User Feedback and Prototyping Design in Escape Game Puzzles

Rickard C. Stureborg
College of Engineering, EECE
Northeastern University
Boston, USA
stureborg.r@husky.neu.edu

***Abstract*— The Koira's "Escaping Time" escape game is the product of iterations of prototypes and user testing. Surveys were designed to gather information on specific changes in prototypes, which were created on the philosophy of rapid prototyping. The target market, undergraduate STEM students at Northeastern, was considered and tailored to in every iteration of the design process. Koira defined success with this target market as being four-part: the escape game needed to actively engage users, have clever and understandable solutions, be fun, and take approximately an hour to solve. The first prototype of the game included 6 puzzles made from mostly paper and tape. Based on the idea of rapid prototyping, tests were conducted quickly and often to redesign these puzzles until their quality aligned with what the team had defined for success. Although quantitative testing was performed in class, it was mainly the qualitative testing which informed design decisions. Three puzzles, chosen based on their clear iterations of prototypes and testing, were discussed. The Pythagoras Puzzle, Timeline Reparation Machine, and Missile Component Puzzles are arguably the three strongest puzzles of the escape game. This paper also briefly covers the impact of testing on the puzzle progression and codex, that is the method in which users are guided through the escape game. The in-class testing revealed faults in the lack of instructions and separation of puzzles, which were addressed between testing. However, the escape game as a whole did not go through rapid prototyping methods.**

Keywords—Escape Game, Intended User, Potential User, Qualitative and Quantitative Testing, Rapid Prototyping, User Feedback, User Experience, Codex.

I. Introduction

An escape room board game is meant to capture the essence of an entire escape room in a single, small box of puzzles. This style of board game is different from many others in that the objectives and "pathway" of the game are not evident until the last puzzle has been solved. To be successful, these board games have to strike a good balance between being difficulty and intuitiveness, all while engaging users through a strong theme and interesting plotline.

The team conducted preliminary research by testing well-established escape room board games to build a framework in which the goals and objectives for the "Escaping Time" game could be developed. The board games that the team tested were from the same creator under a series named "Exit: The Game." The first game had an ancient egyptian theme and was called 'The Pharaoh's Tomb,' and the second had a cabin in the woods theme and was called 'The Abandoned Cabin.' The team also watched several videos on escape rooms in order to establish a well-defined intuition for how these puzzles usually operate.

Since the success of the escape game puzzles is entirely based on the users' experiences, Koira decided that it was important to design the game using user feedback. Small changes to puzzle designs were to be evaluated by users as often as possible, thereby aligning the design process with the wants of the intended users. The team aimed to utilize rapid prototyping methods as often as possible to maximize the design cycle iterations.

Initial meetings with the Koira team were focused on establishing a theme which would enhance the user experience of the escape game. Time travel was identified as a theme which fit well with puzzle games because of its natural relationships to the type of higher-level thinking required in puzzle solving. Time travel, when presented in film, is often the basis for complex storylines and interesting plot-twists. Research was also done on the Grandfather paradox, to inspire the plotline of the escape game in a unique and interesting way. Although large parts of the grandfather paradox plotline were dropped in the final version, this lead to other ideas for puzzles.

II. Target Market and 'Success'

This section will address the factors considered in defining intended users and identifying potential users of the escape game, as well as the analysis for what defines success of the escape game design.

A. Defining a Target Market

The established theme of time travel guided the decision of intended users for the escape game. Since time travel is often presented in science-fiction movies and novels, we assumed the audience best suited for our escape game would also share areas of interest in Science and Technology. As a team of engineers, it was easy to imagine students in STEM majors sharing these interests. To ensure that the team had a strong intuition for the intended users' tendencies and interests, we decided to limit the target market to undergraduate students at Northeastern. Thus, undergraduate students at Northeastern with STEM majors were chosen as the intended users of the final escape game.

Potential users, those which the escape game was not explicitly designed for but whom are likely to interact with it due to proximity and exposure, included all first-year and second-year students at Northeastern. Although not ideal for user-based design, feedback from potential users of the game still gave insight into the likely user experience of intended users. As the puzzles were in the design cycle, the easiest users available for testing were often roommates and friends, who were not always STEM majors. At points, issues arose due to this methodology, such as during testing of the "Schematics of a Missile Component" puzzle, due to its solution being based on technical skills learned in engineering classes. This is discussed further in section III. D.

