

How Knowledge of the Player Character's Alignment Affect Decision Making in a Game

Daniel Svejstrup Christensen

Aalborg University
Rendsburggade 14
9000 Aalborg, DK
dchr12@student.aau.dk

Mette Jakobsen

Aalborg University
Rendsburggade 14
9000 Aalborg, DK
mejako12@student.aau.dk

ABSTRACT

With the still increasing popularity of story rich games, the demand for engaging content is ever present. In a narrative, the protagonist has an central role, and in many games, this becomes a challenge as the protagonist is often times the player character, and as such, is most often in the player's control, not the author's. It is interesting to look into how telling a dramatic story influences the player's engagement into the player character, and how this in turn might affect the player's engagement with the goal of the game. This paper investigates whether adding an internal conflict, which conflicts with the external goal of a game, can make players more likely to abandon their external goal, and ultimately change their in-game behaviour. A test was conducted in which participants played through one of two similar versions of the same visual novel, one in which they were exposed to an internal conflict of the player character, and one in which they were not. The test was conducted online and achieved 467 participants. The data shows that a significant ratio of players changed their behaviour in the game, when exposed to the internal problem of the player character.

Author Keywords

Narrative, Video Games, Player Character, Protagonist, Character Alignment

INTRODUCTION

In games, the player character has the specific role as the character from which the player interacts with the virtual world. The player character can be abstract or complex, depending on their role in the game's narrative[10][19], however, in action-based video games, the player character is often also the protagonist of a story, that revolves around solving an external problem, i.a. a mission or quest. The player's engagement with their character can be divided into goal-related and empathic engagement[12], which means there is a difference between the player being engaged in

the player character's goal, and the player being engaged in the player character, as a character with whom they can feel emphatic towards. In a game like Super Mario Bros.[18], the player does not rescue the princess because they have any romantic feelings for her, but because the goal of the game cleverly correlates with the character Mario's goal of rescue his girlfriend. In this paper we investigate what happens if the opposite is true. If the protagonist's morals conflict with their external goal, will the player then choose the wishes of their character above the goal of the game?

We designed a test, in which players would be presented with a series of choices. Every choice had one option that would help them reach their external goal, and another option that would work against it. In this set-up we predicted that players would prefer to advance their external goal. Another version of the same game was made, where players would be given information about their character's alignment, prior to making a choices, using Lankonski's definition of alignment as the access players have to the character's actions, knowledge, and affects[12]. This is conveyed as an inner monologue, describing the player character's personal thoughts about the choice. At times, the player character's alignment would go against the option that lead to the external goal, and at other times they would correlate. We predicted that the test group, which was allowed entrance into the inner thoughts of the player character, would change their behaviour and become more inclined to refuse the option conflicting with their character's alignment, because their engagement would shift from goal-related towards empathic. We therefore hypothesised that participants in the test group would abandon their external goal in favour of the player character's wishes.

We conducted a test online, in which players would be divided into a test- and control group, each group playing different versions of the game. The choices the players made was logged to a database, and when analysing the data it was found that the test- and control group made significantly different decisions. The test group was significantly more likely to follow their character's wishes, and abandon the goal of the game. Furthermore, in some control choices, where the player character's wishes correlated with the goal, the test group was also significantly more likely to follow the goal, than the control group. From these findings we can conclude that giving players information about their

character's alignment has a significant influence on the way they make decisions within the game.

RELATED WORK

Internal problems are defined in Phillips et al.'s book as problems that occur in the mind, opposed to external problems that are found in the environment[16]. Empowering the player to choose how the protagonist reacts to internal problems poses an interesting problem; can the player get invested in an internal problems that is not their own? In her book[21], Ryan discuss how games can provoke feelings towards the self, but even if the story involves emotions such as love or jealousy, these are not the emotions that motivate the player. The interactivity of games does not easily combine with emotional immersion, because it requires players to feel empathy towards a computer-controlled character. In his paper, Davis[5] includes Fantasy, which is empathy towards a fictional character, as a sub-category in their widely used questionnaire to measure empathy. However, the questionnaire was designed for movies, and when considering games, these characters also serve a purpose in the interaction, sometimes as helpers or hinderers of the player's objective. This means the player has a natural ludic interest in these characters as a means to an end[20]. However, Nitsche[15] argues, that if this ludic interest can be combined with the narrative interest of the character as a person, then it would elevate the player's perception of the character from being a game object, into a human being. According to Lankoski, the player's empathic engagement with the player character is dependant on three different factors; their recognition, alignment, and allegiance[12]. Where recognition is about how the character is presented and interpreted, alignment is the player's access to the character's knowledge and affects, and allegiance is how the character can elicit sympathy by being either a positive or negative morally appealing individual.

Undertale[9] is a game that challenges the player's engagement in what would be both the goal of the game, and the external problem: Escape the kingdom of monsters. In their quest to escape, the player will encounter likeable characters, who serve the purpose as hinders of that goal. The player, who at this point is the protagonist of the story, thereby develops an internal conflict: Not wanting to kill the monsters standing in their way. If the player chooses to change their behaviour and spare the monsters, thereby work directly against their external goal, the game will adapt and the ending will change.

Games handle internal problems differently. Jørgensen[19] analysed two story rich games, Dragon Age: Origins[6] and Mass Effect[1], and found two different types of player characters. In Mass Effect, the player has agency over their character's actions, but not their voice nor their feelings, while in Dragon Age: Origins, the player has complete control over what their character says and does. The result was that in Mass Effect, the player character Commander Shepard was clearly the protagonist of the story, while the Warden in Dragon Age, was so to a lesser degree. In Dragon Age: Origins, the interesting internal problems were instead found

in the non-player character(NPC) cast, whom she argues to be the true protagonists of the game. The reason why developers might want to move the protagonist role of the narrative to a player independent NPC, could be that it is a challenging to give the player power over the central character of the plot, and still have that character not be flat[24], using E.M. Forsters definition of a flat character as being one whom does not change throughout the course of the work[8].

