



# Recreating Believability In NPCs: The Effects Of Visual And Logical Behaviour

Simon Ohrberg

Computer Science

Bachelor

15 Credits

Fall 2019

Supervisor: José Maria Font Fernandez

# Abstract

NPCs are in many games the foundation on which it operates. NPCs create the illusion of inhabitants or assign purpose to the player. Whatever they do, represent a character. Introducing a character has certain margin of error, as a poorly portrayed character may cause more damage than the NPC would add. Without creating a believable environment which would include the NPCs, immersion could be difficult. As such this thesis investigates the effect of different behaviours to NPCs. With the focus on how visual and logical behaviours affect the players' perception of believability. The experiment was conducted in a game of the RTS survival genre with visual behaviours selected from the games Banished[1] and Frostpunk [2]. The logical behaviours were inspired by F.E.A.R [3]. In the experiment testers were used to test three versions of the artifact. The first acted as a default was used as the starting point for the second version which introduced enhanced visual behaviours. The third continued and added a predictive functionality as logical behaviour. The tests concluded that the visual behaviours had a positive effect on perception but no conclusive evidence to suggest the logical behaviour had the same effect.

# 1. Introduction

Modern NPCs have the opportunity to move around in 3D environments and they fulfill many kinds of roles, such as quest-givers, companions or enemies. In some games these NPCs move and behave in manners we can relate to such as the NPCs in Banished [1] and Frostpunk [2].

In the survival RTS Banished, the player is in control of an emerging settlement. It starts off small and it is the task of the player to form this settlement into a thriving community. The player must make sure the villagers do not freeze to death, starve or die from disease. This is done by ordering the construction of relevant buildings and assigning workers to occupations to fulfill the demand by the settlements buildings. The residents of the settlement will be seen visually walking around and not only performing their tasks. But also perform various actions which includes, walking to a market building or their house in order to eat. They can go to a blacksmith to get new tools. While unassigned the residents will become laborers.

Frostpunk works much like Banished in that the player is in control of an emerging settlement. Gameplay wise it offers different kinds of challenges but the two games share a common core of maintaining a population by indirect control. Workers are assigned to buildings but the player has little control in which NPC will be working where. Frostpunk also shares the move around function as Banished however under different circumstances.

In other games they are perceived as making 'smart' choices, as in F.E.A.R [3]. Which is an FPS in which the player progress the map and story one level at the time. In order to progress the player must often defeat enemies along the way and is given ample tools to combat enemies. These enemies will try to flank you if you are behind cover. They will reach for cover themselves and support other enemies as they try to flush the player out by using explosives or manoeuvres. The enemies operate on several levels of AI but at their core they utilize a goal-oriented action planner (GOAP) as explained by Orkin [4]. He also explains that often the enemies in F.E.A.R seem more intelligent than they necessarily are, which is credited in part to the GOAP.

## 1.1 Purpose

This motivation to investigate comes from a previous game assignment in which Nilsson et.al created the game, Oil [5]. The game was inspired by games such as Banished [1] and Frostpunk [2]. The projects' NPCs were perceived as robots by testers, even though design decisions were made to make the NPCs relatable.

The games previously mentioned, contain behaviours which may improve the believability the NPCs in games that use NPCs. By having characteristics they become, to various degrees, believable characters. According to Ermi [6] if the illusion of a believable agent is broken, then so is the immersion.

In certain games such as Banished and Frostpunk the NPCs are always on the move. In short they can be put to work and when they aren't actively working they're moving about their town. Another interesting behaviour can be seen in both of these titles, where the NPCs

need to be at specific locations in order to perform certain actions. The actions differ between the titles but between them they include eating, drinking, warming up, resting, working or delivering resources. These two visual techniques provide the player with some form of visual stimulation when the NPCs walk around and perform their tasks. For context games such as the Warcraft 3 [7] and Starcraft 2 [8] series by Blizzard Inc, do not employ behaviours of this kind. Instead the units in these RTS games only move on command, with some minor exceptions.

Orkin [4] writes about some of the structure behind the NPCs in F.E.A.R [3] and how their decision making is done. F.E.A.R doesn't focus solely on the visual aspect of their NPCs but also adds a layer of decision making.

Salen and Zimmerman [9] talk about 'Design and Meaning' in which meaning can be said to be derived from 'signs'. These indicators aims to make the player associate with a certain concept. This is likely where Oil failed and subsequently the purpose of this research. How could certain 'signs' be introduced in such a way that the NPCs are interpreted as people; believable NPCs.

There are some research surrounding believable NPCs such as Warpefelts' [10] dissertation concerning the believability of NPCs. In which Warpefelt lists three main mechanics which players use to identify which "role" an NPC has.

- Surrounding area and location of the NPC
- Actions taken by the NPC
- The NPCs' attributes and visual presentation

He goes on to identify "affordances" which describe situation context much like the 'signs' described by Salen and Zimmerman. Warpefelts research lays the groundwork for this paper as these mechanics lie at the core of this paper.

Another related paper was written by T Lindstram and A. Svensson [11] in which they try to create "life like" NPCs in a village environment. They use a combination of a behaviour tree and a planner which they conclude can increase the believability in NPC agents. This relates to this paper as they try to utilize methods which may not have been made for that purpose in order to increase the believability of their agents.

The main goal of this thesis is to explore how a visible behavioral implementation to an action, performed by a population of agents, alters the users' perception of their believability. It will also explore if this effect can be amplified with a logical behaviour added to the NPCs, enabling them to make tactical decisions. In particular for this thesis the NPCs are capable of tactically deciding when to perform certain actions. These particular focuses are meant to evaluate how the differences in perception change with the additional 'signs'. Ultimately certain types of signs may be more obvious therefore may serve a restricted developer better to implement into their game than the other options.

It is expected that these changes will improve the believability of the NPCs. It is also expected that the visual improvement will allow for a significantly higher perception of believability to the NPCs. The logical implementation is not expected to be noticed by testers with little to no experience in playing games but is expected to be noticed by experienced players. As such the logical implementation is not expected to have much of an impact on the testers with a low experience in playing games.

This paper will be using the Design science method (DS), described by Peffers [19], to construct the artifact. Then in order to measure the difference in believability user tests will be performed, these tests should be set up in three stages. The first test will not include the

behaviours adapted from Banished and Frostpunk, the second test adds these behaviours. Lastly the third test introduces the ability for the NPCs to predict future needs and tactically eat or rest before walking for long periods of time. The testers will not be told which test they are testing or differences in each test. Afterwards they will be asked to fill out a questionnaire which will be used to provide the conclusion for this paper.

## 1.2 Research Questions

1. How will the visual techniques described in section 1.1 affect the perception of believability of the NPCs in an RTS survival environment?
  - a. How will the perceived believability change when adding a predictive behaviour to these techniques?

## 1.3 Limitations And Risks

A limitation is the constraint to work in a RTS survival genre. It helps keep the scientific rigor, but all results are affiliated to the RTS survival genre. There is an added risk that any results of this thesis are not applicable outside of this genre.

This thesis will aim to expand upon the artifact with each version. This provides a clear path of progression and rigor that can be maintained however it also adds limitations to the evaluation. Such a limitation is that there is no way to evaluate the artifact versions in a standalone manner against the default. This is a risk which may contaminate the results.