B. Defining Success

The other framing definition the team had to make early in the process was to establish a yard-stick for 'success.' Not only did this create a frame for evaluating the success of final products but it helped guide design decisions and objectives of testing. Every prototyping iteration was influenced by this yard-stick as the changes aimed to come closer to meeting the requirements of success. Through meetings with the team, these were the four criteria of a 'successful' game:

1) Even if the solution does not become apparent immediately, *the user should begin getting ideas for how to solve each puzzle almost immediately.* The user should at no point in gameplay sit still for longer than 2 minutes with no idea of how to approach the game.

2) Solutions should seem obviously correct once reached. When a user reaches a solution they should have an 'Aha!' moment, where they know that the solution is correct. If this is true, it is unlikely that they feel there could be a better solution, or that a solution they had assumed existed earlier made more sense.

3) The entire game should take just around an hour to complete with four to five users. This let's the game be accessible to most of the defined target market. Our target market is unlikely to be able to spend much longer than one hour on a game when they are in classes.

4) Upon completion, users should identify the game as being "fun." For the purpose of defining success the exact factors that go into making the game fun are not important. A simple yes or no suffices for this criteria.

III. Rapid Prototyping

This section will address the issues and feedback identified through the quantitative and qualitative testing of three individual puzzles from the Escape Game. For each puzzle, the design changes inspired by these tests are broken down by the number of the prototype iteration.

A. Pythagoras - Beating Physics

The pythagoras puzzle, inspired by the digital trick known as the "Chocolate bar illusion," aimed to make the user believe they had broken a law of physics. The user is given a book with a triangle-shaped frame. In the frame lay 18 puzzle pieces, perfectly lined up to the edges and filling all area. The user is then given an additional square and told to insert this square into the already filled area, without having any puzzle pieces overlap. Although seemingly physically impossible, the solution yields a three digit code which becomes apparent once solved.

Since the focus of early design changes was to make the puzzle pieces fit together better, user testing was not introduced until the third prototype. Earlier testing was conducted by the engineer that designed the puzzle.

1) The first prototype was created with graphing paper cut with an exacto knife, the puzzle pieces were only squares and triangles. The solution *with* the additional piece had been drawn on the graphing paper, and each piece was the cut from there. A frame was drawn on a white letter-sized sheet of paper, where the puzzle could be arranged on top of. The first test was then done on this prototype, yielding the following feedback:

- There is a large gap between the two rightmost pieces in the solution without the additional piece
- The solution without the additional piece seemed too intuitive, although since the engineer knew the solution this was largely dismissed.
- Moving the pieces around was difficult.

2) The second iteration aimed to solve the gap in the two rightmost pieces by first drawing out the solution *without* the additional piece and then cutting each piece out. The same frame and graph paper was used. Testing of this prototype lead to the following feedback:

- Moving the pieces around was difficult without pieces folding over.
- In the solution with the additional piece, a triangle piece stuck out about 1cm from the right edge.

3) The third iteration was created directly from the second. The corners of some triangle pieces were trimmed and two squares were recut to be slightly smaller. A test was conducted with a first-year computer science major. Testing results:

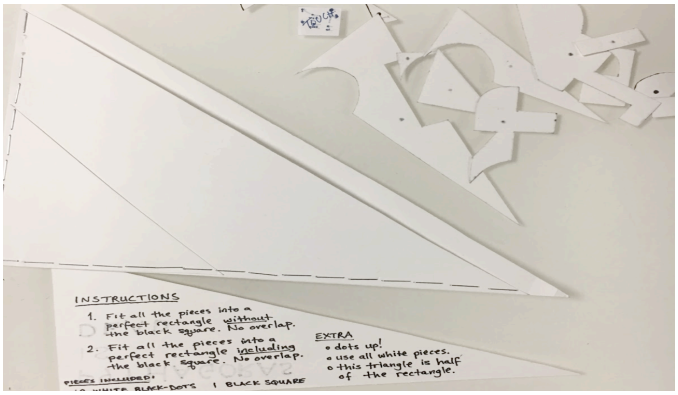
- The objective of the puzzle was not intuitive and the user asked for directions twice.
- The user had difficulty moving the pieces around
- The solution was reached within 2 minutes of directions being given.
- Upon reaching the second solution, the user laughed and asked how the puzzle had been made.
- When asked if the puzzle was fun, the user remarked that although it was fun, it was too easy.

