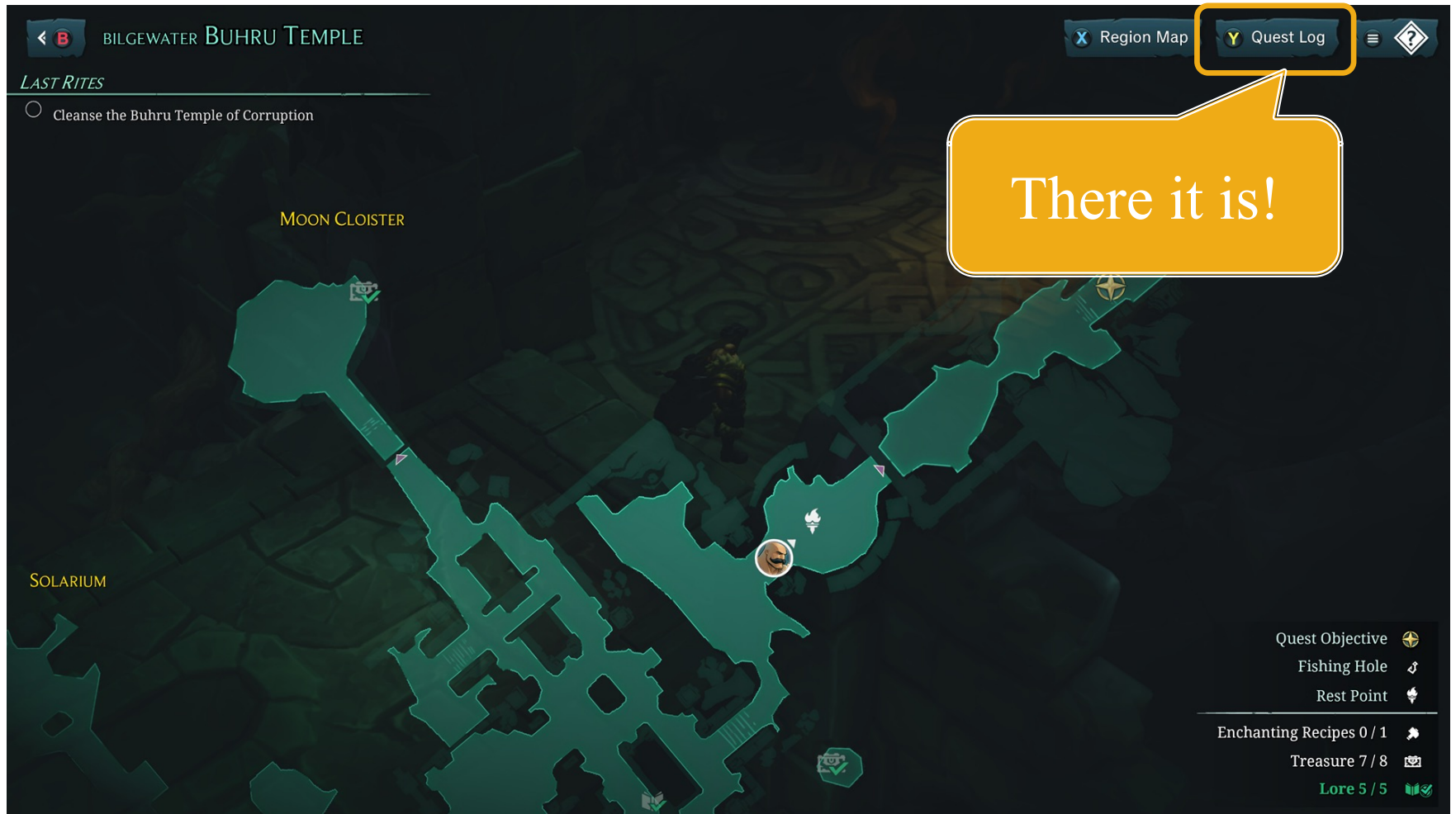
# Quest Log Fail: *Ruined King*



# Quest Log Fail: *Ruined King*



# Quest Log Fail: *Ruined King*





# The User Study

---

- **Participants**

- Who are you studying?