As the artifact version one and two builds upon the baseline version and version two build upon version one. There is a risk concerning the contributions of one version may be perceived by the tester as the product of another version. Alternatively the tester might not notice the change between the versions at all.

Lastly as this thesis is working within the subjective sphere of believability, there is a risk that the testers will have their own biases. Any change may mean something different to any given tester. This applies to implementations done in the artifacts but also bugs and other errors that were not intended, will still carry an impact on the result.

## 1.4 RTS Survival

Real-time strategy games (RTS) like Warcraft 3 [7] or Starcraft 2 [8] is, a subgenre to strategy games, played with an emphasis on strategic thinking and planning in a real-time environment. In the two examples presented players are fighting each other in an attempt to destroy the enemies' bases. To do this they need to collect, and sometimes fight for, resources.

Survival games such as The Forest [12] focus on surviving the elements, and in the case of The Forest this is mostly achieved by crafting and building whilst making sure the avatar doesn't dehydrate or starve.

RTS survival games, as far as this paper is concerned, can be seen as survival games made in a fashion to incorporate the strategic elements of RTS whilst maintaining the survival elements. Frostpunk and Banished are examples of RTS survival games in which RTS and survival are part of the core game loop. RTS survival games, such as the two games mentioned, usually adds on problems such as keeping the population fed or making sure they are protected from the elements. These problems are introduced to the player in part to increase believability and add a layer of immersion to the game.

## 1.5 Believability

Believability as used in RTS survival games and others. Apart from environmental factors believability can be found in what is considered as NPCs, but what an NPC is, is a matter of some debate. Therefore as far as this paper is concerned an NPC contains as described by Warpefelt [10], “characterhood”, a concept of a set of traits which the NPC represents. By adhering to these traits there can be a clear distinction between static objects and agents. As Warpefelt writes, a NPC must be perceived as “rational” and behave with some form of intent. This basically means that the agent must make choices that make sense. In this paper there will priorities on the “rational” and “intentional” traits as per Warpefelt’s research.

If a game simulates a character, then this character needs to be believed in by the player according to Ermi [6]. It is important to take into account that a believable character isn’t always a realistic one. For instance Loyall [13] uses Disney films as an example as to how believable characters, such as Bambi or Aladdin, are believable, but not realistic.

In order for an agent to be perceived as believable, it is required to behave in accordance to the world it inhabits. In turn, an NPC can provide the player with challenge and immersion. Immersion allows the player to be fully absorbed in the game as described by Brown [14]. As long as the game and its agents are believable, immersion will be maintained as described by Lankoski [15] which ties back into Ermis’ research into how the players’ state of mind reflect on how the game experience is perceived.

## 1.6 Immersion

Much like believability, immersion is a concept which is hard to define. However in a study performed by Brown and Cairns [14] they describe how immersion is a vital part of any game. They also break down immersion into three levels of involvement. Which allows for an easier definition and it’s this description which will be used in this thesis.

- **Engagement**
- **Engrossment**
- **Total Immersion**

Through a qualitative study they interviewed seven proclaimed gamers and had them answer their questions about immersion in games. Brown and Cairns identified a set of barriers which separated the different levels of involvement. In their research they define *access* and *attention* as the two barriers which makes players engaged in the game. Engrossment is a result of the previous two barriers combined with *game functions and design* which directly affects the players *emotions*. By becoming engrossed in a game Brown

and Cairns write that players, “[...]are involved with more than just the physical aspects of the game and have, in a sense, suspended their disbelief of the game world.”, Total Immersion is the last level. To reach this level of immersion the player must first get engrossed and through what Brown and Cairn describe as *presence*. This is achieved by the players’ empathic feelings and the atmosphere of the game. The atmosphere as described by Brown and Cairn is a combination of graphics, sounds and plot. Further it is important for the immersion that all game features make sense, much like described by Ermi and Mäyrä [6]. The difference as according to Brown and Cairn is that this increase in engaging game elements raised the attention of the player and in doing so also raises the immersion [13,6]. Murray [16] offers a different description of immersion. She describes it as a feeling of being “surrounded by a completely other reality”, with all our attention and perceptions solely focused on this immersive experience.

## 1.7 Goal-oriented Action Planner

As the implementation of new features in order to adhere to the survival aspects of a RTS survival game gets more complex. Solutions to solve this complexity were investigated.

Orkin [4] writes about the issues F.E.A.R [3] had with complexity which laid the foundation for their decision to implement a GOAP.

Orkin lists three main benefits of using a GOAP. The first is the decoupling of goals and actions allows for a modular code structure which in turn separates game data so adding or removing behaviours or unit types can be done with ease. The second Orkin lists as layering behaviours, he explains will allow the developer to add new features which the GOAP can utilize without the developer directly telling the GOAP it can. Lastly Orkins defines dynamic problem solving as the third major benefit. Which refers to the replanning the GOAP does. After each action has been done a new search to reach the goal is done and because of this the NPC can adapt to changes around it.

There are some downsides to using a GOAP. The first concerns actions which needs to be prewritten. The second concerns how prerequisites must be pre written,by the programmer, to remain logical. It is run realtime by each citizen thereby it could require more resources than the entertainment value is worth. Further there is a lot of resource usage by constantly planning and replanning actions which limits the amount of agents that can use the GOAP according to Orkin [17].

## 1.8 Finite-State Machine

A finite state machine(FSM) as described by Yannakis [18] provides the overhead for logical state transitions and allows states to run their logic and then switch to another based on some parameter. The FSM complexity ranges from a simple switch statement to actual systems acting to isolate logic in different states of transition as described by Yannakakis and Togelius. The downside to using an FSM is mostly tied to projects of greater sizes in which the transitional logic becomes unmanageably complex as described by Orkins [4].

## 2. Method

This paper uses the DS by Peffer [19]. The benefit of using this method is the balance between engineering an artifact whilst maintaining scientific rigor. Through the construction of artifacts to test new or improve upon pre-existing solutions to solve issues as explained in detail by Peffers. Peffers also talks about iteration cycles, which in the case of this thesis can be seen as the versions of the artifact that was constructed.

### 2.1 Research Process

Peffers [19] defines the Design Science process as 6 steps.

- Problem Identification and Motivation
- Define the objectives of the solution
- Design and Development
- Demonstration
- Evaluation
- Communication

#### 2.1.1 Identifying and Motivating the Problem

When the NPC believability is lacking, this lack affects the immersion. According to Browns research [14] this disruption may cause certain players to play these games less. What makes a believable NPC is defined in section 1.5. In order to investigate Salen and Zimmerman [9] wrote as mentioned in 1.1 about how signs act like indicators from which the player will assume some concept. The problem is what signs would be better to focus on, if restrictions are imposed.

#### 2.1.2 Defining the objectives of the solution

This thesis aims to provide an indicator for the which ‘signs’ have the greater effect on the outcome.

This goal has been defined into two research questions in section 1.2. In order to answer these, an artifact with two versions was built with the second version expanding upon the first. For the first version, two visual behaviors were added. These behaviours were chosen partly based on the research of Lankoski [15], Ermi [6], Loyall [13] and Salen [9]. Then partly based on the games which inspired Oils’ design, Banished [1] and Frostpunk [2], the behaviours chosen are present in at least one of these commercial titles. The behaviours were selected for their compatibility with Oil as a framework according to “Context Shapes Interpretation” [9]. These behaviours are the ‘signs’ which are aimed at enhancing the concept that the NPCs are humans, by attempting to increase their believability.