At this point, three of the four characteristics of a succesful puzzle had been met. However, there were still issues. The third component of success had still not been reached: the entire game should take just around an hour to complete. Since the escape game was to have six puzzles, each puzzle should take about 10 minutes. Further, the instructions were not clear, and the pieces were difficult to move around.

4) All these issues were addressed in the fourth prototype. This prototype (Figure 1) was made of thick construction paper from the architecture studio. Furthermore, shapes such as circles and complicated boxes were included as puzzle pieces to complicate the solution. The frame was an exact copy of the triangle the solution would construct. Putting a given solution up to the edge of this triangle thereby made a right rectangle. Instructions were written on this frame. This was then tested on a first-year architecture major, which although not an intended user maintained many of the defining criteria of one. Results:

- No instructions were needed beyond a brief introduction to the puzzle, the user began moving around the puzzle pieces immediately
- The puzzle was solved in approximately 12 minutes
- The user identified the puzzle as very fun
- The user was impressed by the solution and design

FIGURE 1: FOURTH PROTOTYPE OF PYTHAGORAS PUZZLE



This version of the puzzle was presented during both in-class testing sessions. It was deemed successful based on these tests, which were conducted on 24 first-year engineering students at Northeastern. However, feedback on the escape game did reveal issues that could be addressed with a fifth prototype. The results relevant to the Pythagoras puzzle of these two in-class testing sessions are summarized below:

- Two out of three groups said the Pythagoras puzzle was their favorite during the first session
- "Triangle puzzle is very time consuming"
- Average score of 3.3/5 for a Likert-Scale question on "How well put-together was the puzzle aesthetically?"

The testing session in class lasted for only 30 minutes per group, sometimes less depending on the time it took to set up. Therefore, the feedback of the triangle puzzle being very time consuming was deemed irrelevant, as within a 60 minute timeframe there is plenty of time to solve this puzzle. Additionally, this team only had 3 members, as opposed to all the other teams who had 4.

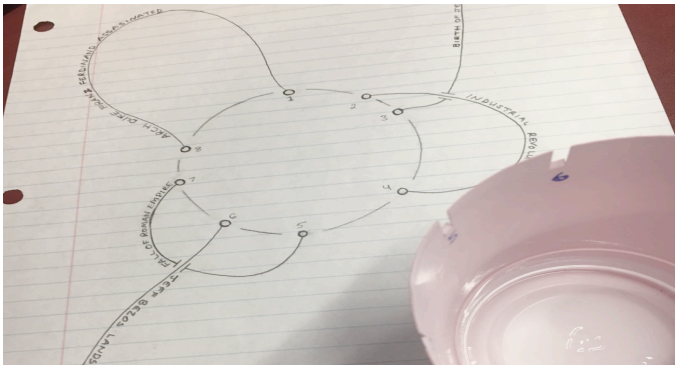
5) Koira did, however, decide to address the aesthetics of the escape game. Aesthetics was an area not considered when defining success, but plays into the perceived quality of the escape game as a whole by the users. Therefore, the fifth prototype was made of two layers of construction paper and thick brown duct tape, giving it a finished look. The frame was set in a book which opened to reveal the puzzle. Instructions were then written on the inside of this book.

B. The Timeline Reparation Machine

This puzzle aimed to go along with the theme of time travel by allowing the player to reconstruct a destroyed timeline in the game. When the timeline is constructed correctly, the user is prompted with the flag of the country which landed on the moon first in the timeline they are "currently in." This flag can be decoded with a page from the book to get a three digit code.

1) Since this puzzle was created before the team had any knowledge of arduino or simple wiring, a representative prototype was created for the first round of testing. This puzzle was created using a solo-cup and string, with the idea being that when the string was arranged correctly on top of the solo-cup a team member would give the user the correct flag. This prototype can be seen in Figure 3.