- **Artifact(s)**

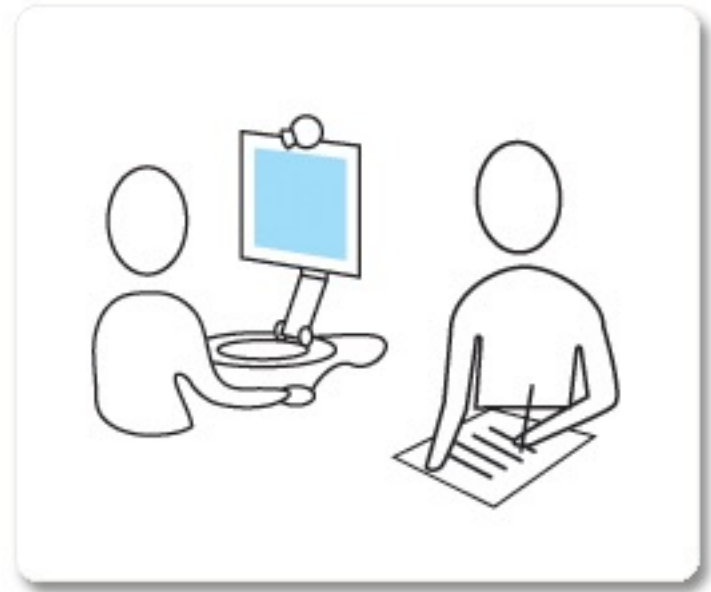
- What are you studying?

- **Tasks and scripts**

- How do you plan to study?

- **Experimenter roles & best practices**

- How are *you* involved in the study?



# Participants: How Many?

---

- People's time is valuable
  - Theirs: how often do they want to play?
  - Yours: you have to administer the test
- Tests yield **diminishing returns**
  - Especially at prototyping stage
  - Rule of 5 (or 3) for qualitative results

# Participants: Who?

---

- Not you (usually)
  - You have privileged knowledge
  - But okay in early stages
- Representative of target audience
  - *Actimates Barney vs. Grand Theft Auto*
  - Back up your concept document claims
- Someone(s) old, someone(s) new

# Ethics and Benevolence

- Remember...
- Your participants are real people
- They are doing you a **favor**



Bug Bash by Hans Bjordahl

<http://www.bugbash.net/>

# Institutional Review Board

---



Cornell University  
Office of Research Integrity and Assurance

# Stanford Prison Experiment

---



# IRB Approval



Cornell University  
Office of  
Research Integrity and Assurance

East Hill Office Building, Suite 320  
395 Pine Tree Road  
Ithaca, NY 14850  
p. 607-254-5162  
f. 607-255-0758  
www.irb.cornell.edu

## Institutional Review Board for Human Participants

### Notice of Course Activity Approval

**To:** Erik Andersen  
**Date:** September 05, 2014  
**Protocol ID#:** 1408004901  
**Protocol Title:** CS-4154

The above referenced Course Activity Project was reviewed by Cornell's Human Research Protection Program (HRPP) and approved for the inclusion of human participants in class assignments. This approval does NOT cover students doing research for theses, dissertations, journal articles, public presentations, or other means of disseminating generalizable knowledge gained from these assignments. Such projects require the individual student to complete his or her own Initial Approval Request form before beginning recruitment and data collection.

You or your students must ensure that the welfare of the research participants is protected and that methods used and information provided to gain subject consent are appropriate to the activity. You and your students should familiarize yourself with and conduct the research in accordance with the ethical standards of the Belmont Report: <http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html>

Please give a photocopy of this approval notice to each student in your class who will be conducting a human participant research project. Acceptance of these terms by students constitutes an understanding that data collection (and allusions to conclusions drawn from these data) from the project(s) covered solely by this approval may never be used for theses, dissertations, articles, or public presentations.

If you have any questions, please contact the IRB office at [irbhp@cornell.edu](mailto:irbhp@cornell.edu) or 254-5162.

# IRB Approval



Cornell University  
Office of  
Research Integrity and Assurance

East Hill Office Building, Suite 320  
395 Pine Tree Road  
Ithaca, NY 14850  
p. 607-254-5162  
f. 607-255-0758  
www.irb.cornell.edu

## Institutional Review Board for Human Participants

### Notice of Course Activity Approval

**To:** Erik Andersen  
**Date:** September 05, 2014  
**Protocol ID#:** 1408004901  
**Protocol Title:** CS-4154

The above referenced Course Activity Project was reviewed by Cornell's Human Research Protection Program (HRPP) and approved for the inclusion of human participants in class assignments. This approval does NOT cover students doing research for theses, dissertations, journal articles, public presentations, or other means of disseminating generalizable knowledge gained from these assignments. Such projects require the individual student to complete his or her own Initial Approval Request form before beginning recruitment and data collection.