Questionnaires are widely used to test for immersion, engagement, and empathy. Jennett et al.[11] and Qin et al.[17] both conducted several iterative experiments to construct their immersion questionnaires, the former by having participants play either an immersive or non-immersive game, and the latter by online testing in which participants had to imagine a familiar game. Lankoski argues, that immersion does not fully describe the relation between player, game, and player character. He proposes the usage of the term engagement instead[12]. Buselle et. al.[3] constructed a questionnaire in which they divide engagement into four different aspects: narrative understanding, attentional focus, emotional engagement, and narrative presence. Here emotional engagement in particular is interesting, as it is derived from a combination of empathy and sympathy. Empathy, which Davis[5] constructed a questionnaire for measuring in the context of film, and for which Lawrence et al.[13] constructed the Empathy Quotient(EQ). It is used to measure a user's general level of empathy, and was constructed by looking at the difference between results of healthy test-participants and participants suffering from Asperger's Syndrome or High-Functioning Autism.

But not all research into player characters have been done by using questionnaires, and it is also interesting to look into how the player behaves when acting through the player character. McCreery et al. did a test on players of World of Warcraft[14]. They were interested in whether the player's perception of their character had an influence on how they behaved in the virtual world. They had players fill out personality-rating scales for both themselves and their avatars, and then observed their virtual world behaviour. They analysed their observations using a behavioural assessment checklist, and their results suggested a relationship between personality and behaviour.

EXPERIMENTAL DESIGN

From the related work, it was found that in games there can be a conflict between the player's engagement in the goal of the game, and their emotional engagement in the player character, which leads us to the problem statement.

Problem Statement:

Does giving the player knowledge about the player character's alignment influence the decisions they make in the game, and cause them to abandon the game's external goal?

From the research it was also found that scientific research done into this field often utilize questionnaires as a way of evaluating participants, and often recordings of play is used instead of a fully functioning prototypes, to ensure participants

get a similar experience. The test design for this paper propose an online prototype in which all players get a similar experience, while still having agency, and while questionnaires will be utilized as well, the bulk of the collected data will consist of observations of player's interaction with the prototype.

Porcelain Hearts

The problem statement eludes to a question about player choice. Do the player advance towards their external goal, or do they neglect it? The test should investigate whether the player character's alignment may sway them in favour of either, it would therefore be necessary to make two versions of the same game. One in which players have insight into the player character's alignment as they progress, and another in which they do not. In order to easily evaluate a player's behavior, it made sense to look at game genres that allow for a binary form of interaction. Large role-playing game(RPG) titles such as Dragon Age: Origins[6], and Mass Effect[1] use dialogue wheels, and are known to present their players with binary choices that affect the cause of the story. Visual novels takes the concept of the dialogue wheel and then create an entire game based on selecting from lists of options. This is a purpose-selective form of interaction which makes it easy to look at players' behaviour, and allows for mostly linear experience.

External Problem

It was a priority to pick an external problem that was not emotional, in order to have a clear contrast between the external and internal problem. It was therefore decided to not use traditional epic plots such as to save the princess, or bring peace to the world, and instead the player's goal became to gather treasure, taking inspiration from games like Super Mario Bros.[18], which uses coins as a collectable reward. It was decided to use an epistemic mystery plot over a dramatic plot, due to the concern that it may be harder to block the influence of an emotional story, because the baseline for emotional engagement would be higher. An epistemic plot focuses on the player's satisfaction of solving a puzzle rather than any emotional immersion[20]. The game's narrative was kept as didascalic as possible, with a very short Alter Audience Distance[2]. The reasoning behind this was that an abstract narrative might cause players to perceive the narrative differently, and thereby create noise in the data.

Crit Choices

Ten crucial moments(Crit Choices) were planned out in a linear story arch, where players had to choose whether to advance towards the goal of the game, i.e. getting treasure, or not. The test group received information about the player character's alignment, in the form of a short text that was written as an inner monologue by the player character. The text described the character's thoughts about the choice they were about to make. In six of the instances, the inner voice was opposed to advancing with the external goal, while in four control instances they agreed to advance. The choices remained the same between test- and control group, however the control group were not given the inner monologue that informed them about their player character's internal problem.

The following list contains the final ten Crit Choices, together with whether the choice is conflicted or not:

1. Take a generic piece of treasure.
2. Break down a door, in order to possibly find more treasure (conflicted).
3. Take an emotional piece of treasure; a wedding ring (conflicted).
4. Your brother goes mission: Ignore it to search for more treasure (conflicted).
5. Take a plot relevant item; a figurine.
6. Approach a potential dangerous individual; a crying ghost child, to gain treasure.
7. Ignore the distress of the ghost child and instead inquire for more treasure (conflicted).
8. Your brother has been possessed by a ghost. Focus on stopping him from destroying treasures, instead of helping him (conflicted).
9. You pacify the ghosts and they turn into a valuable figurine. Take it with you to keep it safe.
10. Leave your brother behind to keep all the treasure for yourself (conflicted).

When the Crit Choices had been planned, the rest of the game was build around them in a flowchart structure. This gave the players an illusion of agency[7], because they were able to make choices which lead to different feedback, according to their actions, but it was ensured that all participants were taken to the ten Crit Choices, in the same order. Furthermore, it was also ensured that all paths in the flowchart would yeild the knowledge players needed to feel like they were uncovering the mystery plot. This was done to make sure the player felt like they made the right choice, and were "winning" the game, which according to Thue et al.[23] made them more likely to feel a sense of agency.

Player Character

In order for the player character to have an alignment and inner thoughts, it had to be a character not a cypher. The player had to have power over the character's actions though, and they should be able to make decisions that conflicted with the wishes of the character, and that was why it was decided to have a fixed background customizable character[10]. The player could customize the character to their liking, decide on their name, gender, and appearance. In addition they made the decisions for the character, most often down to the word they would say. There were, however, parts of the character they could not control. The character is a treasure hunter by profession, they have a brother, and they have a certain moral code which is apparent in the inner monologue.

