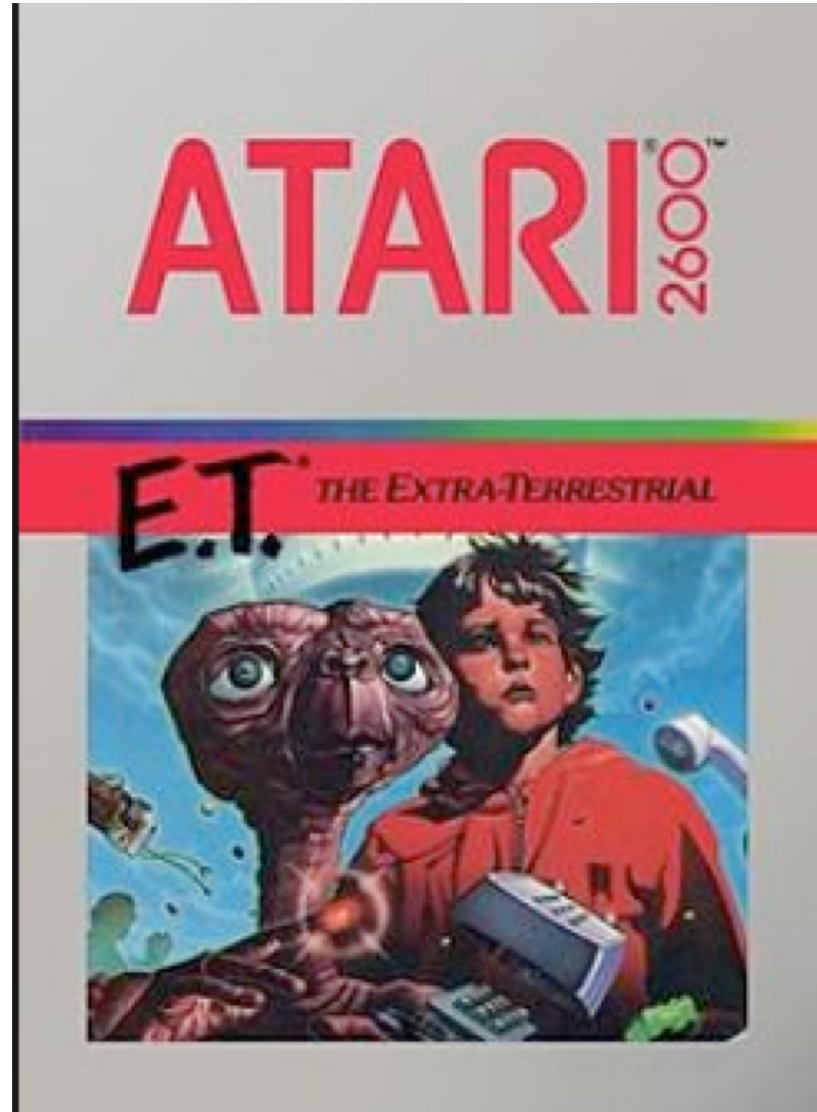


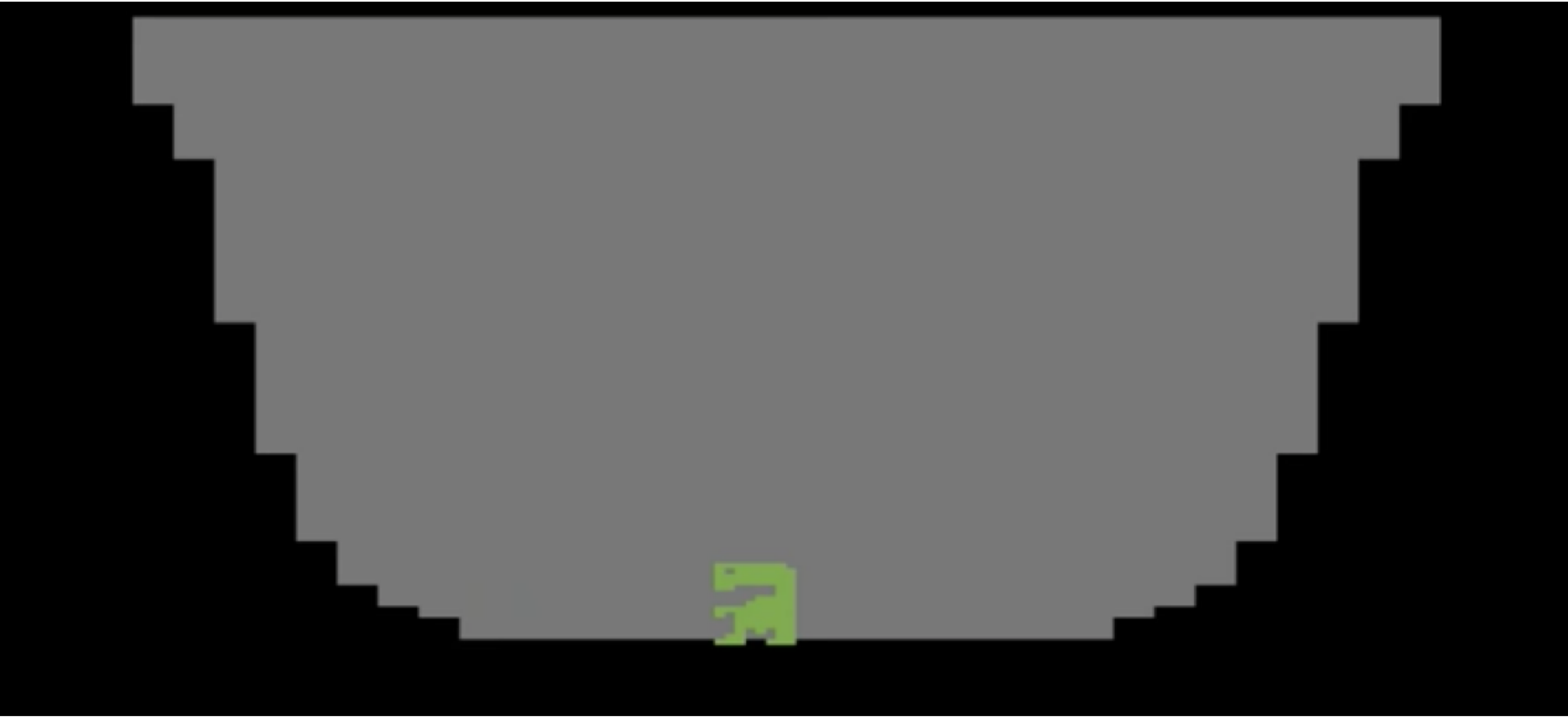
Playtesting

Lecture 23

Playtesting
allows
developers to
understand if
their game is
playable.



“And...I’m in a hole. How the hell do I get out of here?”



Atari video game burial



From Wikipedia, the free encyclopedia

The **Atari video game burial** was a mass burial of unsold [video game cartridges](#), consoles, and computers in a [New Mexico landfill](#) site, undertaken by American [video game](#) and [home computer](#) company [Atari, Inc.](#) in 1983. Up until 2014, the goods buried were rumored to be unsold copies of *[E.T. the Extra-Terrestrial](#)*, one of the [biggest commercial failures in video gaming](#) and often cited as one of, if not, the [worst video game ever released](#), along with the [Atari 2600 port](#) of *[Pac-Man](#)*, which was commercially successful but critically maligned.

Atari video game burial



Atari game cartridges from the burial, as seen during the 2014 excavation

Date	September 26, 1983
-------------	--------------------

The image part with relationship ID 162 was not found in the file.

Today's Outline

- Initial Questions
- Qualitative Usability Metrics
- Conducting a User Study/Playtest
- Data Collection/Analysis

Other people's
time is valuable.

Why conduct the test?

Playable?

Does it
work?

Controls?

Pace?

What will be learned?

Playable? or
Frustrating?

Does it
work?

Controls?

Pace?

What will be learned?

Do players find workarounds?

What unexpected things do they do?



What will be learned?

Do players find workarounds?

What unexpected things do they do?

Levels?

Tutorials?

How will results be used?

Iterate the
design?

Refine art?

Decide about
tutorial need?

What matters in a game?

How do you measure it?

Playtesting can reveal severe limitations (and markets).

Pay attention to accessibility.

The image part with relationship ID r122 was not found in the file.

Game accessibility guidelines

[BASIC](#) [INTERMEDIATE](#) [ADVANCED](#) [FULL LIST](#) [WHY AND HOW](#)

A straightforward reference for inclusive game design

Supporting the industry since 2012, through award winning guidance and examples of how to cater for gamers with disabilities and other impairments

“ I think this web page on making games more accessible is awesome.
Paul Barnett, Senior Creative Director, EA Bioware

“ This should be required reading. Because we exist, and we want to buy your games.
Silas Humphreys, gamer, disabled

The image part with relationship ID r122 was not found in the file.

NanoEmpire is a great example
of a successful accessible game.

Cornell's James Senter

iOS didn't market well, but....

blind users market boomed.

Energy: 44

Followers: 0

Attack: 0

Defense: 0

Production: 2

Production: 0

Production: 0

Production: 0

Create

Choose a location to create in your Energy.

Create

Towers of Wizardry

The ability to manipulate local nanorobots is almost indistinguishable from magic, especially when you employ a system of wands and incantations to regulate followers' access. Black and white towers signify opposite schools of thought.

Create the Towers of Wizardry for 20 Energy?

+1 Energy Production, +1 Defense Production

Yes

No

Towers of Wizardry

Lightning Vortex

Ultimate Race Cour

Temple of the Sta

Nuclear Hatchery



MUDs and early MOOs provided other text-based gaming (and teaching platforms).

Super Awesome Adventure Game

You are about to enter the headquarters of your arch nemesis.
You need a code name for this mission. What will it be?
Lily Snickersee
Your code name is confirmed to be Lily Snickersee. Good luck!