FIGURE 3: REPRESENTATIVE PROTOTYPE OF THE TIMELINE REPARATION MACHINE



Testing of this prototype on twelve first-year engineering students (three groups of four) revealed several issues:

- The instructions for this puzzle were not intuitive and users did not interact with the puzzle for a long time
- When the solution was finally reached, they did not understand the point of it
- The users were frustrated by this puzzle, some did not solve it
- Many users did not realize the lines on the paper represented a timeline

2) The second iteration was done after the team had been familiarized with arduino and breadboarding. This prototype was a cardboard box with holes where wires came through. The RedBoard was powered using a computer. The objective was to attach each wire to the correct event which followed (e.g. Obama became president after the first nuclear nuclear bomb was dropped). When all wires were connected correctly, an LED lit up indicating which flag had been selected (on another page in the book). This was then again tested on three groups of four first-year engineering students:

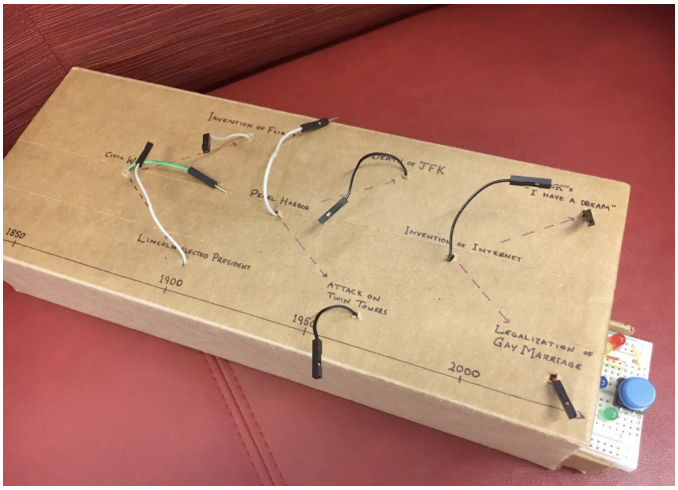
- The two pages in the book were often overlooked, and therefore no three digit code was ever reached.
- One group discovered that every LED could be lit up by connecting the wires in a certain way
- One group got no wires to light
- One group figured out the solution by examining the wiring of the breadboard rather than the timeline
- The power cable came undone multiple times during testing
- Users who did interact with the timeline said the puzzle was fun

3) The last iteration of this puzzle aimed to solve each of these issues. It was again made of cardboard, this time covered in a strong brown tape which gave it a finished look. The team designed the wiring in such a way that only the LED's were visible. The timeline was clearly labeled with years to prompt the user to organize the events by date, and the Redboard was powered using a 9V battery. An image of the final version of the puzzle can be seen in Figure 4. Finally, this version was tested on a first-year engineering student, and a group of mixed major first-year students.

- It took approximately 15 minutes for the individual to solve this puzzle, and 10 minutes for the group
- When the solution was reached, all users identified the three letter code correctly and were confident it was the right answer
- All users identified this puzzle as fun
- The users interacted with the components of this puzzle immediately.

Based on this testing, the final puzzle was defined successful.

FIGURE 4: FINAL PROTOTYPE OF THE TIMELINE REPARATION MACHINE



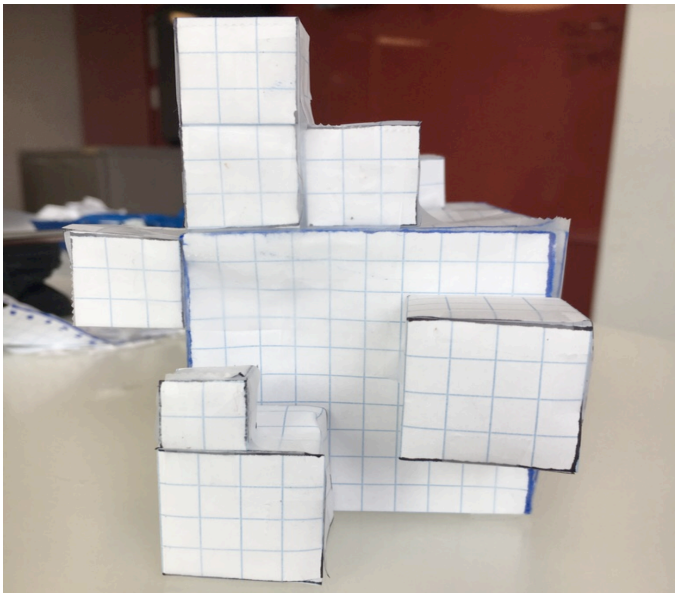
C. *Schematics of a Missile Component*

The Missile Component puzzle requires 3D visualization skill and some knowledge of orthographic projections. The aim is to line up the three correct views of a given object, which are printed on plastic sheets, to get a three digit code.