Building the Game in HTML5

The game was built using HTML5, and could run in any modern internet browser, which allowed the test to be put online for distribution. The game was made with a responsive layout that would fit properly into screens of any resolution and aspect ratio. No sound, or any other auditory feedback, were used for the game, because it was impossible to properly log whether a participant heard the sound. We wanted for participants to have similar experiences, and thus sound was excluded, as it possessed a hard to solve challenge of blocking the influence of sound.

The graphical user interface (GUI) was kept simple with few interactions. The player could click or tap to advance the text or select an option from a menu of choices. Text was displayed at the bottom of the screen, and each text box was situated beneath the character currently speaking, an example of this can be seen on Figure 1.

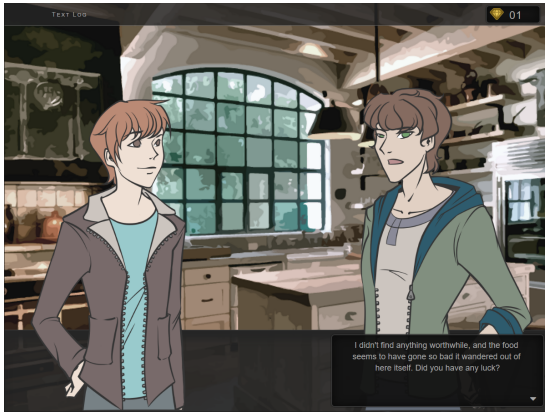


Figure 1. Screenshots from the game, Porcelain Hearts. Is shows a conversation between the player character and the brother character. The text box is situated beneath the speaking character

A pop-up would be shown for 1900 milliseconds whenever the player was awarded a treasure, and the player would be able to close this pop-up prematurely by clicking or tabbing it. Another interaction was added to the game, namely a text log where players could access previous displayed text, allowing players to read text they accidentally skipped. See Figure 2 for an example.

Choices are the most important form of interaction for the prototype. When the player is about to make a choice, the scene desaturates to add focus to the menu buttons. The buttons are placed either at the center of the screen or to the right, depending on whether the space to the right is taken up by a character. An example can be seen on Figure 3.

Data Collection

The prototype was put online to reach a large and diverse sample of players, but it also had other benefits. The game had to be of a certain length, in order to have a proper dramatic arch and to fit in all ten Crit Choices, and having an online test allowed for multiple tests to be run simultaneously without supervision, thereby saving time and ultimately gathering more data than would have been feasible within a laboratory



Figure 2. Screenshots from the game, Porcelain Hearts. Near the top-left corner the button for opening the text log can be seen, while the score of the game is located in the opposite top corner. The treasure pop-up for when the player acquires a treasure is displayed at the bottom center of the screen.

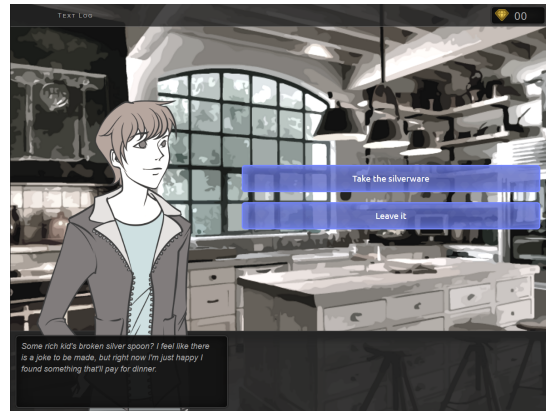


Figure 3. Screenshots from the game, Porcelain Hearts. The scene has been desaturated and the menu buttons are arranged to the right of the centre. This is the first Crit Choice, where the player has to decide whether to take the silverware or to leave it behind.

test setting. In addition, having participants play the game on their own devices, and at their own leisure, made for a better representation of how they would play a game normally.

But online testing also had certain limitations. It was impossible to fully monitor players while they were playing, nor was it possible for them to ask questions during the test, or for the conductor to interview them afterwards. Discourse analysis and other types of observation were also not possible. It was therefore important to log data that could be used to assess the quality of the data, and in addition the test was conducted in a laboratory environment, before it was put online for final testing. Data was logged to an online MySQL database using PHP scripts, and all interactions with the game was logged asynchronously to the database, with each entry holding the data necessary for any play session to be recreated from the database.

Distributing Participants

Several measures were taken to ensure the quality of the collected data. A demographic questionnaire was included which

had to be submitted before participants were taken to the game. The purpose of the questionnaire was to distribute similar participants evenly between test groups. The sorting algorithm first considered gender balance, then size balance, and finally the amount of hours they reported spending on playing video games for an average day.

The first pass of the algorithm would check whether one of the test groups were under represented by the gender of the participant, and if so, the participant would be distributed to the under represented group, to even out the gender balance. The second pass would see if any of the test groups were smaller than the other, and if so, the participant would be put in the smaller group to even out the size difference. For the third pass, the participant would be put in whatever group would bring the mean value for hours spent playing games on an average day between test groups closer together. Meaning if the reported time by the participant was larger than the mean of both groups, it would be put into the group with the lowest mean. Likewise, if the reported time by the participant was smaller than both means, the participant would be put into the group with the highest mean. If the time reported by the participant was between the two means, the reported time between groups would even out no matter which group the participant is placed in, but in this case the participant would be put in the group which had the highest difference between mean and the time reported by the participant. This way the new entry would have the highest impact on the difference between means, and thus even out the difference between test groups the most. Gaming experience was chosen above other features, because of an observation done by Thue et al.[23], when they did their experiment, and found that player experience had an influence on their data.

