V.R. Narrative Research Project

Robert Quinn, Macquarie University, Australia.

Research Proposal

Working Title

The effectiveness of responding to user focus in virtual reality games.

Background

An in-depth investigation will be conducted into how a game designer can determine and respond to a user's focus in-game in order to more effectively present VR narratives without losing a player's attention, to ensure that key narrative elements are not lost.

Determining a user's attention to software and the effect this ultimately has on the use of said software, has been a popular topic of research in the fields of computing, game theory and behavioural psychology. For example, an experiment was conducted using eyeball tracking devices in order to determine what users were and were not paying attention to in a gaming environment and how this affects their information retention skills (Polinio, Di Guida and Goricelli, 2014.) This found that there was a significant if unsurprising link between information retained by a player and the parts of virtual scenes that they paid the most attention to.

Game theorists Ala Avoyan and Andrew Schotter attempted to measure attention and develop ways to improve a player's attentiveness in a game-setting, though this was in the field of thought games only, and did not surround video games nor the intense audio-visual stimuli that comes with them. Their findings were that the ways of maximising a user's attentiveness were through maximising the possible payoff of particular games, whilst not minimising risk beyond the point of threat. (Avoyan and Schotter, 2016.)

Early research has been conducted on using Virtual Reality to measure and test attention amongst children with ADHD, and this laid much of the groundwork in the construct empirical methods with which to test and compare focus and attention within VR. This paper did not begin to examine whether or not virtual reality could be used to improve these outcomes amongst participants, however, and focused instead on the different results amongst children with and without ADHD. (Neguț, Jurma and David, 2016.)

Due to the recency of Virtual Reality devices' widespread commercial availability and mainstream popularity, there is no commonly accepted design research methodology. By incorporating research into the development of an overall games research methodology, a fair study with generally applicable and statistically relevant results will be conducted. Of particular note is a chapter from Petri

Lankoski and Stefan Bjork's Game Research Methods entitled *Experimental Game Design*, which outlines the steps that must be taken in study design to ensure this. (Lankoski and Bjork, 2015)

Method

The experiment will be conducted as an A/B Prototype study, with two prototypes created and compared in order to determine effectiveness. A sample size of at least 15 unique participants for each prototype will be targeted.

Prototypes

Both prototypes will run through a scripted, single-path narrative whereby the player is having a conversation with a non-player character inside a virtual reality room. The player will be surrounded by a variety of distractions, such as toys and visually interesting objects.

Control prototype

The non-player character continues talking regardless of whether the player is paying attention to them.

Investigation prototype

The non-player character pauses and waits for the player to have them within their field of view before continuing to talk.

At the conclusion of the conversation with this character, a gameplay 'test' will ensue, where the player must, for example, place a ball into a bucket with a particular symbol, or type a code into a safe before a time limit expires. The correct answer to this will have been revealed during the player's conversation with the non-player character. A success in this gameplay test this will reward the player with a small prize, such as a chocolate bar.

At the conclusion of the VR experience, the player will be asked a series of questions through a survey. These questions will both test their recollection of key details in the scene (e.g. what was the name of the character you were having a conversation with, what colour shirt were they wearing, what time did the clock on the wall say it was.) as well as ask them how invested they felt in particular elements of the scene through a 5 point Likert scale. This will provide a mixture of both quantitative and qualitative data for analysis.

Data Analyses

The total number of successes and failures in the gameplay test at the end of the experiment will be collected. These proportions will be tested using a two proportion z-test to determine if there exists any significant difference.

From the survey questions, each participant will be graded based on the number of multiple choice factual questions they correctly answer. A two sample t-test will be used to determine if there is a significant difference in the mean grade between the two samples.

Finally, a χ^2 test of proportions will be used to determine if there is any significant difference between the samples of personal responses to the qualitative questions, in order to determine if the participants themselves believed that the prototype they tested impacted their attentiveness.

From the results of these three statistical tests, conclusions will be able to be drawn to determine whether or not the prototype's responsiveness to player focus

Possible Extensions and Necessary Research

Dependent on time limitations, this experiment should be expanded to better test the impact of this focus on gameplay elements of VR design, rather than focusing as heavily on narrative elements as it does in its current state. Currently, the only gameplay element is only loosely tied to that of the narrative, and so is perhaps not the best measure of how much attention the user has been paying. Better methods must be researched of how to link the gameplay and narrative elements in a more natural way.

More research, specifically into behavioural psychology articles, should be conducted on best-practice ways of measuring participant focus and attentiveness. If this determines that there are far more tried and tested ways to do so than the way planned, then the experiment design must be changed accordingly.

References

Polonio, L., Di Guida, S. and Coricelli, G. (2014). *Strategic sophistication and attention in games: An eye-tracking study*. Science Direct.

Avoyan, A. and Schotter, A. (2016). Attention in Games: An Experimental Study. *Stanford University Press.*

Neguț, A., Jurma, A. and David, D. (2016). Virtual-reality-based attention assessment of ADHD: ClinicaVR: Classroom-CPT versus a traditional continuous performance test. *Child Neuropsychology*.

Lankoski, P. and Bjork, S. (2015). Game research methods. *ETC Press*.

Design Document

High Concept

In this research project, I aimed to investigate ways to best handle dialogue systems in Virtual Reality games. In particular, I wanted to focus on how to deal with the fact that players must retain free movement during dialogue. In traditionally controlled first person games, a developer can take control of the player's movement and camera position, to ensure that they are looking where they 'should be.' In virtual reality, however, such tactics would not work effectively. As such, alternative measures must be taken to attract and maintain a user's focus when a non-player character is talking. Investigating the effectiveness of such measures was the main aim of this project, and a prototype was created to do so.

