Software Verification Games:
Designing Xylem, The Code of Plants

Heather Logas
Jim Whitehead
Michael Mateas
Richard Vallejos
Lauren Scott
John Murray
Kate Compton

Joseph Osborn
Orlando Salvatore
Dan Shapiro
Zhongpeng Lin
Huascar Sanchez
Michael Shavlovsky
Chris Lewis

Center for Games and Playable Media
University of California, Santa Cruz

Presentation by Jim Whitehead, slides by Heather Logas

Formal Software Verification

- Formal software verification is a process by which software systems can be proven to have specific qualities or be free of certain kinds of bugs
- Done by experts (~1000 in the United States)
- Time-consuming
- Expensive

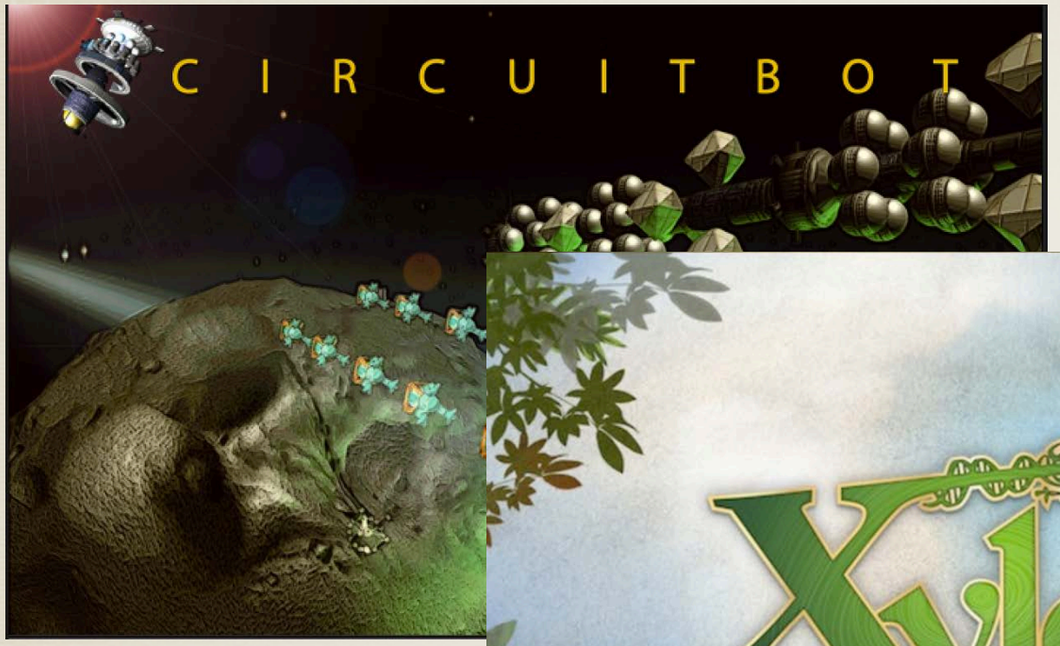
Automating Formal Software Verification

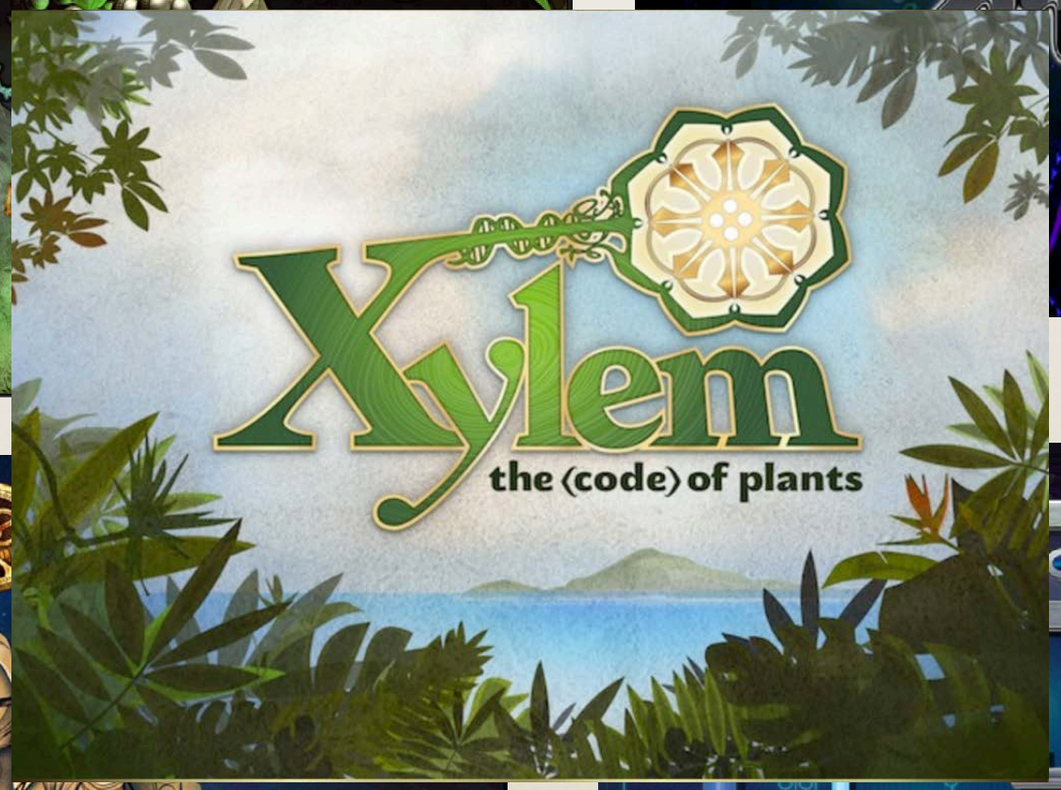
- Hard to scale verification efforts with such a limited pool of experts
- Automating Formal Software Verification is a rich area of research with no solid answers
- What if… non-experts could help software verification in a non-trivial way?

Crowd Sourced Formal Verification (CSFV)

- Goal is to scale formal verification capabilities via game based crowd sourcing
  - Games with a purpose
- Have human game players do tasks that computers currently can’t do well

Crowd Sourced Formal Verification (CSFV)





Our Focus: Loop Invariants

- Loop Invariant: a statement that describes the behavior of a loop (for, while, do)
  - Expressed in a way that is always true
    - Without use of “transfer equations”
    - e.g. x is always < 10, or x + y = 25
- Some loop invariant synthesis can be automated, but not all, and results tend to be very imprecise

