Aalborg University Copenhagen

Semester: Med10 – Master Thesis

Title: OPIL – A Model Proposal for Evaluation and Design of Productive Interaction in Digital Media

Project Period: Spring 2018

Semester Theme: Master Thesis

Aalborg University Copenhagen

Frederikskaj 12,

DK-2450 Copenhagen SV

Semester Coordinator: Stefania Serafin

Secretary: Lisbeth Kirstine Nykjær

Supervisor(s):
Luis Emilio Bruni
leb@create.aau.dk

Author:
Steffen Hansen
sthans13@student.aau.dk

Anders Damkjær Nielsen
adni13@student.aau.dk

Abstract:
In this thesis we present a framework for investigating the interplay between user and system, that we call productive interaction. The framework is built into a model for both design and evaluation of productive interaction of digital media. The model, OPIL, is investigated through testing an implementation of a digital game, designed on the requirements of the framework. Results show that the model must be used in conjunction with considerations of target group and potentially flow measures and that it can identify and categorise productive interaction.

Keywords:
Emergent Narrative, Narrative Paradox, Productive Interaction, Agency, Chatbot, OPIL, Flow, Interaction.

OPIL - A Model Proposal For Evaluation and Design of Productive Interaction In Digital Media

Anders Damkjær Nielsen Steffen Hansen

Thursday 31st May, 2018

Contents

1 Abbreviations

- **ABL** A Behavior Language
- **AAU** Aalborg University
- **AI** Artificial Intelligence
- **AIML** Artificial Intelligence Markup Language
- **AAD** Author-Audience Distance
- **CIS** Character-based Interactive Storyteller
- **CPH** Copenhagen
- **EN** Emergent Narrative
- **NLP** Natural Language Processing
- **NLU** Natural Language Understanding
- **NPC** Non-Player Character
- **OPIL** Observing Production Interaction in Levels (The model devised from this thesis)
- **VE** Virtual Environment
- **VPA** Virtual Personal Assistant
- **UI** User Interface
- **SDK** Source Development Kit
- **SAPI** Microsoft's Speech API library

2 Introduction

Productive interaction describes the unique interplay between system and user, which is useful for both design and evaluation of digital systems. Productive interaction allows for developers to locate points of interest in their systems and proves useful as a way of quantifying the quality of interaction of a given system. The categorization and delimitation of previous research were focused on the product rather than the process, and in doing so the high level of complexity had been lost. The process cannot be described purely as player action or system response but requires interplay, which many frameworks struggle to adequately account for. The OPIL model redefines the concept of productive interaction and sets three key requirements for the evaluation and design of productive systems. The first requirement is that the system must be of adequate complexity, which demands the player's attention and forces players to evaluate their possible choices. The second requirement, that of quasi-unique progression, furthers the concept that productive interaction is a process and not an outcome. The third requirement, that of understandable consequences and outcomes, assures that the player is given agency by their actions. Furthermore, the framework presents two ways to focus usage of the model. In the present effort, a social game prototype was designed based on the requirements derived from the OPIL model. Twelve male participants in the twenties were tested through the combined study of a flow questionnaire, logs of their interaction, and interview data regarding their thoughts of the system. The system was then evaluated based on the original requirements for the OPIL model. This thesis provides: Theory of narratives stemming from the initial problem, the framework for usage of the OPIL model, the design requirements used for the prototype, a start to end design and implementation of the prototype, an evaluation and discussion of the results from the test, and thoughts on the usage of OPIL derived from the test.

3 Analysis

We engage with the initial problem statement:

How can we satisfy the user's desire for interaction during a digital dialogue in a virtual environment, while still maintaining a high degree of control for the author?

Which we can then narrow down to a more manageable problem. It would be prudent to consider the elements of the initial problem statement, and why they have been included: *satisfy the user's desire for interaction* refers to the Narrative Paradox, which will be investigated shortly, as it can be viewed as a problem in need of solving; *digital dialogue* refers to the ability to interact with virtual agents; *virtual environments* (VEs) describes the interaction occurring in a digitally rendered environment; *degree of control with the author* also refers to the Narrative Paradox.

3.1 Narrative Paradox

The Narrative Paradox (aka. the interactive paradox, the interactive dilemma) is the conflict of pre-authored narratives in conjunction with interactive elements. Aylett (2000) formulated it: "*how to reconcile the needs of the user who is now potentially a participant rather than a spectator with the idea of narrative coherence - that for an experience to count as a story it must have some kind of satisfying structure.*" It is often seen in relation to the combinatorial explosion, that claims that choices in narrative creates exponentially more possible paths in which the narrative could develop (Stern, 2008). Through the lens of narrative theory regarding authority and the relationship between author and audience, the narrative paradox can be described as the audience usurping authorship from the author through interaction (Louchart & Aylett, 2003). For when a user takes any action, will that not change the narrative as intended by the author? And if the author only allows limited interaction, is it then (meaningful) interaction at all? Aylett (2000) describes this as two issues, the first being how much narrative can be relaxed and stray from its pre-determined nature, the second how much it is possible for the user to participate in a narrative. Furthermore, the two issues are considered related as a wholly pre-determined narrative also determines the degree of interactivity of the user. This relation can be described by the continuum that goes from wholly-scripted narrative on one side, and "improvisation" on the other Aylett (2000). The improvisation heavy side is also described as emergent narrative, the situation in which users' actions make the narrative. As such, the paradox itself makes it impossible to reconcile both the needs of the user and author simultaneously, though Louchart and

Aylett (2003) argue in their paper that roleplaying games partially "solve" the narrative paradox. They claim that "[roleplaying games] *seems to offer a good compromise between the freedom exercised and experienced by the user/player and the narrative control necessary for the development and unfolding of interesting stories* [...]" However, as this is partly due to the presence of a gamemaster, who oversees the making and adapting of the plot based on the player's action, the difficulty of implementing it digitally yet remains. While there have been attempts at story generation (See Talespin (Meehan, 1977) and Universe (Lebowitz, 1983)), as a digital implementation capable of providing the narrative control, their output is often rather simplistic, or their stories of questionable integrity.

3.2 Emergent Narratives and Embedded Narrative

Salen and Zimmerman (2004) identifies Marc Leblanc as the first to propose the combination of the terms *emergent narrative* (EN) and *embedded narrative*. They define embedded narratives as being the pre-generated narrative content of a game, which is scripted by the authors and the player experiences as story context. Embedded narrative provides the major story arc for the game, as well as providing meaning for the player's actions, structuring their interaction and movement in the game world in a meaningful way (Salen & Zimmerman, 2004). Opposite of embedded, they define emergent narrative as the narrative that arises from the interaction with the game system, in mostly unexpected ways. This is however only one of many definitions of emergent narrative, as Louchart, Swartjes, Kriegel, and Aylett (2008) argues that emergent narratives are conveyed through characters specifically. Others (Bevensee & Schoenau-Fog, 2013; Jenkins, 2004) hold to the definition that the narrative material is provided through a rich environment, as well as intelligent characters, and that interaction with these elements allows the user to construct her own understanding of the story. We consider that emergent narrative can occur from other elements than interaction with characters specifically but consider many of Louchart et al.'s points to hold true regardless. Louchart et al. (2008) dubs the user *the interactor*, who must take on a role and responsibilities regarding the quality of the interaction, where the relation between the interactor and the story then becomes an aspect of the emergent narrative system. They argue that the narrative development and interaction must be flexible enough to accommodate each other, however must keep a definite amount of options available to the interactor. In this model, the interactor can gradually shape and reshape the spectrum of actions available for meaningful and purposeful experience (Louchart et al., 2008). Louchart et al. (2008) illustrate this through Laurel's "flying wedge" figure (See Figure 1. (Laurel, 1991)) where an interactor within an emergent narrative system determines the direction of the narrative development by engaging in certain interactions, and in turn the narrative development constrains the probable future interactions. If Laurel's "flying wedge" model is used in an emergent narrative with multiple characters, each of the characters would then have their own wedge in which the interactor can choose actions (Louchart et al., 2008).

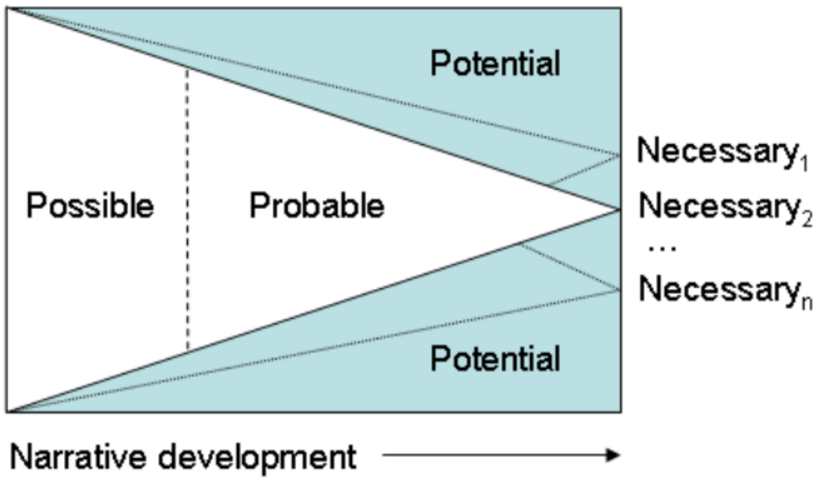

Figure 1: Laurel's "flying wedge" (Laurel, 1991): "EN interpretation of Laurel's "flying wedge" (interactive version). Interactors' choices determine the direction of the narrative development and in turn, narrative development constrains the range of probable future interactions." (Louchart et al., 2008)

We need to consider the author's possibility to think and work within the emergent narrative system as proposed by Louchart et al. (2008). They approach this problem through a metaphor of a story landscape in which an author can create interactive experiences, to get an overview of the possible stories the author is creating.