Prototype Concept

The Case of the Bundersnippet Diamond is a short single-player virtual reality mystery game designed for SteamVR compliant devices, created in Unity 3D. Players explore the Bundersnippet mansion, searching for clues and talking to the three suspects in the robbery of the enigmatic Lord Bundersnippet's prized treasure: The Bundersnippet Diamond. Once they are confident in their reasoning, they can choose to accuse one of the three suspects. If correct, they are rewarded by finding the diamond. If not, then they've failed in their task. The world features a variety of physics objects with which the user can interact, as well as some visually interesting props to keep them engaged.

The game was developed in order to determine the effectiveness of waiting upon user focus before proceeding with narrative elements in virtual reality games. As such, two distinct prototypes were created. In the experimental prototype, characters only speak to the player when they remain within their field of vision. In the control prototype, however, characters simply begin talking when the player moves within range, and continue until finished. Target Genre, Audience and Platform

The Case of the Bundersnippet Diamond is narrative-focused exploration game targeted for the HTC Vive, Oculus Rift and other Steam-VR devices. As the game was developed for data-gathering, research-based purposes rather than mainstream profitability, it targets a young adult, 'hardcore gamer' audience, as it was reasoned that the majority of research participants who would be taking part in the trial would fall into this description.

Some familiarity with Virtual Reality devices may aid players in the game, but is not required due to no advanced controls being required.

The virtual reality, clue-gathering based gameplay places it in a similar space to games such as *The Gallery*, *Obduction*, and *Conductor*.

Design Goals

In order to collect data that could be best generalised for virtual reality game design, the prototype had to be fun and engaging enough for players to feel compelled to solve the mystery. To achieve this, three specific kinds of fun from Le Blanc's *8 Kinds of Fun* were targeted.

Fantasy

The prototype allows players to live out the fantasy of being a classical detective, in the mould of characters from Agatha Christie style golden-age cosy crime stories. This immersion is built through both setting, in a stereotypical enigmatic millionaire's mansion, and through the virtual reality gameplay, which allows for the player to interact with the scene and other characters in a somewhat more natural way than a typical mouse-look first person controller would.

By targeting this fantasy, I hoped to ensure that players would be encouraged to be deductive and pay attention to clues, so that they did not simply guess a suspect at random.

Narrative

The game needed to be engaging enough for users to want to solve the mystery and pay attention to clues, for any data to be useful. As such, a dramatic arc was created so that the more clues and discussions with suspects a player has, the more tense the scene gets.

The narrative of the game is told both through dialogue and clues hidden around the mansion. By juxtaposing information gained through dialogue with tangible evidence, I aimed for players to be required to use deductive reasoning in order to solve the mystery. Only by paying attention to what is said versus what they can see with their own eyes, can they be certain of the true suspect. For example: the butler informs the player that he was cleaning the library all of the night of the crime. If the player looks in the library themselves, it is clearly a mess, refuting the detective's claims.

Finally, I attempted to create an overall tension and uneasiness in the game. This is achieved through ominous non-diegetic music, and antagonistic characters that are rude to players not 'paying attention', or indeed to players even daring to question them. As such, tension builds over the scene, leading to either relief when the correct suspect is ascertained, or frustration if incorrect. To alleviate this frustration, players can keep guessing until they get the right suspect, in order to tie up loose ends. The game, however, only records the user's first guess in its output data.

Discovery

The main gameplay beats of the prototype come through discovery. When the player initially begins the experiment, they are in uncharted territory, and must piece together the solution. To do so, they explore the many rooms of the strange Bundersnippet Mansion, searching for clues and revealing more about the characters than dialogue alone would. By interacting with clues in a physical sense, and manipulating, rotating, and moving them, they can immerse themselves in this discovery in a way that virtual reality can best allow.

Mechanics

Controls

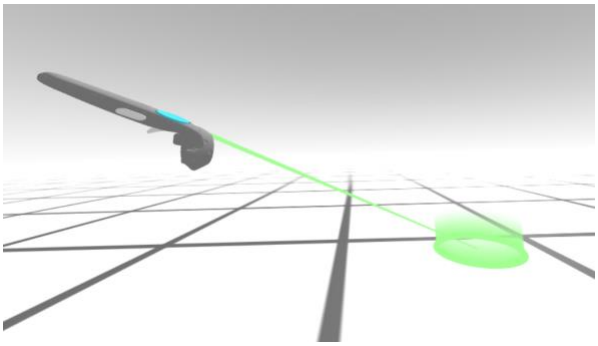

Fig. 2: A visualisation of how the teleportation system works (Serrano, 2016)

The player can control their movement in two ways. Firstly, they can physically move around the game's playspace by walking in the real world, as it is a room-scale VR experience. This, however, is limited to a 'safe-area' (typically approximately 5 metres by 5 metres) dependent on where the prototype has been set up.