1) The first prototype was an object made of paper and tape (Figure 5). It was created by making several boxes of different sizes and taping these together. The aligned orthographic projections were drawn onto cut-out pieces of Ziplock bags. This puzzle was tested an a very large amount of intended users, sourced mainly in the common room of the fourth floor of East Village at Northeastern. The resulting feedback is summarized below:

- Some users found the correct orthographic projections but put the plastic upside down
- Almost all users said the puzzle was fun
- All users solved the puzzle within about 5-15 minutes
- Many users said the plastic was difficult to sift through because it was too thin.
- One user suggested making more projections so that you could vary the difficulty
- The poor build quality of the object was mentioned many times
- There was no confusion regarding the instructions and users began interacting with it immediately

FIGURE 5: OBJECT FOR MISSILE COMPONENT PUZZLE



2) This puzzle only took two prototype iterations to reach a final product. The final object was 3D printed in plastic, and the orthographic projections were printed on overhead projector plastic film. The printing was done is such a way which made clear what side was up and what was down. Two crosses were also printed on each page to indicate how the pages should be lined up. This version can be seen in Figures 6 and 7. Since the previous prototype had already been deemed successful based on the four criteria, changes were mainly addressing cosmetic changes necessary for the look and feel of the escape game. No further testing of this puzzle was done until the final version of the escape game was put together.

FIGURE 6: 3D-PRINTED OBJECT FOR MISSILE PUZZLE

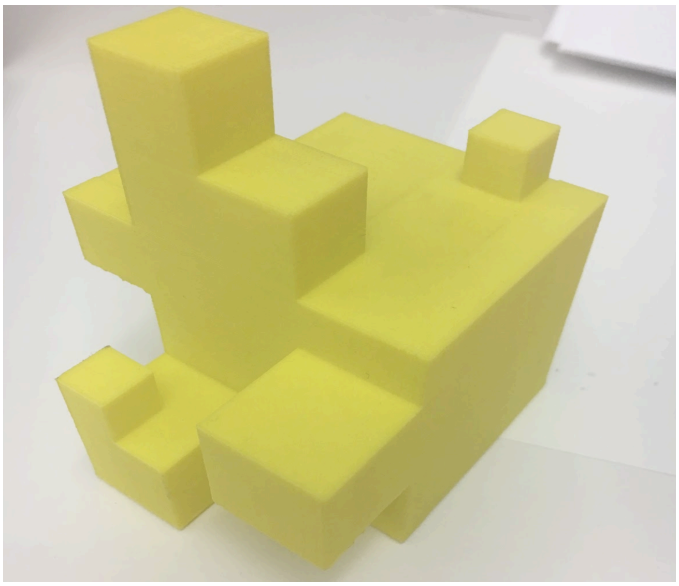
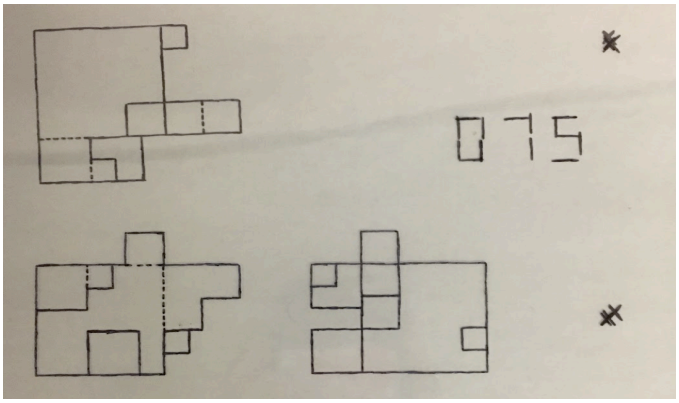


FIGURE 7: ALIGNED ORTHOGRAPHIC PROJECTIONS



It was the frequent testing of small design changes which enabled the team to go through this rapid protyping method with most puzzles.

IV. CODEX AND THEMATIC DECISIONS

This section breaks down the impact of testing on the puzzle progression and playability of the escape game as a whole. Decisions regarding instructions and theme were derived from this methodology.