You or your students must ensure that the welfare of the research participants is protected and that methods used and information provided to gain subject consent are appropriate to the activity. You and your students should familiarize yourself with and conduct the research in accordance with the ethical standards of the Belmont Report: <http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html>

Please give a photocopy of this approval notice to each student in your class who will be conducting a human participant research project. Acceptance of these terms by students constitutes an understanding that data collection (and allusions to conclusions drawn from these data) from the project(s) covered solely by this approval may never be used for theses, dissertations, articles, or public presentations.

If you have any questions, please contact the IRB office at [irbhp@cornell.edu](mailto:irbhp@cornell.edu) or 254-5162.



# Artifacts: What is Tested?

---

- Working with incomplete product/prototype
  - Some stuff is clearly not finished
  - Do not want comments on unfinished bits
  - Be very clear of the scope of your test
- The Mechanical Turk
  - Can hide unfinished details with hacks or tricks
  - Great for AI in games (replace AI with human)
- Test should focus on **finished** details

# Tasks & Scripts: Direction

---

- “Climb up to the blue square”
- “The game has put information in your codex. You may want to read it before continuing”
- “What do you think should happen if you go here, touch this, hit that?”
  - Example of **pre-interactive** direction
- “Just try things out, explore”
  - Sometimes **no direction** is direction

# Tasks & Scripts: Design

---

- Directedness of tasks depends on **goal**
  - Do *you* know what you are looking for?
  - Are some goals more important than others?
- How long should tasks be?
  - When should you mercy-rule them?
- How long is the test/how many tasks?
  - User attention wanes over time
  - Do you want to order by priority?

# Some Great Resources

---

- <http://www.usability.gov>
  - Standard government usability guidelines
- <http://www.irb.cornell.edu>
  - Ethical guidelines for usability testing
  - Covers all “human experiments” at Cornell
  - Professors need approval before research

# Roles & Best Practices

---

- At least two testers
  - **Experimenter**: run the show
  - **Observer**(s): record what happens
- Be unobtrusive as possible
  - Will you be there when they play?
  - Your input will bias participants
  - But do not frustrate the user



# The Study: Data Collection

---

- Notes
  - What did they say, what did they do?
- Video & Audio
  - To capture what you might have missed
  - Audio okay as long as the player thinks aloud
- Game state logs
  - Log the state of the game to a file
  - Gets even more information than screenshot

# Think Aloud Method

---

- While you shut up, they should talk
  - About **everything** going on in their head
  - Gives you “inside the head” data
- Complements and explains observations
  - Separate player **failures** from **frustration**
  - Think of games like *Super Meat Boy*
- A little unnatural, may need reminders