To move further than this area, the user can hold down the centre button on their Vive controller to send out a 'teleporter beam.' This allows them to choose an area within a certain range for them to teleport to. If the area they have selected is appropriate (i.e. floor, or another object that a player should be able to walk on) then releasing the button will instantly teleport the player to the new location. This method of locomotion was selected both due to the ease of implementation, and due to Costas Boletsis' 2017 literature review of a range of research into various VR movement solutions finding that a common consensus amongst academic research into the subject was that whilst teleportation could be slightly jarring due to being non-continuous motion, it was easy to use and easier for new players to understand than many other forms of player control such as gesture based controls, arm swinging controls, or headlook controls. (Boletsis, 2017)

The player is also able to interact with physical objects, which is key for investigating the crime scene. By holding the trigger on the Vive controller, any rigid-body object which the controller is currently colliding with is 'gripped.' The player can then move the object around freely. When they release the trigger, the item is no longer held. If players combine this release with moving their hands at speed, then an item can be thrown around the scene.

Clues

Fig. 2: An example of a clue in the game

Several objects that reveal information about the mystery are hidden around the scene. These include a first class aeroplane ticket booked by the Butler, love letters between Harrison and the Lady D'estitute, and rubbish throughout the supposedly clean library. Players can find and physically interact with these clues to examine them.

Dialogue

Fig. 3: The Butler, giving his alibi.

All non-player characters in the scene have several lines of dialogue to deliver to their player. The manner in which they are delivered varies based on which prototype the player is being tested on. The SALSA Lip-Sync add-on for Unity was used to incorporate basic facial and lip-sync animations in time with this dialogue.

Prototype A (Control)

When a player stands within a certain range of a non-player character, the character begins to recite their dialogue. They continue until the audio clip is finished, regardless of if the player is there to hear it. The dialogue can only be played once.

Prototype B (Experimental)

When the player looks within a certain range of the non-player character's eyes, the character begins their dialogue. If the player looks away from this range for more than 3 seconds, the dialogue is paused, and the character interjects, asking the player if they are still listening to them. They then remain silent until the player looks at them again, and only continue when this 'eye-contact' is resumed. As the HTC Vive does not have any form of eyeball tracking, this is done in a simplistic manner by ray-tracing from the centre of the camera. If this ray intercepts a spherical collider approximately 2 metres in diameter around the head of the character, then 'eye-contact' is established.

Early prototypes had the character immediately stop talking when this contact was broken. Testing showed this to be extremely frustrating due to the inaccuracy of the tracking method, and so the three second leniency was added.

Accusers

Fig. 4: The "Accusers"

To submit their decision as to which suspect to accuse, the player must push one of three buttons on large 'accuser' machines in the centre of the mansion's atrium. To do so, they must grip the button and push it down using the controls that they will by this point have learned by interacting with other objects around the scene.

When this button is pushed down, if it corresponds with the correct suspect, a 'success' noise plays, and the Bundersnippet Diamond appears in-front of the player, and the crime is solved. If it was the wrong button, however, an unsuccessful tone sounds, and the player has 'lost.'

Art Style

Due to time constraints and the fact that this was an individual project, the majority of art-assets were sourced from third-party sources, both free and paid. As such, the prototype did not have an

especially coherent art style. However, care was taken to ensure that all characters had at least some expressive ability to their faces, and that lips, eyes, and facial muscles were all animated when a character was talking. This was done so that players did have some kind of motion to look at when in 'conversations' with these characters, rather than mostly static screens.

The house itself was designed with inspiration from clichés of the Golden Age of detective fiction, with lavish gold trims on all furniture, and an excessively large chandelier as a centrepiece of the atrium. Rooms were clearly labelled so that players knew where they were in the mansion at all times, so that they would best know which clues lined up with which character. Initial playtesting did not have these labels, and I discovered that many players did not know which room the library was when the Butler claimed he had cleaned it, resulting on them not realising the significance of this clue. Hence, signs were added above all doorways to fix this.

Audio

All dialogue in the prototype is fully voice acted. Characters have an accent and vocal style corresponding to tropes of their character within Golden Age detective fiction. For example, the detective has a sneery, nasally, high-pitched British accent. The layabout son, Harrison, sounds groggy and tired, and has a hoarse voice. The Lady D'Estitute has a 'French' accent, and a smokey voice meant to reflect many so-called 'Femme Fatales' of the genre. The Lord Bundersnippet himself is loud, eccentric and posh, and sounds particularly 'snobby'. These voices were designed to give some more human characteristics to the somewhat robotic 3D models in the scene, as the limited facial animations and rigidity of the characters' poses lead to a lack of any possible immersion.

Although the voice acting was not professional quality, it was still ultimately necessary in order to do any tests at all involving dialogue.

Smooth, dark, jazz music (Kevin Macleod's *I Knew a Guy*) also plays in the background, also playing into the genre conventions of 'cosy' crime fiction. This further builds the atmosphere and immersion of the prototype.

Data-Collection

When the player submits their first guess, several key pieces of data are immediately recorded.

- Which prototype the player participated in
- The time taken in the prototype before making the guess
- Who they accused
- How many of the characters they spoke to before making their guess

Results of the analysis of this data will be included in this project's play-testing document.

Dynamics

All players, regardless of prototype, spent the first minute or so of play experimenting with the controls, and simply walking around and familiarising themselves with the teleportation system.

Once players learned how to pick up physics objects, they typically began to throw objects about the scene as often as they could. Most players would often attempt to throw objects at the non-player characters in the scene.

The first time that dialogue was playing, players in both control and experimental prototypes would usually get distracted by physics objects in the surrounding area.

However, the first time that a player in the experimental prototype was ‘told off’ for not paying attention usually resulted in that player stopping and listening to all other dialogue for the rest of their playtime.