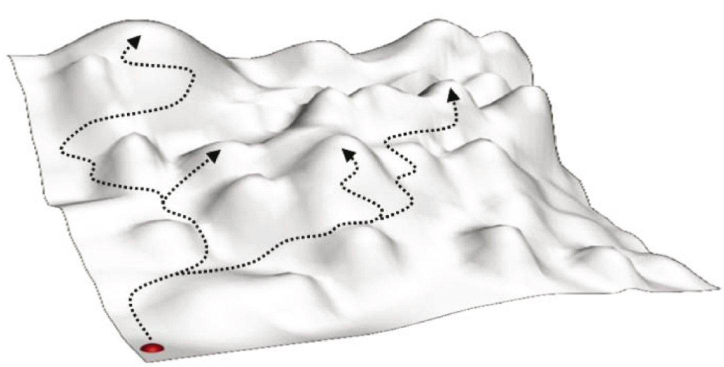

Figure 2: The story landscape of an emergent narrative, with multiple possible paths sketched (Louchart et al., 2008)

Louchart et al. (2008) argues that within the story landscape it is possible to contextualize the interactions as goals, actions, and emotions and under which conditions they occur (character's autonomy). Additionally it can be used to track the interactor's 'movement' across the landscape. The landscape represents possibility for emergent narrative and dramatic necessity. The possible emergent narrative is represented as points on the landscape in which the user chooses to move across the landscape. The dramatic necessity is represented as the hills and valleys on the landscape and is based on the emotions and intentions of the character controlled by the user. As a user ascends a hill the necessity for the dramatic increases and as they descend into a valley the dramatic necessity decreases (Louchart et al., 2008). The character of the user can influence the user's behaviour and how the user chooses the narrative path throughout the landscape. Due to this story landscape it is possible to guide the local interactions of the user based on character and boundaries of the landscape if the emergent narrative is a character-centric approach. Boundaries is what separates the story landscape from the rest of the universe as Louchart et al. (2008) puts it "[...] *the sea around the story landscape* [...]". This is necessary for an emergent narrative not only because of the technical infeasibility of simulating an unconfined world but also because the boundaries help define the topic, scenario, and message of the emergent narrative. Louchart et al. (2008) notes that the boundaries are quite abstract and can be setup in many different ways, exemplified as: Spatial boundaries ("*given by the locations where the story takes place*"), contextual boundaries ("*e.g. the bullying context in FearNot!*"[1]) and interaction boundaries ("*limiting the ways of how the user can interact with the world*"). When putting up the boundaries for the story landscape a factor to consider is the critical mass for emergence, which describes the concept of a limited area being crowded with the number of available paths, which end up damaging the interactive narrative (Louchart et al., 2008). Nonetheless, the reverse problem exists as well: If there is a vast amount of space, but little narrative interactivity. Once the author begins to create the narrative and starts to branch out to different storylines within the landscape it is important to avoid dead ends. Dead ends in the context of emergent narratives refer to the points where there is no possibility for further narrative development (Louchart et al., 2008). As an example, dead ends can occur because of the author's inability to overcome the combinatorial explosion, where the author cannot possibly cover all possible paths. When the interactor encounters a dead end in a narrative, it may leave them without a way of continuing or returning to other paths, something Louchart et al. (2008) argues against.

[1] FearNot! is a game example used by Louchart et al. (2008), stemming from their prior research in (Aylett et al., 2006).

3.2.1 Mediation of Emergent Narratives

Design based on Laurel's flying wedge can be argued to be partially emergent and partially classically structured narrative. Through the number of potential paths, the author can construct a narrative that appears emergent to the user, but in which every path has been preconceived. However, this does not answer who is actually telling the story, which Louchart et al. (2008) consider through terms of mediation. *The receiver* is the recipient of the message and in most cases the same as the interactor. *The sender* is in a traditional sense the author that writes the narrative and delivers the questions and story to the receiver in a unidirectional relationship. However, with the introduction of emergent narrative, the author still writes the story in advance but the responsibility of narratorship becomes shared between system and interactor (Louchart et al., 2008). It is the interactor that raises the questions and sets out to answer them, and in doing so establishes an emergent narrative as a dialogue between system and interactor (Louchart et al., 2008). *The message* in a traditional narrative is the story's moral, sometimes referred to as substance or fabula. In the traditional narrative the message is unidirectional, from author to audience, whereas in the emergent narrative Louchart et al. (2008) argues it becomes more nuanced. The nuance stems from the interactor's responsibility to construct their own message based on the author's template of the emergent narrative, which it turn makes the message personal. When it comes to shared narratorship Louchart et al. (2008) present three implications on the interactor's role within an emergent narrative:

1. "*Interactors do not have to be able to predict the consequences of their actions in terms of story outcome in order to be able to experience agency;*"
2. "[...] *the interactor can predict the consequences of a certain course of action, it should not be expected that they make the choices that they would if it were real life.*"
3. "[...] *Emergent Narrative presupposes a willingness to play within the formal constraints of a role. These formal constraints might be partially defined at the start of an emergent narrative but also establish themselves further during play in the form of offers.*"

Louchart et al. (2008) argue that without consideration for these implications an emergent narrative could be thought to cater to a vast array of actions the user might want to do, and to ensure all these actions have consequences in the story world. To decrease the burden of the author, Louchart et al. (2008) suggest that the notion of agency or meaningful action must be considered. Ideally, the interactor can be limited to meaningful choices, so that the author does not need to implement every possible interaction.

As a supporting theory to Louchart et al. (2008) and an alternative to the attempt by Louchart and Aylett (2003) to solve the narrative paradox, Bruni and Baceviciute (2013) present a framework for use in considering and using the narrative paradox, claiming that perhaps it is not solvable. To understand the concept of intentionality (how the author tries to get a user to correctly understand a message) in narrative communication between an interactive system and a user, Bruni and Baceviciute (2013) argue that it becomes relevant to consider the Author Audience Distance (AAD), by looking into the rich semiotic tradition which frames the issue in terms of interpretation, coding and decoding processes. Due to the nature and limitation of natural language or multimodal representation, it is accepted that in any narrative communication act there is a so called *interpretation gap* between sender and receiver. This can be expressed as the distance between sender and receiver, depending on the communicational context. Bruni and Baceviciute (2013) use the concepts of aberrant decoding and preferred decoding from Eco's seminar paper (1981) to explain the interpretation gap. Aberrant decoding happens when a message can be interpreted differently from what was originally intended (preferred) by the sender (Eco, n.d.). This is due to the sender and receiver not sharing properly the coding system, which makes the receiver deviate from the preferred decoding intended by the sender. The AAD illustrates the continuum (the interpretation gap) that goes from complete aberrant decoding to preferred decoding, depending on how defective the sharing of the coding and decoding system between author and audience is (Bruni & Baceviciute, 2013). Bruni and Baceviciute (2013) also note that the complexity of the interpretation gap increases with the introduction of immersive-interactive media as it is no longer linear communication, but also includes degrees of agency given to the audience, so they can realize their intent. However, Louchart et al. (2008) argues that interactivity can be limited and guided by the author through the emergent narrative, meaning that despite the potential divergence between intents the AAD should be controlled by the goal of the system. Bruni and Baceviciute (2013) introduce the notions of *abstract* and *didascalic* narratives as parameters of the AAD, which relay how abstract or descriptive a narrative work is. They approach it by analysing the level of abstraction inherent to the manifestations of narrative content. Bruni and Baceviciute (2013) do this by defining a continuum that goes from abstract narrative to didascalic narrative. An abstract narrative would only have intrinsic form with little or no attempt at pictorial, figurative, or explicit representation, but with little potential to elicit a degree of narrativity (Bruni & Baceviciute, 2013). Where a didascalic narrative would be a very didactic, explicit, obvious, and self-explanatory message. Bruni and Baceviciute (2013) then propose a model in which one can represent the interpretation gap by taking the AAD and the abstract-didascalic continuum into account.

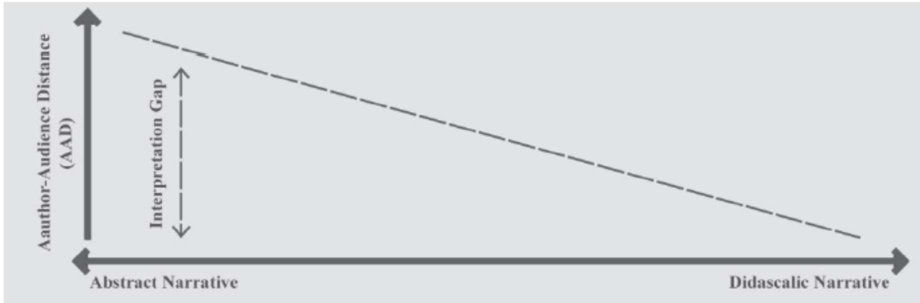

Figure 3: Author-Audience Distance and Abstract-Didascalic continuum function (Bruni & Baceviciute, 2013)

To understand the AAD as a function of the abstract-didascalic continuum, we investigate the terms of narrative intelligibility and narrative closure in Bruni and Baceviciute (2013). *Narrative intelligibility* is defined as "*The understanding of the substance of a narrative very closely in the way it was intended by its author or creator. The level of intelligibility defines the author-audience distance.*" (Bruni & Baceviciute, 2013). Where *narrative closure* is defined as "*The experience of coherence and completeness of understanding after having experienced a narrative, even though the narrative's substance is not understood very closely to the way it was intended by the author or creator.*" (Bruni & Baceviciute, 2013). The model by Bruni and Baceviciute (2013) goes into more detail as to also explain narrative intelligibility and closure with respect to each other.