The first version contains two added visual behaviours, the first is idling movement whilst ‘resting’. The citizens will idle around a central structure if no house is provided to them. If they do have a home, they will idle around that structure instead. The second



behaviour is location based actions. In which the citizens has to move to a certain building or location before it can proceed with the action specific to that location. These are the visual signs.

The second version concerns a 'hidden' behavioural trait which allows the NPC to plan ahead of time to assess future needs that may interrupt the work process. This is the logical sign.

Orkin [4] writes how F.E.A.R [3] solves their complexity issues by implementing a planner. This thesis will produce an artifact which will receive more 'signs' and functionality as the versions increase. Potentially creating a lot of complexity, which will make the development harder and may also affect performance. In order to effectively incorporate these behaviours a planner together with a FSM was created inspired by Orkin's paper. The planner and FSM combo should provide the modularity needed with as little complexity as possible.

By implementing this artifact, it is possible to test how these techniques function as 'signs' to increase the perception of the citizens in Oil as people. Additionally this will test how the citizens' behaviour affect the immersion of a game. Lastly it will also test how visual, as opposed to logical, implementations affect this perception of believability.

### 2.1.3 The Design and Development Process

Based on the research mention in section 1.5 and 1.6. The artifact was introduced into the game Oil [5], due to the pre-existing graphics and atmosphere. The artifact was written in C# using visual studio 2017 whilst working with the Unity Engine 2017. For more information on Oil as a testbed see section 3.2. As Peffers model is iterative in nature, the original idea and design was revisited many times. Each revisit had minor alterations done to the design as problems were identified. This iterative process, explained by Peffer, would however suggest a redesign and redefinition of the objectives. However this thesis only contains one such iteration. In the future work section 7.1 it is suggested another iteration could be used to expand upon the results of this thesis.

### 2.1.4 Demonstration Process

As an evaluation of the design the artifact was tested using testers, whom then filled out a ranking questionnaire [19,21]. The testers will be playing the game as the game is normally played. They will be asked to pay attention to the NPCs and decide which test they find the NPCs to be the most believable.

### 2.1.5 Evaluation Process

This paper is using a ranking system [20], which will suggest better or worse techniques for implementing more believable NPC behaviours. It will also suggest if artifact version two, which contains the 'hidden' behaviour affects the testers and by doing so allowing for a discussion concerning visible techniques and invisible ones. The testers were asked to play three versions, where the first version is a baseline version with none of the behaviours of this paper present, of the game for 10 minutes each. The testers were advised to pay attention and examine the believability of the NPCs but were not told in which ways the NPCs were changed. The order of the versions tested were altered from tester to tester. The

purpose of using a different order to the tests were to prevent inexperience to the game affecting the perception. The tests were all supervised, for more information see section 2.2.

### 2.1.6 Communication of Results

The results of this paper are communicated through sections 4, 5 and 6 being the result, discussion and conclusion sections respectively.

## 2.2 Game Test Protocol

The game tests will be divided into two parts the first being the game tests. And the second, a questionnaire part which is designed based on Braces' [21] book on questionnaire design. There will be three different tests which are to last for 10 minutes each. However Brace refers to a concept called 'telescoping' which can affect the results of the tests performed in this thesis. In order to avoid telescoping, the tester will play two tests and then fill out the first part of the questionnaire, demonstrated in appendix A. It needs to be two tests, as the questionnaire will be using Martinez [20] research concerning ranking instead of rating. After the tester has completed the relevant portion of the questionnaire the tester will proceed with the rest of the test. After which the tester will fill out the second part of the questionnaire, see appendix B. The order of the sessions will be decided randomly at the test session and the tester will not be made aware of the order. The supervisor will be close by to answer any questions posed by the tester. The supervisor will also, discreetly, analyse the behaviour of the tester, both physically and in-game. This analysis will provide the grounds for discussing immersion in relation to believability.

The game test will be divided into three game sessions. One of the sessions will have no alterations to the agents compared to the original Oil implementation. This will serve as a baseline to test against. The second session will have agents which operate on a GOAP but without the ability to predict future needs beforehand. The third test will use the GOAP agents with the ability to predict future needs and make 'tactical' decisions.

The questionnaire design is fairly straightforward using the before mentioned ranking style of questionnaire. It focuses heavily on ranking the three tests against each other. It also allows the tester to pick which test they believe to be the most believable option. Further it asks the tester to explain which differences the tester saw between the tests. This is to allow data to answer the RQs. There are three forms for each respective test sequence in order to keep track of the data when the testers start with different versions. These are referred to simple as 'Questionnaire version 1' for the questionnaire which starts the test sequence with the baseline version of the artifact, which will be referred to as version one. The 'Questionnaire version 2' starts the test sequence with the version one of the artifact, which will be referred to as version two. Lastly 'Questionnaire version 3' starts its' sequence with the artifact version 2, which will be referred to as version three.

## 3. Solution

### 3.1 Engine

This project was implemented in Unity 2018 version 2.8 [22] which is a developer environment designed to assist developers in game creation.

### 3.2 Software

The artifact is a behavioral alteration to examine the perception of realism. Brown [14] establishes the need for a certain amount of atmosphere, such as graphics, sounds and music to allow the users to immerse themselves in the experience. As mentioned before, there is a connection between believability and immersion. Therefore the artifact was implemented in an RTS survival game called Oil [5], to utilize the atmosphere already present. For the purpose of answering the questions posed in this thesis certain functions of Oil were disabled as to not distract the testers.