Players in the control prototype, however, would usually continue walking around whilst dialogue was playing, after briefly examining the character with whom they were talking.

Players in both prototypes often missed clues by throwing them away as soon as they picked them up, which suggests that allowing full physical control of the objects in this manner may reduce the effectiveness of physical clues, as players found throwing objects as fast as possible to be the most enjoyable of the interaction allowed within the prototype, often resulting in clues unintentionally ending up hidden behind furniture.

References

- *The Gallery*, Cloudhead Games, 2016
- *Obduction*, Cyan Inc., 2016
- *Conductor*, Overflow, 2017
- *Developing an A-Frame Teleport Component*, Fernando Serrano, 2016
- *The New Era of Virtual Reality Locomotion: A Systematic Literature Review of Techniques and a Proposed Typology*, Costas Boletsis, MDPI, 2017.
- *I Knew A Guy*, Kevin Macleod, 2006

Playtesting Report

Initial Notes

During the main playtesting session for this project, a Vive Base Station was damaged, meaning that no further participants could be studied. As such, sample sizes for each experiment were much smaller than originally intended. The limited sample sizes must be taken into account when observing all results contained in this report, as any conclusions will unfortunately require significant extrapolation.

Investigation 1

To establish whether waiting on user focus in dialogue made a player spend longer inside a Virtual Reality prototype. Two candidate prototypes were created:

1. A control prototype, where dialogue simply plays out when a user enters a character's proximity.
2. A prototype where characters only recite dialogue when players are continuing to look in their general direction.

The goal of this question was to find out whether 'requiring' users to pay attention to dialogue in a VR scene:

1. Lead to them spending longer in the prototype
2. Did not only extend play-time in an enforced and frustrating manner, but was at least equally as immersive.

Prototype

Two prototypes were constructed with identical stories, clue placement, and controls. The only difference came in the methods with which dialogue was delivered.

In the control prototype, dialogue would start when a player came within 4 world-units of any particular character. It would continue until finished, regardless of the player's positioning.

In the experimental prototype, the character begins their dialogue, when a ray that is traced from the centre of the player's camera intercepts a large spherical collider, 2 world-units in diameter around the player's head. If the player looks away from this range for more than 3 seconds, the dialogue is paused, and the character interjects, asking the player if they are still listening to them. They then remain silent until the player looks at them again, and only continue when this approximation of 'eye-contact' is resumed.

Each prototype recorded a player's time taken when they submitted their answer using the 'accuser' machines in the centre of the scene.

Experiment

10 volunteers (all university students or recent graduates, in a variety of fields, 9 males, 1 female, aged 20-24) were randomly split into two groups of 5. The first group attempted the control prototype, and the second group the experimental. Participants were asked to not talk about details of the narrative of the game with other participants either during nor after the prototype, until all participants had finished, to prevent this influencing other user's results.

Participants who had not yet attempted the prototype were prevented from seeing my observer screen, so that they could not get a head-start on the case in this way either.

Shortly after completing the prototype, players were asked to fill out a short survey by rating their responses to several prompt questions on a 5-point Likert scale (from 1 = Strongly Disagree to 5 = Strongly Agree.) The prompts they were asked to rate, relevant to this investigation were:

Q1. *I felt the characters were talking to me.*
Q2. *I felt like I wanted to solve the crime.*
Q3. *I found the dialogue boring.*

Additional general observations were recorded and included below.

Results

The results of the time recordings, in seconds, are collated in the table below. For the purposes of these analyses, Prototype A refers to the **Control** prototype, and Prototype B refers to the **Experimental** prototype.

	Prototype A	**Prototype B**
Sample Size	5	5
Mean	316.40	354.00
Median	321.00	312.00
Standard Deviation	62.16	97.09
95% Confidence Interval for True Mean	(239.22, 393.57)	(233.45, 474.54)

These times were first examined using boxplots.

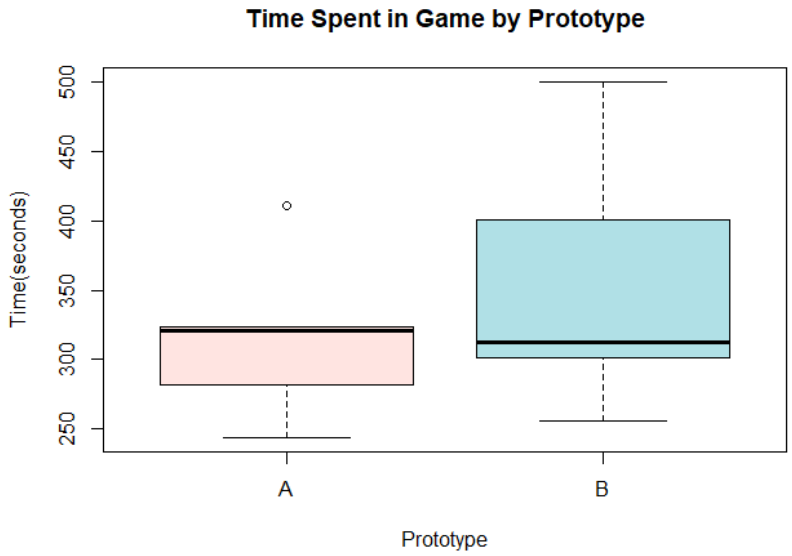

Fig. 1: Boxplots of time spent in game, separated by prototype.

From these boxplots, there was no clear difference in mean times in either prototype, though the spread of times was visibly slightly larger in Prototype B, supported by the larger standard deviation. Prototype A was heavily influenced by one outlier who took 411 seconds.