Narrative Intelligibility

The understanding of the substance of a narrative very closely in the way it was intended by its author or creator. The level of intelligibility defines the author-audience distance.

Narrative Closure

The experience of coherence and completeness of understanding after having experienced a narrative, even though the narrative's substance is not understood very closely to the way it was intended by the author or creator.

A successful narrative communication would ideally entail both narrative intelligibility and closure. However this ideal may vary according to the intentions of the author-designer.

Narrative Intelligibility without Narrative Closure

This can theoretically happen in a case in which the author intended to transmit a specific narrative substance but his narrative construction lacks coherence and completeness, and this is faithfully transmitted to the receiving subject who then experiences the original lack of closure.

Narrative Closure without Narrative Intelligibility

The recepient can experience a sense of coherence and completeness without necessarily understanding the narrative as intended by the author (or sender), or when the author does not intend to communicate a particular narrative substance but is interested in producing the experience of narrative closure (as in the case of emerging narratives).

Box 1. The Relation between Narrative Intelligibility and Narrative Closure

Figure 4: The relation between narrative intelligibility and narrative closure (Bruni & Baceviciute, 2013)

J. O. Ryan, Mateas, and Wardrip-Fruin (2015) mention the concept of state space, which is the different possible states for a narrative. By having an interactive narrative, the amount of possible states grows, and thus the state space enlarges. To cover the massive potential state space of an emergent narrative J. O. Ryan et al. argue that it must have Modular Content, which they describe as: "*with relatively little effort, a human should be able to produce many units of content that each express specific aspects of underlying state*". For the truly emergent narrative, they argue that the game should be able to cover as much state space as possible, which is made possible by reusing lines of dialogue. However, by making the lines generic enough to be reused, they also lose expressive content (J. O. Ryan et al., 2015).

3.3 Agency

The prior chapter leaves us with a reasonable toolkit for designing from the author's perspective, but as stated, we need to appease the user's desire for interaction. Stern (2008) argues that the priority of any game wishing to achieve its potential as a new narrative form must provide the player with true agency. Agency is defined by a variety of different scholars, who seems to agree that its related to interaction. In the Oxford dictionary, agency is simply "*Action or intervention producing a particular effect.*"[2]. However, some scholars consider agency as the sense a user experiences when taking meaningful actions. One such example is Murray (2000), who defines it "*the satisfying power to take meaningful action and see the results of our decisions and choices.*". So more than just interaction, or a sense of interaction, agency is the sense of taking meaningful actions, and to see the results of those decisions.

Mallon (2008) seeks to expand on the definition of agency: "*Player agency provides an illusion of player authorship: that the active and creative realization of the game text is accomplish by the player.*" The illusion Mallon (2008) refers to can be related to emergent narrative, or the fulfilment of an embedded narrative. We can consider whether it can be true agency if the player merely believes to have an effect, even though in reality everything has been carefully laid out by an author (as in the case of Laurel's flying wedge). Can it be said that the actions matter if they just follow the scripted paths? Mallon (2008) argues that the player does not see this distinction and that they cannot see the difference between perceived and real opportunities. Instead, Mallon (2008) suggests that a product's ability to cultivate the user's suspension of disbelief provides enjoyment. So, the user can have agency in embedded narratives, assuming the product is well-crafted enough to participate in the fiction. This reasoning seems to be behind the definitions of agency as used by Emirbayer and Mische (1998) and Andreen (2017) that agency is the capacity to act independently and to be free to make one's own choices within the presented structure. This definition may focus more on

[2] Definition of agency in English by Oxford Dictionaries, `https://en.oxforddictionaries.com/definition/agency`, Accessed: 2018-02-26

the mechanics of interaction and delimitations of the structure than the definition by Mallon (2008).

3.3.1 Provoking Agency

In this chapter we review the game aspects as presented by Mallon (2008), originally derived from Mallon's focus group studies. The intention is to fuel our considerations for appeasing the user's desire for interaction:

1. "*The idea that successful accomplishment of game tasks gives players the sense that they are the central protagonists that they are responsible for achieving the goals, is fundamental to the players' sense of agency in game-play.*"
2. "*It was not just the number of variables under the players' command that influenced their enjoyment, it was also the degree of precision with which they controlled them.*"
3. "*The lack of adequate sensory response affected the feeling of control.*"
4. Mallon presents that the skill ceiling may demotivate the user, and points to Flow theory (Czisikszentmihalyi, n.d.) for a different study with a similar conclusion. The reasoning is that if the player cannot contribute to the tasks and goals, they lose the ability to influence the game, and thus their agency.

These suggestions are considered for the design requirements listed at the end of this chapter. To make designs based on these rules, we can also use Mallon's evaluation of the quality of interaction. Mallon (2008) divides interaction into 5 types; (1) Partial but limited control, in which the agency is only in pointing the tool in the right direction and activating it; (2) Attributes of tools, where the user has manual control of the tools; (3) Refined, subtle, and careful action-motor skill, which requires skill to work and provides feedback to the player; (4) Possibility to get better, which relates to Czisikszentmihalyi (n.d.) Flow theory, in that the difficulty of interaction must be in alignment with the user's skill level. As the skill level is raised through play, the difficulty must rise with it; (5) Multiple combined skills, which is the ability for the player to take strategic decisions based on their understanding of the game, and possibly requiring elevated level action-motor skills. The above are examples of how to create agency through game mechanics and interaction, whereas Mallon (2008) also specifies how character relations can create agency, which is likely to prove more beneficial to this project, as the ideal dialogue system will allow for simultaneous agency and author control. Mallon (2008) argues that agency can be made through characterization through; (6) Making intelligent, moral or attitudinal choices, notably with different consequences; (7)

That action consequences are proportionate with players' intentions and motivations; (8) Through building psychological relationships with the game characters; (9) That long-term relationships are remembered, as in that the player's actions and relations to characters are referenced and remembered at later points; (10) Providing the player with the possibility to design their own character. Mallon (2008) mentions abilities, skills, and weaponry, which are all elements that have an exact impact on the gameplay, but also argues that agency could be enhanced or diminished by how well their own character was drawn. It should therefore be considered whether the visual and auditory aesthetics may aid or hinder our prototype, and if the customization allows for any actual impact on the game, this will be considered in the design chapter. While a more extensive chapter could be written on the subject it falls outside of the scope of this thesis. Returning to interaction and how it can provoke agency, Andreen (2017) argues that the link between choice (we call it interaction) and agency is identity. Identity in video games, as Andreen (2017) describes it, is "*The learning experience of engaging with the choice-response-consequence structures in video games informs a player's understanding of the game world.*" He proceeds to quote Salen and Zimmerman (2004) and Gee (2003) for building the framework for identity as an amalgamation of their terms of sensory input, player output, internal player cognition, and Gee's classifications of identity: virtual, real, and projective. Salen and Zimmerman (2004) refers to the choices and experiences the game provides as the sensory input, whereas how the player reacts upon the sensory input is the player output, and how the player develops based on the response to their output is the internal cognition. Andreen (2017) then proposes that this loop creates what he and Gee (2003) refers to as identities. Following Gee's framework, we see that this loop likely influences the projective identity. The projective identity addresses the relationship between the virtual and real identities, the former being the avatar or character's identity, the latter being the identity (or part of the identity) of the real-life player (Gee, 2003). Specifically, Gee (2003) defines the projective identity in two concurrent definitions:

1. "*To project one's values and desires onto the virtual character*"
2. "*Seeing the virtual character as one's own project in the making, a creature whom [he imbues] with a certain trajectory through time defined by [his] aspirations for what [he wants] that character to be and become (within the limitations of her capacities)*"

It should be said that these models by Andreen (2017) and Gee (2003) struggle to explain characterless games, unless one hypothesizes that the player creates a game-identity regardless. Like if the player is playing Tetris, where the player could be theorized to make a temporary identity that is only concerned with winning the game. Gee (2003) investigates this problem further through his real identities and how they can conflict, exemplifying how his real identity of gamer is subdued by his real identity of parent. Further analysis in this area falls outside the scope of this project. In relation to our project, we can accept that the player must be given the opportunity to form a character, but Andreen (2017) does argue that there is no great distinction between having a pre-made character or being allowed to create one themselves, as players should have no trouble assuming the identity of pre-made characters.

3.4 Story Generation

We refer to systems that generate narrative or story worlds with characters and persistence as story generation systems. Persistence is the system's ability to "remember" elements and their context. These systems are also commonly defined by creating the whole narrative on their own, and as such have little concern for interaction between user and system. Here we review two early examples of story generation systems to understand generative content as a proposed solution to the narrative paradox.