# Think-alouds

---



# Think-alouds

---

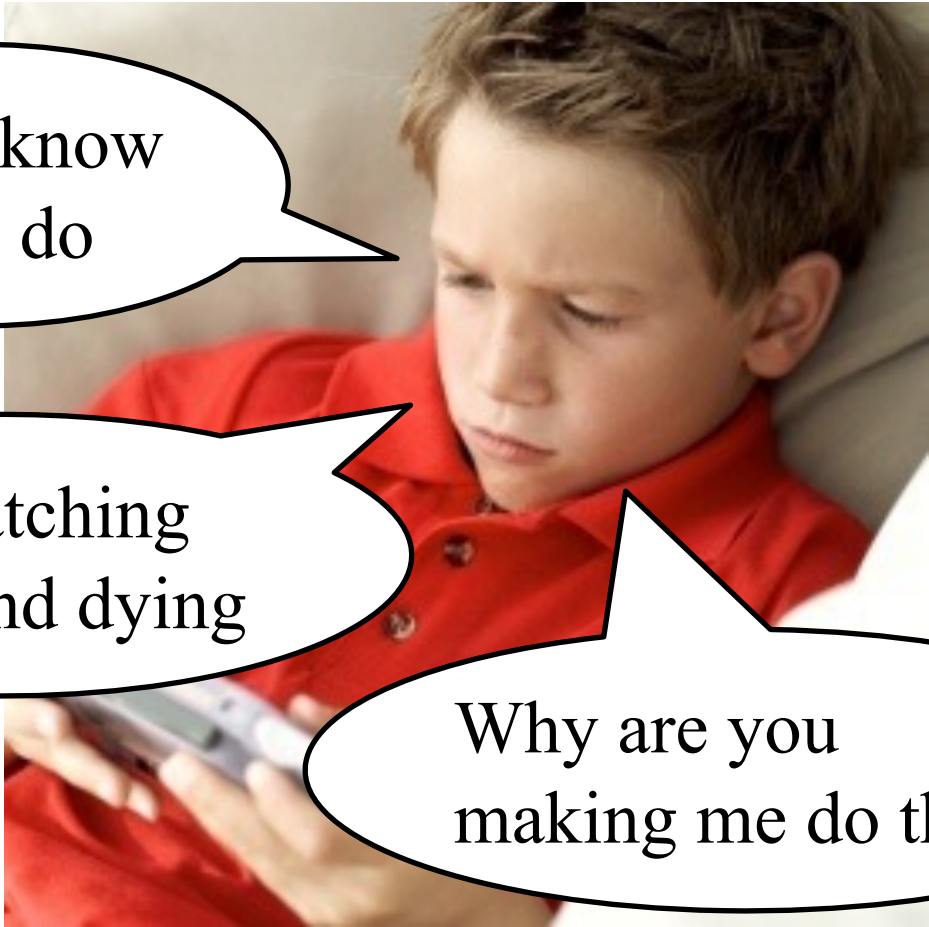


I don't know  
what to do

I keep catching  
on fire and dying

# Think-alouds

---



I don't know  
what to do

I keep catching  
on fire and dying

Why are you  
making me do this

# The Study: Data Collection

---

- **Questionnaires**

- When? Is it a pretest or posttest?
- Multiple choice vs. open-ended questions

- **Interviews**

- Again, before or after?
- Don't have to be super-formal


# Questions and Answers

---



# Questions and Answers

---



What happened when you went through the portal?

I can freeze water now

# Do Not Take it Personally

---

- People will criticize the interface
- It can be sad watching people fail
- But **do not** coach them
  - This will bias your results
  - If you mercy rule them, move on

# Post-Test Team Debrief

---

- Fresh observations taste better
  - Do not wait too long to debrief
  - But do not debrief with player in room
- Talk about each session post-session
  - Comparing results to previous sessions
  - But ignore sessions that are too old
- Talk about general issues every day



# Problem Chart Spreadsheets

---

- **General Format:**
  - Statement of the problem
  - Observation(s) that prompted it
  - Estimated importance
  - Ease of fixing
- This allows you to prioritize
  - And also define “problem”
  - In games, some things should be hard.

# Problem Chart Spreadsheets

---

- **General Format:**

- Statement of the problem
- Observation(s) that prompted it
- Estimated importance
- Ease of fix

- **Things to consider when prioritizing**

- And also define “problem”
- In games, some things should be hard.

Not unlike bug tracking!

# Analyzing with purpose

---

- **What** is the *biggest problem* with our game
- **Why** is it wrong
- **How** might we improve it
- What is the *second biggest* problem
- Repeat as needed

# A/B test analysis

---

- The *biggest problem* with our game is X
- *We're not sure why it's wrong*
- Therefore we tried two conditions: A and B
- It turns out that A does better
- We speculate that A is better because \_\_\_\_\_
- Therefore we'll stick with A