To determine whether there was a significant difference between the average times spent in each prototype, a Student's Independent Sample Two Sample T-Test at a 95% significance level was conducted on the data.

$$\mathbf{H_0}: \mu_A = \mu_B$$

$$\mathbf{H_1}: \mu_A \neq \mu_B$$

Where μ_A is the true mean of Times Spent in Prototype A and μ_B is the true mean of Times Spent in Prototype B.

The variances were deemed close enough from observing the boxplots. To assess whether it was fair to assume that the times in prototype were approximately normally distributed, histograms of the frequencies of both samples were created.

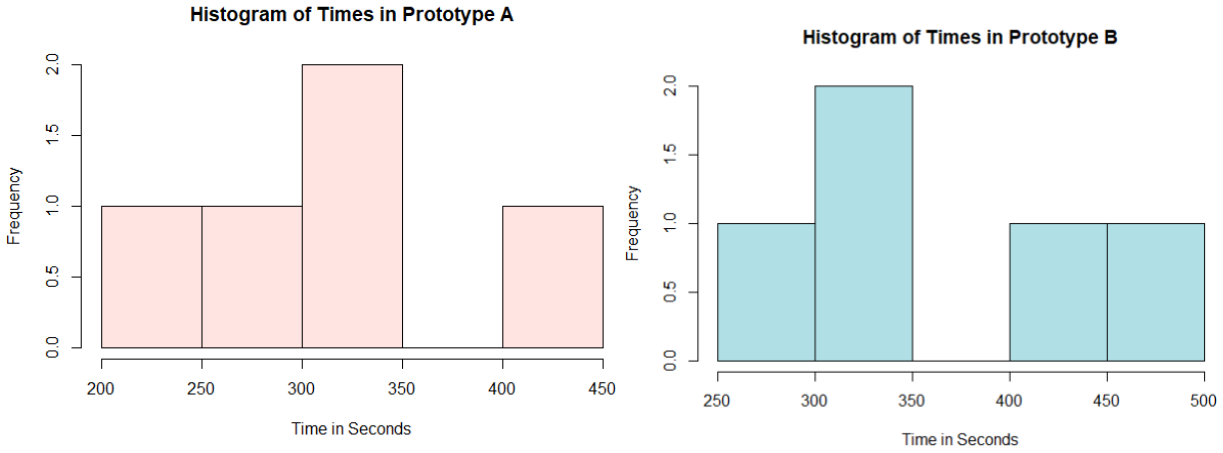

Fig. 2: Distribution of Frequency of times in each prototype

Due to the small sample size of 5 in each group, these histograms showed enough evidence of a normal distribution in their shape to proceed with the test.

From the above data, the following test statistics were calculated:

$t = -0.72934$, $df = 8$

95% Confidence Interval for Mean Difference: (-156.48, 81.28)

$$\boldsymbol{p = 0.4866}$$

Therefore, there is no evidence with which to reject H_0. There is nothing in our sample to suggest any significant difference between the average time players spent in either Prototype A or Prototype B.

This result must be interpreted with the knowledge that both sample sizes were very small, and as such any generalisation from them is ill-advised.

To determine whether players found either prototype more or less immersive, the results of relevant survey questions were summarised and examined, showing mean Likert values with their standard deviations.

	Prototype A	Prototype B
Q1. I felt the characters were talking to me	**1.7** ± 0.4	**2.4** ± 1.1
Q2. I felt like I wanted to solve the crime	**3.6** ± 0.5	**3.6** ± 0.5
Q3. I found the dialogue boring	**3.3** ± 1.2	**3.6** ± 0.5

The question with the only significant difference in response was the first, asking whether or not players felt that the characters were actually talking to them. Although the means of both scores are still very low, the marked improvement in Prototype B suggests that it was a success in this regard, and that users did indeed feel more immersed in dialogue with characters when they responded to focus.

The change in prototypes appears to have made no difference to a player being compelled to succeed in the prototype or not, with very similar average scores and variances between these two response groups.

Finally, although most users ultimately found the game's dialogue boring, that this did not increase dramatically between prototypes suggests that perhaps as hoped, 'forcing' the user to pay attention to the dialogue did not make the prototype significantly more boring.

General Observations

- Players in both prototype spend most dialogue time either throwing objects at characters or trying to pick up the characters themselves.
- Players in the experimental prototype tended to try and stand as still as possible, perhaps thinking that they would be reprimanded for changing position rather than where they were looking.
- Most players in the control prototype did stay in the close area of the character currently talking to them, only two walked off whilst characters will still talking.

Conclusion

The results collected here, whilst severely limited by sample size, suggest that requiring user focus to continue dialogue makes no significant impact on the average playtime of a virtual reality experience.

Such a technique also does not seem to impact a player's engagement with a virtual reality experience, nor their desire to 'do well' at the game.

The most significant difference between the two prototypes that can be deduced from this subset of data is that it did appear to improve upon user's perception of whether or not non-player characters were communicating directly to them or not.

Investigation 2

To determine whether the previously described focus-based system lead to players paying more attention to key narrative points, therefore leading them to choosing the correct suspect in a simple mystery story. The same two prototypes were used:

1. A control prototype, where dialogue simply plays out when a user enters a character's proximity.
2. A prototype where characters only recite dialogue when players are continuing to look in their general direction.