3.4.1 Talespin

Talespin is a program that simulates rational behavior through characters in a world Meehan (1977). It has three active components; 1) *The problem solver* that when given a goal, produces other goals (sub-goals) and actual events. 2) *The assertion mechanism* that is in charge of "memorizing" events and contexts of the current instance. What the physical world looks like at that moment, what social relationships exist between characters at that moment, etc. 3) *The inference mechanism* that produces the consequences of an event. So when an event is asserted its consequences are calculated and asserted, and likewise for any new consequences that arise thereby creating a cycle. Talespin works on distinct levels like *Sigma-states* which are the top-level goals such as: What is the story about? Hunger, thirst and so on. *Relationships* is where Talespin gives the characters relations to each other in terms of competition, dominance, familiarity, affection, trust, deceit, and indebtedness. Though a bit limited in variety it is a good handler for creating interactivity between characters and possible preconditions for sub-goals. Furthermore, Talespin uses *Personalities* to describe character traits, by giving the characters attributes, degrees of kindness, vanity, honesty, and intelligence. These character traits is what Talespin utilizes in order to create proper reactions and are part of the inference mechanism for storytelling. But in order for a story to make sense, such as when one character wants to get closer to another a *physical space* must exist for the characters to exist and interact. This is done, through abstract maps in which the characters will ‘travel’ to their waypoints in order to progress the goal. In the case of Talespin, English is used to commune the story meaning. Meehan (1977) needed an *English generator* for these actions. The generator does not use grammar and is straightforward as every action is simply translated into one sentence stating the action. Talespin provides us with an approach for story-world handling, in three aspects, the problem solver, the assertion mechanism, and the inference mechanism that we can consider for generated content in design.

3.4.2 Universe

Lebowitz (1983) presents an alternative method towards creating a story by focusing on maintaining *consistency* and *coherence* throughout the story generation. Lebowitz (1983) achieves this through the motivations of the characters, such that all the actions the characters take, can and should be logically derivable, at least in retrospect from the information available to the reader. It ensures the reader only has the necessary level of information available as it is not desirable to provide too much information to the reader because it could lead to the reader predicting every event that would occur in the story. Lebowitz (1983) also agrees that actions should be based on the backgrounds and personalities of the characters just like in Talespin (Meehan, 1977). This is to maintain a consistency in the story universe, for example when new characters join a universe populated by characters that persisted since the beginning of the story. Lebowitz (1983) refers to this character consistency as a person frame that holds the characters personality traits, interpersonal relations and to some extend goals. For the characters to make believable stories, Lebowitz generalizes by given them a stereotype (examples: doctor, warden, klutz, socialite, nasty-person, movie-fiend, etc.) that acts as a template for the characters’ personality and is then expanded upon as the story progress.

3.4.3 Story Generation Considerations

While story generation systems provide theoretical potential for creating a narrative with coherency and consistency, these systems are still far from writing complete, sensical, and interesting stories. Furthermore, they tell us nothing of how the user would be allowed interaction in such a context, and yet another concern is how generative content usurps control from the author, which makes it a poor tool for an author who wishes to tell a particular story. While story generation systems might lack the capacity for considering and overcoming the narrative paradox, it can still be theorized that their development may lessen the impact of the combinatorial explosion.

3.5 Interactive Storytelling

Where story generation systems mostly generate their content from random input or pre-experience input, interactive storytelling focuses on creating human to system interaction with an understandable and satisfying outcome. Examples Facade and the Character-based Interactive Storyteller are investigated below.

3.5.1 Facade

Mateas and Stern (2003) attempt to create a real-time 3D animated experience akin to being on stage with live actors that are motivated to make a dramatic situation happen. Rather than provide a long and action-packed game, they provide a smaller emotionally intense, unified, and dramatic experience. They argue that the experience is varied enough to support replayability for about 6 to 7 times, due to the branching nature of the plot's development (Mateas & Stern, 2003). Their system takes a pre-experience input of name and gender, which the two characters in the game will use when talking to the player. Interestingly, the agents are based on a system of what Mateas and Stern (2003) refer to as *story beats*, which is essentially a series of timed events based on current context. The player can interfere with the system by the interactions of moving, writing input, and interacting with objects, giving the player an impact on the development of the narrative. The beat-system and freedom of the player means that instead of a plot that provides a small number of obvious choices, the plot is smoothly mutable, varying in response to a global state that is defined through many small actions performed by the player throughout the experience (Mateas & Stern, 2003). Even though the player's input can be almost unlimited in creativity, there are limits both as to what the game can understand, what can influence the plot, and the virtual space of the game, which prevents the narrative from developing in a totally unexpected direction. Furthermore, the beat-system pulls the narrative back on track, should it veer too far outside of what the game can encompass. Some of the requirements Mateas and Stern (2003) needed to fulfil in order to create Facade was to achieve believable agents through A Behaviour Language (ABL) (see Figure 5), the player agent, natural language processing (NLP) and the beat-system and dramatic performance in a 3D story world. The believable agents needed the capability to do several intelligent activities in parallel, such as gaze, speak, walk, use objects, gesture, etc. (Mateas & Stern, 2003). They created the agents Trip and Grace which are comprised of an ample collection of parallel, sequential, and joint behaviours written using ABL. The ABL activities such as walking to the player or speaking a line of dialog is presented as a goal and these goals are supplied with one or more behaviours to accomplish the task. The behaviours do not have to occur parallel but can also occur sequentially and if the goal is a success all the behaviours to the goal also absolves with it. However, if the behaviour fails in its steps towards completing the goal it will find a different behaviour to accomplish the task instead (Mateas & Stern, 2003).

Features of the ABL language:

- WMEs – working memory elements, which are data structures that can hold information (ie, variables), equipped for match tests in conditionals such as preconditions
- preconditions – an optional test attached to a behavior to determine if the behavior is available to accomplish a goal in the current context
- context conditions – an optional test that will fail a behavior if the test becomes false
- success tests – an optional test that will succeed a goal if the test becomes true
- acts – a step in a behavior that takes some sort of action in the world, typically an animated body action such as taking a walk step, gazing at an object, speaking dialog
- mental acts – a step that does some arbitrary internal computation, in raw Java code
- subgoal and spawngoal – activate a new goal as a child of the current behavior, or as a child of a some other specified behavior
- joint with teammembers – when subgoaling, synchronously start the same behavior in multiple believable agents, if possible
- priority – determines which parallel behaviors run before others
- persistence – a way to retry a goal if it succeeds or fails
- ignore failure – if a goal fails, do not fail its parent
- effect only – do not require this goal to succeed for its parent to succeed
- atomic behaviors – during this behavior, do not allow any other behaviors to run (shut off parallelism)
- specificity – specify the order in which to attempt different behaviors for the same goal
- number needed for success – how many children of a parallel behavior are required to succeed for the parent to succeed
- sensors – a way to have WMEs get automatically updated information about the 3D story world
- conflicts – a way to prevent certain parallel behaviors from running at the same time, if needed
- demons – a technique of creating a behavior that waits in parallel to all other behaviors for a condition to become true, and then takes action
- meta-abl – a way to reflect upon and affect the currently active behaviors themselves (kept track of in the ABT), to allow a behavior to directly alter, succeed, or fail other behaviors

Figure 5: Features of the ABL (Mateas & Stern, 2003)

The player also holds an ABL agent known as the player agent, although it does not take any actions on behalf of the player it acts more as a supplier of information for the agents, such as when the player is making significant movements around the room or when the player has been looking at an object for a significant amount of time (Mateas & Stern, 2003). Allowing the believable agents to act upon indirect player actions helps feed the drama manager. The drama manager uses a method Mateas and Stern (2003) calls beat sequencing language: When an activity occurs either prompted by the player or the believable agents a story beat becomes active. Once active the behaviours inside this beat will be tailored to focus the activity of the characters towards a particular narrative direction, while still keeping them broadly reactive to other narrative directions. Meaning that when a beat is active the believable agents will try to coax the player towards resolving the goal and progress the narrative. However, if the player refuses to return to the beats intended activity, the beat may abort, and the drama manager moves on to a different beat, hopefully aligned more towards the player's interactions. For the believable agents to understand the player, Mateas and Stern (2003) created a simple natural language processor (NLP) that utilizes a custom natural language understanding (NLU) template. This non-general, a-theoretical, author-intensive technique can draw basic context out of the typed text and determine which reaction the believable agents should take (Mateas & Stern, 2003). This is done through two phases; Phase 1 involves recognizing which action to take out from the typed text by using simple rules if certain keywords are recognized in the typed text. Phase 2 is about choosing the potential reaction based on the recognized words from phase 1, these reactions have been authored for each beat written in a Reaction Decider language. Through these interactions the drama manager is capable of understanding if a beat has been "collected" or lost and will progress the narrative accordingly (Mateas & Stern, 2003).

3.5.2 Character-based Interactive Storytelling

The interactive storyteller approach proposed by Cavazza and Charles (2005) is a Character-based Interactive Storyteller (CIS). In CIS each character has their own role and affinities to other characters and topics which influences their dialogue when communicating with or about other characters. Due to the structure of the interactive storyteller the scenario is structured as a sitcom in which character communicate with each other. Cavazza and Charles (2005) took this approach as it provides a manageable context for narrative generation while also allowing for assessing the validity of the technical approach and evaluating scalability.