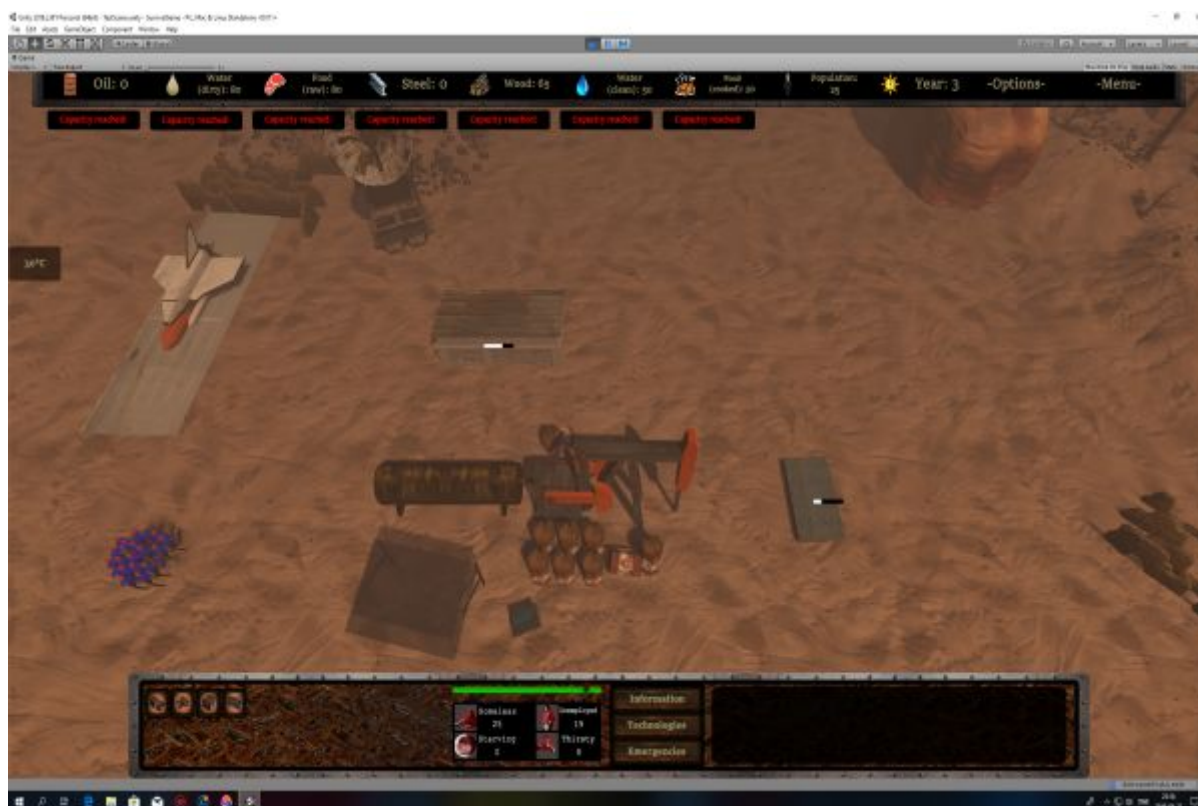


Figure 1 Depicting the Oil game

Oil is an RTS Survival game made by Nilsson et.al as a course assignment [5]. The game is written in C# using the Unity Game Engine version 2018.2.8. It revolves around constructing

and maintaining a town by gathering and managing resources using the population (citizens). The player has no direct control over these citizens. They are managed indirectly by assigning workers to various buildings. These buildings are purchased by the player for a given number of resources. The player then places the structure at a valid location where it will become available for use after a given construction time. The goal of the game is to survive for long enough for the player to progress to the point when a spaceship can be launched with the population in it. The game is set in an apocalyptic world setting caused by global warming. To oppose the player there are survival elements such as heat and cold which are managed by researching new technologies. The players need to manage the food and water supply by acquiring and refining raw food and dirty water into cooked food and clean water. To add an element of randomness there are events which pop up as the game goes along. These events force the player to do quick text-based decisions that affect the world in some way. For instance some of these events can damage buildings or kill citizens.

### 3.2.1 Buildings

In this section the buildings that are available to the player will be listed together with the purpose of the building. All buildings which can be upgraded most first have their upgrades researched in the research section. Note that some buildings have been disabled from the scope of this thesis and will therefore not be listed.

- **Oil Pump.** The Oil Pump is the starter building and at the center of the town. It produces the Oil resource. By upgrading the oil Pump the player gets access to higher tier buildings.
- **Resource Depot.** The Resource depot (RD) provides the player with a set amount of maximum resources that can be stored at the same time, but will increase with more RDs. It is also from this building the player manages citizens whom are supposed to gather resources.
- **House.** A house can have 10 residents associated with living in it. If the citizen has a home, they will receive a 'moral' bonus. They will also enter the house if in a resting state in the baseline version.
- **Canteen.** A Canteen is used to convert dirty water into clean water and raw food into cooked food. The rate of conversion increases the more citizens assigned to the Canteen.
- **Water Well.** A water well produces dirty water. The production rate increases the more citizens are added to the building.

### 3.2.2 Resources

This section will list the resources available in the game as well as the purpose of the resource.

- **Oil.** Oil is used to reduce heat and power advanced buildings.
- **Dirty Water.** Dirty water is transformed into clean water in the Canteen.
- **Raw Food.** Raw food is transformed into cooked food in the Canteen.
- **Steel.** Steel is used in the construction of buildings.
- **Wood.** Wood is used in the construction of buildings.
- **Clean Water.** Is consumed by the citizens.
- **Cooked Food.** Is consumed by the citizens.

### 3.2.3 citizen



*Figure 2 Depicting the citizen model*

The main objective of the citizen is as a production resource for the player. As such it has to work and keep this as its main task. The citizen also has certain needs which affect what is named “moral”. These needs include thirst and hunger, but also comfort which is based on activity or heat exposure. As these needs decline so does the “moral” of the citizen and as the “moral” declines the citizen will first stop working and if the measure reaches a certain threshold the citizen will die. The citizens have a natural life cycle in which they start as children and after a certain amount of time they reach adulthood and eventually die of natural causes. There is an “Emergency” game mechanic which allows the player to utilize child labor to push through hard times but with a moral penalty to all the citizens.

Warpefelt [10] writes about NPCs’ “characteristics” as a definition of NPCs, these characteristics can be translated into the citizens’ needs. The player decides if a building should be utilized and which number of citizens should occupy the building. However the player has no control concerning which citizen will be working at said building or when the citizen will take a break. The sole role of the player is planning and maintaining of the town through high level decisions. Therefore these citizens, even though their goals are selected by the player, are still by Warpefelts’ definition, NPCs.

## 3.3 State Machine

The state machine operates as explained by Yannakis and Togelius in [17] on a set of predefined states. It contains the transition logic to switch between states as well as the logic to run the states. There will always be one state running. The state machine lost the function to handle state transitions after the GOAP was implemented, as the states were selected outside the states and the state machine. See section 1.8 for more information.

### 3.3.1 States

For all version one states, they contain the transition logic to switch states and no pre-requirements or outcomes are defined. For all the other versions, no transition logic is defined in the states and the pre-requirements and outcomes are defined.

State Name	Prerequisite	Outcome	Effect
Rest	Tired	Not Tired	Increases comfort meter whilst active.
Eat	Hungry	Not Hungry	Increases hunger meter on success, lowers comfort meter whilst active.
Drink	Thirsty	Not Thirsty	Increases thirst meter on success, lowers comfort meter whilst active.
Work	Has Job	Is Working	lowers comfort considerably whilst active.
Worksite	Has Job	Is Working	lowers comfort considerably whilst active.

*Figure 3 Depicting the state prerequisites and outcomes of performing the action. As well as the effect of being in each state.*

**Rest(version one)** is the state performed by the citizen after her moral attribute reaches a certain value or no job is assigned to them. The state will check if the citizen has a home and move there. If the citizen has no house it will instead walk to an idle point which is fixed in the gameworld. In the baseline version the citizen will once it arrives, enter the house. The modified **Rest(version two)** state adds random movement around the position of the house if any, or the oil building if homeless. The citizen will not enter the house in any other version than the baseline version.

**Eat(version one)** will allow the citizen to eat if the cooked food resource  $> 1$ . The modified **Eat(version two)** state will move the citizen towards a canteen, if any, otherwise the citizen will move to a fixed position in the gameworld. At the canteen or point the citizen will attempt to consume a cooked food resource.

**Drink(version one)** will allow the citizen to drink if the clean water resource  $> 1$ . The modified **Drink(version two)** state will move the citizen towards a well, if any, otherwise the citizen will move to a fixed position in the gameworld. At the well or point the citizen will attempt to consume a clean water resource.

**Work** state will send the citizen to their workplace. They will enter the workplace and when the needs decay below some threshold the citizen will leave the workplace to fulfill the need. When the need is satisfied the citizen will return to the workplace. This pattern will continue until cancelled.

**Worksite** state will send the citizen to their gathering site. They will enter the gathering site and after a timer they will return to the resource depot and deliver an amount of resources.