The goal of this question was to find out whether 'requiring' users to pay attention to dialogue in a VR scene:

1. Resulted in them paying more attention to dialogue, and as a result retaining more important information, therefore being more likely to correctly pick the guilty suspect.
2. Made them more confident in their choice of suspect at the end, reducing the necessity for guesswork.

Prototype

The same prototypes described in Investigation 1 were used to simultaneously run both investigations. For this investigation, however, different key data was output at the end of each prototype. Namely:

- Which suspect was selected
- How many characters the player had spoken to before making their guess

Experiment

The same group of 10 volunteers (all university students or recent graduates, in a variety of fields, 9 males, 1 female, aged 20-24) was used, as both investigations ran simultaneously. At the conclusion of the experiment, participants were asked additional questions to gain qualitative data on their perception of each prototype. Participants were asked to rate the following statements on a 5-point Likert scale: (from 1 = Strongly Disagree to 5 = Strongly Agree)

Q1. *I was sure of the suspect that I chose at the end.*

Q2. *I found the plot simple and clear to follow.*

Players were also asked to select one response from this multiple choice question:

Q3. *My choice of suspect was most influenced by:*

- Physical clues
- Dialogue
- Character appearance / voice
- Guesswork

As before, Prototype A refers to the **Control** prototype, and Prototype B refers to the **Experimental** prototype.

Results

Every player's accusations were taken, and converted to a binary correct/incorrect response, and the results summarised below.

	Prototype A	**Prototype B**
Correct	**3**	**4**
Incorrect	**2**	**1**

Although Prototype B did have more successful participants, it was only one more. The limited sample sizes of this data makes it hard to take any significant meaning from the summary alone. From this summary, a Pearson's χ^2 test of independence was conducted at a 95% significance level in order to determine whether there was a statistically significant link between the type of prototype and number of successful candidates.

Due to the small sample sizes, such a test can only give an approximate summary and can not be deemed statistically significant, as the model can not be fully validated.

H_0: There is no association between Prototype and Success Rate
H_1: There is an association between Prototype and Success Rate

From the above data, the following test statistics were calculated:

$$\chi^2 = 0.48,\ df = 1$$
$$\boldsymbol{p = 0.49}$$

With a p-value significantly higher than 0.05, this data can not be seen as statistically significant evidence suggesting any link between the type of prototype a player experiences and their likelihood of success. However, it should also not be seen as proof that there is no link. This

output suggests that in a longer term, expanding the sample size could give significant and interesting output.

Participants responses to the qualitative questions are summarised below. Scores shown for the Likert response questions are the mean response from 1 to 5 along with its standard deviation.

	Prototype A	**Prototype B**
Q1. I was sure of the suspect that I chose at the end.	**3.3** ± 1.2	**3.6** ± 0.5
Q2. I found the plot simple and clear to follow.	**2.6** ± 2.2	**3.2** ± 0.4

Q3. My choice of suspect was most influenced by:

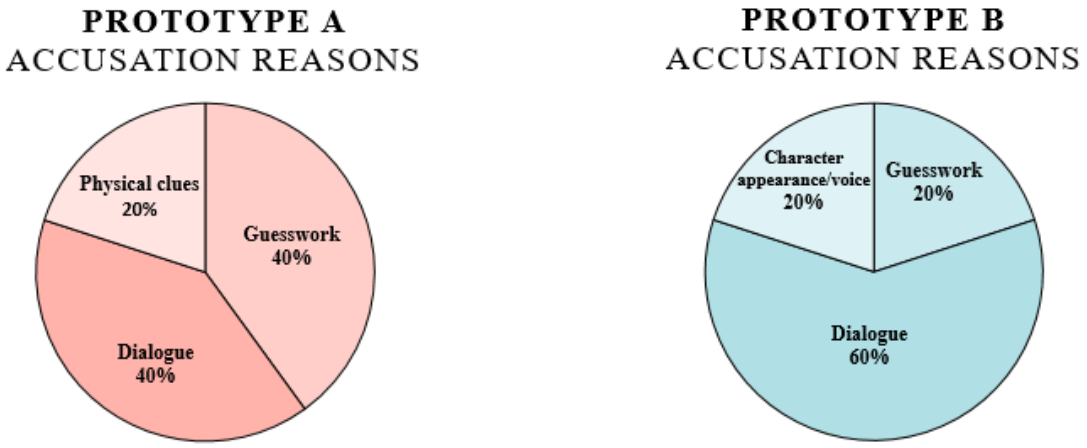

Fig. 3: Percentages of key reasons selected by players

The responses to the first question were very similar in both prototypes, so there is no evidence to suggest any improvement or decline in a player's confidence in their choice. Responses to the second question, however, varied significantly.

Again, with such a limited sample size, every individual response does have a significant influence on the overall average, and the high standard deviation of Prototype A's responses here suggests that players in this group tended to give either 5's or 1's as their response. Interestingly, examining the data closer showed exactly this, with 2 responses of 5 and 3 responses of 1 to this question in the Prototype A data.

This shows, perhaps, that in this control prototype that does not require user focus, players are likely to either force themselves to pay attention or not - which leads them to either understand the plot in detail, or not understand it at all.

Finally, the change in proportions of key evidence behind accusations did not change too dramatically. Prototype B did, however, have an increase of 20% in user's choosing 'Dialogue' being the key reason for accusations. Since this is a change of just one participant, however, it is not significant evidence. Interestingly, no players in Prototype B chose physical clues as their reason for their choice in suspect, instead mainly selecting dialogue or a character's appearance / voice.