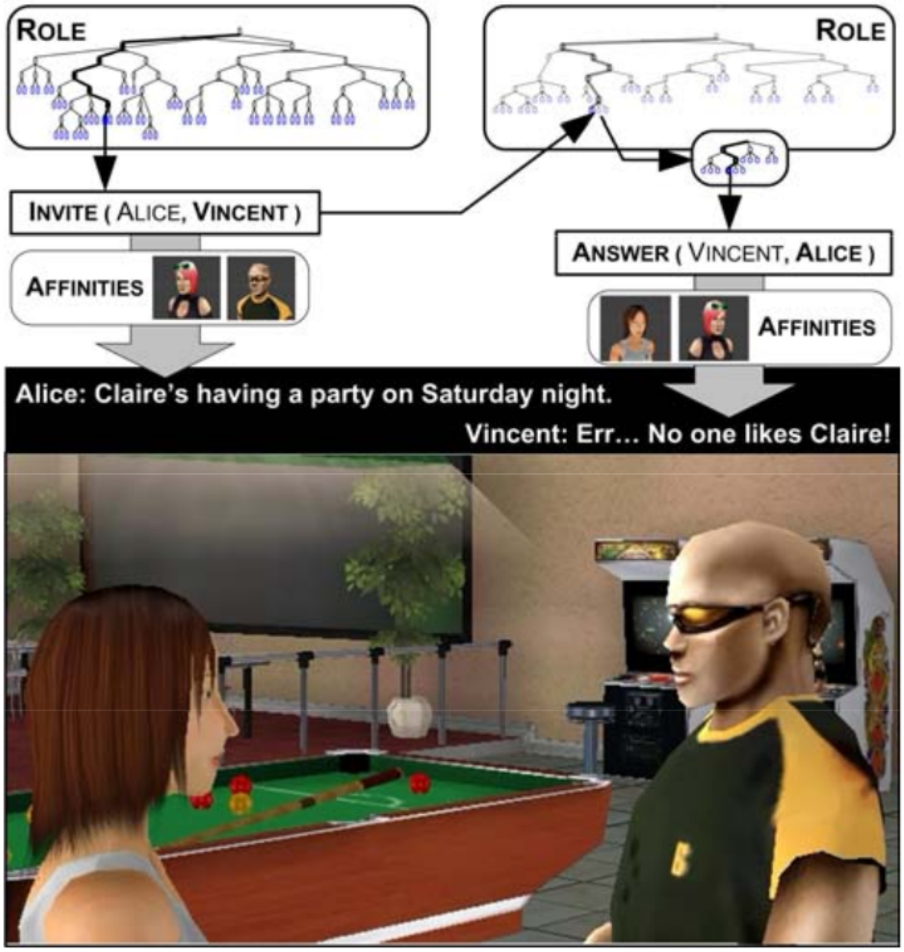

Figure 6: An example of character's roles and the influence of their affinities towards other characters (Cavazza & Charles, 2005)

It is with this interaction between characters that the narrative develops, so to make it interesting for the reader the linguistic expressions should display the properties of real dramatic dialogue. While one of the characters will take on the role of the influencer, who can take different rhetoric functions such as advice, threat, and seduce to express the influence the character seeks (Cavazza & Charles, 2005).

3.5.3 Interactive Storytelling Considerations

The interactive storyteller type provides plenty of emergent activities the player can influence and partake. Facade is more self-driven which means that the believable agents themselves can provide the necessary steps towards completion of the goals set out by the author, prompting rather than forcing the player to partake in the events. These dialogue systems may be capable of providing the agency we seek. The interactive storyteller method does presume that an existing story world exist and that there is a goal to accomplish, but how they accomplish it can vary as seen in Facade and CIS.

3.6 Chatbots

Chatbots are a more generalized usage of dialogue systems as their purpose is in closing the gap between human and machine communication on a natural language basis. There are different varieties that tackle different problems, examples of such are Cleverscript and ELIZA.

3.6.1 Cleverscript

Hill, Ford, and Farreras (2015) does a study on Cleverscripts variant known as Cleverbot[3], where they study the capabilities of the chatbot's ability to construct believable sentences. Since Cleverbot learns from every conversation with humans its knowledge on many topics keeps expanding. It has potential to learn how to chain single words or do proper messages using internet 'slang' and knowing when to use emoticons in a proper context. However, due to the nature of how Cleverbot learns, and that it has no restrictions on what it can learn, a side effect is that it learns profanity and uses it in the same contexts as it learned it. These are all factors that helped Cleverbot in passing the Turing-test (in 2011) as it can great nuance of vocabulary (Hill et al., 2015).

3.6.2 ELIZA

ELIZA was an early communication attempt between man and machine using natural language (Weizenbaum, 1983). The concept was to have a human write a sentence and ELIZA had to respond while understanding what was said by the human. ELIZA can understand pre-taught keywords and learning new phrases from the user but is still limited in the response ELIZA can provide. This is due to the minimal understanding of the natural language ELIZA can comprehend due to the complexity of natural language. ELIZA had a ranking system that would favour some keywords over others, as scripted by the *teacher* (Weizenbaum, 1983). Even with these limitations the ELIZA program still managed to convince users that they were talking with an actual intelligence and thus ELIZA was one of the first

[3]A chatbot first launched on the internet by Rollo Carpenter (1997).

chatterbots that passed the Turing-test. ELIZA uses a YOU ME structure to construct a conversation, meaning that ELIZA would wait for the user to input a response before replying. However, the YOU ME still allowed for a simple conversation branching thanks to the pre-scripted categories using keywords and a rule list structure. The rule list structure is a tree of possible conversation progressions ELIZA can consider, but if a user input would not match her keywords, she would switch to the closest fitting branch based on a "pattern match" algorithm (Weizenbaum, 1983).

3.6.3 Virtual Personal Assistant (Siri/Google Assistant/Cortana)

A virtual personal assistant (VPA) is a dialogue system constructed with the purpose of providing a purposeful response to the user that can provide them with the necessary information or function desired (Zubairm, Bhat, & Lone, 2017). Examples of such usages can range from getting weather forecasts, setting up reminders, telling jokes, sending emails, finding files, searching the internet, etc. (Zubairm et al., 2017). All of this while still having the capability to adapt to the user's preferences thus learning how to respond in the future without even needing to ask the user. The difference here unlike previous mentioned chatbots is that most virtual personal assistants utilized on a phone has the capability of analysing speech. In this example we will cover Cortana and the inner workings of that VPA. This is due to the similar nature of the other known VPAs such as Google Assistant and Siri, with only slight differences being how they handle their online data (Zubairm et al., 2017). There are five steps before the VPA can provide a proper response from the user's request. 1) Signal processing is what enables the VPA to extract analogue signal to digital, by utilizing filter or the gain control to extract the information within signals to translate it into recognizable words. 2) Speech recognition is where the VPA processes the actual recognition. This is done through the feature vector sequence by decoding the sequence of words that through the algorithm Dynamic Time Warping is split up into a sequence of word, by measuring the similarity between two temporal sequences that may vary in speed during the time series analysis (Zubairm et al., 2017). In Cortana this is used specifically to cope with different speaking speeds. 3) Semantic interpretation is the step where the VPA checks the phrase for grammar, and to see if there is a combination of words that makes sense. If the VPA finds a semantic property that the VPA understands, it will respond to the user accordingly and execute the proper command. 4) Dialogue management is the next step after understanding the semantics where it sends a list of instructions to other parts of the dialogue system in which the semantic interpretation is converted to human language by the natural language generation component. 5) Response generation once the task is performed the response of the result of that task is generated. Here the VPA decides whether the response is desired in speech or text and replies to the user in an understandable language (Zubairm et al., 2017).

3.7 Building a Model for Evaluating Interaction

In this chapter we investigate how to understand and classify interaction that we may develop a model that can aid in the analysis and design of interactive systems. In this section we seek to establish a foundation of prior work and a vocabulary to help us define the various concepts for the model.

3.7.1 Levels of Interaction

In her study from 2000, Aylett presents three hierarchical levels of a narrative, pictured below in Figure 7.

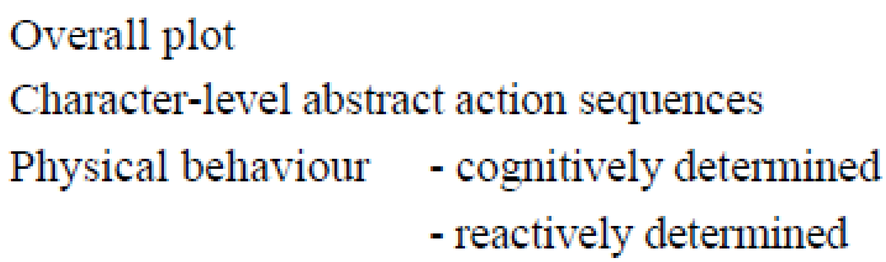

Figure 7: Aylett (2000) suggestion for hierarchical ordering of a narrative.