At the front of the imposing building you see a weathered old man with a cart.
As you near, you see the cart is filled with what looks like mostly junk and
only a few useful items. All you have on you is piece of a chalk.
You offer it to him, and he says he'll trade a flashlight or an umbrella for it.

To choose type either 'flashlight' or 'umbrella'.
Lily Snickersee what is your choice?

Colorblindness affects about
10% of player market.

Coblis — Color Blindness Simulator

If you are not suffering from a color vision deficiency it is very hard to imagine how it looks like to be colorblind. The **C**olor **BL**indness **S**imulator can close this gap for you. Just play around with it and get a feeling of how it is to have a color vision handicap.

Colorblindness Fail: The Witness

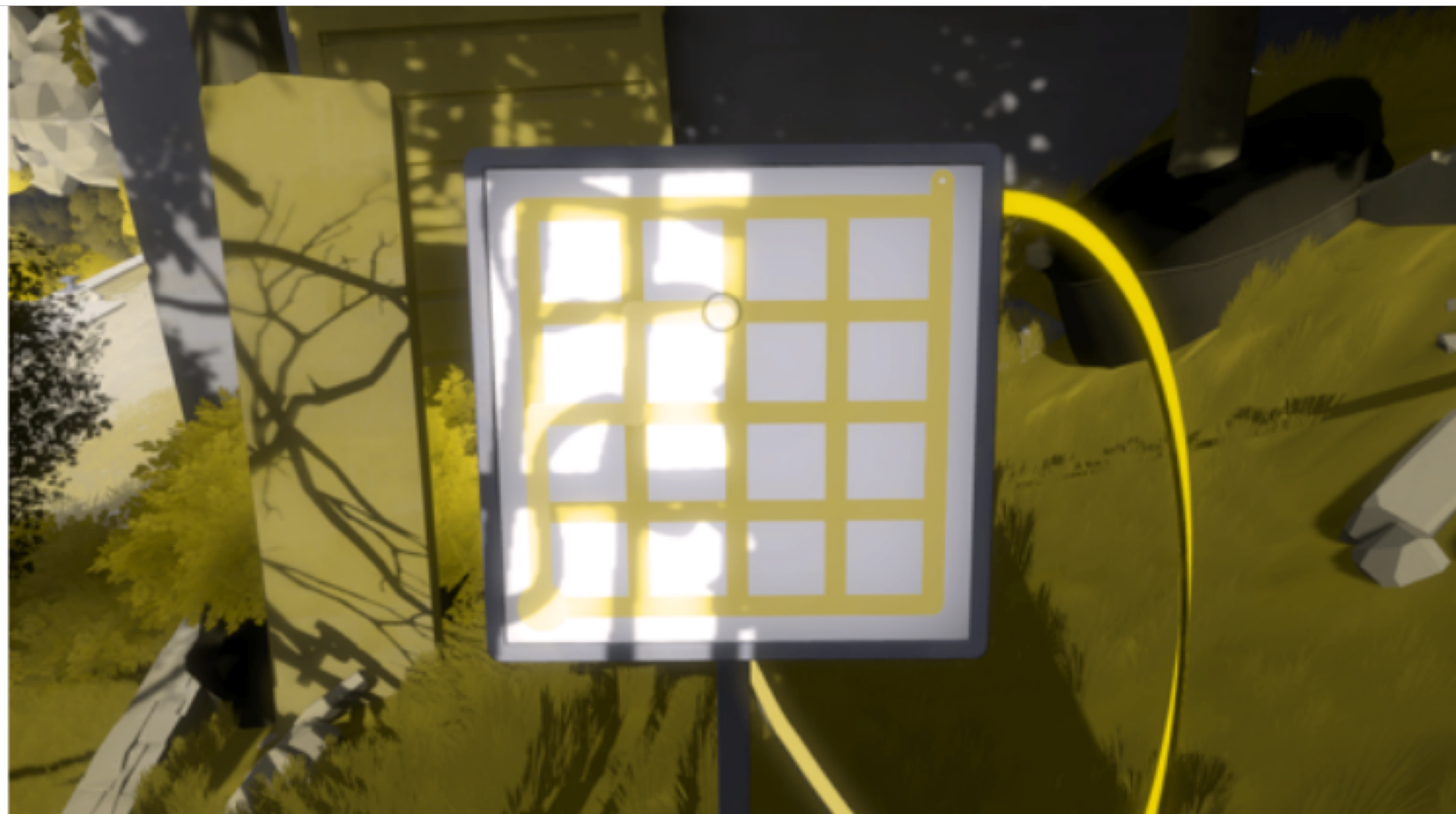
(original view)



The image part with relationship ID 1162 was not found in the file.

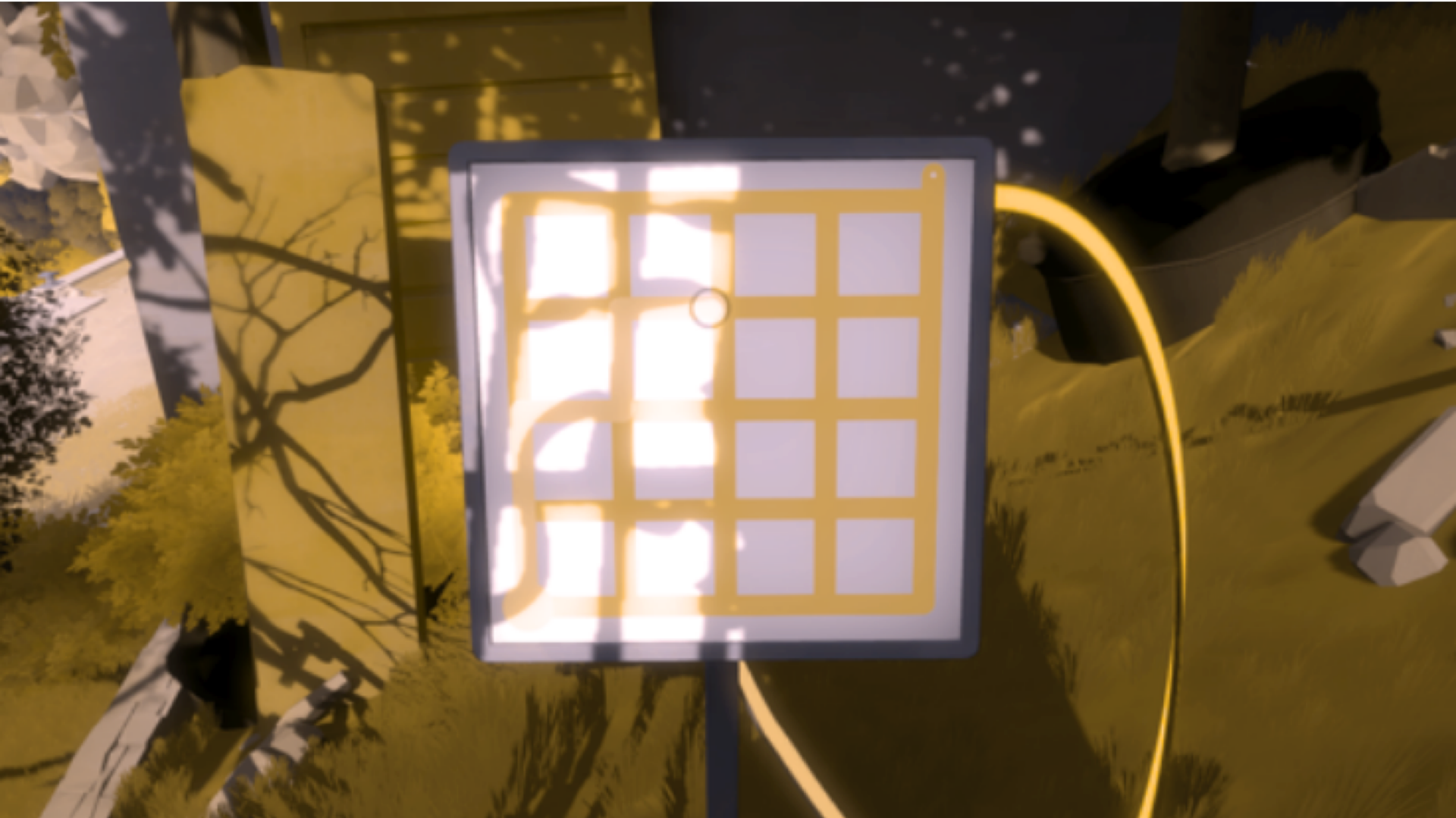
Colorblindness Fail: The Witness

(test view 1)



Colorblindness Fail: The Witness

(test view 2)