This pattern will repeat until cancelled or when the needs decay below some threshold the citizen will leave the workplace to fulfill the need.

### 3.4 Planner

The artifact contains four main parts; the first is the foundation of it all, the GOAP. It consists of a single class which looks at the prerequisites and outcomes in a pool of available actions. Then it puts together viable sequence of actions to satisfy a predefined goal. See section 1.7 for more information.

The second part is the goal machine which handles the assignment of goals to the GOAP. By looking at the current needs of each individual citizen and then determining the most urgent need and if this need requires to be handled. If not it will analyse if and which job the human is assigned to. In version three this is done through an activation variable which represents the measure of the distance and time it takes to complete, for instance, the resource gathering sequence. By measuring this sequence a citizen can conclude if a need will fall below the threshold before it completes the sequence. The citizen can therefore have a tactical dinner before it leaves to gather the resource. The future need is calculated as follows.

$$T1 = (V2 - V1)/S$$

$$T2 = (V2 - V3)/S$$

$$FN = R * (T1 + T2) + CN$$

$T1$  is the total time it takes to go from the current position  $V1$  to the occupation buildings' position  $V2$  with the speed  $S$ .  $T2$  similarly calculates the time to move from the occupation building to the respective building  $V3$  which would satisfy a particular need. The future need  $FN$  is then calculated by the rate of decline ( $R$ ) multiplied by the sum of times to move with the current need ( $CN$ ) added lastly.

Thirdly the FSM handles the logic in each action. The transition logic is handled by the GOAP with the exception of version one. Which means that the FSM simply fires the actions. The last part concerns the actions. These are made up of two parts, the first is a prerequisite and outcome functionality which the GOAP communicates to. The second part is the logic to handle the citizen if the FSM fires this particular action. The states are explained in section 3.3.1.

## 4. Result

This section contains all the data from the questionnaires. For simplicity the data from all the versions have been merged into a single graph. The 'version 1' represents the baseline version with no parts of the artifact introduced. 'Version 2' contains the visual behavioural implementations only. The 'version 3' contains the logical behaviour embedded with the visual behaviours.

The tests began with the test supervisor giving a brief explanation of the game along with a demonstration of how the game is played. Along with some additional information concerning the build they were testing and some typos in the questionnaire. The testers were told the focus of the study was to study the believability of the NPCs and therefore the

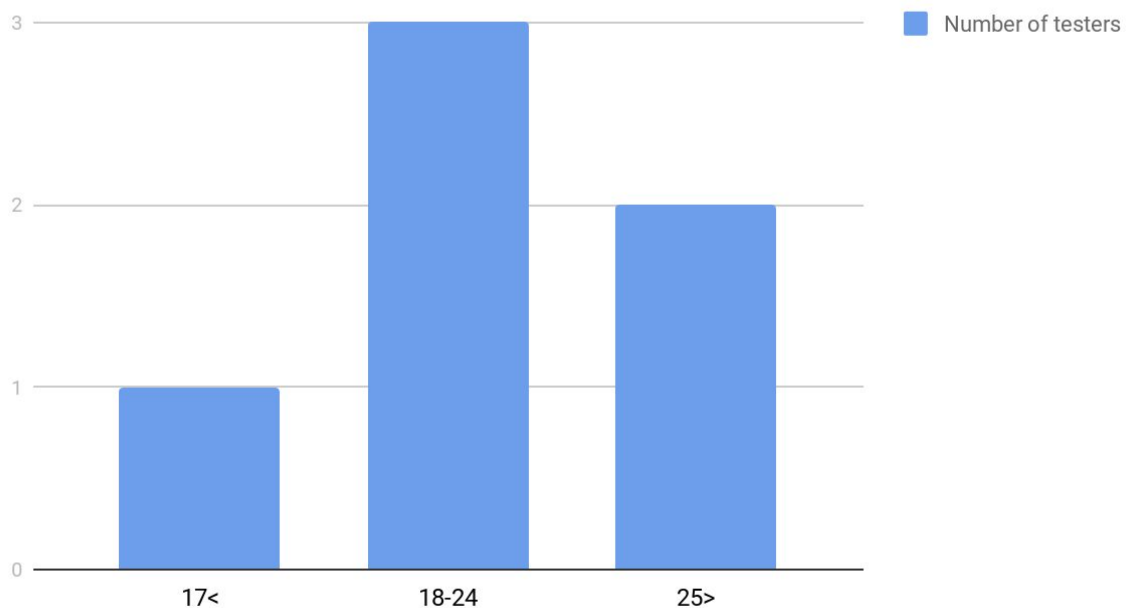
testers should pay extra attention to them. The test supervisor would remain close by and available for questions but would otherwise silently observe the testers' gameplay and physical behaviour.

Version Number	version 1	version 2	version 3
Testers per version	2	2	2

*Figure 4 shows the number of testers starting with which version*

6 different testers tested the artifact starting with different versions. The number of testers starting at each version was evenly spread across all versions.

### Year range of participants



*Figure 5 shows the age range of the testers*

The graph shows that there was a spread between the ages of the testers but the majority remained in the ages of 18-24.



## Time spent playing games

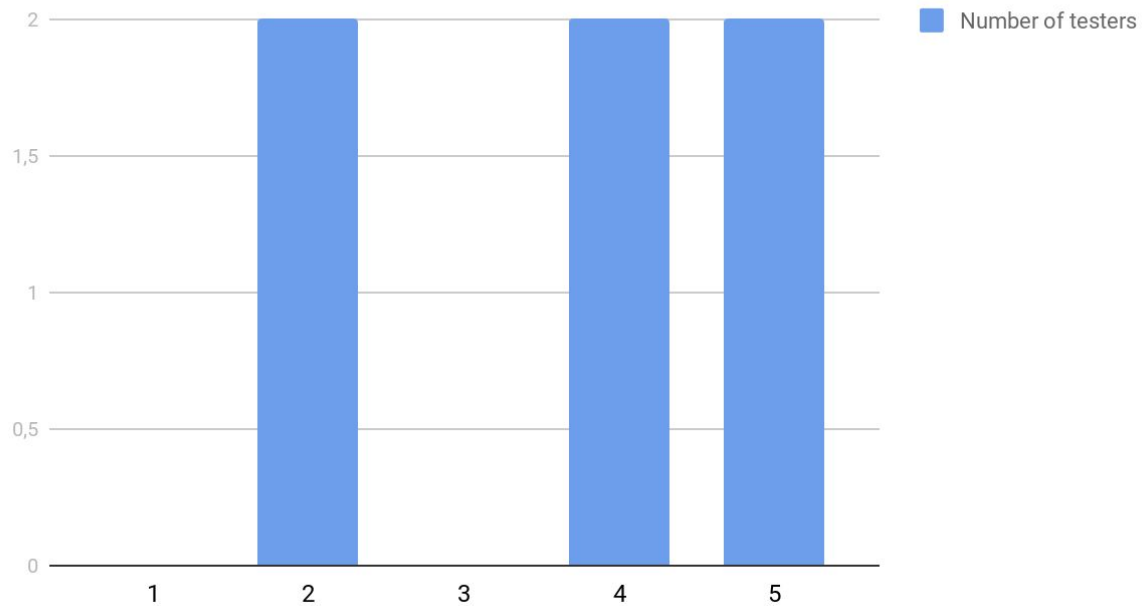


Figure 6 shows the approximate time spent playing games by the testers. Where 1 is none and 5 is 'all the time' (which can be considered as 'very frequently', instead of the literal meaning).