The topmost level is the overarching plot, Aylett (2000) presents as how one would quickly describe a movie or novel without going into details. Aylett (2000) proposes a generic example of "*boy meets girl, boy loses girl, boy performs heroic feat, boy regains girl*" to explain the topmost level. Even in interactive narratives, this level is almost always controlled purely by the author, like in the attempt by Louchart and Aylett (2003) to solve the narrative paradox through inspiration from tabletop roleplaying, in which they conclude that the game master oversees the plot. Though the player's actions may seemingly change the plot, it is ultimately the game master's role to use the tools at her disposal to develop the plot based on the player's interaction (or to draw them back into the original plot by railroading[4]. The level below known as character-level abstract action sequences (or simply abstract actions),

[4]Term borrowed from Flowers, Magerko, and Mishra (2006) and the role-playing community). Refers to the game master forcing the players to follow a specific plot.

is described by Aylett (2000) as the abstract actions taken by characters. In the formerly mentioned example she proposes that the first part of the plot could be described in abstract actions as "*come into the room, walk up and say hello*", "*creep up behind the character and say 'boo'*" or "*stand close to the character near others and join in an existing conversation*". The point being that while the plot remains the same, the general approach and actions of characters can vary greatly. The lowest level describes the specific physical behaviours, which she deems to divide in cognitively determined and reactive, which is the exact words and movements of characters, be they intended or unintended by the character in question. We suggest changing the level of abstract to *approach*, and the level of physical to *concrete*, the intended difference being to distinguish between overall and specific interactions. In relation to gameplay approach describes the possible paths a player can and does choose to take, though it should be considered that the approach can have intended or unintended consequences. So, while the player might choose to sneak in or go in guns blazing, the potential for failure at the concrete level, ex. being discovered or running out of ammunition, can force a change of approach.

Overall plot
Approach
Concrete
- cognitively determined
- reactively determined

Figure 8: An updated version of Aylett (2000) suggestion for hierarchical ordering of a narrative.

Less explicit Aylett (2000) identifies interaction and author control in these levels going from the extreme of narratives scripted down to exact details to the opposite extreme of heightened levels of improvisation, and she ponders whether improvisation is at all possible on the topmost level. Further along, she suggests that narrative of the higher levels can be defined by improvisation at the lower levels and refers to such as emergent narrative (2000). As such she concludes that a bottom-up approach to the model can theoretically create improvisation on the plot level (Aylett, 2000; J. O. Ryan et al., 2015). Thus, we argue that the three levels can each be interacted with as either scripting or improvisation. We use the former to refer to the author's creation of the narrative and application, and the latter to refer to the user's interaction. An interesting aspect of these two words is that they both involve an element of time, scripting being pre-experience and improvisation being during-experience. While it would be interesting to view the author's possibilities as during-experience (to some extent like a tabletop game master), or improvisation as pre-experience (ex. a game where the player decides all actions before starting the simulation), it is not the focus of this thesis, and scripting therefore only regards author's control before game start, and improvisation only covers user's interaction after game start. A narrative in an application with low scripting moves towards the emergent narrative, while a narrative in a highly scripted application can be said to move away from the interactive. An interactive application has a degree of both scripting and improvisation, but we argue that levels likely have semi-dependent degrees of both, outside of the extremes of wholly scripted and wholly improvised. To illustrate, an RPG-stealth game like Dishonored[5] provides the player with interaction at the character level (concrete and approach levels, (Louchart & Aylett, 2003)), in that they can choose their approach to each mission, and that they control the main characters concrete actions through the game controls. However, while the player can choose approach, the different approaches have been scripted by the level design of the developers, giving them as authors control of how the player proceeds. On the concrete level, the character may control when the main character swings his sword, but the attack animation and hitbox are scripted, meaning that the player could ex. never choose to make a vertical upwards swing, unless scripted to do so. Special for Dishonored, the player also gets a chance to influence the plot level, as the developers implemented a system that changes the game world and ending based on the number of NPCs the player kills (*Dishonored developer Harvey Smith details "Chaos" system and Dunwall. Kernel Description*, n.d.). Returning to our definitions of scripting and improvisation, they can help us further define the relation between the author's control and the player's interaction in a given level. However, as we do not know an exact correlation between scripting and improvisation, we can at best propose scenarios with high scripting and low improvisation or vice versa. It is also possible for the scenarios to be wholly scripted or improvised, but as the former lacks interactivity, and the latter lacks an author, it does not fall within the problem area. Theoretically we can also consider whether it is possible for a level to have half-half scripting and improvisation, but must ask: How would one quantify either adequately? Instead we propose that it is only currently possible to consider which of the parameters of authority is dominant, and which is not. Thus, we can consider interaction in levels with either (a) High improvisation and low scripting, (b) High scripting and low improvisation. The concepts of improvisation and scripting allows us to describe overall or on the different levels, the player's capacity for interacting with a system. However, these

[5]Game developed by Arkane Studios, published by Bethesda Softworks, 2012.

terms do not allow us to say much above the types and details of the interaction, reducing it to mere levels of capacity. Therefore, we investigate types of interaction below.

3.7.2 Ryan's Types of Interaction

M.-L. Ryan notably have two models considering the types of interaction, both of which are prudent to consider. In M.-L. Ryan (2001) she considers the terms of reactive, random selective, purposeful selective and productive interaction, found in Table 1 below. While in her work from 2006 and notably the revision from 2015, she leaves out the prior model in favour of one based on her principles of external-internal and exploratory-ontological, rephrased in Table 2 and the four combinations presented by M.-L. Ryan (2006) reproduced in Table 3.

Table 1: Ryan's types of interactivity provided for reference (2001)

Type of Interaction	Ryan's Definition
Reactive	"*[...] which does not involve any kind of deliberate action on the part of the appreciator.*"
Random selective	"*When the user takes action deliberately but cannot foresee the consequence of his actions, the purpose of interactivity is to keep the textual machine running so that the text may unfold its potential and actualize its virtuality.*"
Purposeful selective	"*[...] may be offered a choice between two paths, one of which leads to success and the other to failure, and the game may cue the player as to which path is the good one.*"
Productive	"*In the fullest type of interactivity, finally, the user's involvement is a productive action that leaves a durable mark on the textual world, either by adding objects to its landscape or by writing its history.*"

Table 2: Ryan's categorization of interactivity provided for reference (M.-L. Ryan, 2015).

Internal	**External**
The player is in a character in the world,	The player is in a god-like state
Exploratory	**Ontological**
The player's actions have no lasting influence.	The player's actions have lasting influence

Table 3: Ryan's terms for defining a level of interactivity, paraphrased (M.-L. Ryan, 2006, 2015)

Internal-Exploratory	**External-Exploratory**
The user is given a virtual body, but their actions have no bearing on the world	The user is external to both time and space, interactivity is limited to choose routes through the textual space, disconnected from the physical space of the narrative.
Internal-Ontological	**External-Ontological**
The player is given a virtual body with which to have lasting influence on the world, such as the possibility for failure related to the fate of the world.	The user plays god to a virtual world, influencing entities and changing the world.

M.-L. Ryan's two models have distinct differences, though some of the aspects have commonalities. Productive interactivity and the ontological term are both partly defined by the interactor having a lasting impact. A point could be made that the 2001 model in Table 1 describes instances of interaction, multiple of which could be present within the same interactive application, whereas the later model describes the overall interactivity of an application or narrative. Although one could argue against this point, by hypothesizing that the player's interaction can change on the explorative-ontological axis over the course of gameplay. An example being that navigating the map might carry no consequences but killing specific NPCs will change the outcome of the narrative (and killing them might be a lasting change to the world). To obfuscate the distinction even further, games like Rise and Fall: Civilizations at War[6] and Dungeon Keeper 2 [7] allow the player to play the strategy element from a godlike perspective, but to enter individual characters during the game, thus muddling the use of

[6]Game developed by Stainless Steel Studios and Midway Games, released 2006

[7]Game developed by Bullfrog Productions, and published by Electronic Arts in 1999

the internal-external axis on the application as a whole. Returning to our terms of scripting and improvisation, we can compare them to explorative and ontological interactivity. As exploratory allows for little change (as any lasting change would be ontological), it could be argued to be the result of high scripting, whereas ontological would then be a result of low scripting. However, this claim only stands if lasting change which is scripted to be possible does not count as ontological interaction. Seeing as the narrative and game world changes due to the player's actions in games such as Assassin's Creed[8], yet are scripted to respond in very specific ways, the former claim depends on whether ontological change is still ontological even if heavily scripted, or whether the pre-made nature of it makes it inherently exploratory. Mallon (2008) however argues the players cannot see the difference between "*perceived versus real opportunities for the user to input into the program*", which makes the conundrum a debate for academics and developers, more so than a problem to solve. We consider exploratory and ontological as terms too vague and too dependent on the level of detail to be useful without it. They do have use in conjunction with other terms, such as improvisation and scripting, or M.-L. Ryan's external and internal terms, but even then we consider that they must always be viewed with the level of interaction in mind[9].

Table 4: Using M.-L. Ryan's terms for exploratory and ontological interaction, we combine them with our terms for improvisation and scripting. We consider that when evaluating a system with these terms, the level of detail must be accounted for, as an application can have different combinations on the different levels described by Aylett.

High Improvisation-Exploratory	**High Scripting-Exploratory**
The player has high freedom to interact, but with no impact on the game world or narrative.	The player has little possibility for interaction, and no influence on the game world or narrative.
High Improvisation-Ontological	**High Scripting-Ontological**
High player freedom to interact, and noticeable non-scripted impact on the game world and/or narrative.	The player has little possibility for interaction, but it leads to prescripted conclusions and changes to the game world and/or narrative.

[8]Game developed by Ubisoft Montreal and published by Ubisoft, 2007

[9]As an example: Assassin's Creed appear exploratory on the plot level. No matter how the player proceeds, Altaïr will slay Al-Mualim at the end of the game and take over the order of assassins. Failure is not even an option, as upon Altaïr's death, the game pretends the real-world protagonist was desynchronized from the memories of Altaïr, and resumes the game before Altaïr's death. Though, throughout the game the player is relatively free to choose their approach for how to assassinate the various targets, but the assassination itself is scripted to play the same cutscene every time.