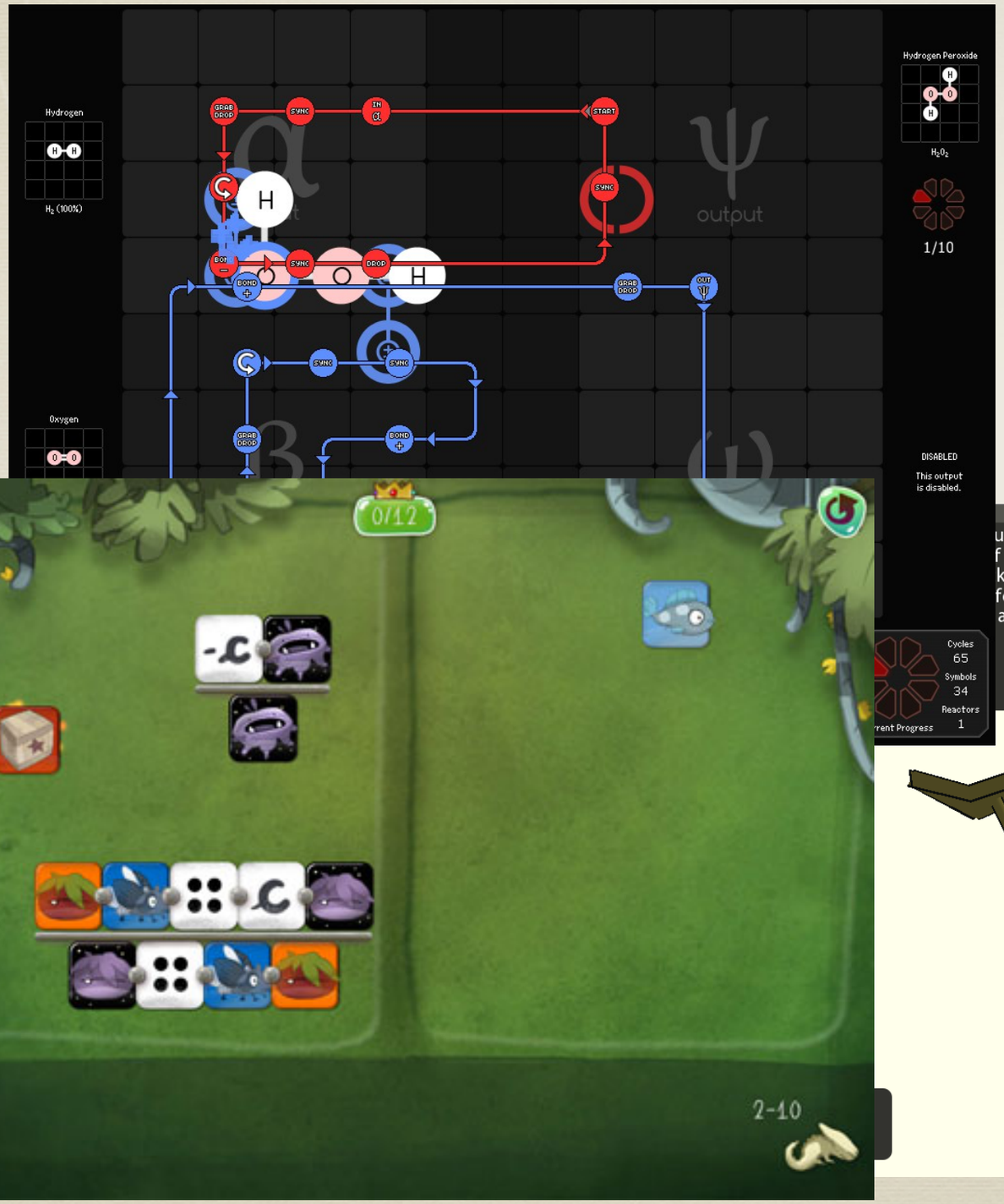
Xylem

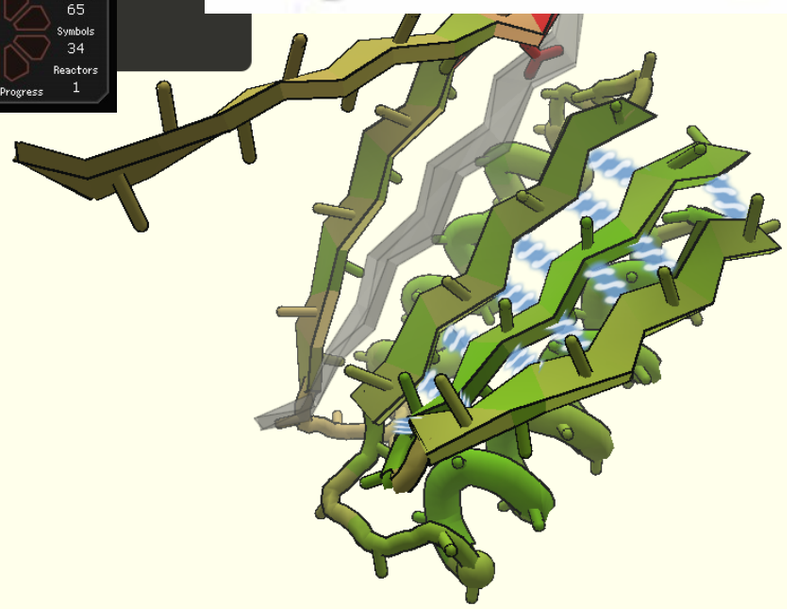
- Science goal: Create a game which allows players to find loop invariants from source software.
- Easy, right? Extra challenge factor:
  - Players are not allowed to see the source code of the loop
  - How can you understand the semantics of a loop without source?