Questionnaires

Research done in the field of player characters often utilize questionnaires as data, and since the platform is online and seen through a browser, it made sense to look into this and use it for the test as additional data. A pre- and post-test questionnaire were made to be an integrated part of the prototype. The appearance of the questionnaires were designed to resemble the game, also using assets in the representation. This was done to mitigate the amount of people leaving prematurely, due to an otherwise unappealing questionnaire being their first impression of the prototype.

The demographics of the players were gathered using a demographic questionnaire, before they were taken to the game. In addition to the demographic questionnaire being used to distribute participants, it was also used to get an overview over the demographic of participants in the test, and to assess individual participants to see if the same person appeared in the test data twice. The demographic questionnaire contained the following items, in the order they are presented:

1. Gender
2. Age
3. Country

4. Occupation
5. Which platform are you watching this page from?
6. How many hours do you spend playing games on an average day?
7. You are willing to read in games
8. Is English your first language?
9. Do you suffer from any condition which could affect your reading speed?

Item # 1 required the player to pick between Male, Female, and Other, and had a default value set to Male. For item # 2 the participant had to fill in their own age, and the field restricted the input to numbers between and including 1 to 99. Item # 3 had a list containing all nations for the participant to choose from, while item # 4 had a list of generic occupations to pick from. Item # 4 also had an optional comment field for participants to give further information about their occupation. Item # 5 had the participant pick between PC / Mac, Tablet / iPad, Smartphone / iPhone, and Other. Item # 5 had a default value set to PC / Mac, and an optional comment field underneath Other, for participants to tell which platform they were using in case of Other. Item # 6 instructed participants to assess and report the amount of hours they spent playing games on an average day, and the input was restricted to numbers between including 0 to 24. It may seem illogical to have allowed participants to report 24 hours, but there was no logical maximum restriction to this number, besides the amount of hours in the day. A 5-point Likert scale was used for item # 7, with points labelled from left to right: Strongly agree, Disagree, Neutral, Agree, Strongly Agree. Item # 8 had the participant choose between a simple Yes and No, and had a default value set to Yes, as it was expected that the average participant would be native English speaking. Finally, item # 9 had the player choose between Don't Know in addition to Yes and No, and had a default value set to No. Item # 9 also had an optional comment field for participants to tell the cause in case of Yes.

After the game had ended, the players were instructed to fill out an evaluation questionnaire. The evaluation questionnaire consisted of four items, which were all 5-point Likert scale questions. Each were taken from Busselle et al.'s paper[3], and item # 1 and # 2 measure for emotional engagement, while item # 3 and # 4 measure for empathy towards characters in the setting. These were chosen due to them being relevant for the player's relationship with the player character.

1. The story affected me emotionally (Emotional engagement)
2. During the story, when a character succeeded, I felt happy, and when they suffered in some way, I felt sad (Emotional engagement)
3. I felt sorry for some of the characters in the story (Empathy)
4. I was worried for some of the characters in the story (Empathy)

Security

Security is an important aspect of anything which is accessible online, and it was also important to protect the quality of the data stored for this test. Without proper security, data can be lost or altered, whether deliberately caused by an attacker or because of an error in the system.

For a MySQL database, data entries are stored in tables, and one security measure taken was to validate any posted entry online, before the entry was placed in a table. If the entry contained improper data, the entry was ignored. Furthermore, the system had no entry points for altering or dropping entries from the database, to ensure the protection of stored data.

This leaves the system vulnerable to proper but false data posted to the database from outside the game. Timestamps and IDs were given to entries by the server, before they were placed in a table, and having these timestamps and IDs would make it possible to separate otherwise similar entries from one another, and see when the entry was added to the server. The timestamp was given by the server to ensure its legitimacy, and the time of submission is crucial to reconstruct a play session, which can be used to validate the likelihood of an entry having been submitted by the game.

Upon being taken to the game, a new play session would be logged to the database, and the local-site part of the system, i.e. the game, would be given a 18 character long authentication key, that had to be passed to the server-site system together with anything, that should be logged to the database for that play session. The server-site system ran using PHP, and compared the parsed authentication key to the one logged in the database for the play session. If the two keys were not the same, the input would be rejected, because the validation of the play session failed.

The authentication key has $1.83253e32$ different combinations, which protected participants' data from being corrupted by other people. However, the system was still vulnerable to false data being posted for a participant's own play session, but this was a much easier problem to handle. The database had a table containing all play sessions, and all other tables which hold play session data, i.a. what choices participants have made, had the respective session ID assigned for all entries, so the session it belonged to could always be found, and play sessions could be recreated, without getting data from other sessions mixed in. No play session data could be logged to the database without a successful validation of the play session, and it would be fairly easy to reconstruct play sessions for validation, if there were any reason to believe someone had attacked the database. False play session data could therefore easily be excluded by removing the corrupted play sessions from the test.

LABORATORY TESTS

Before the prototype was distributed online for testing, two local tests were conducted to assess the quality of the prototype, starting with a pilot test of the prototype while it was still being developed.

Pilot Test

The purpose of the pilot test was to assess the functionality of the prototype, players understanding of the GUI, the quality of the narrative, and to develop and practice the test procedure for the preliminary test. The pilot test consisted of half the game content, the demographic questionnaire, and the evaluation questionnaire.

To test the quality of the prototype, another post-test questionnaire was added for participants to evaluate the narrative. All items in the questionnaire used a 5-point Likert scale. The first two items were questions related to the quality of the game, and the player's understanding of the narrative. Item 3-6 were from a study about the illusion of agency[7]. Item 7-10 were about Narrative Understanding, which is a part of Narrative Engagement[3].

In addition, there were a two-item, moral choice questionnaire, used for evaluating each Crit Choice. One item was used to rate the difficulty of the choice, and the other for rating how emotional the choice had been, both using a 5-point Likert scale. The items in the moral choice questionnaires were taken from a paper which investigated the influence of stress on moral decision making[22].

A semi-structured interview was conducted after each test session. This was mainly used to find issues with the GUI and narrative, and allowed participants to share any thoughts they might have had about the test and prototype.

