Virtual Agents

HCI Capstone 2016
Mid-Fidelity Report

March 27, 2016

Su Baykal
Researcher
sab@andrew.cmu.edu

Ben Boesel
Designer
benboesel@gmail.com

Ryan Donegan
Team Lead
rdonegan@andrew.cmu.edu

Sam Gao
Designer
samanthg@andrew.cmu.edu

Ian Go
User Test Coordinator
iigo@andrew.cmu.edu

Contents

Introduction

The Articulab at Carnegie Mellon is working to improve the understanding of education and children, as well as how technology can be implemented to augment learning spaces. The SCIPR project - a multi-year endeavor headed by Justine Cassell - is working to bridge the gap between virtual agents, education, and the physical world. By investigating how virtual agents can extend beyond their digital confines to affect their environment in tangible ways, SCIPR seeks to understand how collaboration and curiosity in learning can be improved.

Our HCI Capstone project specifically focuses on building a foundation for this physical intervention to occur. Thus, we seek to understand more about the educational space as well as how to spur curiosity and engagement in learning. Through background research, interviews, and user testing, we have explored creative and practical ways to teach powerful educational concepts through technology.

Our early low-fidelity prototypes broadly explored strategies for enabling collaboration through physics-based activities. Now, in our mid-fidelity testing, we aim to focus more exclusively on embodiment. Namely, how does the embodiment and role of a virtual agent impact curiosity and collaboration? This report details our efforts, testing performed, and the results we derived.

Low-Fi Summary

In our low-fidelity testing, we tested two unique activities that made use of both 2D and 3D game interfaces: a digital physics game and a 3D modeling activity.

We experimented with embodiment in two ways across each experiment: the virtual agent on a screen, and virtual agent with partner, BeepBoop.

In all, we tested four participants across three conditions: Digital Physics, BeepBoop, and Digital Physics with BeepBoop. All conditions were simulated using a screencast of a team member portraying the virtual agent, and another team member moving a block as BeepBoop when necessary.

In general, we found that BeepBoop brought interactions with an agent to the table, and was overall likable, and capable of invoking curiosity. However, interviews revealed that BeepBoop lacked a distinct function within the playspace. Moreover, the success of BeepBoop in attracting attention may also have been attributed to its novelty.

In terms of representing the virtual agent in each scenario, we found that the agent was particularly involved in the digital physics activities. The agent's utility could be due to the greater difficulty in the game, requiring increased reliance on collaboration between players and agents.

However, the agent's helpfulness might have extended too far, as some subjects reported the agent acted almost as an advisor or aid. Rather than a symmetrical partner, the agent was viewed almost as an instructor due to its assumed knowledge of the game.

Mid-Fi Testing Overview

The goal of mid-fidelity testing is to expand upon the results and analysis derived from low-fidelity testing. The research gained from this stage, we hope, will provide information on how to further refine and focus our prototyping concept. Whereas our low-fidelity testing phase featured a broader investigation of game types and interactions, our aim is to identify the strategies that are most successful in inspiring collaboration and curiosity with participants.

The focus of mid-fidelity testing is embodiment: how can we better represent a virtual agent in this physical world? For this phase of experimentation, we'll test three specific embodiment types:

1. Virtual agent on the screen
2. Virtual agent and proxy
3. Virtual agent as the proxy

Adaptations from Low-Fidelity

While our first phase of testing was more general, looking at overall how individuals interacted with virtual agents in play situations, the goal now is to investigate physicality. Specifically, we aim to understand how the way a virtual agent actuates its environment is interpreted by humans.

In the "virtual agent on screen" scenario, we test how participants to react to an agent that is able to move physical objects seemingly by will without physically reaching out. This could be realized in the future by use of magnets underneath a game table, used to move pieces discreetly in accord with a task.