A. Puzzle Progression and Codex

Many criticisms of the in-class testing had to do with the Codex. This only became apparent after the first set of data, collected using standard surveys provided by the course instructor. The data that alerted us to this being an issue is sampled below:

- "unsure of the theme"
- "it had lots of parts and it wasn't clear where to start or how to progress"
- "The individual puzzles were interesting and fun, but we couldn't figure out what the point was or what we were trying to achieve"
- "There was no clear instructions or backstory and no obvious connections between puzzles"
- "It wasn't clear where to start"

See complete survey data in Appendix B, Table 1.

Based on this data, multiple design decisions were made. First, a codex was developed where puzzle solutions yielded three digit numbers. These numbers in turn give a puzzle card which tells you to pick another card. This mechanism acted as a two-part verification to avoid accidental solutions based on three digit codes that yield the same puzzle card number. Further, instructions were written for every puzzle, and these were put in envelopes in an effort to hide information the user doesn't need yet - and therefore reduce any possible overflow of information that might occur. A plotline was written complete with factual information from history to add a feeling of authenticity.

Surveys, aimed to evaluate the success of these changes, were prepared for the second round of testing. To specifically address the changes in codex design, the following questions were asked:

1) What could use clarification in the rules/instructions?

2) Do you think the order of the puzzles makes sense? Why or why not? Is there anything you would change?

3) Please rate how easy the 'rules' of this game were to comprehend (1-5, 5 being very easy).

4) What was the worst part of the game?

Based on the second round of testing, the escape game had yet to meet the criteria for success in regard to the codex. The relevant answers in response to this specifically designed survey was that although the order of the puzzles was logical the decoder's symbols weren't intuitive and there was no way to get help if you were stuck. Specifically, the data said:

- "Which symbols do we count?"
- "which symbol goes with which puzzle?"
- "No hint cards!"
- "Lack of hints"
- "Consider creating hint cards"
- When rating how easy the 'rules' of the game were to comprehend, the scores were 3,1,3 out of 5. An average of 2.33.

Since this feedback was repeated several times, the final design addressed just this. Three hint cards were created for every puzzle in the escape game. The first hint let the users know what the object of the puzzle was, the second let them know how they should approach solving the puzzle, and the last showed them the solution. Secondly, the symbols on the decoder were redesigned to be clearer. The decoder was then cut using a laser cutter, and therefore the symbols of the decoder became very clear. The combination of mechanisms which guide the user through the game was thereby entirely created based on user feedback. User feedback was the strongest asset of the team's methods in reaching success.

B. Theme and Story

Decisions on the thematic choices were largely made before the design process even began, and often framed decisions and inspirations for designs. However, in Koira's continued effort to lean of user input and testing as heavily as possible, surveys addressed these issues directly as well. From the first standard survey given, it was discovered that one group was "unsure of the theme" but complimented the "interesting colors and [how the puzzles] looked well made" and some thought the storyline was boring or hard to follow.

The next iteration, as expected, was designed to address these issues. In Appendix A.2 to A.7 are the pages regarding theme and storyline which were planted in several envelopes in the final version of the game. However, since the users who were testing the game knew the Koira team personally, we worried that written feedback would be biased to only include nice comments. Therefore, the second survey relied more heavily on Likert scale questions where the users could simply circle an answer rather than write it down. The teams logic was that this would remove the tester's voice from the feedback and therefore it may be easier to give unbiased feedback. The following data is that which we collected in three different areas:

Question Group:	A	B	C
How fun was the game?	4	4	3
How nice did the game look?	3	4	3
How difficult was the game?	3	4	4

This data then informed the decision to create a coherent look for every puzzle to tie the entire aesthetics together. To tie the aesthethics together the team bought robust brown tape, and taped the outsides of all puzzles with it. Not only did this give each puzzle the same (neutral) visual color, but it also improved the texture of the puzzles. This proved to be a cost and time effective method for creating coherent color and texture, as opposed to the previous version which ranged from light brown plywood to bright purple paper. Furthermore, the colors in the book were amended to be slightly darker and brownish in tones.

ACKNOWLEDGMENTS

I would like to sincerely thank my group members for all their hard work this semester in not only finishing this escape room, but for maintaining a commitment to excellence in regard to work ethic and work quality. Individually, I thank Justin Sargunas for the countless time spent together working on individual puzzles in between official Koira meetings. I would also like to extend a warm thank you to all those who participated in user testing of individual puzzles and the escape game as a whole. In particular: Elias Hirschfeld, Joshua Solomon-Culp, Mark Lewis, and Reed Kellet. Thank you.