Xylem Approach

- Only provide values of variables inside a loop
  - Player can see values for multiple iterations of the loop
  - Static value analysis generates values
- Inductive reasoning task
  - Relies on pattern observation
  - Humans can do this!

Design Inspirations



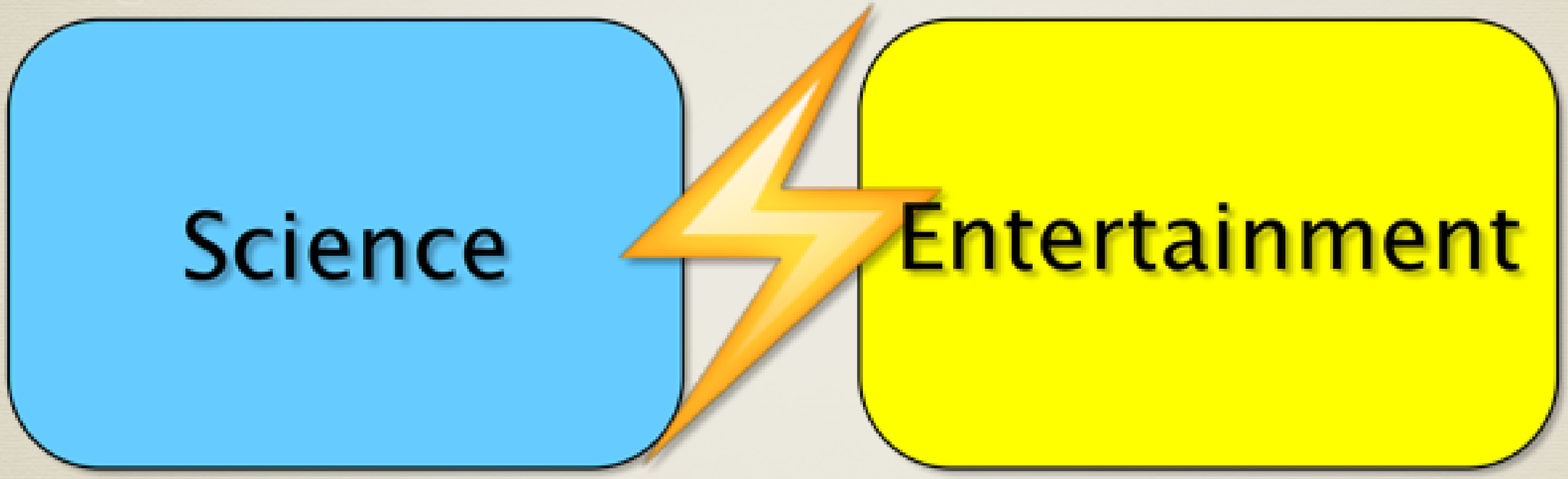


Xylem Game Footage

Challenge: Balancing Verification and Experience

- Specific design considerations
  - Visually representing code
  - Mathematical gameplay
  - Reward system
  - Creating appropriate difficulty curve
  - Appealing to a broad audience
- Lessons Learned

Navigating the Constraints



Intended Audience

Science: Collect *a big crowd*

Game Design: *Math gameplay is inherently non-“casual”*

- Shoot for as *casual an audience as possible*
  - Make math parts interesting and accessible
  - Design and implement features that support motivation and retention in casual audiences

Visualizing Data

Science: Allow crowd to discern loop invariants without showing actual code

Game Design: Provide a visual metaphor that helps players see patterns

- Appealing (or at least not off-putting) to intended audience
- Maps well to domain
- Flexible enough to represent a range of programmatic data structures

Plants!