The "virtual agent and proxy" and "virtual agent as proxy" utilize similar physical mechanisms. In both scenarios, objects in the gamespace are manipulated by the agent by way of a mobilized device (e.g. a tractor, car, pet, etc.). The device can push items, thereby contributing to gameplay

or making suggestions to players. The difference is that in "virtual agent as proxy", the agent is considered a part of whatever is causing this manipulation. In "virtual agent and proxy", the agent is seemingly in control of a third party proxy, though not necessarily a part of it.

To further emphasize the physicality of our research, we've adapted the gameplay for each scenario to be largely the same. Unlike in low-fidelity testing in which we tested a variety of games/activities with participants, we plan to test the same game for each scenario. The game itself is a simplistic physics-based activity in which players work collaboratively to move a ball from point A to point B on a table. Unlike our low-fidelity testing, all pieces and play are physical, rather than the digital gameplay we tested prior.

Participants

To emphasize the importance of collaboration, all tests were performed on at least two participants at once. This forced participants to work together, as well as with the simulated virtual agent.

We aimed to test each of the three scenarios twice, with at least two participants performing each trial. Thus, a total of 6 trials were run, with 12 unique participants.

Recording Results

Each pair of participants was tested on one condition through all the levels of gameplay designed. After the final round of gameplay, a brief interview was given. Later, a survey judging engagement was administered.

Building the Prototype

The physical nature of our project required some variations in how we developed our prototypes from low to mid-fidelity. For this phase, we designed and created custom foam blocks and pieces specifically catered to the physics gameplay we designed (e.g. ramps, towers, adjustable panels). Pieces were designed to allow for a greater degree of flexibility in exploration which may have been limited in previous explorations where we simulated some actions.

The medium for the game itself is tabletop surface. We used a sheet of acrylic and a sheet of paper under it where we drew all necessary elements such as starting point and ending point.

The Game: Marble Run

We intended our game - dubbed Marble Run - to be both straightforward and exploratory. The game consisted of a tabletop board marked with start and end locations, a marble, and foam pieces of various heights and types. Players were instructed to use the pieces to build a path to move the marble from its starting perch to one of three endpoints (each endpoint varied with the game level, and was indicated at level onset by a teammember).

Marble Run takes advantage of physics in 2.5 dimensions. That is, players are able to pick up and move blocks on the board, but pieces cannot be stacked. This was done for the sake of simplicity, and to coincide with more advanced sensing technologies.

Despite this minor restriction, Marble Run encourages exploration of physics in a friendly environment. The pieces themselves vary in form from paths, to ramps, to corner turns. This affords the creation of innumerable paths and trajectories as players explore momentum and velocity of the marble as they try to gauge its path. The game could also easily be made more difficult or involved with the inclusion of new pieces (e.g. those representing simple machines), requiring the agent to move specific objects, or introducing new levels of play.

Procedure

Pre-Experiment

Paper and acrylic board placed on top of a table to represent the gamespace. The paper was labeled with a 'starting' zone and three other zones to represent the levels of the game. For each round of testing, the starting piece was preset on the board, along with the endpiece in the corresponding zone (starting with 1, then progressing up through 3 as each experiment was completed). Other pieces were laid off to the side of the board.

We mounted a camera on a tripod to record each session of testing.

During Experiment

Participants were seated at table facing each other. The virtual agent - which we referred to as "Alex" - was positioned at one end of the table on a screen, or on the table when testing the agent as proxy.

Instructions outlining the rules of Marble Run were described to the players, and the role of the virtual agent was described (see Appendix for script). Players were instructed to collaborate with each other and with the agent to solve each level of the game.

Players were then free to attempt playing the game as they saw fit. After each level was completed, a member of our team would reset the board and instruct the players to try to solve the next level.

If any level took longer than 10 minutes to solve, we planned to instantly advance the players to the next level for the sake of time. This case never arose during testing.

After Experiment

Following the experiment, we administered a survey to each participant (see Appendix) and conducted a brief interview.

Condition 1: Virtual Agent on Screen

Setup

A laptop was placed at the end of the table surrounding the gamespace. Initially, one team member portraying the agent sat behind a wall, away from the game, and communicated with players through Google hangouts as displayed on the laptop.