**Comparing the first test with the second, which would you say had the most believable NPCs.**

## Summarized rankings of version 1 and version 2

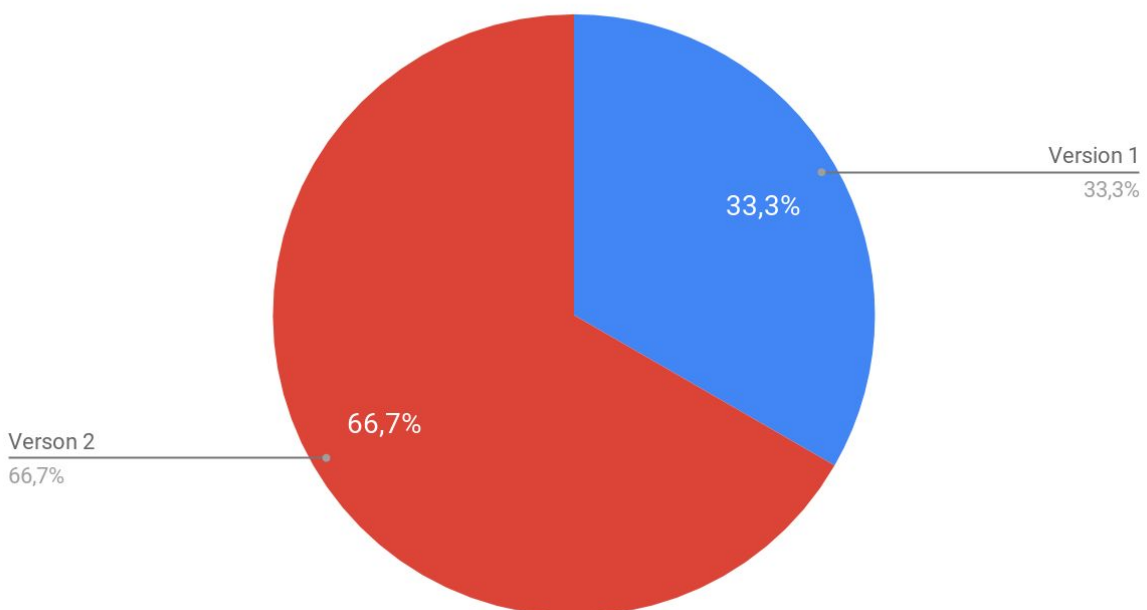
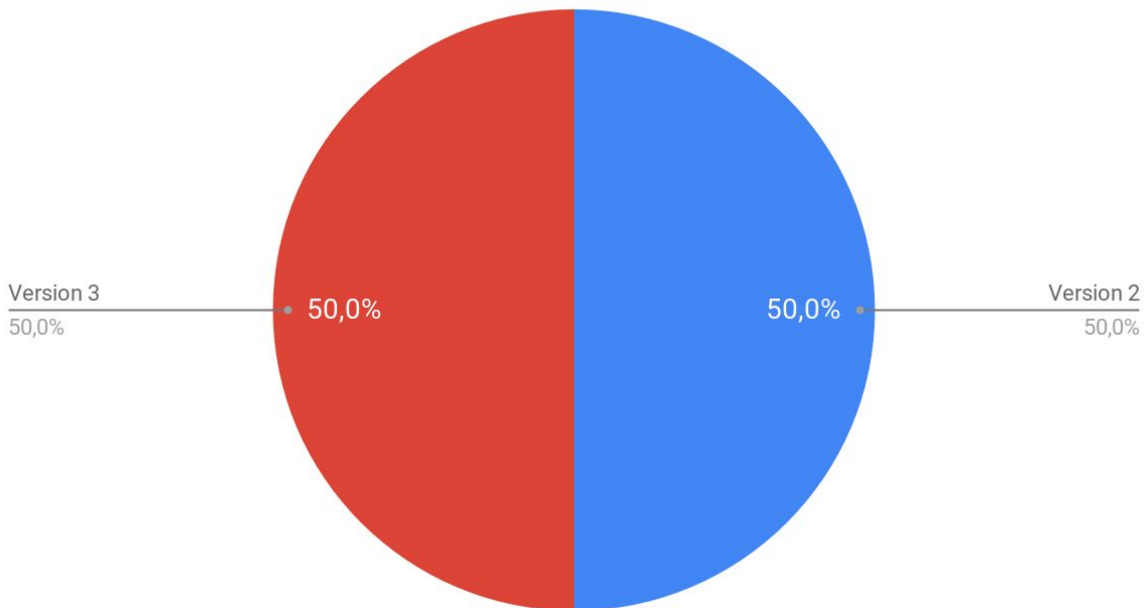


Figure 7 depicts graph data from questionnaire concerning the preferences of the testers when asked to rank the baseline version (version 1) with the artifact version one (version 2).

*It shows a majority of people found the GOAP version to have the most believable NPCs out of the two versions.*

**Comparing the second test with the third, which would you say had the most believable NPCs.**

Summarized rankings of version 2 and version 3



*Figure 8 depicts graph data from questionnaire concerning the preferences of the testers when asked to rank the artifact version one, containing the GOAP (version 2) with the artifact version two, containing both the GOAP and the predictive behaviour (version 3).*

**Comparing the third test with the first, which would you say had the most believable NPCs.**

## Summarized rankings of version 3 and version 1

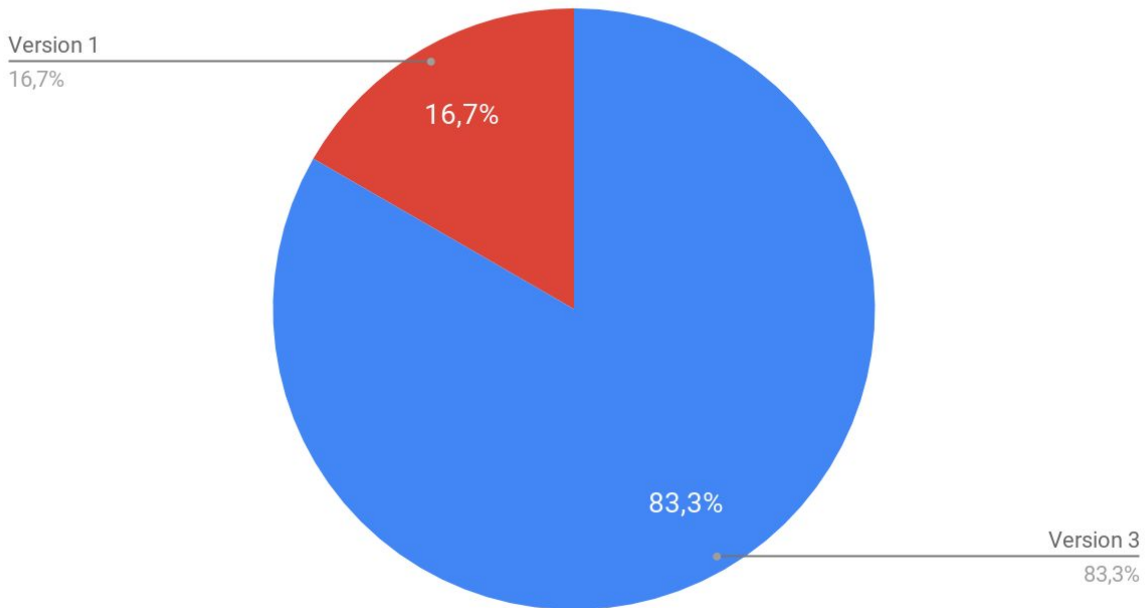


Figure 9 depicts graph data from questionnaire concerning the preferences of the testers when asked to rank artifact version two (version 3), containing the GOAP with predictive behaviour and the baseline version (version 1).