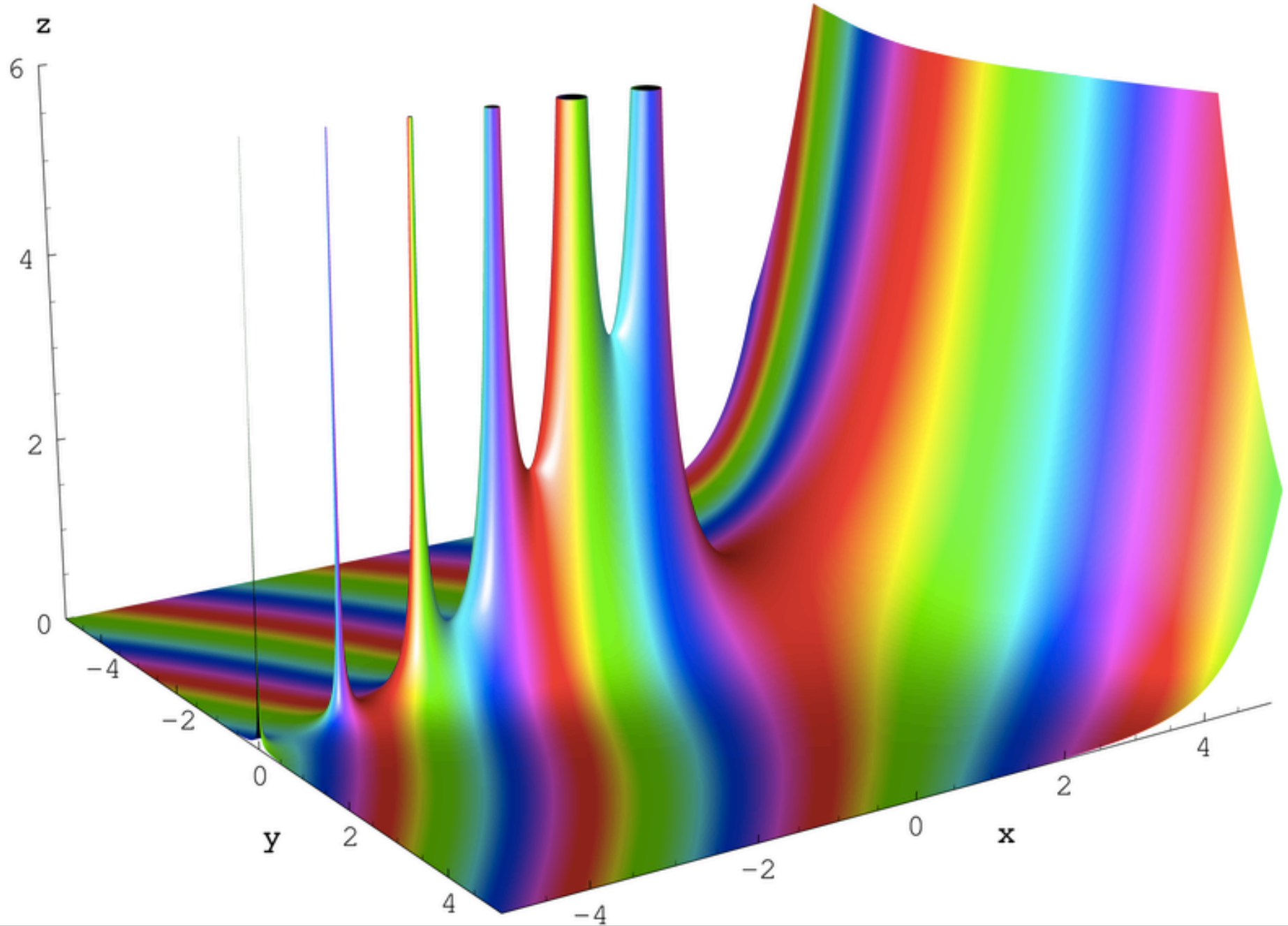
Colorblindness Win: *Witcher 3*



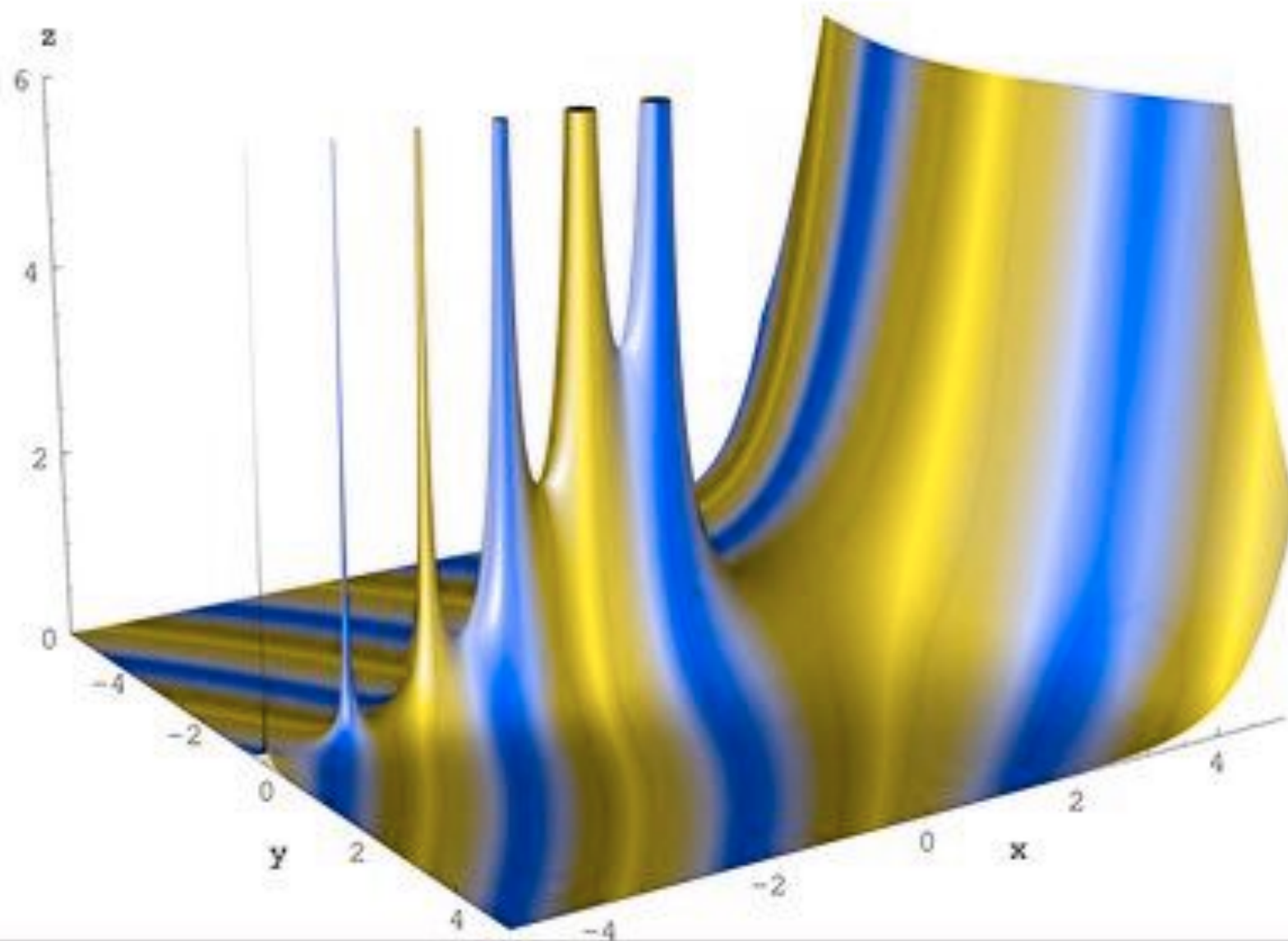
Highlights are traditional hues



Highlights are a much brighter palette



Deuteranope Simulation



Other issues can arise during playtesting.

Inventory management in RPGs can be a problem.

Truths:

Playtesting is needed because your team is too familiar with the game.

Good management of inventory is critical.

Bad UI is game breaking.

Case study: Baldur's Gate

Arrows were ranged weapons that use ammo.

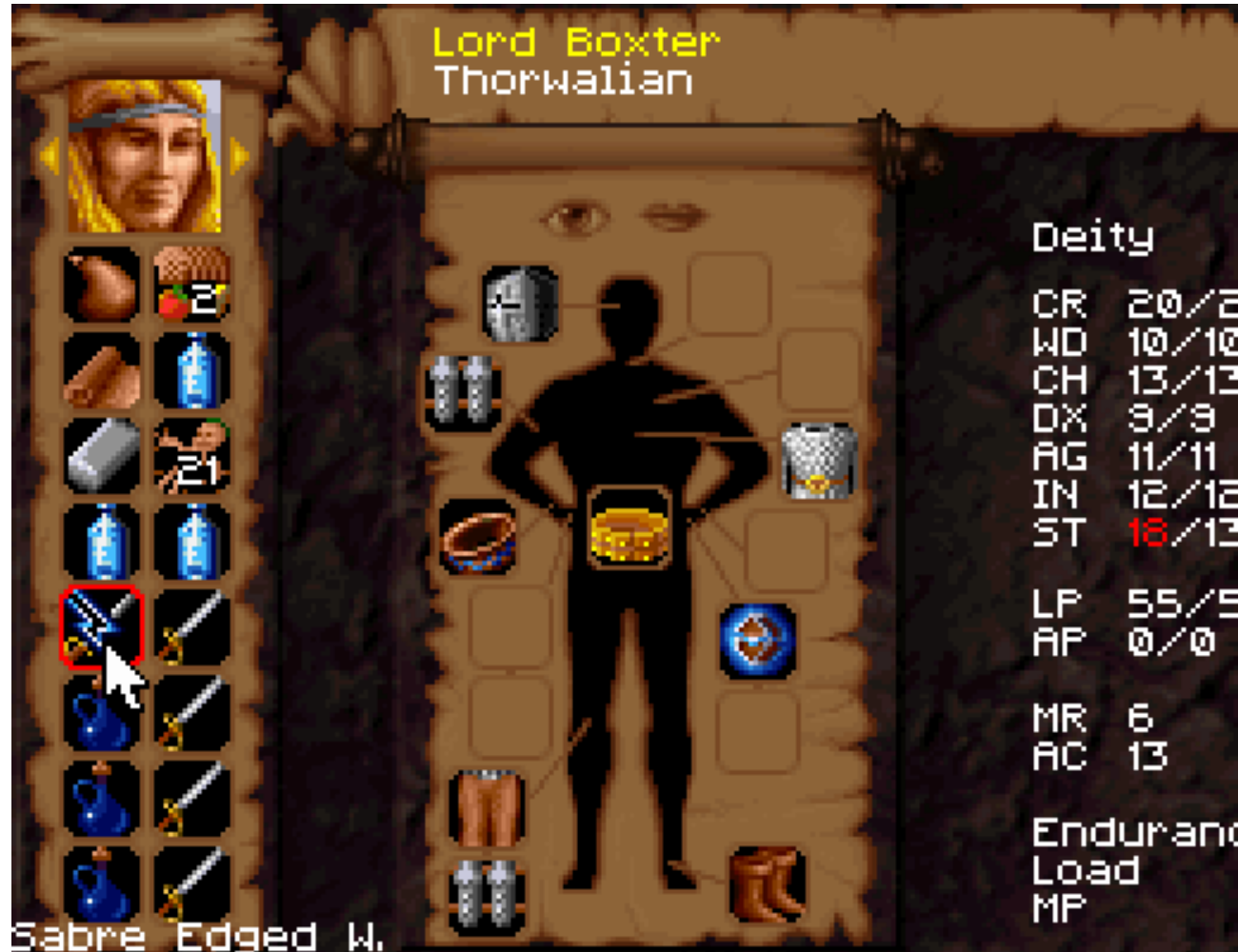
Players could not type amount.