A coordinating team member acted as the agent's influence on the game, and moved pieces on the gameboard.

Roles

Virtual Agent: communicated verbally and visually (on monitor) with players.

Actuator: played with test subjects, coordinated with agent's influence on the game.

Gamerunner/Interviewer

Players interacted with the virtual agent, displayed on a monitor. Another teammate moved blocks to represent "the force", or the actuation of the agent.

Condition 2: Virtual Agent and Proxy

Setup

The setup for the virtual agent and proxy scenario was similar to the virtual agent on screen. However, for this scenario, instead of moving pieces by hand, the agent's actuator pushed an object meant to represent the proxy. We named the piece BeepBoop and gave it a face to suggest agency. The actuator communicated with the agent on screen by making 'beep boop' sounds.

Roles

Virtual Agent: on screen, can communicated with people and its helper, BeepBoop.

BeepBoop: controlled by a teammember. BeepBoop can push blocks around the gamespace to draw users' attention and assist in gameplay.

Gamerunner/Interviewer

The virtual remains on the monitor while another teammate pushes proxy, BeepBoop, across the table.

Condition 3: Virtual Agent as Proxy

Setup

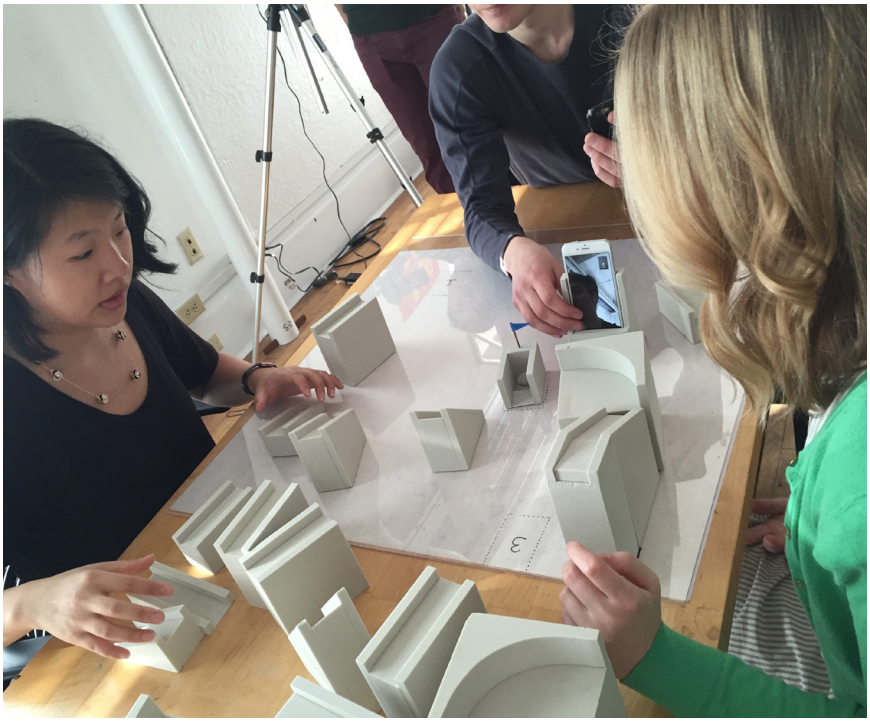
In this scenario, agent and proxy were played by the same teammember. For setup, we placed a mobile phone on a wheeled platform that could be pushed around the gamespace, similar to BeepBoop. During each test, the team member portraying the agent would Facetime with the phone so that their face appeared in place of the proxy. The person playing the agent would also push the phone apparatus around the gamespace simultaneously to simulate actuation.

Roles

Virtual Agent: visualized on phone screen and given the ability to move objects around the space by way of pushing.

Gamerunner/Interviewer

Players interacted with the virtual agent, who was displayed on a mobile phone pushed across the board.