3.7.3 The Case of Productive Interaction

Returning to M.-L. Ryan's term of productive interaction, we here intend to further develop the concept, and discuss some of the issues of her model. To explain a key issue, we would like to present the metaphor of a jigsaw puzzle: A jigsaw puzzle has only one correct outcome, as the user has no room to improvise on the outcome. While the jigsaw might lack a narrative plot, it does have a well-defined conclusion. However, the user's approach to making the jigsaw can change drastically. One approach is to choose one random piece, and then in turn test every other piece next to it, in every possible solution. We argue that this becomes permutations of selective interaction, as the user simply tries different options. Yet, few would solve an extensive jigsaw like this, due to the tediousness of trying every single combination. Instead most users rely on self-invented heuristics, systems based on colour, shape, gathering the corner pieces first, and possibly others. We consider that the same effect may occur in interaction be it analogue or digital. Here we argue is a key flaw in M.-L. Ryan's (2001) model. She argues that productive interaction must leave a durable impact, but fails to specify durable: Is the order in which jigsaw pieces are connected durable? The end results are the same, yet the interaction has been guided by the user's desire to solve a system. One could argue that her explorative-ontological model seems to rectify this, but we propose and argue that the key to productive interaction lies not in the result, but in the process. While the word 'productive' carries with it a requirement for production of something, we see that it can either be the production of the user's mental model, or because of it, the production of something in an interactive system. We propose that the durable outcome is not inherent to productive interaction, but nonetheless possible. One could argue that productive interaction is the more complicated form of the cognitive process of decision making, in that it occurs when a choice is not simple to make. That is not to say that it must be hard, but that the complexity of the task must be sufficient to require prolonged consideration. To exemplify, a player must take a decision for purposeful selective interaction to occur. This decision requires for them to consider the various factors to make an informed choice. Thus, the player engages in a form of quasi-productive interaction, at the very least they can be said to construct a mental model, albeit simple, to solve the task at hand. When the information needed to take an informed choice exceeds a certain threshold, or when repeated instances of interaction makes a significant memory requirement of the user, the interaction nears our definition of productive. We also consider however that the user must be willing to enter into the state of being productive interactive. For example, a tired or inattentive user might try random combinations when laying a jigsaw puzzle, using random selective interaction as a learning tool or merely as a least possible effort attempt to solve the system. As such, the user has the possibility to make any interaction random by not allowing for the mental model required for productive interaction to take shape. We can also consider whether the user can raise the interactivity to a higher level and argue that if the player can create meaning in an otherwise selective interaction, they can approach productive interaction, as they can almost always fall back on the complicated system of their own mind and desires. However, even if the cognitive processes of the human mind are considered productive, we stress that productive interaction is a higher level of multiple selective interactions combined with thought process meant to understand the system. As such selective interaction and choices in general are using the same principles of productive interaction, but on a lower level. Thus, we argue to make the distinction that selective interaction is the term used when the interaction is 'easily' understandable and 'easy' to carry out. We can consider, but not currently elaborate on, whether there is a noteworthy difference between the productive interaction that stem from systems that a inherently productive, and interaction with simpler systems the user upgrades to be productive through meaning-making and intentionality. Another consideration that is worth bearing in mind at this point is the similar concept of flow, or flow theory, coined in 1975 by Mihaly Csikszentmihalyi. Flow is described as the optimal experience, the sense and situation in which the user forgets their surroundings in favour of devoting their full attention to the task at hand (Nakamura & Csikszentmihalyi, 2009). We see a correlation between flow theory and productive interactivity because both rely on the relation between the user's skill and the challenge of the task, but flow seems the likely response to productive interaction. We cannot claim that only productive interaction can create flow and are aware that if the everyday cognitive processes of people are considered productive, the two terms approach a state of being indistinguishable. Therefore, we would like to present productive interactivity as the possible goal of a system, which when designed for should ideally create flow. As such, we can also evaluate productive interaction with a system by its capacity for inducing and maintaining flow. A crucial and complex problem we may draw from flow, is that a productive system must involve a challenge fitting to the user, and it must grow in difficulty, so the user's development of skill never supersedes the challenge. Opposite, but equally important the system must not be too complicated for the user or must include ways for the user to regulate the difficulty (Nakamura & Csikszentmihalyi, 2009). Thus, based on the jigsaw example, we redefined the requirements for productive interactivity that; (1) The interaction must be with a system of adequate complexity, or with sufficient contextual material for the user to raise the complexity of interaction themselves; (2) there must be the possibility for understandable quasi-unique progression through the system; (3) which leads to understandable and defined consequences and outcome.

To exemplify the use of these requirements, we consider the games of Ludo, Shadespire, and

Assassin's Creed. The two formers are analogue games of varying complexity, while the last is a digital game.

3.7.4 Ludo

In the well-known game Ludo, players take turns moving around one of four pieces a number of steps equal to a dice roll, while landing on specific spaces on the game board will have additional effects. The game ends when a player gets all four pieces through the entire board. Ludo only ever provides the player with a maximum of four options: Which of their pieces to move, and there is no underlying context that heightens the complexity of the choices. Thus, Ludo fails the first of the requirements. The progression through the system, the process of understanding how the system works is also extremely limited, and the second requirement is likely exhausted within a few minutes of play. The outcome of both individual actions and finishing the game are both understandable and defined. This leads us to the theoretical conclusion that Ludo is inefficient as a system for productive interaction.

3.7.5 Shadespire

Shadespire is a game developed by Games Workshop[10], which has each player control a group of figures with varied rules on a hexagon map, while playing cards to either influence their own or the enemy models, and cards to score points during the three game rounds. Each round each player has four activations to use on their models as they wish, with certain repeated actions or combinations of actions restricted, and some actions requiring the roll of specialized dice to succeed. Considering the movement aspect of the game, each of the players' 3 to 7 models have the possibility to move a number of hexagons in any direction. Furthermore, they can attack enemy models, with support from either cards in their owner's hand, or nearby models affecting the combat. A part of the game is then the outmaneuvering of the enemy and bringing the player's models into combat with the best possible match-ups, but without leaving room for the enemy to retaliate with other models. The sheer amount of possibilities makes for a system of adequate complexity, but for the player to understand these interactions, they must know the rules before playing. This could be theorized to remove the aspect of productive interaction that is learning the system, which is why we do not consider learning the rules a requirement for productive interaction. Alternatively, future research could consider if the learning aspect of a game is truly exhausted when the

[10]British miniature wargaming manufacturing company, owning IPs such as Warhammer 40K and Age of Sigmar.

player has general knowledge of the rules, or whether there exists a 'hard to master' aspect, which potentially also provides productive interaction. The use of dice to resolve certain actions, and the amount of actions available, combined with the random cards chosen, means that two games of Shadespire are unlikely to be the same. The system is also quite clear on what happens when, to makes the consequences understandable even when players are new to the rules. Thus, we theorize that Shadespire as a system has the potential for productive interaction, as it fulfils all three requirements.

3.7.6 Assassin's Creed

Considering triple A games as systems for productive interaction is a complicated matter. There are some of the underlying systems that we cannot fit in this brief example, so instead we will focus on what sets the digital apart from the analogue. First, the ability of the computer to run aspects of the game without the player's involvement opens for artificially controlled characters and the world reacting to the player. Assassin's Creed uses such a system to use non-controllable characters as obstacles, making it more difficult for the player to escape on the street level. It also uses such a system to control the characters who serve as enemies, who use different approaches to attack or hinder the player character. How these enemies react can be hard to predict, as they are almost as free to move around as the player. Digital systems also make it possible to use real-time movement, like children's make-believe, but unlike board games which are almost always turned based. The inclusion of time increases the likelihood of a quasi-unique approach, while also increasing the cognitive burden on the player's perception, as they must repeatedly evaluate and re-evaluate their approach. The systems presented up to now are used in Assassin's Creed to make the sub-system of movement into productive interaction. Combat, on the other hand, is restricted to attacking, blocking and counter-attacking. Attacking is done by repeatedly pressing the mouse button, usually until the enemy is dead. Blocking is holding down the other mouse button, and counter-attacking is done by pressing the attack button, while blocking when an enemy attack. While the player has the possibility to attack, the game is scripted such that counter-attacks will immediately dispatch most enemies, making it a more efficient and safe form of combat. While choosing when and who to attack could provide a quasi-unique approach, waiting around for the enemy to attack does not. If counter-attacking is viewed on its own, it fails to provide productive interaction, but the choice of preparing for a counter-attack instead of attacking might be enough of a consideration for the player to experience it as a choice, at least until the novelty of the choice has been exhausted. Thus, we note that digital systems are hard to define, as some must be viewed as separate systems for productive interaction.