### **How would you describe the difference, if you noticed any, concerning the behaviour of the NPCs between all three tests**

This question allowed the testers to describe which differences in the behaviour of the population they had noticed. Some testers spotted a part of the visual behaviour accurately, however none noticed the logical behaviour of the artifact version 2. The following are all the answers provided by testers. Sorted by the questionnaire versions for clarity.

1. Questionnaire version 1.
  - a. *“Det var lättare att se vart de olika arbetarna befann sig i försök 3”*(It was easier to see where the different workers were in try 3)
2. Questionnaire version 2.
  - a. “The people in the 3rd test were sometimes walking on a spot. Also in the 3rd game two persons were walking opposite of each other and could not move from the spot until a third person came through them. However, during the second and first test the people were usually not able to go close to the resource building, they "bounced off".”
  - b. “One of the tests (third) had really strict NPCs, doing nothing but their assigned task. Another test (first) had the NPCs stroll around when not assigned to a task and seemingly tended to group up more with fellow NPCs compared to the second test where they seemed to prefer solitude.”
3. Questionnaire version 3

- a. "First was too erratic running around. Second was better due to not distracting from the game. Third was best, similar to first but the NPCs weren't running around as much"
- b. "I noticed no difference between the first and third test. The one major difference I perceived between test 2 compared to test 1&3 was how the houses were used. During test 2 all children seemed to go straight inside the houses and never came out for the entire duration of the game. As a contrast, the children in test 1&3 hung around in the town and never entered the houses. The workers in test 2 also seemed to move in a straight line to all locations which wasn't always the case in test 1 &3. In test 1&3 some workers also seemed to hang around outside the buildings they were assigned to instead of disappearing into them completely."

**Which test would you say had the most believable NPCs of this experiment?**

Game Version	1	2	3
Number of testers	1	1	4

*Figure 10 shows the rating value for by each tester. This question was intended to show if the tester would rate the same version, as they had implicitly ranked, the most believable.*

**Which test would you say immersed you the most.**

Game Version	1	2	3
Number of testers	1	1	2

*Figure 11 shows the immersion to the respective game version. Only 4 testers filled out this question. This question was intended to affirm the connection between immersion and believability.*

## 5. Analysis

The results will be analysed in this section by first analysing the rankings of the game versions and the flat rate value and how this affected the perceived immersion of the game. It will also analyse the testers' response to the difference between the versions. Finally evaluating the rating and how this correlates to the ranking values and the immersiveness predicted in section 1.6.

## 5.1 Results Analysis

The experiment had 6 testers, 1 out of 6 were less than 17 years old. 4 out of 6 were 18-24 years old and 1 out of 6 were above the age of 24.

When asked to rate the amount of time they play games 2 out of 6 testers rated themselves as 2. Another 2 out of 6 testers rated themselves as 4 and the last 2 testers rated themselves at 5.

When asked to rank version 1 and 2, 4 out of 6 testers felt that the artifact version one had better NPC believability. With a majority in favor for the visual behaviour modification. This critique goes both ways as other testers found version 2 to deliver a way to spot workers easier. Some of the testers managed to partially notice the added visual behaviour to the NPCs, between version 1 and 2. However none of the testers noticed the location based action behaviour.

The ranking between version 2 and 3 shows a split of 50% for either version. None of the testers managed to figure out the predictive behaviour in version 3. With one tester noticed the increased movement of the NPCs due to them predictions. One tester wrote that no difference between version 2 and version 3 could be found. Even though only one tester wrote it down, this was a concern among most testers.

Finally the comparison of version 3 and version 1 shows that 5 out of 6 testers preferred version 3. The testers did not seem to notice all of the differences between version 2 and 3. But they did almost completely agree that the addition of the behaviours gave the NPCs a better sense of believability.

When asked if the testers could identify the changes between the versions the answers suggest that most noticed a difference in NPC movement. With one tester identifying the movement while idle trait, but not the location based behaviour. Another tester did notice a correlation between objects, in this case houses, with how the NPCs moved. Which is fair to say that the location based movement was implicitly discovered. However the answers also suggest that most couldn't identify the different behaviour in the same way as the two aforementioned testers did. Instead they identified either bugs or more subjective notions as the changes between versions. None of the testers could identify even remotely the predictive behaviour however one tester makes a remark on extra movement in version 3.

Though the testers were indecisive when asked to rank version 2 and version 3 if asked to select which test they found the NPCs to be the most believable in. 4 out of 6 picked version 3 as the one they found to be the most believable. 1 out of 6 chose version two and another 1 out of 6 chose version one. It is interesting to see the expected result if the testers were asked to rate the versions but not if when they were asked to rank the versions.

When asked which version immersed the tester the most 2 out of 4 chose version 3, 1 out of 4 chose version 2 and one chose version 1.

## 6. Discussion

In this discussion the reasoning behind using a planner for the artifact versions 1 and 2, but not in the baseline version will be explained. Further discussing the results and the analysis in sections 4 and 5 in relation to the visual and logical components.

### 6.1 Planner

The baseline version doesn't use a planner. It has all the transitions connected with the FSM. The artifact version one and two uses a planner. This may interfere with the results however, the choice of using a planner goes back to Orkin [4] and his papers on F.E.A.R [3] and how they handled complex transitions. The baseline version simply doesn't have complex transition logic. However as new behaviours are added, specifically with the location based actions this complexity rises. With the ability to 'look ahead' this complexity increases further. The planner replaces the transition logic but doesn't alter the gameplay shown to the player other than the changes that were intended. Therefore it was decided that the risk of contaminating the experiment was lower than the benefits of allowing the planner to handle the transitions were high enough to justify the risk.

### 6.2 Visual and Logical Behaviours

In this section the different results will be discussed.

#### 6.2.1 The Visual Component

It became evident in the analysis in section 5.1 and by the results in section 4. That the testers had a more mixed reaction to the baseline version and the visual behaviours with a tendency to favour the artifact version one. However this tendency then becomes overwhelmingly one sided in favor of the version two of the artifact, when compared to the baseline version. This result would suggest the artifact version one which was made based on Banished [1] and Frostpunk [2] works. As such I do believe that it is possible to, by extension, increase the immersion of RTS survival games by implementing these behaviours. See sections 1.5 and 1.6 on the connection between believability and immersion.

#### 6.2.2 The Logical Component

No tester has accurately noticed the difference between artifact version one and two. In fact some tester commented on this "I noticed no difference between the first and third test.". The tester is referring to the difference between the artifact version version one and two. Because the tester started at the test version 2. It may be that the testing needed to be altered to increase the duration of playtests and the quantity of testers. This unfortunately results in the inconclusive result, predicted in the Limitations and Risks section 2.3. In F.E.A.R [3] we do not see this issue and I suspect it is because we expect certain behaviour in F.E.A.R, as per Loyalls' [13] research. The player would expect the behaviour in F.E.A.R however this natural expectation is not as present in the context of this thesis. I believe

F.E.A.R communicates the intent of their NPCs in a much clearer manner through both audio and visual queues. If such 'signs' were added to the version two of the artifact we may have seen a clearer distinction between the versions of the artifact.