The pilot test had 10 participants, and the results revealed some shortcomings with the GUI, and in the test design, that was hence corrected. The results from the moral choice and post-test questionnaires were promising, and the items were therefore kept for the preliminary test.

Preliminary Test

The test was designed to be a replica of the final test, but performed in a local environment, where it was possible to observe the participants and perform interviews. The goal of the preliminary test was to assess the quality of the prototype and locate errors in the program, before final testing.

When looking at the database after the preliminary test, there seemed to be no data missing, nor flaws in the stored data. Two errors with the game system were observed during the test, and was corrected before the prototype went online. The interviews revealed a good understanding of the GUI elements, and the narrative.

The demographic questionnaire revealed a participant who had reported 24 hours for the item about time spent playing games on an average day, an amount which is not feasible, and therefore the participant must have deliberately reported false data.

The results from the post-test questionnaire all leaned towards the end of the scale that was expected, though less so than they had for the pilot test. The reason behind this was investigated, and when looking at the demographic questionnaire, for how willing participants were to read text in games, the median value was 4 (Agree) for the pilot test and 3 (Neutral) for the

preliminary test. It is possible this may have caused the results from the post-test questionnaire to center more around Neutral for the preliminary test, as the lack of willingness to read may have caused the participants to be less engaged with the prototype.

The participants in the preliminary test did play through the whole game, while the participants in the pilot test only played through half of the game, and as such it is conceivable it may simply be the quality of the latter content of the prototype causing this, or the added length of the game. The participants of the preliminary test did, however, not mention anything during the interview when asked about the quality of the narrative that could lead to believe the latter part of the game was somehow inferior to the early half.

The results the moral choice questionnaires showed no difference in the median value for difficulty or emotional impact for Crit Choices between test groups. All Likert scales have a median of 2 (1: Not at all, 5: Extremely) which suggested the average choice was not too difficult or emotional, but this was not necessarily a problem. The results for individual choices and participants ranged between 1 and 5 for both difficulty and emotional impact, and there were cases of choices being ranked on each opposite ends of the scale, and being ranked similarly on either end of the scale. The results suggested that participants experienced a diverse level of difficulty and emotional impact in the Crit Choices, both between participants for the same choice, and between choices for the same participant.

The demographic- and evaluation questionnaire remained unchanged through these tests, and the Crit Choices were unaltered since the preliminary test. The game and narrative remained unaltered, aside from a few grammar and system corrections, before online testing.

RESULTS

The test was online from 12th of April to 2nd of May, for a total of approx. 19 full days. By then the test had accumulated 467 completed play sessions. During this time, the game was distributed through social media, and promoted at local game related events; a game jam and a board game event.

Target Group

The test was distributed online, which resulted in a fairly diverse target group. However, it must be recognized that approaching people through social media sets some limitations to who might be reached, and that different groups would not be equally represented in the data, because of this.

Noise Reduction

After having accumulated the data, several measures were then taken to reduce noise, by removing participants with improper data entries.

Multiple Sessions

Duplicate sessions by the same IP address were removed, to diminish the chance of one player appearing in the sample twice. Only completed sessions, that were also the first session by any given IP, were considered. This was done to to exclude

players, who had played parts of the game, and then gone back to the start, and created another session. This would also exclude players who may have observed someone else play the game, prior to their own play session, if they played on the same device.

Going through the demographic submissions, two sessions on different IP addresses were found to likely be the same person, either returning on another device or after having received another IP address, and the latter session was therefore removed from the sample. This gave a starting sample size of 446 sessions.

Missing Crit Choices

Data entries were missing for one or more Crit Choices for 12 participants in the sample. These then had to be removed due to them being incomplete data entries. This technical error happened due to the fact player actions were posted to the database asynchronously, so it would not disrupt play sessions, but it also invited the possibility of data being lost if the call failed. It is possible that more calls were lost, but for the remaining samples, all important data was present in the database.

Willingness to Read

Participants reported in the demographic questionnaire, how willing they were to read in games, on a 5-point Likert scale. Participants who had reported an unwillingness to read in games were excluded from the sample, because it was found during prior testing that this could cause participants to be less invested with the game, and have a worse experience. 18 sessions were removed from the sample. See Table 1 for participants by their willingness to read in games, before exclusion.

Table 1. Participants by willingness to read, when ranking themselves according to the phrase: You are willing to read in games

Willing to read	Participants (Percent of Total)
Strongly disagree	3 (<1 %)
Disagree	15 (3 %)
Neutral	31 (7 %)
Agree	164 (38 %)
Strongly agree	221 (51 %)

Questionable Demographic Data

In prior testing, a participant was observed to report a number that was not feasible for the item in the demographic questionnaire about how many hours they spent playing games on an average day. Therefore John Tukey's outlier filter, with an IQR of 1.5, was performed to find outliers for that item, and 9 sessions, by participants who reported they spend more than 8.5 hours a day playing video games, were removed. It was decided to remove outliers because it is likely they either did not take the test seriously, or were unable to make a proper self assessment, which were qualities that could harm the rest of the sample.

Improper Play

It was important to exclude sessions by participants who had not played the game properly, as players who did not, could harm the test results. The test was reliant on participants having read the text given to them, and for them to have

properly assessed the choices they were presented with. Using The Model Human Processor[4], three cases that could merit exclusion from the test was formulated:

1. The completion time of the participant was faster than what was physiologically possible for an average person, without skim reading.
2. The time spent deciding on a Crit Choice was less than what was physiologically possible for even the fastest of people.
3. The participant was an outlier in the amount of content skipped, i.e. the number of times the participant used less time reading a text box, or making a choice, than is physiologically possible for an average person, without skim reading.

Using examples from the article, models for estimating the minimum required time for reading text and making a choice was constructed, see Table 2. The model for estimating the minimum required time for making a choice expected participants to have at least read two options of a choice, for them to have made a proper decision. When using the model, it was assume players read choices from the top down, so the word count for this estimation was taken from the two first options from the top.