Observations

Condition 1: Agent on Screen

1. Players found Marble Run activity to be enjoyable and collaborative.

2. The WoZ functionality split between two different teammates was difficult to achieve.

3. Players saw the virtual agent more as an "Assistant".

4. The virtual agent's helpfulness in playing the game and offering advice was limited.

Condition 2: Agent with BeepBoop Proxy

1. Some players found BeepBoop to be an enjoyable, fun presence in the game.

2. There was a clear disconnect as to the role of BeepBoop and its relationship with the agent (e.g. "Is BeepBoop helping independent of Alex?").

3. Players were unsure what the purpose was of having both the virtual agent and BeepBoop present.

Condition 3: Agent as a Robot

1. Players tended to work in pairs, ignoring the virtual Agent.

2. It was difficult for the robot to move across the table while maneuvering around blocks and pieces.

3. The agent's ability to draw attention and actuate objects was very limited.

Post-Trial Analysis

Method

Each participant was asked to take a survey after the activity. The survey asked about the following:

- perceived agency of the virtual agent
- feelings of collaboration
- curiosity
- engagement levels

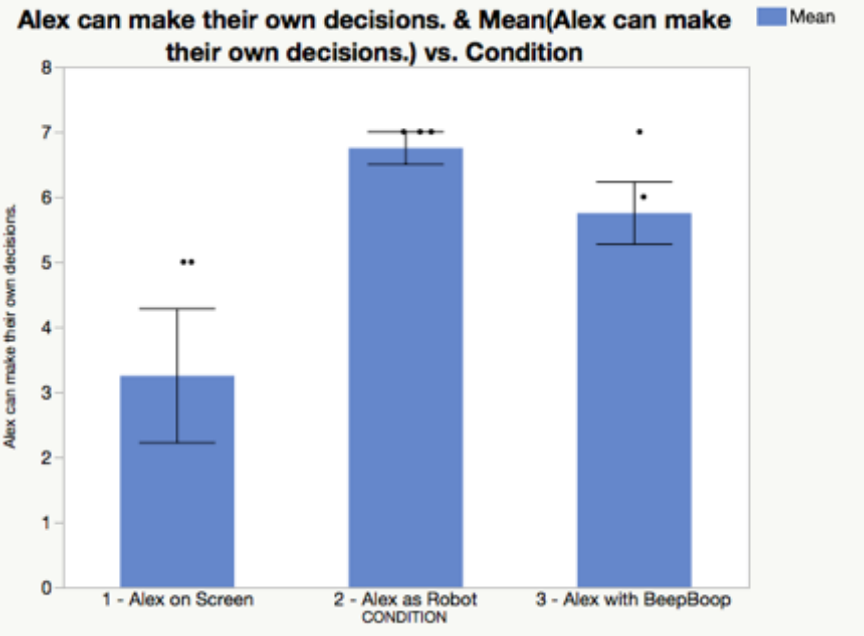
We also interviewed participants in pairs to probe them on how collaborative the activity felt and the role of Alex in activities.

Final Results

Based on the survey, we found a statistically significant difference between conditions for perceived agency of Alex ($p = 0.0136$).

Results of survey question "Alex Can Make Their Own Decisions." See Appendix for full list of survey questions.

Condition	Agent Can Make Their Own Decisions
Agent on Screen	3.25
Agent with BeepBoop	6.75
Agent as Robot	5.75



There were no statistically significant differences between conditions for collaboration, curiosity, or engagement.

Interviews revealed that the virtual agent functioned more as a *mediator* and *not a collaborator*. Moreover, players noted that the virtual agent's presence and its effects on the physical world were both important.

Players attributed this to several common issues regarding their interaction with the virtual agent over the course of the experiment. Their chief concerns are summarized as follows:

Alex did not usefully participate enough

"It was mostly us [the players] occasionally dropping in to make points."

There was a disconnect between virtual presence and effects on the virtual world

"This person was making decisions but for some reason, the activity was coming from here...there was a disconnect."

"It's different when the physical presence isn't there because we were dealing with something physical….Alex couldn't really pick up a piece."