### 6.2.3 Deviations and Causes

Curiously the rankings suggest no clear winner between version one and two of the artifact, however, when the testers were asked to select the best version in a rating fashion, they predominantly selected artifact version two.

As mentioned when testers were asked to specify the changes, one such mention is *"However, during the second and first test the people were usually not able to go close to the resource building, they "bounced off".*" This tester is clearly referring to a rouge variable, a bug, as a change in behaviour. Which technically it is, however not the intended one. This factor was predicted in section 2.3 which is now unfortunately confirmed.

## 7. Conclusion

For the main research question "How will the visual techniques described in section 1.1 affect the perception of believability, of the NPCs in an RTS survival environment." Although there is no clear answer from this thesis, the results suggest an increase in the perceived NPC believability by the test group. Most testers noticed part of the visual behaviour but completely missed the second part. However when the artifact version two was compared with the baseline version the majority of testers preferred the artifact version two. I would like to add to the discussion in section 6.2, that I believe the possibility of increasing immersion is not confined to these particular behaviours, but any behaviour that can be rationally explained and seen.

The second research question posed in this paper "How will the perceived believability change when adding a predictive behaviour to these techniques." According to the test data there is a 50% chance the tester would rank the artifact version 2 higher than the artifact version one. With comments by testers such as "I noticed no difference between the first and third test." referring to the difference between the artifact versions and subsequently the research questions. It is mentioned in section 6.2 that I believe the testers needed more time. As such there is a margin of improvement and an emergent importance of which logical behaviour should be implemented. As the predictive behaviour used in this thesis did not seem to increase the perception of believability as the testers did not notice the increased intelligence of the NPCs.

### 7.1 Future Work

There are many avenues for continuing this research. The most direct one would be to continue with the iterations as described by Peffers [19] to receive a more conclusive answer to the questions posed in this thesis. Another such avenue could be to try this experiment in another genre of games as there are some games which are heavily dependant on the rationale of their NPCs. Such a result could be used in comparison to the results of this thesis.

Another angle which could be explored is which visual behaviours would increase the believability of the NPCs the most and which logical behaviours would have similar effects. Such an investigation could have an impact on future game design concerning an increase in NPC believability.



# References

- [1] Shining Rock Software LLC, “Banished”, 2014
  
- [2] 11 bit studios, “Frostpunk”, 2018
  
- [3] Monolith Productions. “F.E.A.R.: First Encounter Assault Recon”. Sierra Entertainment, Inc. 2005.
  
- [4] J. Orkin, “Three States and a Plan: The AI of F.E.A.R”, *Proceedings of the Game Developer’s Conference (GDC)*, 2006
  
- [5] Ohrberg. S, Björkman. A, Mossberg. M, Terins. D, Nilsson D, Pieslinger. H. “Oil”, Faculty of Technology and Society, Malmö University, 2018.
  
- [6] L. Ermi and F. Mäyrä, “Fundamental components of the gameplay experience: Analysing immersion” in *Changing Views: Worlds in Play*, Proceedings of the 2005 DiGRA International Conference, 2005.
  
- [7] Blizzard Entertainment, “Warcraft 3”, 2002
  
- [8] Blizzard Entertainment, “Starcraft 2”, 2010
  
- [9] K. Salen and E. Zimmerman, *Rules of play: game design fundamentals*. Cambridge, MA: The MIT Press, 2010.
  
- [10] H. Warpefelt, ‘The Non-Player Character: Exploring the believability of NPC presentation and behavior’, PhD dissertation, Department of Computer and Systems Sciences, Stockholm University, Stockholm, 2016.
  
- [11] T. Lindstam, A. Svensson, ‘Behavior Based Artificial Intelligence in a Village Environment’, bachelor thesis, Faculty of Technology and Society, Malmö University, Malmö 2017. Available at: <http://muep.mau.se/handle/2043/23212> [Accessed: February 25, 2019]

- [12] Endnight Games Ltd, "The Forest", 2018
- [13] Bryan, A, Loyall, "Believable Agents: Building Interactive Personalities", PhD thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA, 1997
- [14] E. Brown and P. Cairns, "A grounded investigation of game immersion", *Extended abstracts of the 2004 conference on Human factors and computing systems - CHI 04*, 2004.
- [15] P. Lankoski, "Character Design Fundamentals for Role-Playing Games," in *Beyond Role and Play: Tools, Toys and Theory for Harnessing the Imagination*, Ropecon RY, 2002, pp. 139–148.
- [16] J. H. Murray, *Hamlet on the holodeck the future of narrative in cyberspace*, 2nd ed. Cambridge, MA: The MIT Press, 2017.
- [17] J. Orkins, "Agent Architecture Considerations for Real-Time Planning in Games", *AIIDE* , 2005.
- [18] G. N. Yannakakis and J. Togelius, "Finite State Machines," in *Artificial Intelligence and Games*, Springer International Publishing AG, 2018, p. 33.
- [19] K. Peffers, T. Tuunanen, M.A. Rothenberger, S. Chatterjee. "A Design Science Research Methodology for Information Systems Research", *Journal of Management Information Systems*, vol. 24 Issue 3, Winter 2007-8, pp. 45-78.
- [20] H.P. Martinez, G. N. Yannakakis, Member, IEEE, and J.Hallam, "Don't Classify Ratings of Affect; Rank Them!", *IEEE Transactions on affective computing*, vol. 5, no. 3, July-September, 2014
- [21] I. Brace, *Questionnaire design: How to plan, structure and write survey material for effective market research*. London: KoganPage, 2018
- [22] Unity, Unity Technologies, 2005

# Appendix

Appendix A (Questionnaire first part) \*Note the typo in the ranking questions' second alternative were explained by the supervisor

## Questionnaire For Thesis v1

This questionnaire will ask questions related to a performed experiment to be answered by the experiment participant. The questions will large try and rank the different tests based on their perception of the believability of the NPCs in the artifact

Name?

Ditt svar

Gender?

Female

Male

Prefer not to say

Övrigt:

Age?

Ditt svar

How often do you play games?

	1	2	3	4	5	
None	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	All the time

Comparing the first test with the second, which would you say had the most believable NPCs.

The First test had more believable NPCs than the second

The Second test had more believable NPCs than the second

Appendix B (Questionnaire second part)

Comparing the second test with the third, which would you say had the most believable NPCs.

- The Second test had more believable NPCs than the third
- The Third test had more believable NPCs than the second

Comparing the third test with the first, which would you say had the most believable NPCs.

- The third test had more believable NPCs than the first
- The first test had more believable NPCs than the third

How would you describe the difference, if you noticed any, concerning the behaviour of the NPCs between all three tests

Ditt svar

---

Which test would you say had the most believable NPCs of this experiment?

- Test 1 had the most believable NPCs
- Test 2 had the most believable NPCs
- Test 3 had the most believable NPCs

Which test would you say immersed you the most

- Test 1
- Test 2
- Test 3