Table 2. Models for estimating the minimum required time for reading text and making a choice.

Reading a Text Box	$words \cdot ms/word + 2T_p + 2T_c + T_m$
Making a Choice	$words \cdot ms/word + T_p + 2T_c + T_m$

The maximum amount of words one could read within a minute, without skim reading, was 652[4], limited by eye movement, given that it would take the person a single move of the eye for each word. This gave the expected minimum time to read a word to be approx. 92 milliseconds.

The model human processor is composed of the perceptual processor, the cognitive processor, and the motor processor, and the time it takes for a pass of each is denoted T_p , T_c , and T_m , respectively.

Processing time

- $T_p = 100[50 \sim 200]milliseconds$
- $T_c = 70[25 \sim 170]milliseconds$
- $T_m = 70[30 \sim 100]milliseconds$

From these numbers, we can estimate the minimum required time for each session and case, by reconstructing play sessions from the database.

Case 1

For the first case, the completion time of the session is compared to the cumulative minimum time required to read the text and decide on the choices given to the player. If the completion time is faster than the minimum required time, the session is removed from the sample. It is possible some text box or choice entries may have been missing in the database, which would cause the estimated minimum required time to be smaller for sessions with missing entries, benefiting the

chances for the session to be kept in the sample for this case. For this case, the average numbers of T_p , T_c , and T_m were used, due to it looking at overall performance.

Case 2

For the second case, the minimum required time for each Crit Choice is estimated, and if a participant decided on a Crit Choice faster than the minimum require time, the Crit Choice was not made properly, and would have to be removed from the test. In the case of any Crit Choice being removed from the test, the whole participant will have to be excluded from the sample, as they would have incomplete data. For this case, the smallest numbers of T_p , T_c , and T_m were used, because it is conceivable a participant could be particularly fast for any single choice, and to give each individual choice the best chance of being kept in the test.

Case 3

For the third case, the minimum required time for each choice and text box was estimated. The times a participant used less time on a choice or text box, than the estimated minimum, was counted as skipped content. Using an IQR of 1.5, outliers were found to be participants who had skipped on content more than 105 times. For this case, the average number of T_p , T_c , and T_m were used, since the case is looking at the time for all text boxes and choices.

The total failed cases were 42, but some sessions failed several cases, and the 42 failed cases were shared among 38 sessions, which were removed from the sample. After having excluded the entries which could not reject these cases, 369 entries remained.

Analysis

Before noise reduction, the number of completed play sessions by each gender were equally distributed between test groups, with 162 male participants and 69 female participants in each. After sessions were removed, the demographic had become a bit skewed between test groups, because more male participants had been removed from the test group, while more female participants had been removed from the control group. The size of both test groups remained similar. A summary of the demographics can be found in the Table 3, and Figure 4 and 5 for the spread of participants between nations and platforms, respectively.

Table 3. The demographics of the remaining entries after the data was put through filters.

	Test Group	Control Group
Gender (Male/Female/Other)	120M/60F/2O	133M/54F/0O
Mean Age +/- SD	25.01+/- 4.80	24.65 +/- 4.17
Mean Game Experience +/- SD	2.65 +/- 1.48	2.58 +/- 1.41
Total Participants	182	187

Crit Choices

The ten Crit Choices were tested across test groups, using a two-sample t-test. The test yielded Crit Choice # 1, # 2, # 3, # 5, # 8, and # 9 to be significantly different across test

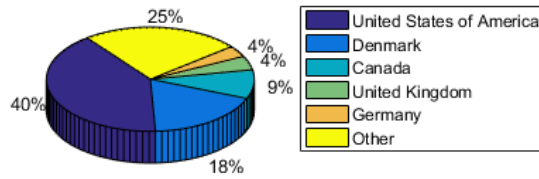


Figure 4. A 3-D pie chart showing the spread of participant across nation.

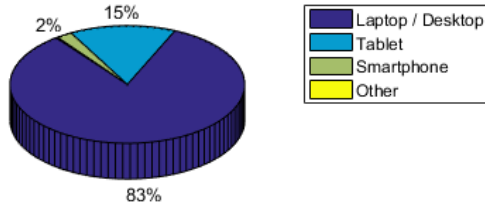


Figure 5. A 3-D pie chart showing the spread of participant across platform. Only one participant (<1 %) reported other as their platform.

groups, see Table 4. Even when considering the Bonferroni correction for 10 hypotheses, all except Crit Choice # 2 remained significant.

When looking at the ratio for each choice across test groups, it could be seen that for Crit Choice # 2, # 3, and # 8, participants in the test group pursued the goal of the game to a lesser degree, see Figure 8, 7, and 10, respectively. For Crit Choice # 1, # 5, and # 9, participants in the test group pursued the goal of the game to a higher degree, see Figure 6, 9, and 11, respectively. For all six significant Crit Choices, participants in the test group aligned their decision with the alignment of their character to a higher degree.

Table 4. A list consisting of the p-values when running a two-sample t-test on each individual Crit Choice, across test groups.

Crit Choice	p-value	Crit Choice	p-value
#1	0.0031	#6	0.5722
#2	0.0238	#7	0.1204
#3	2.9992e-06	#8	1.2890e-08
#4	0.4723	#9	0.0002
#5	8.0789e-13	#10	0.7406

Evaluation Questionnaire

The items in the evaluation questionnaire were tested using a Mann-Whitney U test, in order to find a significant difference between test groups. The test yielded no significant results for any of the four items, see Table 5.