You had to set number of arrows with up/down keys!



Case study: Realms of Arkania

Food had to be
dragged to the
avatar's lips!



Today's Outline

- ~~• Initial Questions~~
- Qualitative Usability Metrics
- Conducting a User Study/Playtest
- Data Collection/Analysis

Quantitative metrics can be easy to capture.

- 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

But qualitative metrics must be gathered intentionally.

- What does the user say?

But qualitative metrics must be gathered intentionally.

- What does the user say?
- Where/how do they run into trouble?

But qualitative metrics must be gathered intentionally.

- What does the user say?
- Where/how do they run into trouble?
- What's the first reaction/impression?

But qualitative metrics must be gathered intentionally.

- What does the user say?
- Where/how do they run into trouble?
- What's the first reaction/impression?
- How would they describe the gameplay?

But qualitative metrics must be gathered intentionally.

- 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?

But qualitative metrics must be gathered intentionally.

- 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

Generate testing queries that gather true data.

Playable? or
Frustrating?

LISTEN. And record.

Playable? or
Frustrating?



The image part with relationship ID 162 was not found in the file.

LISTEN. And record.

Playable? or
Frustrating?

What the...????

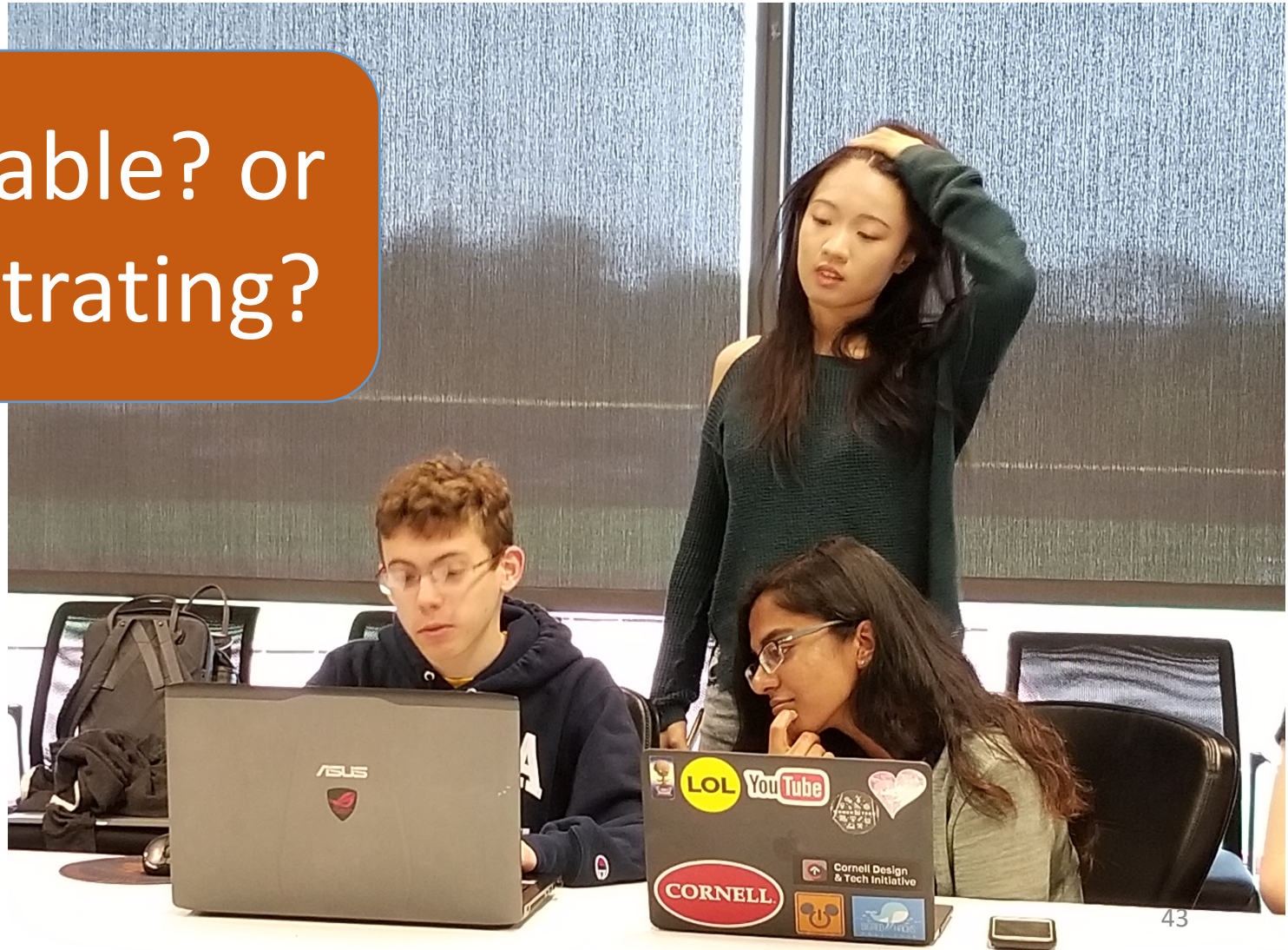
This
makes
me dizzy.

I have no
idea what
to do right
now...

Oh! That was
awesome.

WATCH (and record OR take notes).

Playable? or
Frustrating?

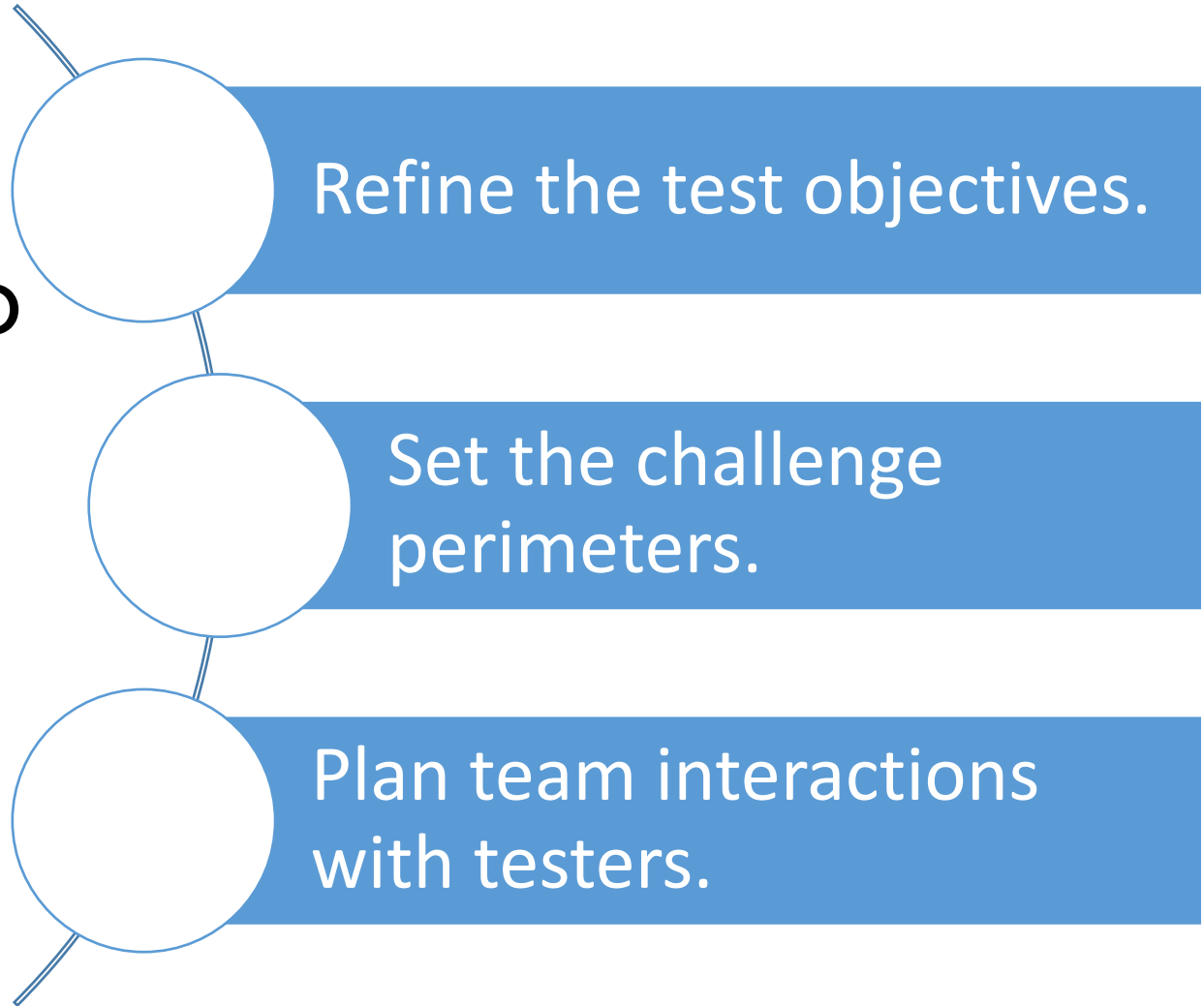


Today's Outline

- ~~Initial Questions~~
- ~~Qualitative Usability Metrics~~
- Conducting a User Study/Playtest
- Data Collection/Analysis

You can't go
in willy-nilly.