# Implementing A/B Testing

---

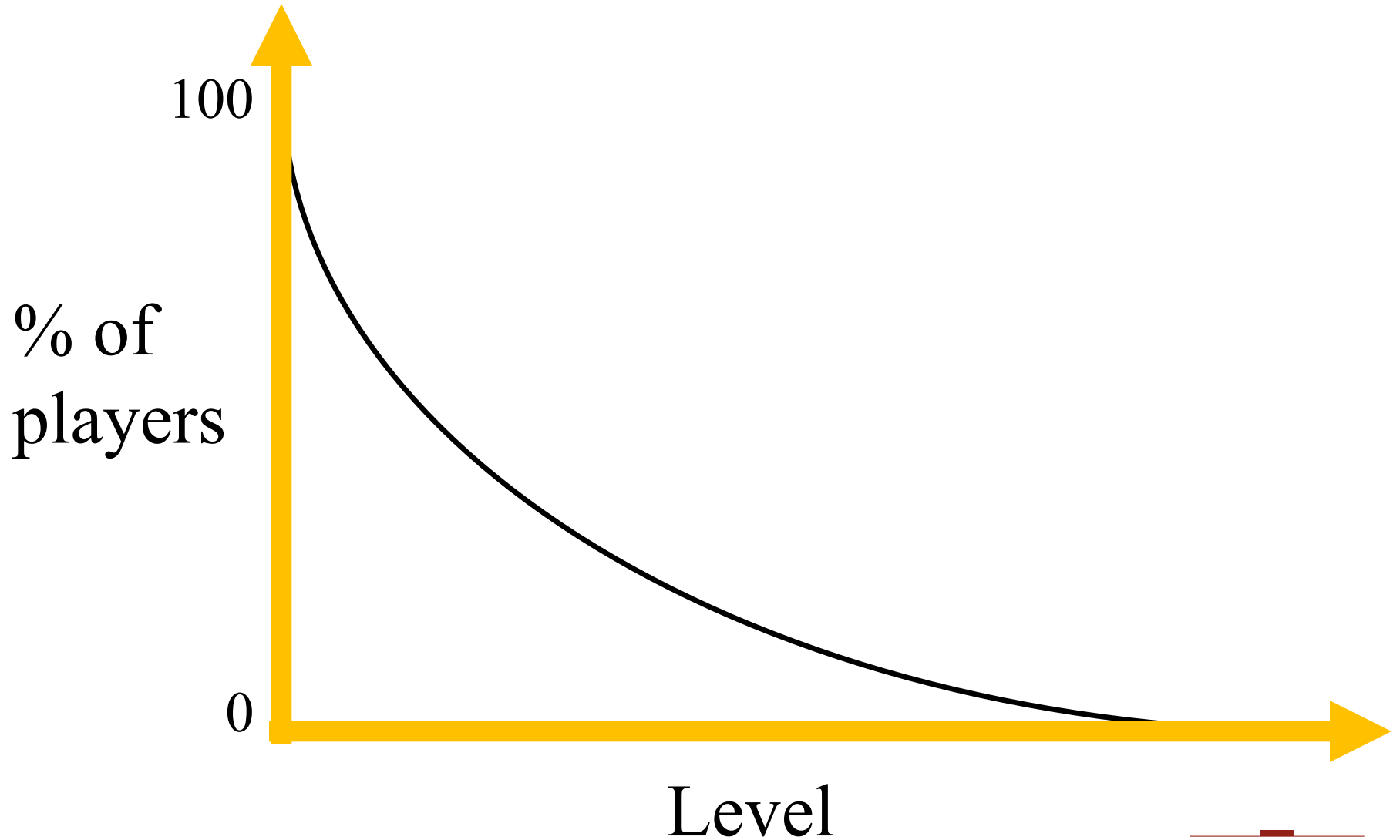
- Have two settings: one for A and one for B
  - Should be modular enough to support both
  - Often a matter of swapping out a controller
- Randomly choose which one for each player
- Record the results of the playtest
  - Works best with quantitative measurements
  - Examples: engagement, player success, etc.
- Compare the two random samples