Takeaways

Perspective is Important

Perspective was shown to be critical in establishing the relationship between players and virtual agent. In almost all tasks, we observed the virtual agent as mostly dismissed by the players. When we interviewed players about this after each trial, they responded that the agent seemed disconnected due to its lack of physical presence and mostly-vocal impact.

The lack of connection players felt with the agent was likely the result of the design of our WoZ study. Players exhibited this by asking questions directly to the human playing as the virtual agent when visible, and not the screen itself. Before we began one trial, a participant asked if he should move the agent's monitor so they would be more at eye-level with the players - reinforcing the importance of perspective.

One player also commented on how it was difficult to interact with the agent because they were unsure of what the agent could actually see. During gameplay, the player held up a piece to the agent's screen to display it, though later admitted he wasn't sure if that would do any good.

Players Move Fast

This was observed mainly from the perspective of teammates simulating actuation. In the conditions involving a proxy or the agent on the table, it was difficult to keep up with human players. Whereas humans could adeptly reach across a table and pick up one or multiple pieces at a time, a proxy takes considerably more time to navigate a table in order to nudge a block.

Players commented on the speed deficit in interviews. Many noted the difficulty the agent or BeepBoop had in keeping up with them, thus limiting their impact on the game.

The Virtual Agent Needs To Be Involved

In interviews, some participants noted that the virtual agent really wasn't necessary in order to complete the scenario. One player described Marble Run as being potentially for individual gameplay, thus excluding the necessity of the virtual agent. Most players felt Marble Run was collaborative between the two human players, but the role of the virtual agent wasn't critical.

Some players suggested making the task more difficult, or more reliant upon the virtual agent in order to improve collaboration. More restrictions on the rules of the games and constraints involving gameplay were cited as other potential ways to improve the virtual agent's utility.

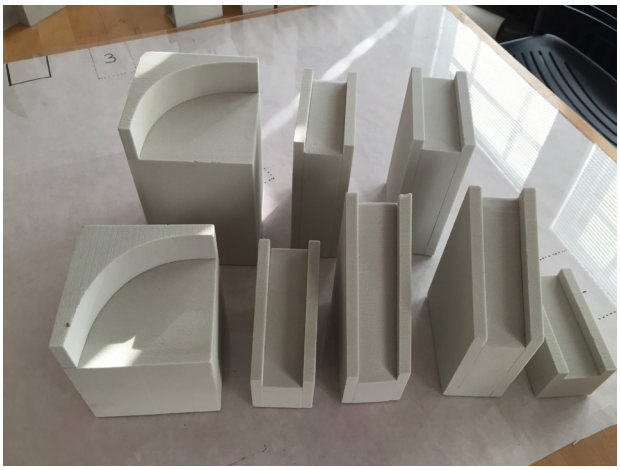
Next Steps

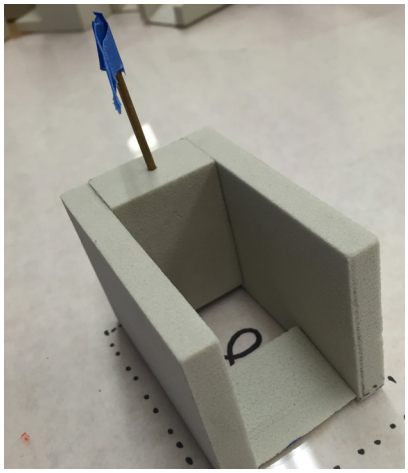
A clear obstacle in our most recent round of testing was embodying agents believably. While we utilized a Wizard of Oz approach, it was difficult for participants to suspend disbelief enough to accept that a virtual agent was controlling the gamespace, and not a human. This flaw in testing may have affected results, and reduced the impact of having a virtual agent interact in game.