PLAN.



Example: Does our game's feature of "fast travel" work?

Example: Does our game's
feature of "fast travel" work? =
YES

Looking around in
here is great!

Whoa...wait...oh, I can
go over here?
Awesome.

Oh, cool...where
does this world
end?

Example: Does our game's feature of "fast travel" work? = **NO**

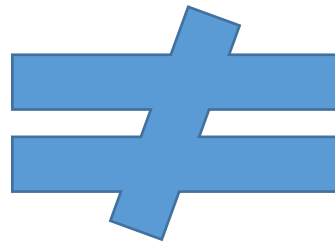
I don't know where I'm supposed to go. I can't decide so fast.

Why can't I stop and build here?

Is it just the same landscape in all parts?

Side note: Educational games often do not spend enough time defining and refining pedagogical frameworks and measurable outcomes.

Game



Learning

Create the test deliberately.

They are doing YOU a favor.
Their ROI is basically nothing.



Define playtester population and goals.



52 was not found in the file.

Refine chosen players.

- Who will be testing?
(Kids? Peers? 3152ers? Random strangers?)
- Why did you choose that body of testers?
- What is to be specifically gained from that pool?
- How many? (3-5 is great)

Define time, place, setup.

- Will you test all players at once?
- Where will this happen?
- By what day will this happen?
- Who will recruit playtesters?
- How much of their time will you take?
- From the team, who will be there to administer and record?
- Where will results be housed for the whole team to see?

Deploy best practices.

1. Have at least two testers.

- **Experimenter**: runs the show
- **Observer**(s): records what happens

Deploy best practices.

2. Be as unobtrusive as possible

- Will you be there when they play? (likely, but where will you BE?)
- Will your input will bias participants or skew results? (likely, but in what ways?)
- Will your presence frustrate the player?

Decide team actions in advance.

At what point can a tester step in on a specific task?

How long should a playtest be?

How do you get them to stop?

What questions do you ask after they stop playing?

For large scale, more formal testing, you should get clearances from Institutional Review Boards.



Cornell University
Office of Research Integrity and Assurance

Refine artifacts to test.

It's ok to test incomplete versions....

BUT you must inform the playtesters of this truth.

THEN, have them focus on finished bits.

Craft test script for the team.

Intro (spoken version)

“We are asking you to test this game for us. It’s not entirely done, and so we are only looking for feedback on X, Y, and Z.”

Craft test script for the team.

Intro (spoken version)

“As you play, we ask that you talk out loud about what you are doing, what is cool, and what isn’t working for you. Just say it...there’s no need to be formal or anything.”

Craft test script for the team.

Intro (spoken version)

“We will be taking notes and maybe recording what you are doing on the screen, too. But it’s really important that we hear you, so we might remind you now and again to talk it out.”

Craft test script for the team.

Intro (spoken version)

“So, we will open up the game for you, but then kind of pretend we aren’t here. We are going to try to not give any hints because we are testing how well a player can learn the game alone.”

You can supplement with
specific inquiries.

“What do you think should happen if you go over here?”

“When you go through this door, what do you expect will happen?”

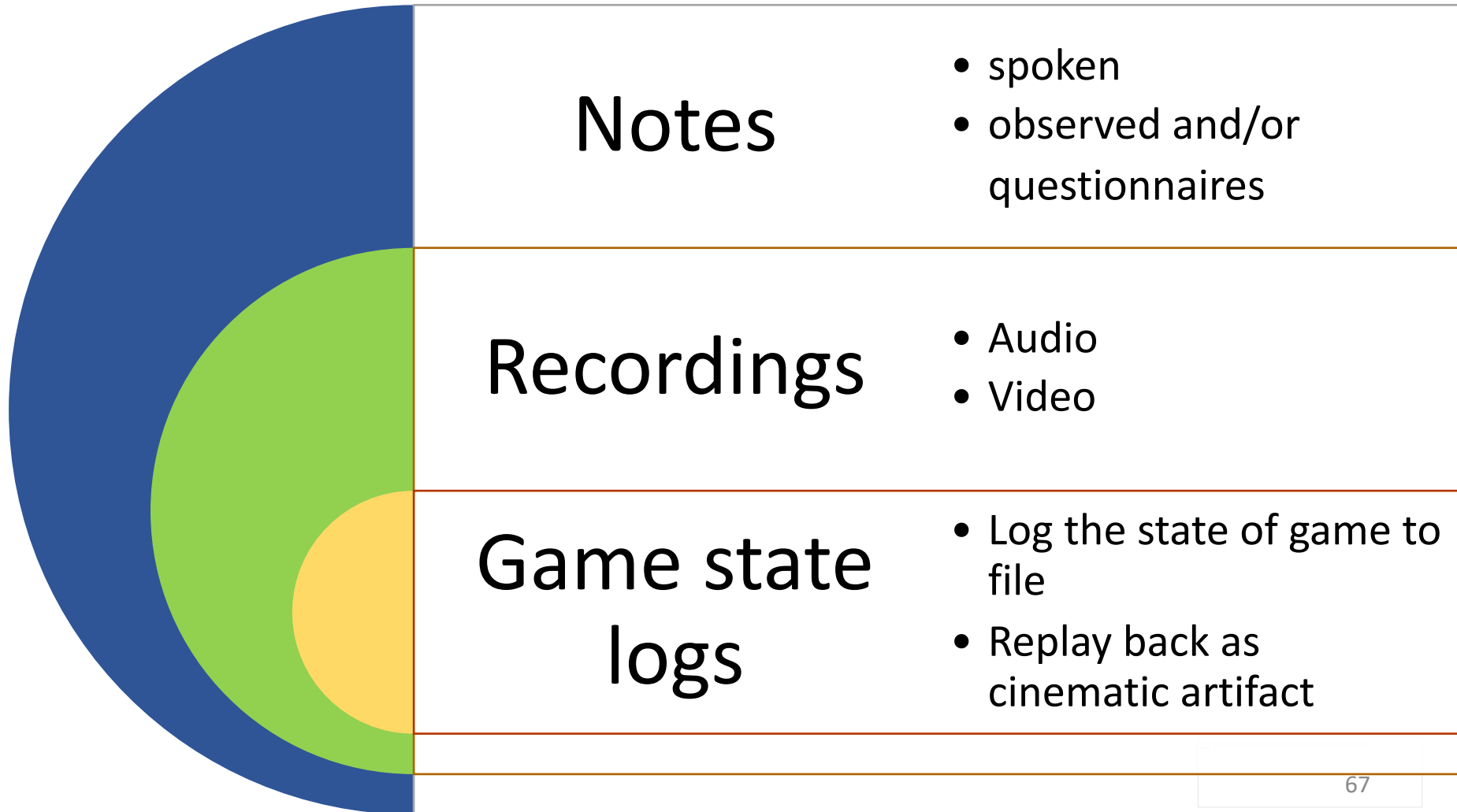
“Does that look like an enemy to you?”

“Do you know what you are looking for?”

“What’s your main goal?”



Data collection comes from several channels.



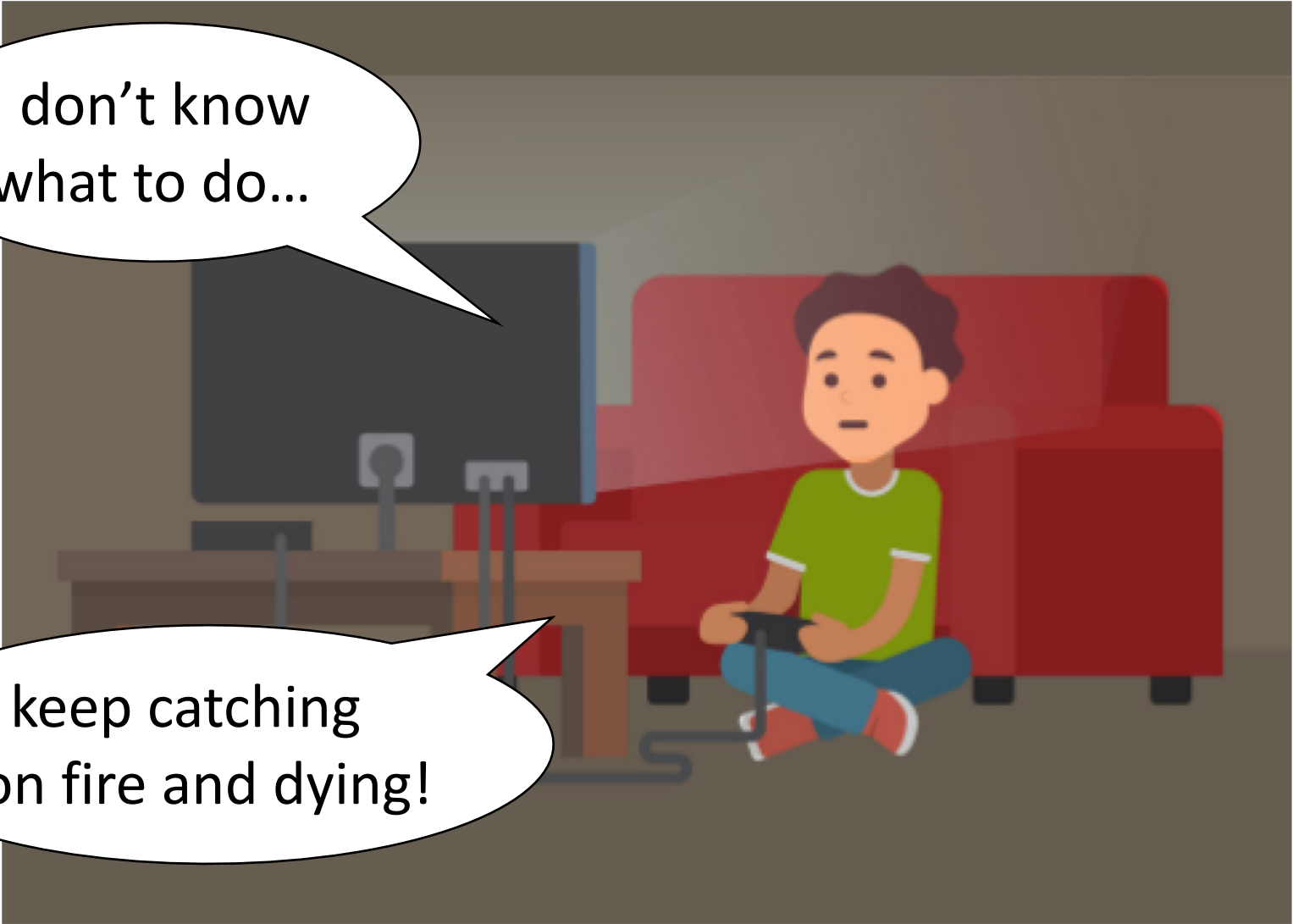
Questionnaires can be effective.