# A/B Testing: Burndown Chart

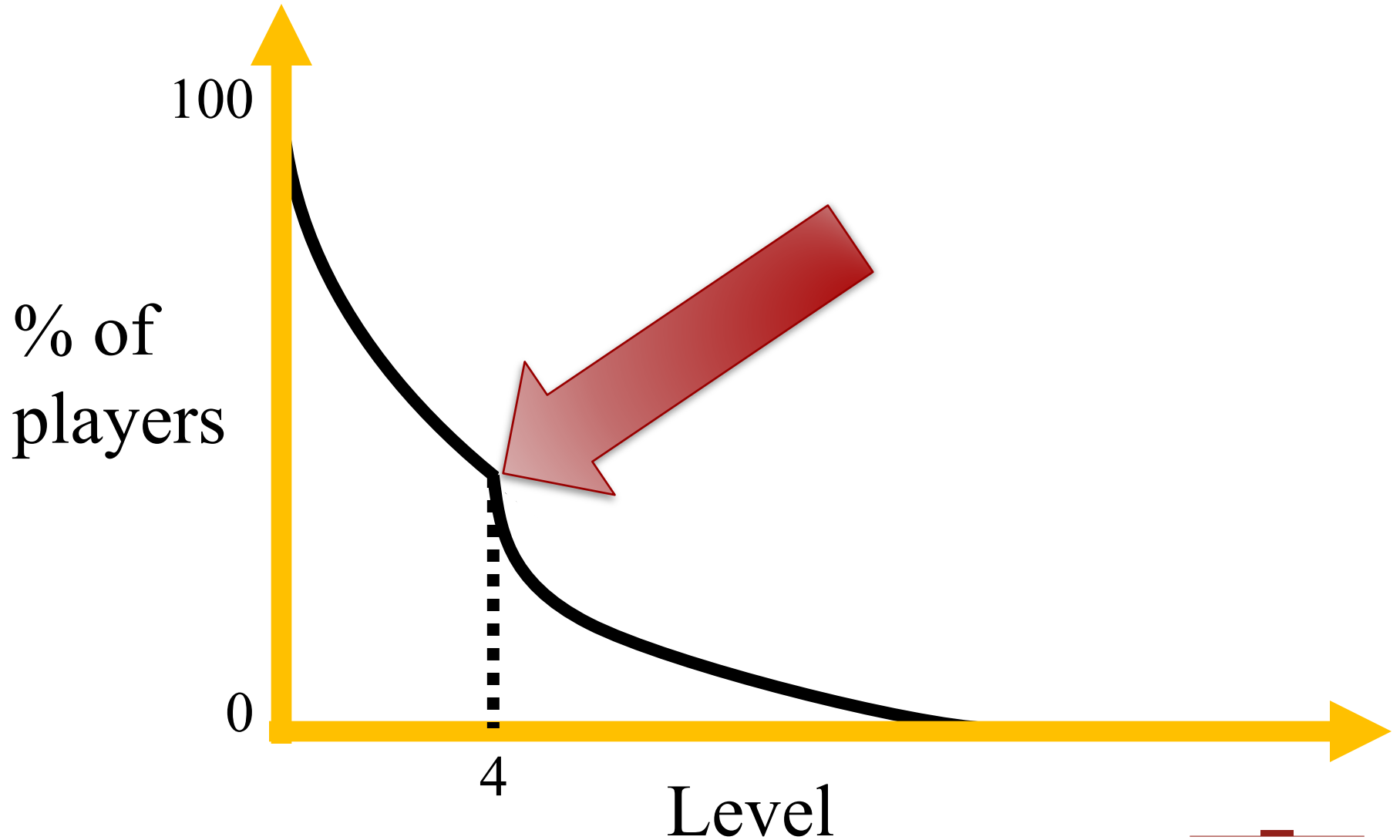
---

After  $x$  levels/seconds, how many people are playing?