Table 5. p-value by evaluation questionnaire item

Item	p-value
1	0.9459
2	0.0973
3	0.6535
4	0.9865

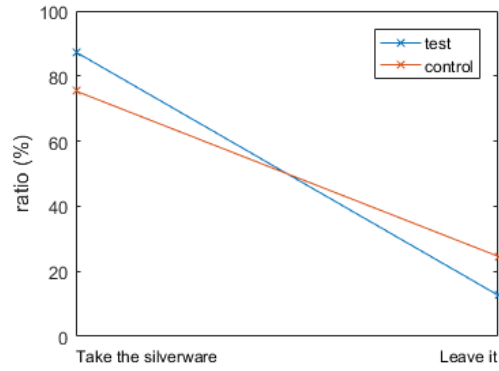


Figure 6. Crit Choice 1: Here the player must choose whether to pick up some abandoned silverware. The PC agrees with taking the treasure.

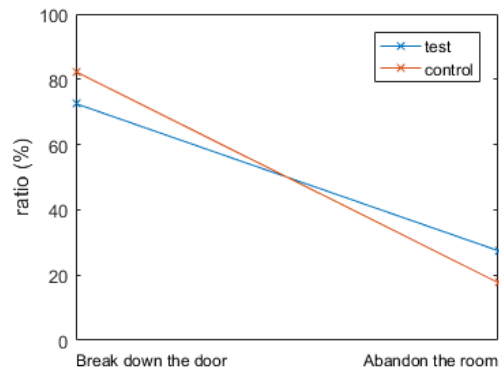


Figure 7. Crit Choice 2: Here the player must choose whether to break down a door to search for more treasure. The PC disagrees with taking the treasure, i.e. abandon the room.

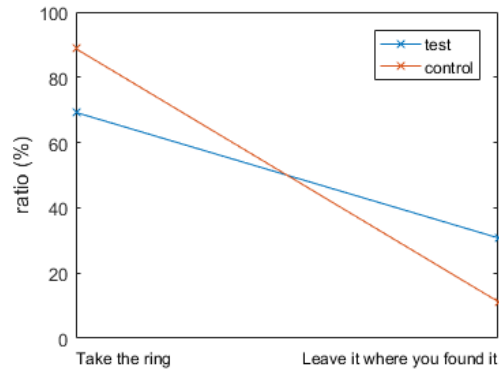


Figure 8. Crit Choice 3: Here the player must choose whether to pick up a wedding ring. The PC disagrees with taking the treasure.

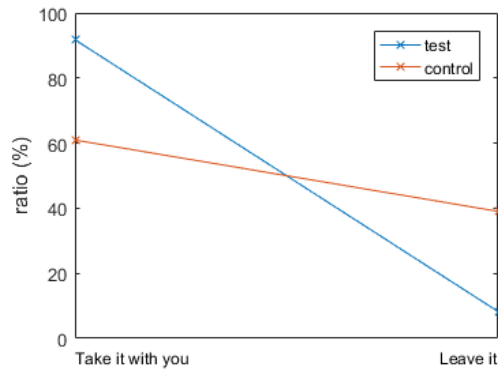


Figure 9. Crit Choice 5: Here the player must choose whether to pick up an abandoned figurine. The PC agrees with taking the treasure.

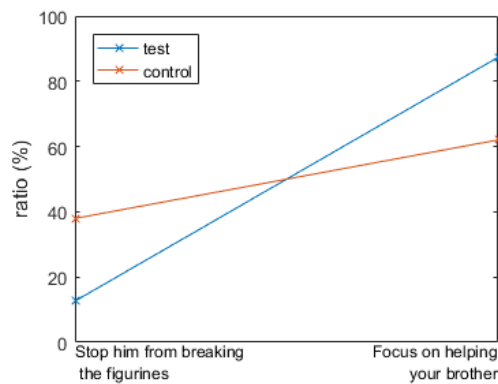


Figure 10. Crit Choice 8: Here the player must choose between helping their brother, or preventing the destruction of treasure. The PC disagrees with taking the treasure, i.e. helping their brother.

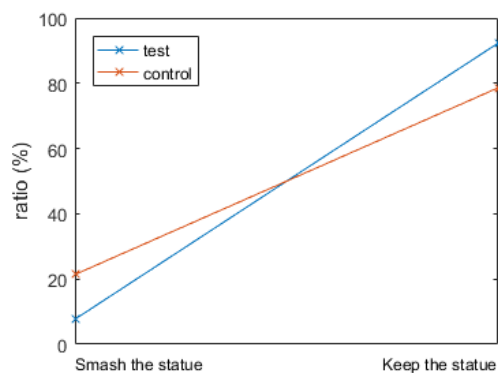


Figure 11. Crit Choice 9: Here the player must choose between destroying the figurine of the two ghost children, or to bring it with them. The PC agrees with taking the treasure.

DISCUSSION

From the results it can be seen that giving the player knowledge about the player character's alignment significantly impacted some of the choices they made while playing. In these instances, the players are significantly inclined to make choices based on what their character wishes, even going so

far as to being reassured to take treasure when the control group was hesitant.

There were some instances in which there was no significant difference. Crit Choice # 10, in which the player had to choose between more treasure or their brother, the vast majority, over 95% in both test groups, chose to help the brother. The reason for this could be because interaction with the brother character had caused players to already be emotionally attached, or that the innate sibling relationship between the player character and the brother character made the player unwilling to abandon the brother character at any rate, and thus the inner monologue had little effect on the outcome. The same could be said for Crit Choice # 4, where the player had to decide whether to look for more treasure or their missing brother.

For Crit Choice # 6, players had to decide whether to approach the ghost sitting on a treasure chest, or to leave her and the treasure be, and for this choice both test- and control group chose to approach the ghost, which was the predicted outcome since the choice was not conflicted. However, it is interesting in context of Crit Choice # 7, where they can either ask about treasure or to the ghost's crying, which was a conflicted Crit Choice. Here over 90% of both test- and control group choose to inquire why the ghost is crying, instead of asking about the whereabouts of more treasure. Similar to Crit Choice # 10, the emotional value of a crying child might be what provoked players to abandon their goal, and the effect of the inner monologue was therefore not apparent.