- Is it a pretest or posttest? Both?
- Will you use multiple choice vs. open-ended questions?
- Paper or online? (Google Forms, Kahoot, whatever)

Interviews can also help.

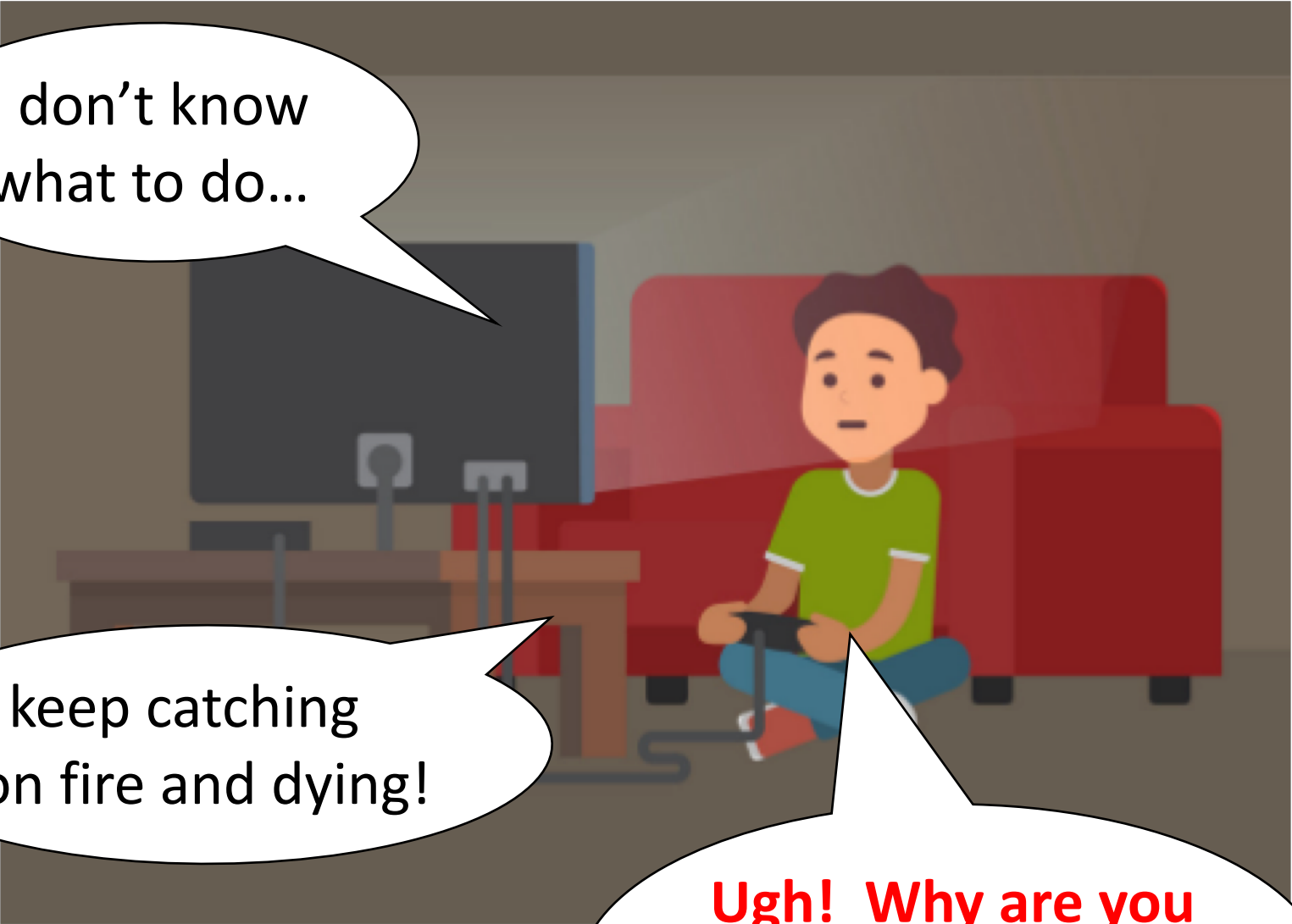
- Again, before or after?
- They do not need to be formal.

The think-aloud method is widely used.



I don't know
what to do...

I keep catching
on fire and dying!

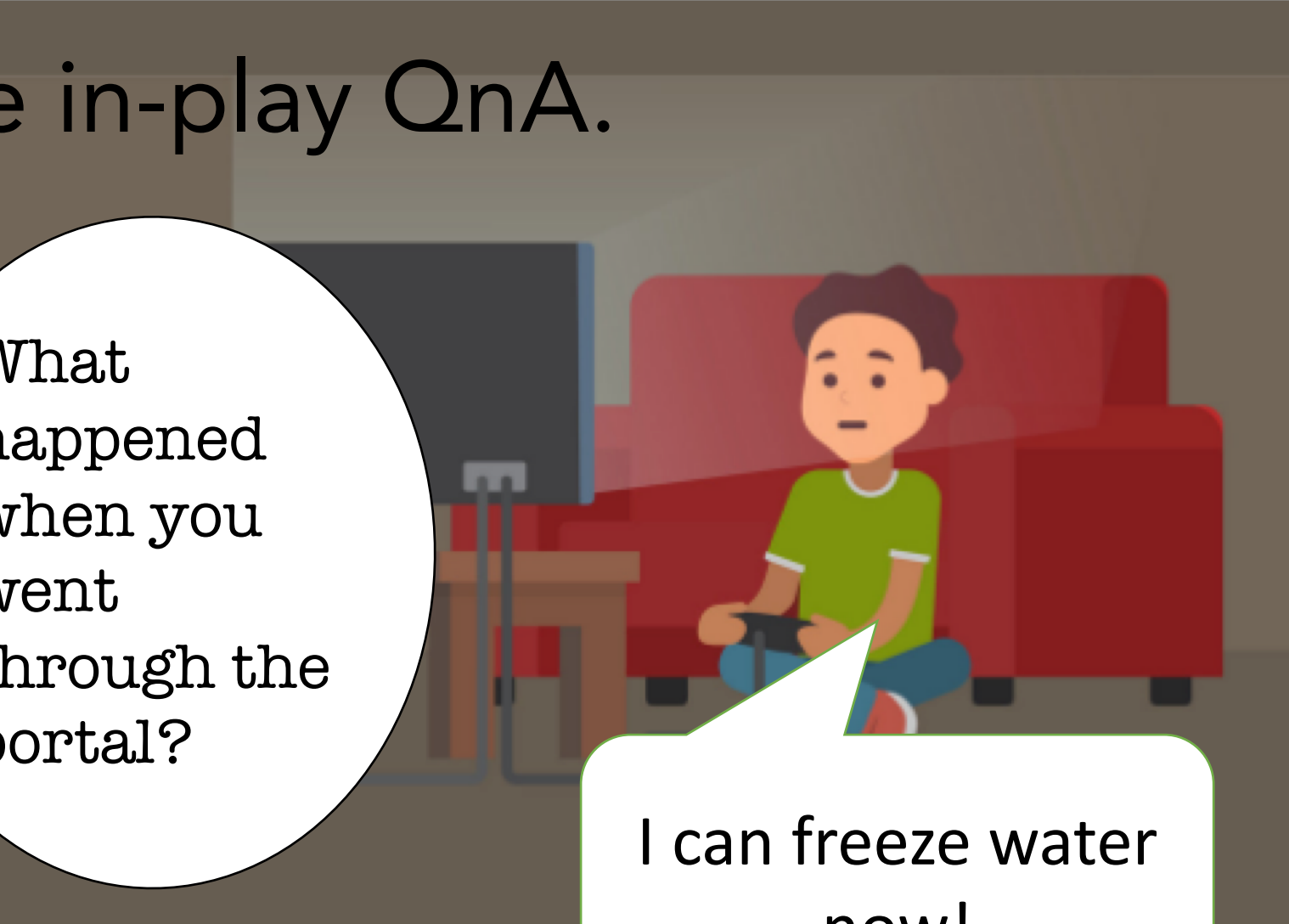


I don't know
what to do...

I keep catching
on fire and dying!

**Ugh! Why are you
making me do
this??**

Use in-play QnA.



What happened when you went through the portal?

I can freeze water now!

A worthy goal...

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

Do not take feedback personally!