# Burndown Chart

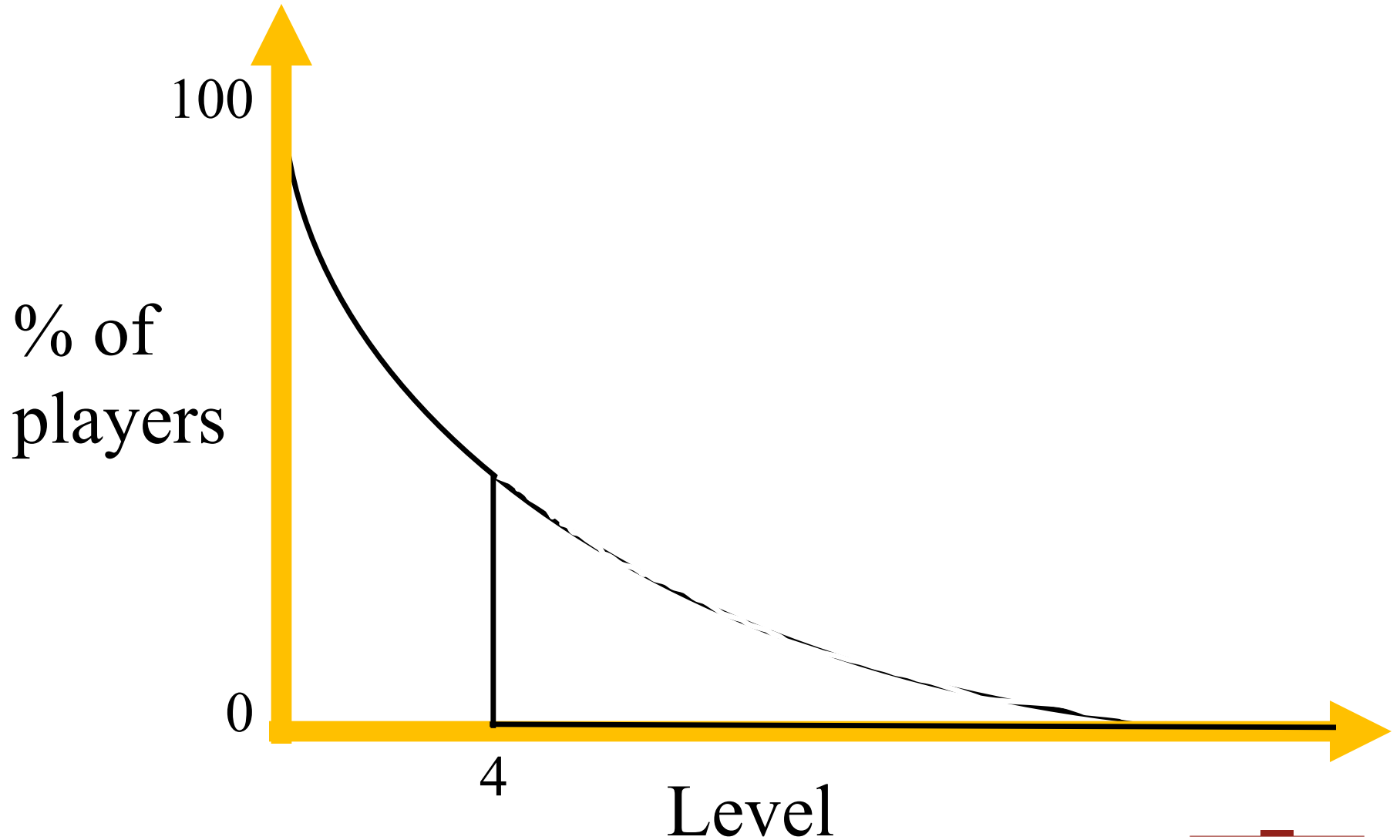


# Unexpected Drop





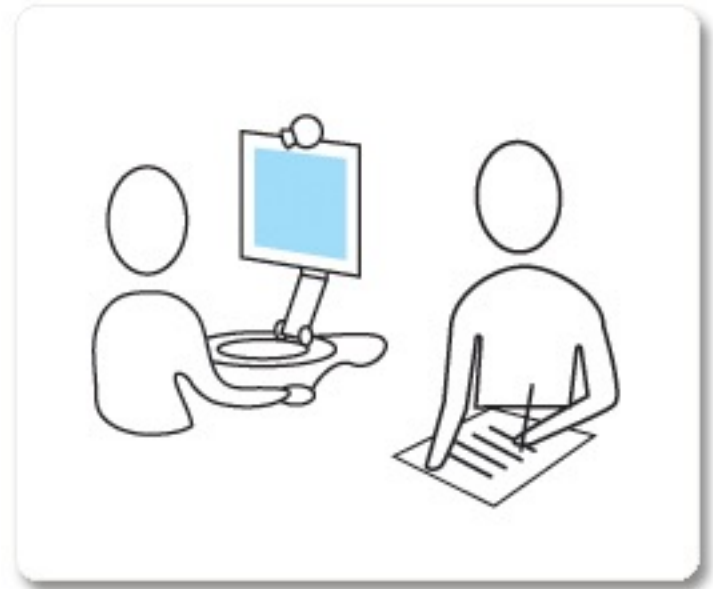
# Unexpected Drop



# Summary

---

- Find representative users
- Have a plan for your test
- Let the player play
- Observe and notice
- Summarize and act
- Rinse and repeat: **frequently**



# Our Goal for Beta

---

- Make a test plan (5-10 minutes)
  - Your artifacts
  - Welcome script
  - Task or two
  - A couple of questions to ask
- Discuss with another group
  - Swap members for testing once or twice
  - Debrief as a group, and with both groups
  - About specific game, about testing overall