- Plants are familiar, living and aesthetically pleasing entities
- Plants grow and change over time (makes sense to show iterations)
- Plant features can be used as indicators of variables
- Plant Kingdom is very diverse
- Allows hooks for narrative framing

0 1 2 3 4 5

Eugene's Notepad

Back.

Growth Phase: 0

= 0
= 0
= 0
= 1
= 1

Growth Phase: 1

= 0
= 0
= 0
= 4
= 2

Growth Pha

> ≥ = < ≤ # + − × ÷ (#)

2 =

Transmit.

Mathematical Equations = Gameplay

- ***Science: Allow for a wide range of possible solutions; take best advantage of the crowd***
- ***Game Design: Keep game accessible to target audience***

* In support of a “mathy” approach:
  * Wider range of invariants can be submitted
  * Multiple solutions possible for each puzzle

CLASS OF SERVICE

This is a fast message to be conveyed by carrier pigeon. Message may be delayed by severe weather conditions.

PACIFIC UNION

PIGEON MESSAGE

Operated by CHEKOFV & Co.

SENT from THE UNIVERSITY

RECIEVED at The Island of MIRAFLORA

The filing time shown in the date line on domestic messages is STANDARD TIME at point of origin. Time of receipt is STANDARD TIME at point of destination.

CONTRIBUTION RECEIVED, RE SPECIES 9E313.
FINDINGS COMPARED TO OTHER FIELD BOTANISTS-

Novel Solution Found!

Number of Flower Types Used: 2

Number of Bonus Tiles Used: 0 | 1 | 2

YOU EARNED A ★

FOR USING AN =

Each ✿ represents 10% of Botanists.

○ denotes your group.

CONGRATULATIONS!

TOTAL - ★ ★ ★

THE COMPANY WILL APPRECIATE SUGGESTIONS FROM ITS PATRONS CONCERNING ITS SERVICE

Re-examine

Move On

Rewarding the Player

- **Science: There is no established way to know how "good" a solution is without forcing player to wait for a counter-example. Would like to steer players towards submitting "good" solutions.**
- **Game Design: Need to give player immediate feedback . How do we steer towards "good" solutions if we have no way to know what a "good" solution is?**

Rewarding the Player

- Use what we know!
- “Generally Speaking”
  - Using more variables is better
  - Utilizing the data from the time zero iteration of loop is better (bonus tiles)
  - Using equivalency is better
  - A variety of invariants (from different players) is better