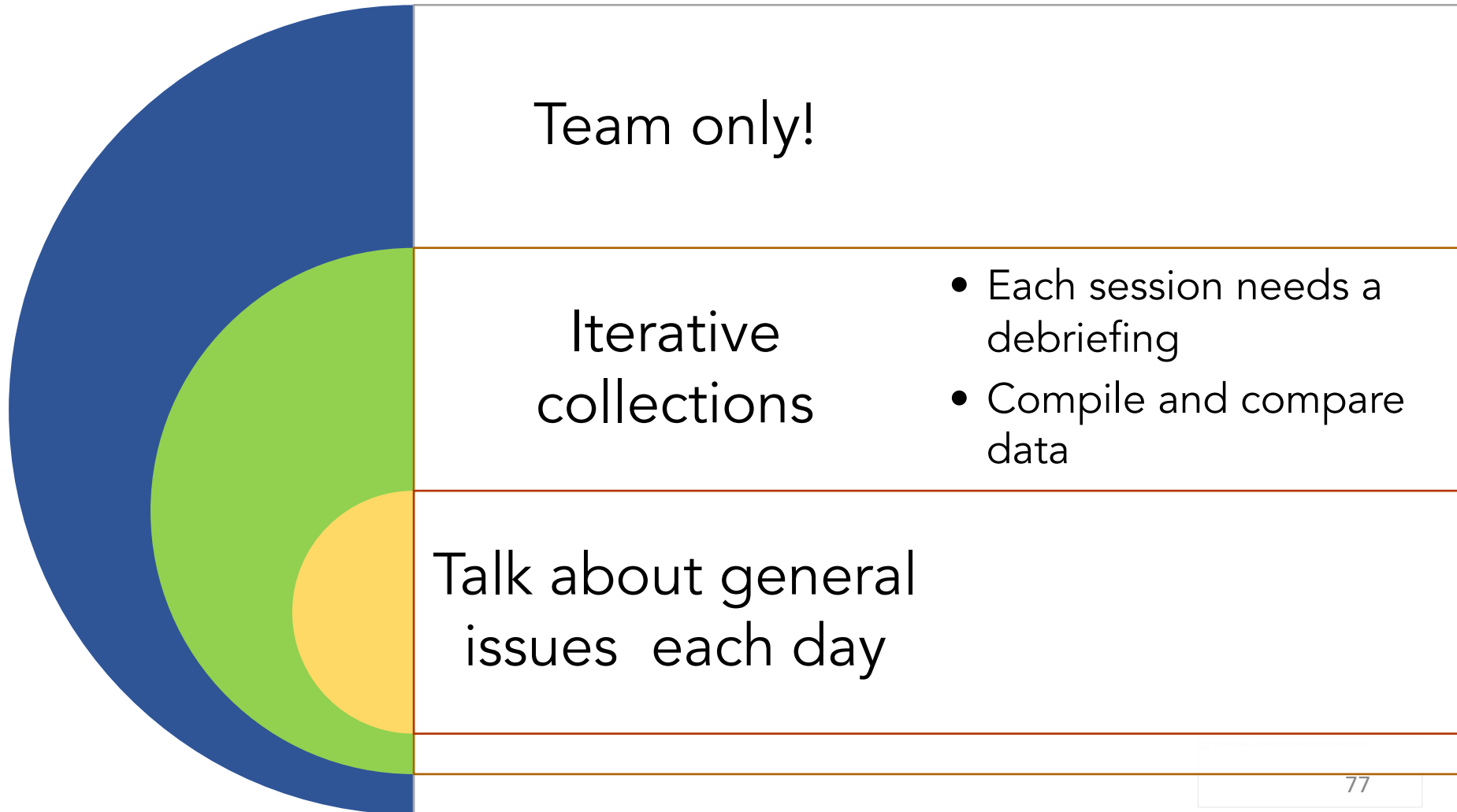


© iD H52 was not found in the file.

Today's Outline

- ~~Initial Questions~~
- ~~Qualitative Usability Metrics~~
- ~~Conducting a User Study/Playtest~~
- Data Collection/Analysis

Conduct a quick post-test debriefing.



You can use spreadsheets.

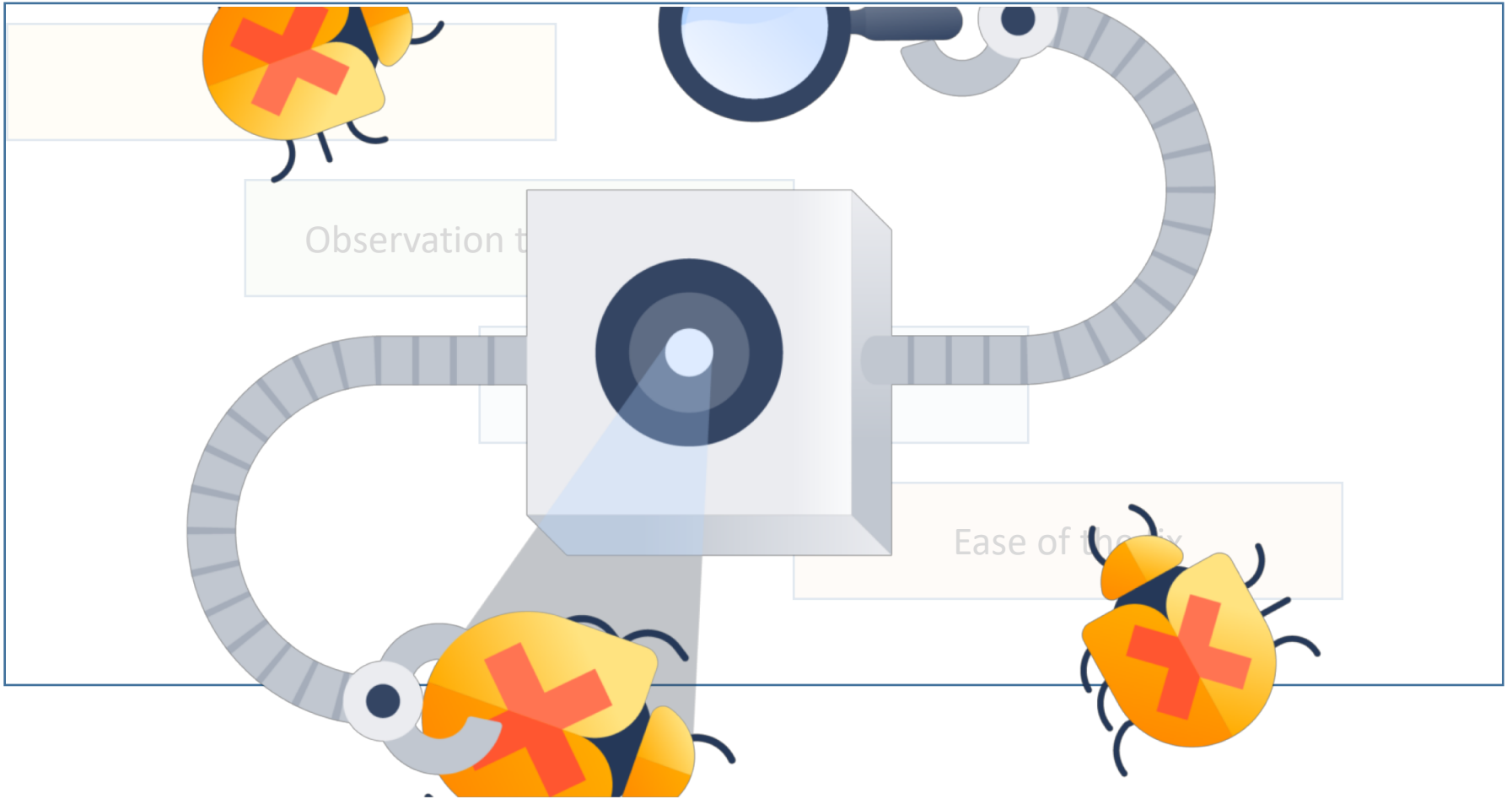
Problem

Observation that caught it

Importance level

Ease of the fix

It's like bug tracking!



Importance level: CRITICAL

What is the
biggest
problem?

Why is it
wrong?

What can we
do?

Importance level:
LESS CRITICAL

What is the
problem?

Why is it
wrong?

What can we
do?

Summary

1. Find representative users
2. Have a plan for your test
3. Let the player play
4. Observe and notice
5. Summarize and act
6. Repeat **frequently**