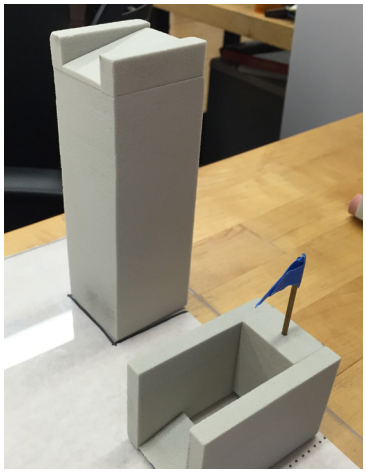
In our next steps, we hope to build and test a more believable Wizard of Oz test. We plan to investigate and implement a form of interaction that doesn't involve visible human interaction. This will allow us to test more reliably. Additionally, we also hope to implement a pseudo-sensing system that makes it easier for the tester playing as the virtual agent to comment on and express their presence in activity.

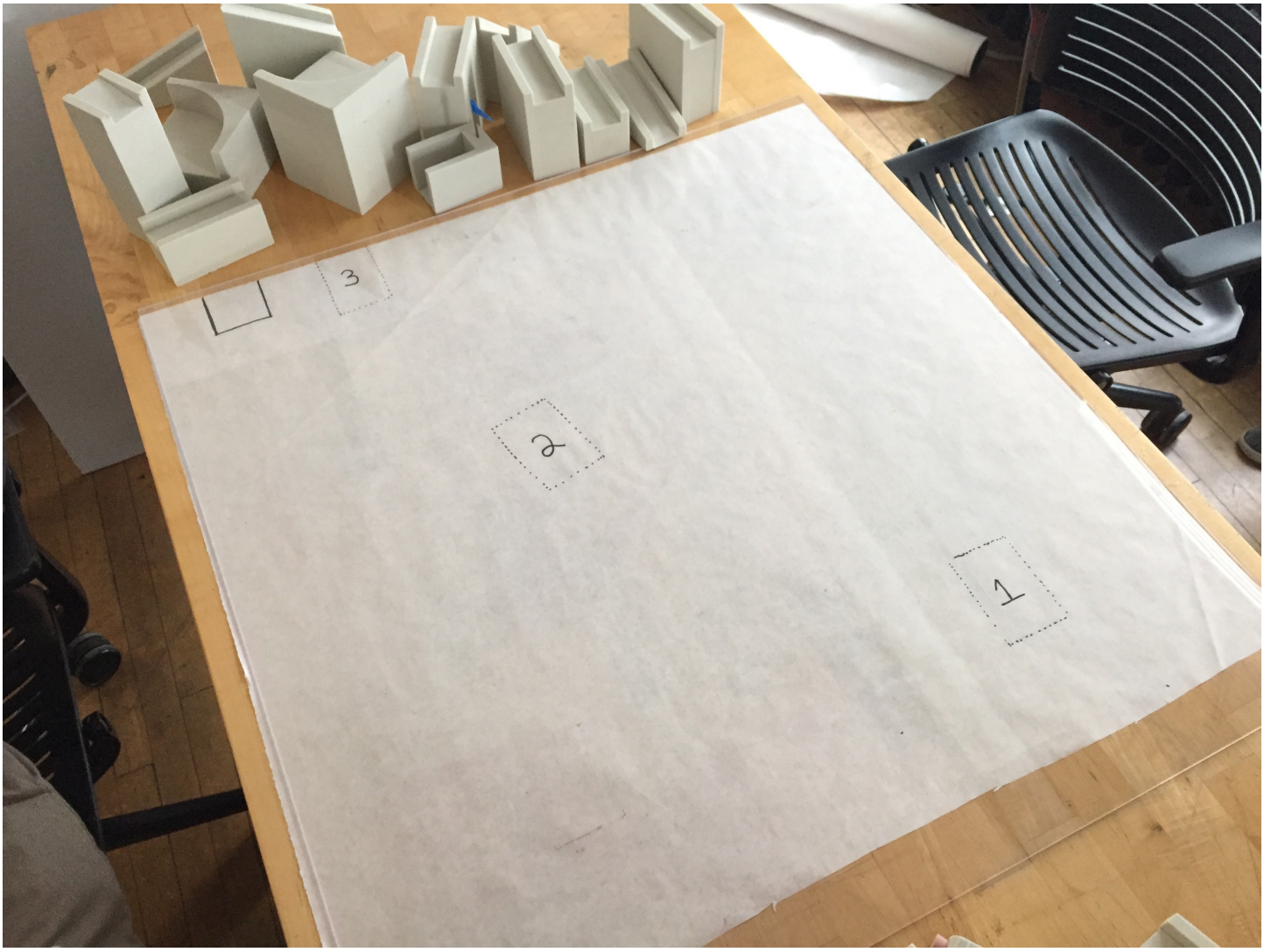
Appendix

Marble Run



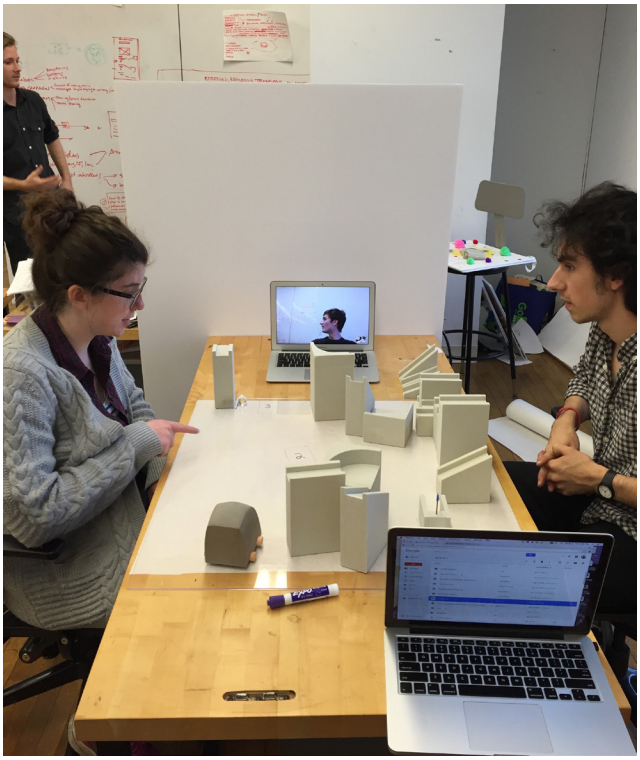


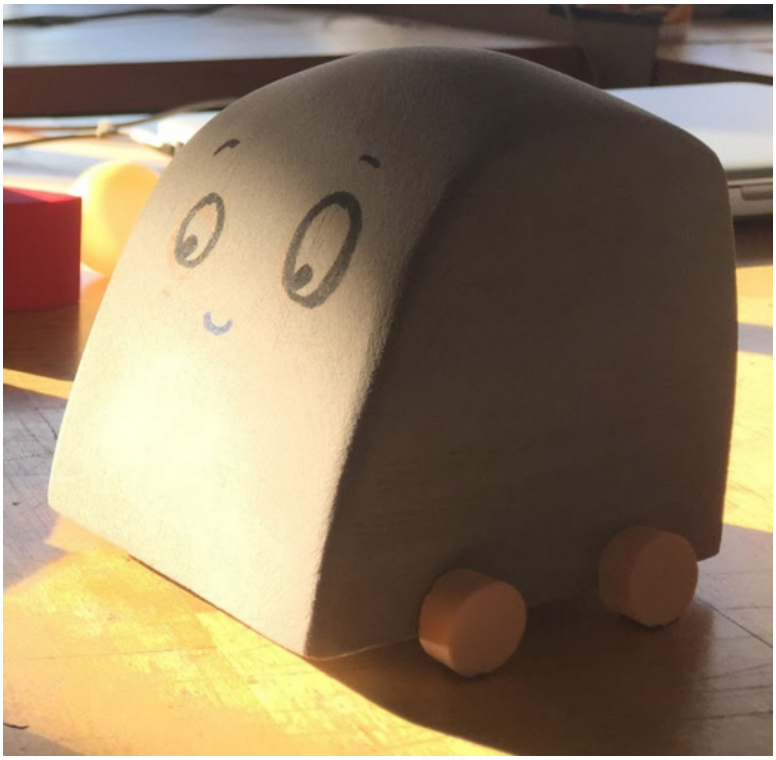




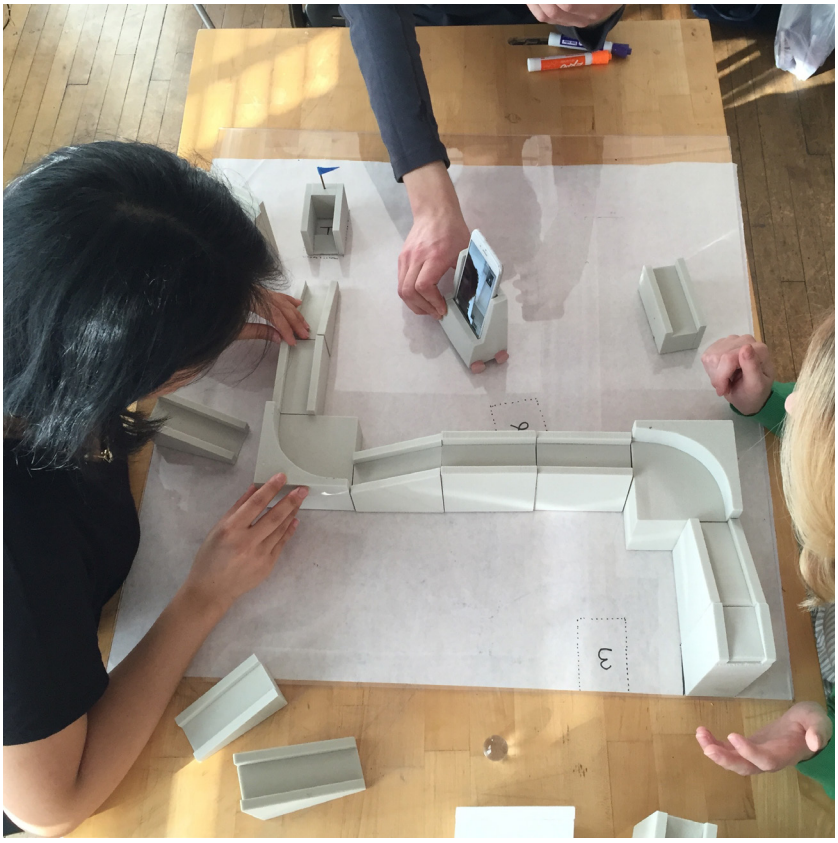
Condition 1: Alex on Screen

Condition 2: Alex and BeepBoop





Condition 3: Alex as Proxy



Raw Video of Testing

https://drive.google.com/folderview?id=0Bx5PifVr3kn5aVRFTmw5bmpsbGM&usp=sharing

Testing Script

Thank you for participating in this research study! Before we begin I'm going to read some information to make sure everything is covered.

Today you will be collaborating to play a game called Marble Run (Physical Physics). The point of this game is to get the marble ball from the starting ramp to the goal with the flag. Start moving the ball by dropping it on to the starting ramp. After you drop the ball you can't touch the ball again, you'll have to restart. To change the path of the ball you will be using different ramps and blocks. Feel free to try as many times until you reach the goal! Your first level might be easy, but later levels can get harder!

You will be collaborating with a few other peers that don't know the solution: the person next you, and Alex. Please take time now to introduce yourselves to each other.

Condition 3: (Have Alex introduce BeepBoop.)

Alex: Hi everyone, also my friend BeepBoop is going to play with us too! Beepboop is going to help me push things around.

Let me make it clear right away that this study is evaluating the game, not your performance or what you can do. There is no right or wrong, we just want to see what happens.

Great do you have any last questions before we start?

This is the first level, good luck!

Survey Questions

1. Alex can make their own decisions.

2. Alex has a personality.

3. Alex can recognize my emotions.

4. This activity was collaborative.

5. I wanted to know more about what was happening during the activity.

6. I was intrigued by what was happening during the activity.

7. I felt absorbed in what I was doing.