However…

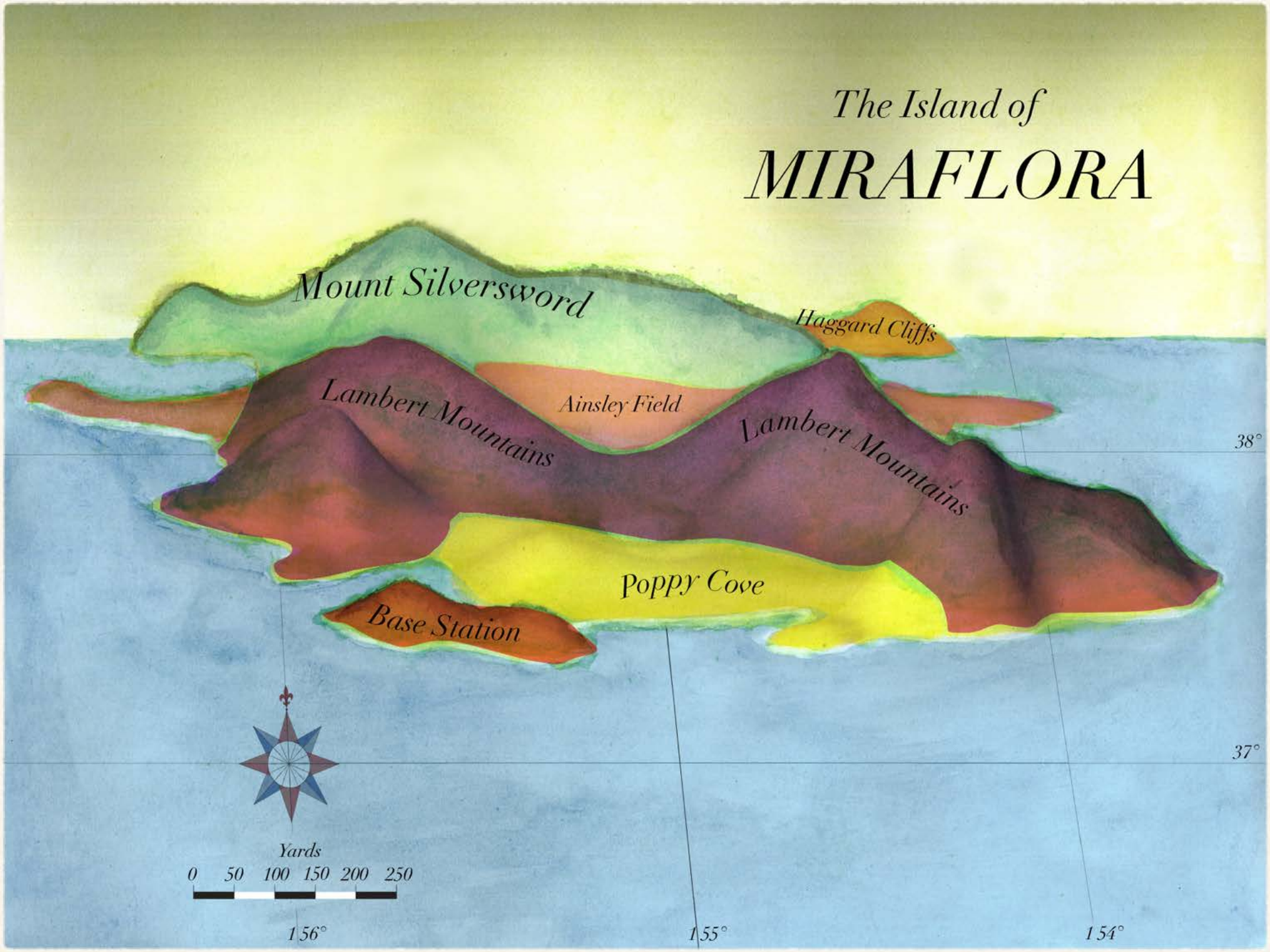
- One more wrinkle!
- We have no way of knowing how many variables or bonus tiles are POSSIBLE to use for any given problem
- Strategy: Compare player progress to other players in order to score

The difficulty of difficulty

- ***Science: There is no established way to know how difficult any given problem is (without just playing a problem and/or guessing)***
- ***Game Design: Game should have a nice difficulty curve -- challenges become gradually harder as players gain competence***