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

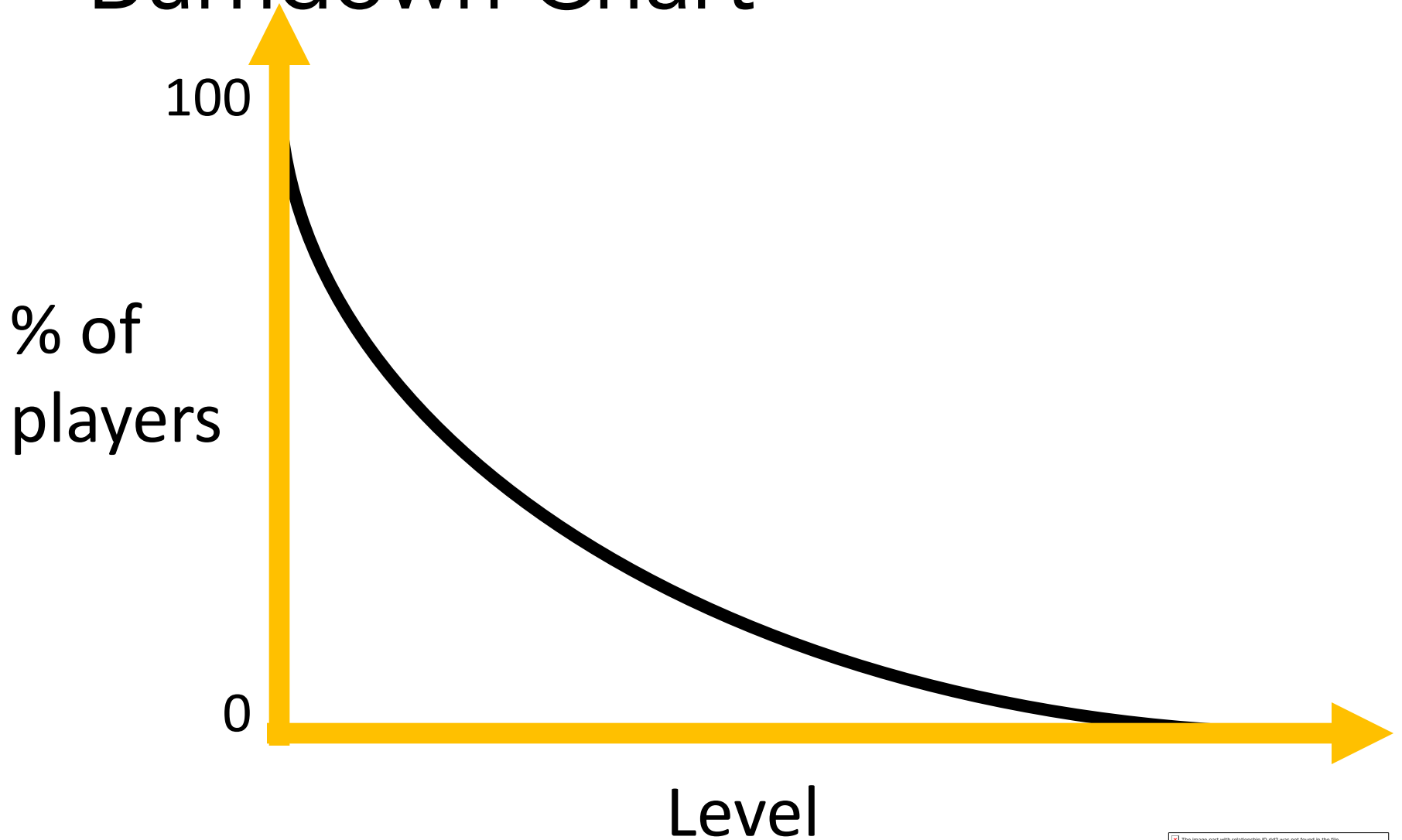
Implementing A/B Testing

- 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

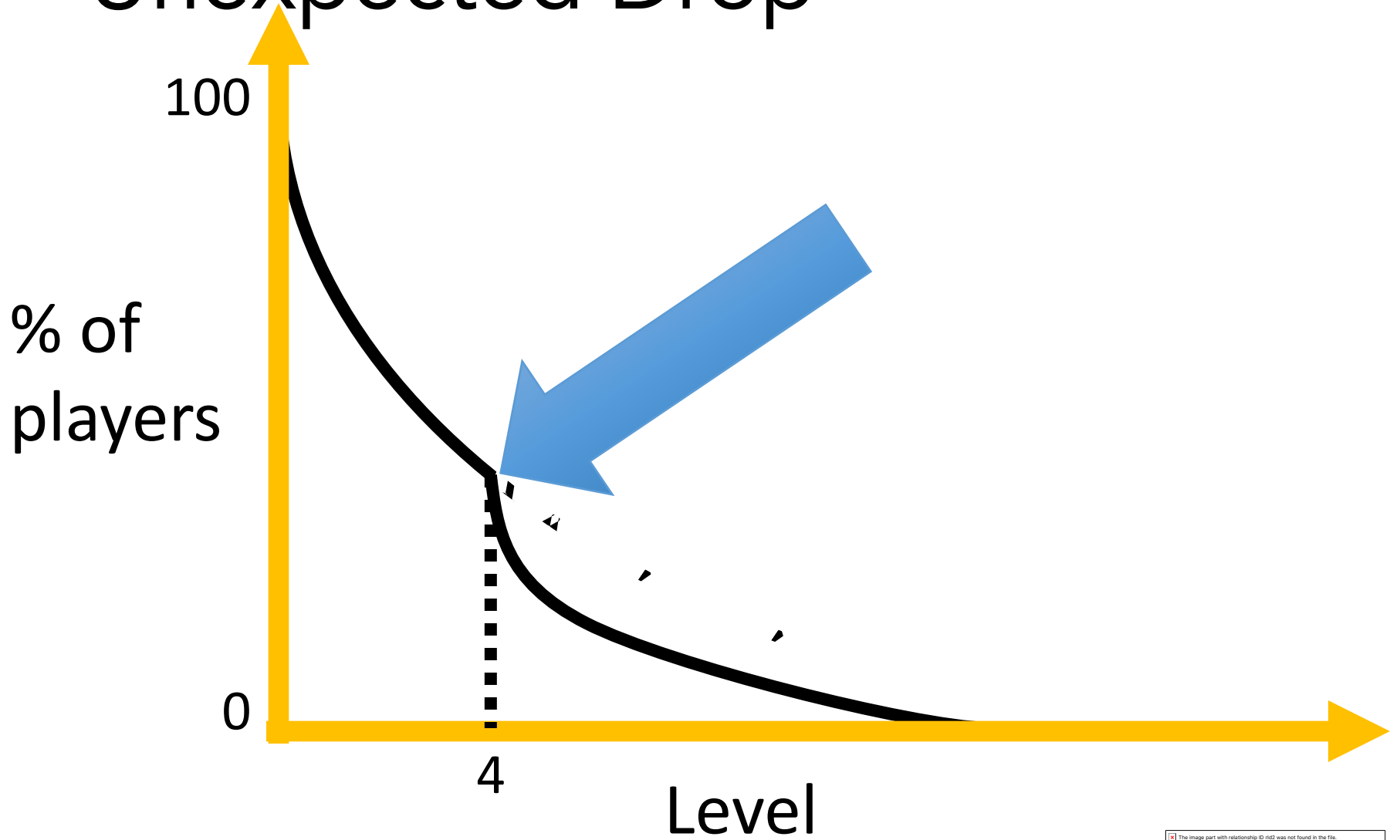
A/B Testing: Burndown Chart

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

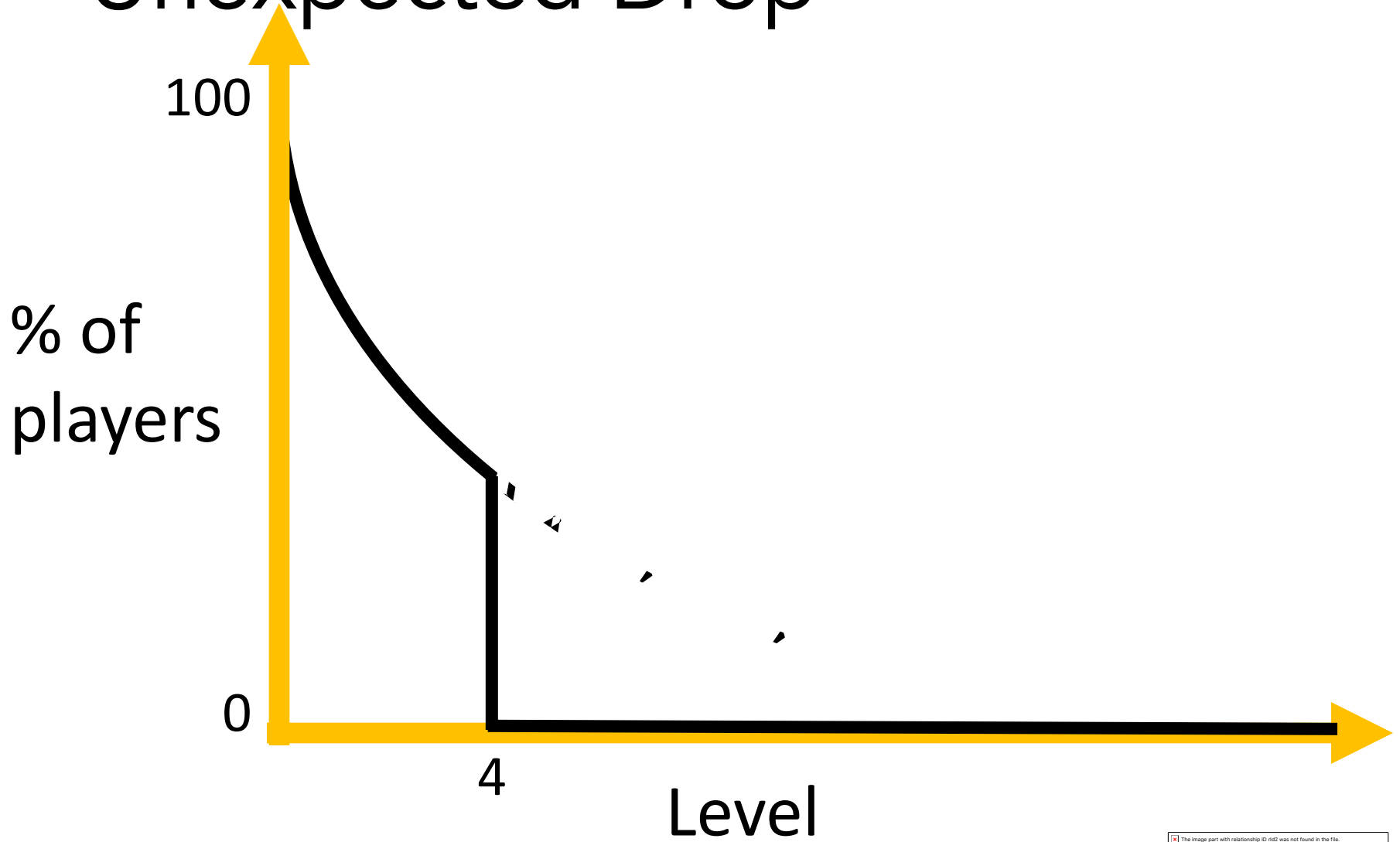
Burndown Chart



Unexpected Drop



Unexpected Drop



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