The results from the evaluation questionnaire yielded no significant results. Either the test group did not feel more emotionally engaged nor empathic towards the character, or perhaps the items were not focused enough to catch the difference. It may have been better to test the quality of more items during the laboratory tests, and then keep the few which showed the most promise for the online test, to increase the quality of the evaluation questionnaire without adding more items to the final test. Measuring differences in questionnaire results in a between-subject test can be difficult, and would arguably be better suited for a within-subject test.

CONCLUSION AND FUTURE WORK

From the data gathered, we can conclude that giving the player knowledge about their player character's alignment had a significant influence on the way they behaved in our prototype. Player in the test group aligned their behaviour more to the wishes of the player character, when their wishes were made apparent. We cannot, however, conclude on our evaluation questionnaires, that the test group felt more emotionally engaged or empathic.

The next step for the research would be to look into whether the change we found between test groups is also apparent in other game genres. The visual novel genre is by its definition a genre that weights story highly, and it would therefore be interesting to make a similar test in a genre like platformer or puzzle, which is usually less focused on story, and see if those players are willing to abandon their goal for the sake of the story.

Another important aspect would be to look into other methods

for revealing the player character's alignment, and see if they foster the same change in behaviour. Since different types of player characters might reveal their alignment in different ways, it is not necessarily attractive to use inner monologues in all types of video games.

Our game did not include auditive feedback, but for a game that those, it would be interesting to see if e.g. the sound of an increased heartbeat, would make players reluctant to take a piece of treasure.

We saw that players in both test groups were willing to abandon their goal for the sake of a crying child, and it could therefore be interesting to investigate the emotional influence of the crying attribute alone.

With the knowledge gained here, designers can reveal knowledge of the player character's alignment, to make players abandon the goal of the game to pursue their character's wishes, and thereby provoke a certain behaviour in a game.

REFERENCES

1. BioWare, Edge of Reality, Demiurge Studios, and Straight Right. 2007. Mass Effect. Microsoft Game Studios (2007) and Electronic Arts (2008 to present). (2007).
2. Luis Emilio Bruni and Sarune Baceviciute. 2013. Narrative Intelligibility and Closure in Interactive Systems. *ICIDS* (2013), 13–24.
3. Rick Busselle and Helena Bilandzic. 2009. Measuring Narrative Engagement. *Media Psychology* 12, 4 (December 2009), 321–347.
4. Stuart K. Card, Thomas P. Moran, and Allen Newell. 1986. The Model Human Processor. *Handbook of Perception and Human Performance* (1986).
5. Mark H. Davis. 1980. A Multidimensional Approach to Individual Differences in Empathy. *JSAS* 10, 85 (1980).
6. BioWare Edmonton and Edge of Reality. 2009. Dragon Age: Origins. Electronic Arts. (2009).
7. Matthew William Fendt, Brent Harrison, Stephen G. Ware, Rogelio, Cardona-Rivera, and David L. Roberts. 2012. Achieving the Illusion of Agency. *ICIDS* (2012), 114–125.
8. Edward Morgan Forster. 1995. *Aspects of the Novel*. Harcourt Brace.
9. Toby Fox. 2015. Undertale. (2015).
10. Tobias Heussner, Toiya Kristen Finley, and Jennifer Brandes Hepler. 2015. *The Game Narrative Toolbox*. Focal Press.
11. Charlene Jennetta, Anna L. Coxa, Paul Cairnsb, Samira Dhopareec, Andrew Eppsc, Tim Tijds, and Alison Waltond. 2008. Measuring and defining the experience of immersion in games. *International journal of human-computer studies* 66, 9 (2008), 641–661.
12. Petri Lankoski. 2011. Player Character Engagement in Computer Games. *Games and Culture* 6, 4 (2011), 291–311.
13. E. J. Lawrence, P. Shaw, D. Baker, S. Baron-Cohen, and A. S. David. 2004. Measuring empathy: reliability and validity of the Empathy Quotient. *Psychological Medicine* 34, 5 (2004), 911–920.
14. Michael P. McCreery, S. Kathleen Krach, P.G. Schrader, and Randy Boone. 2012. Defining the virtual self: Personality, behavior, and the psychology of embodiment. *Computers in Human Behavior* 28, 3 (2012), 976–983.
15. Michael Nitsche. 2008. *Video Game Space*. MIT Press.
16. Melanie Anne Phillips and Chris Huntley. 2004. *Dramatica* (10 ed.). Write Brothers.
17. Hua Qin, Pei-Luen Patrick Rau, and Gavriel Salvendy. 2009. Measuring Player Immersion in the Computer Game Narrative. *International Journal of Human-Computer Interaction* 25, 2 (2009), 107–133.
18. Nintendo R&D4. 1985. Super Mario Bros. Nintendo Entertainment System. (1985).
19. Kristine Jørgensen. 2010. Game Characters as Narrative Devices. A Comparative Analysis of Dragon Age: Origins and Mass Effect 2. *Journal for Computer Game Culture* 4, 2 (2010), 315–331.
20. Marie-Laure Ryan. 2008. Interactive Narrative, Plot Types, and Interpersonal Relations. *ICIDS* (2008), 6–13.
21. Marie-Laure Ryan. 2015. *Narrative as Virtual Reality 2: Revisiting Immersion and Interactivity in Literature and Electronic Media*. Johns Hopkins University Press Baltimore.
22. Katrin Starcke, Christin Polzer, Oliver T. Wolf, and Matthias Brand. 2011. Does stress alter everyday moral decision-making? *Psychoneuroendocrinology* 36, 2 (2011), 210–219.
23. David Thue, Vadim Bulitko, Marcia Spetch, and Trevon Romanuik. 2011. A Computational Model of Perceived Agency in Video Games. In *Proceedings of the Seventh AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*.
24. Daniel Vella. 2014. Modeling the Semiotic Structure of Game Characters. *DiGRA* (August 2014), 3–6.