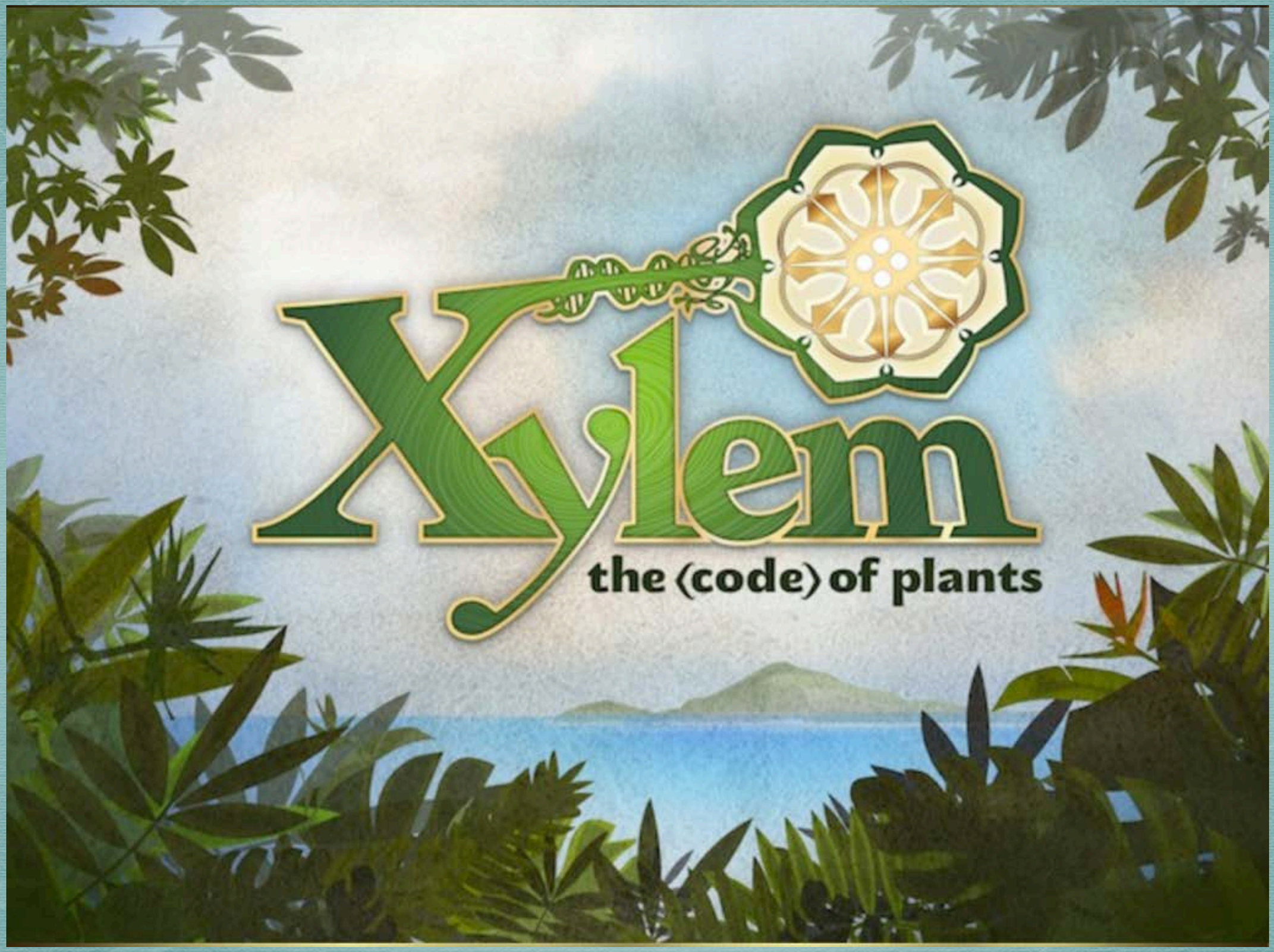
Difficulty Level

- With no systematic way to determine difficulty, we derived heuristics based on play testing
  - Integer Problems are less difficult than Array Problems
  - More variables = harder
  - Larger numbers = harder


The Island of
MIRAFLORA
Mount Silversword
Haggard Cliffs
Ainsley Field
Lambert Mountains
Lambert Mountains
Poppy Cove
Base Station
38°
37°
Yards
0 50 100 150 200 250
156°
155°
154°

Outcomes

- 1750 Downloads
- Close to 8000 solutions have been submitted by players
- Audience that has been attracted to game is more advanced math-savvy puzzle audience
  - Despite best efforts, bridging less mathy players has proven very challenging
- Retention for game is not where we'd like it to be
- Audience we have doesn't want features we designed for other people



www.verigames.com – Available for iPad in Apple App Store