Since both of these options focus on the characters appearance, voice, or dialogue content, this does perhaps suggest that user's in Prototype B were paying more attention to the non-player characters than those in Prototype A.

General Observations

- Players in Prototype A seemed to take longer after approaching the Accuser machines than players in Prototype B, often hovering between multiple choices before making their selection.
- Players in Prototype A often went back to characters at the end, hoping for dialogue to repeat, which it would not. This only happened once in the Prototype B group.

Conclusion

There is no clear evidence to suggest either style of dialogue delivery improves outcomes in terms of successful accusations in this prototype. Significantly more data needs to be collected in order to determine whether or not this is the case.

The data collected does suggest, however, that waiting for user's focus to deliver dialogue can improve other outcomes for players, specifically ensuring that players do pay attention to a character's appearance, voice, and key dialogue.

It also suggests that implementing such a system may improve a player's understanding of the plot of a narrative-based game.

Overall, the experiments contained within this report do show enough to indicate that conducting additional experiments with much larger sample sizes and a more detailed prototype could reveal potentially useful tools for constructing a Virtual Reality game design methodology.

Postmortem

Design

The design of my overall project and prototype was partially successful. The initial concept for the research project was interesting, and the prototype that I planned was of an appropriate scale to collect preliminary data to investigate such a topic. There were, however, certainly issues with the design for the project which ultimately reduced the quality of the project as a whole.

What Went Well

The base mechanics and controls of the prototype were simple and clearly understandable. Players who had no experience with Virtual Reality devices had full control of the prototype within a matter of minutes playing.

The narrative of the prototype was engaging enough for the length of the experience, and players found the voices of the characters amusing, and told me that they enjoyed piecing together these characters' various interrelations and motivations.

The music for the game is appropriate, and does a good job at building the atmosphere and theme in the prototype.

The physical controls of objects in the scene feels good, players clearly enjoyed throwing objects around the scene and manipulating clues with their hands, and reported that this did a great job at immersing them in a detective role.

Eventually, the design of the two prototypes was an effective way to explore the overall focus of the project, namely whether or not waiting for player focus improves the delivery of narrative elements in a Virtual Reality game.

What Went Badly

There was not enough gameplay in the original design for the project. I focused too much on the narrative sides of a Virtual Reality game, without ensuring that it had any actual game elements.

The 'accuser' devices used by players to submit their choice of suspect was a poor quality solution, and negatively impacted any sense of immersion in the scene. I should have spent more time implementing a diegetic way of doing this rather than bright coloured boxes standing out in the middle of the scene. Since the game employed motion controls, a method such as pointing at a suspect would have fit much better.

The lack of game elements remained an issue by the completion of the project. This meant that extending any analysis of the prototype that I created, to analysis of Virtual Reality games in general, is a significant stretch.

I should have spent a lot longer in the preliminary stages of my prototype thinking about how I could have designed gameplay mechanics that reflected my topic area, instead of making a prototype devoid of any substantial gameplay and then ‘tacking some on’.

Overall, I should have made earlier, simpler prototypes throughout the project, to see if the idea was working at all, instead of working on the final prototype from the start. If I had tested the design of the project early and often, I would have discovered how crucial this need for more gameplay depth was.

Lessons for Research

If I want to do research in game design, it’s extremely important to design games. Different elements of game design, such as the narrative aspect I wanted to investigate in this project, can not be investigated in isolation. Games are a combination of a massive combination of factors, and to only look at one of those in a prototype ignores the significant interaction between these factors.

Designing the entirety of a project before writing a single line of code is crucial. I found often throughout the semester that I would be guilty of modifying my design to fit the prototype that I had constructed, a fatal flaw.

If I had instead ironed out a concrete design for a prototype that incorporated significant gameplay elements and better reflected what I had learned from my initial research, then I would have finished up with a prototype that players actually enjoyed participating in.

Doing so would have given me more interesting results that could be better used to learn lessons for Virtual Reality game development as a whole.

Research prototypes may be limited in scope and scale, but they must still reflect the quality and substance of the larger style of game that you are using them to try to investigate.

Development

Although much of the development of this prototype was successful, several key mistakes in this part of the project lead to diminished results and a slightly lower quality final product than could have been achieved.

Weeks 1 and 2

Summary

The first weeks of the project were spent collating and reviewing literature in similar fields to the focus area of my project. This culminated in the writing of a project proposal for submission in week 3, wherein I set out what I intended to investigate, and how.

Lessons

My research proposal was very broad, with little consideration of the technical side of what I intended to implement. It also lacked any concrete gameplay in my description of what I wanted to test. If I had been more detailed and precise at this stage, I would've been able to spend less of the remaining weeks changing my ideas and refining the overall basis of the prototype.

I should've also made more of an attempt to integrate my findings from other research into my final project, as I was somewhat guilty of merely stating what had already been done, rather than taking any lessons from that.

Weeks 3 and 4

Summary

After refining and clarifying my project idea, and ultimately determining a prototype that could be created, these weeks were spent learning the basics of V.R. development in the Unity environment. By the end of these weeks, I had a rough demo in which a player could walk around a small environment and pick up a physics object.

Lessons

Because the initial stages of VR development in Unity were simple, and were ultimately a matter of dragging Steam-VR prefabs into a scene, I assumed that the rest of the prototype's development would be similarly simple.

I also did not set up any sort of version control during these early weeks, which was a mistake as by the time I decided to do so, the project was so large that setting up a repository was overly complicated.

Because of poor decisions made in these weeks, my file organisation for the rest of the project remained chaotic for the duration of the semester.

Weeks 5 and 6

Summary

These weeks were spent collecting and learning how to use various third party libraries for Unity that I believed would be useful. These included basic V.R. interaction tools, human character animators and lip sync and facial animation tools.

Lessons

I spent too long trying to make these assets fit my project, instead of thinking what they would actually add to the prototype or how they would help the research. With SALSA lip-sync in particular, I spent almost an entire week going through various tutorials and reading documentation in an attempt to improve the results of this add-on in my project, to no avail.

Week 7 and Mid-semester Break

Summary

The majority of the base of both prototypes was created in these weeks. This consisted of:

- Using ProBuilder and third party 3D models to model the entirety of the mansion.
- Testing in VR to ensure that object's scales felt correct.
- Finding and/or purchasing appropriate character models for each character.
- Finding some way of giving these characters somewhat human animations.
- Writing and voicing the script for the narrative component of the prototype. including recruiting additional voice actors.

Lessons

In these weeks I was guilty of promising to do too much at once, in my desire to have the prototype as finished as possible by the end of the Mid-semester Break. Because of this, I added a whole lot of things very quickly to the prototype, and didn't afford these additions the testing time they needed. This meant that many large bugs (such as being able to teleport out of the mansion, or being able to pick up and throw characters) went undetected at this stage, whereas any sort of thorough testing would've immediately highlighted both of these.

Weeks 8 and 9

Summary

These weeks were spent preparing for the first official playtesting session for the project, held on the 9th of May. This meant that I aimed to finish the two separate prototypes for use in the experiment, meaning that I developed the two different ways of characters delivering dialogue that differentiated the

Lessons

If I had cleaner code and project hierarchies in the initial builds of my prototype, then modifying them to run as I had designed the two prototypes would've been much simpler.

However, because of the Mid-semester Break rush described previously, my project was extremely messy and much of this milestone was spent cleaning that up before I could even begin implementing the two different prototypes.

Weeks 10 and 11

Summary

The first playtesting session did not ultimately get me any useable data, due to significant bugs in the prototype that I could not fix in time. As such, these two weeks were spent working out what went wrong and preparing for Week 12's playtesting session so that I could be more sure of collecting appropriate data. This meant bug-fixing and tweaking some of the dialogue to be clearer. As well as this, I created added additional clues to the world, and signs labelling each of the rooms in the mansion so that players would be less confused.

Lessons

I should have done a lot more of my own testing before the first playtesting session, so that I could have recorded useful data then, instead of relying purely upon Week 12's session.

Weeks 12 and 13

Summary

Week 12's playtesting session went very poorly. Due to my setting up of the base stations for the Vive in a very haphazard way, one of the two base stations was irreparably damaged. This meant that for the remainder of the playtesting session, I was trying to fix this issue, and ultimately only managed to collect two samples, both of whom were constrained to standing still as two base stations are required for a room-scale experience. As such, these weeks were spent trying to organise an alternative way of collecting data, which I ultimately managed to organise with a Newcastle based V.R. arcade. After managing to collect some data, (a total of only ten

samples,) the remainder of the project was spent writing up the documentation for final submission.

Lessons

Working with periphery hardware was always going to be risky. I should have had alternative data collection methods in mind from earlier stages of the project, so that one setback in Week 12 didn't compromise the entire project as much as it ultimately did.

What Went Well

Due to weekly milestones and progress meetings, regular progress was made on this project. There was no possible way of leaving all the work to the last minute, which meant that I was working on it throughout the semester. This was ultimately a success.

Unity proved to be an excellent choice of software to develop the project, as my familiarity with the software, combined with the tools that are included for V.R. development, combined with the wealth of online support and documentation meant that developing the technical side of the prototype was relatively simple.

Sourcing third party art assets was also crucial, as if I had decided to try and do all the models myself, that would have ended up taking the majority of the duration of the project.

What Went Badly

I should have planned out each week well in advance, rather than only thinking on a week-by-week basis of what I would be doing for each milestone. Because I did not do this, I ended up spending far too much time on relatively less important components of the project, and far too little time on major, vital components.

I also should have developed redundancies for things going wrong in playtesting sessions, instead of naively assuming that these sessions would go fine and that I would be able to collect all the data I need in just two small, two hour sessions. Doing this meant that I had to spend the remainder of the time on the project playing catch-up, and ultimately vastly diminished the quality of research performed and the usefulness of my results.

Lessons for Research

Working with a team is incredibly useful. Throughout the duration of my undergraduate degree, I often found this annoying and unhelpful. When conducting this research project, and trying to not only complete all of the design components simultaneously to doing all of the developing, I realised that working with additional people to whom different duties can be delegated is crucial for large-scale, high-quality projects.

As I knew from the start that the technical side of development was not my strongest skill, it was helpful that I chose a project idea that was not too technically demanding. It is important to know one's strengths and weaknesses when trying to choose an idea.

Planning and documentation is vital. If I had conducted these components with more care earlier in the project, then I would have not had the issues with scheduling described previously, and would have ultimately had a more directed and useful prototype.

Finally, bug-testing and polishing is vital even for research prototypes. The excuse I often found myself using of "it doesn't matter, it's only a rough prototype" does not hold up, as it is still important for such prototypes to be a high quality, so that results from their analysis can be extended beyond such simple cases.