3.7.7 Andreen's Types of Interaction

Andreen (2017) presents interactions in an ordered fashion based on how the player can carry out the interaction and what consequences it can result in. Like the earlier levels by Aylett (2000), Andreen's considerations allow us to consider when and at what levels the player can experience productive interaction. Below follows a review of Andreen's hierarchy of ordered choices and their relation to productive interaction. **First-Order choices** are purely mechanical and commonly instinctive, and cover the simple actions of pressing a button, moving a mouse, or increasing an attribute. Andreen (2017) subdivides First-Order choices in spatial, interactive, and statistical choices. These he describes: "*Spatial determines where a player moves, interactive determines what a player can do once they arrive at a destination, and statistical determines to what extent they can enact a decision.*" (Andreen, 2017). Combining First-Order choices can make for more complex interactions, an example being first person shooter games that require the player to simultaneously look around, move, and shoot. On the base level, first-order choices lack the complexity necessary for productive interaction, but when combined as in the example above productive interaction can occur. This kind of productive interaction is inherent to the system, assuming the player allows themselves to partake in it. **Second-Order choices** implement narrative elements that are explicitly presented mechanically, which are further subdivided in binary choice, multi-variable choice, and false choice. Binary choice is defined by being two options with distinctive outcomes, and multi-variable is similar, but allows for more than two options. False choice is defined by pretending to be a choice but having the same conclusion regardless of the player's decision. Second-order choices only result in productive interaction when the player must carefully consider the context, as the explicit nature of the choice removes the possibility for a quasi-unique approach from the act of the interaction. False choice eliminates the possibility for productive interaction, as it violates either the first or the second requirement: If the user understands that their choice does not matter, the system loses complexity, or if the user's actions to not results in the expected consequence, the outcome cannot be understandable. **Third-Order choice** is the combination and rearrangement of first- and second-order choices into larger structures. Essentially it would seem third-order choices describe the branching narrative structures discussed by M.-L. Ryan (2001) and many others. Andreen (2017) separate them in simple branching choice, complex branching choice, and implicit choice, where many of the branching narrative structures M.-L. Ryan (2001) presents could be taken as specific examples of these. However, Andreen's model visualizes something not inherently visible in M.-L. Ryan's models. Andreen's simple branching choice covers all narrative structures that are defined by multiple binary and multi-variable choices that lead to different endings, but otherwise does not change the material. Complex branching choices on the other hand delimits parts of the narrative content based on other unlocked pieces of content. Thus, after solving one part it might not be possible for the player to complete other branches. The interconnectedness of the branches explains a depth that two-dimensional models are hard pressed to relay as the amount of material grows, as size and complexity of the model makes it difficult to grasp in its entirety. M.-L. Ryan's narrative structures cannot encompass this concept without an addition, like for example 'locks' that visualize the relation between completion of one branch and another branch being removed from the model. Example in Figure 9 choosing paths a or b first will delimit the player's access to content, but taking path c first, then path a, and finally the b path will only lock content after it has been completed. Finally, for third-order choice, Andreen (2017) presents implicit choice: choices that matter, but with consequences that are not readily transparent. In Figure 9, this could simply mean starting path a or b did not warn the player in any way that other content would be locked upon completion of the path.

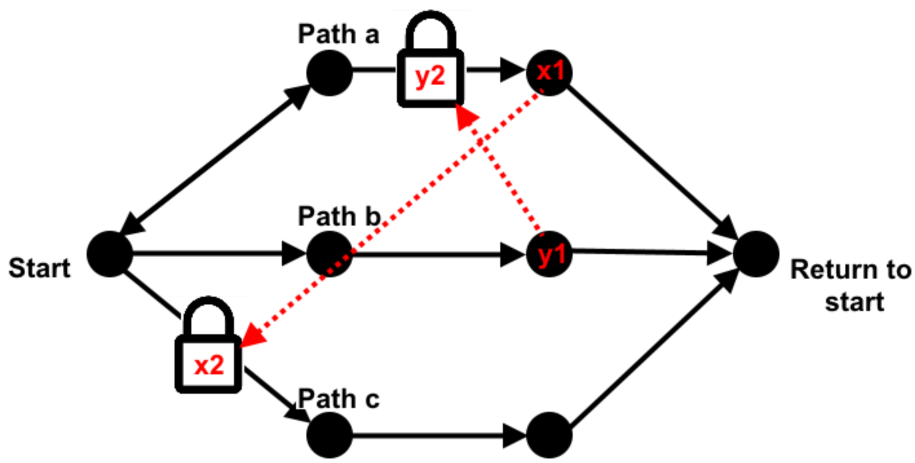

Figure 9: Branching narrative structure with content limitations. Reaching x1 will lock the path marked with the x2 lock and reaching y1 will delimit the path marked with the y2 lock.

Considering productive interaction at the level of third-order choice is made difficult by the difficulty of digital games to provide interaction at this level. The complexity of the narrative structure and where it can lead, combined with the lack of interaction at this level, means that the player is rarely expected to have productive interaction in this area, though it cannot be denied possible. We expect purely theoretically that users might only have the understanding necessary to partake in productive interaction at this level upon replaying a branching narrative. The knowledge of their prior choices gives them an understandable outcome, and the knowledge of which actions will take them there. This knowledge can be used to choose a different path, or to replay the original path, or it can be disregarded, depending on the player's desire to partake in productive interaction at this level. **Fourth-order choices** rarely provide any design considerations, because they are defined as the player taking actions outside of what the developer anticipated and designed for (Andreen, 2017). To some extent it could be argued that the fourth-order choices observed in other games could influence design choices in later games, but if the developer deliberately implements it, it becomes a lower level order choice. Andreen (2017) separates fourth-order choices into exploratory; which is where and why the player go where they go, constructive; in which the player constructs something not predefined, and interpersonal; which is when players interact with other players. Interpersonal can further be subdivided into cooperation, competitive, and incidental, as different ways the players may interact with each other. Fourth-order choices depend on the player to make the system that they are trying to solve, in the attempt to infer productive interaction. This can be compared to the make-believe games of children, which may be among the strongest examples of productive interaction, as it requires the participants to simultaneously create and understand the system that they are building. A note could be made, and in the future researched, as to how productive interaction relates to the concepts of ludus and paidia, as coined by Caillois (1961). The transformation of the free paidia to the rule bound lupus in children's play could be theorized to be a prime example of productive interaction at work, where the outcome is the game system, the approach is everything the participants agree on, and the complicated system is based on the desires of the players and social systems of cooperation.

3.7.8 A Model for Observing Productive Interaction in Levels (OPIL)

Our model considers three requirements for productive interaction, and as they can be used during design, we also argue that they can serve as the foundation for any test seeking to determine the productive nature of an interaction.

1. The interaction must be with a system of adequate complexity, or with sufficient contextual material for the user to raise the complexity of interaction themselves.
2. There must be the possibility for understandable quasi-unique progression through the system.
3. The approach through the system, must lead to understandable and defined consequences and outcome.

To apply it efficiently, especially in large digital systems, the researcher must consider how to observe and investigate the system, and when to split it into multiple systems. As the example with Assassin's Creed formerly presented suggests, some aspects of a system may present productive interaction, while others fail or were never designed to do so. Here, we would note that while productive interactivity and the flow that it brings is generally considered positive, not necessarily all elements of a system must induce it. First, it can make the system too complicated, which may reduce the flow gain and frustrate new users. Second, we imagine that the cognitive strain of productive interaction is not something which can be maintained without the occasional pause, though some pauses will come naturally when a productive system reaches its natural end, or when a part of it is successfully solved. To consider how to split a system for closer examination, we presented Aylett's model of levels and Andreen's ordered choices. While applying productive interaction in Aylett's levels, which we renamed Plot, Approach and Concrete, the researcher can map the player's freedom to interact with a system, intending to show patterns of where and possibly why productive interaction emerges. Using Andreen's ordered choices, however, allows the researcher to identify what kind of productive interaction is occurring, if it is *system inherent*, *context based*, *meta dependant*, or *player inherent*. System inherent refers to the productive interaction specifically designed for in the rules of a system, where context based depends on the player's understanding of contexts (such as narratives, morals, personal desires, etc.). Meta dependant occurs when the user seeks to understand and use the meta subsystems of a system, and player inherent occurs when the player forges productive interaction outside of the material provided by the system. They are considered to match Andreen's ordered choices as follows:

1. First-order choices can produce *system inherent* productive interaction.
2. Second-order choices can produce *context based* productive interaction.
3. Third-order choices can produce *meta dependant* productive interaction.
4. Fourth-order choices can produce *player inherent* productive interaction.

For systems and subsystems that do not appear to be productive, the researcher can consider them in M.-L. Ryan (2001)'s terms of reactive, and random and purposeful selective. These terms allow the researcher to explain lower level interactions, be they independent or part of a productive system. As mentioned we also consider flow a key outcome of productive interaction, and thus any researcher aiming to use the model should consider the various methods of evaluating flow. As the researcher's goal should influence the choice of flow measurement in conjunction with the OPIL model, we suggest the paper by Moneta (2012) on the measurement of flow, as it provides overview and suggestions as to the use of various measurements of flow since its origin in 1975.

4 Final Problem Statement

To which extent is the OPIL model a suitable framework for designing and evaluating productive interaction?

4.1 Design Requirements

The design requirements are made from what we consider the key elements of OPIL. The requirements provide the foundation of the design chapter, in which we strive to make an example of productive interaction. They are also used in the evaluation chapter to evaluate the outcome of the design. We are aware that designing and evaluating based on the same principles may create a circular logic but consider that the model is the only sufficient measure of productive interaction. Thus, its use for evaluation is utilized to validate its application for design.

- The prototype should strive for a system of adequate complexity. Adequate complexity requires the players to be challenged, but without deterring them from progressing.
- The prototype must allow for a quasi-unique approach for progression, through a form of interaction on the character levels (approach and concrete), leaving the plot level mainly for the authors to control.
- The prototype should include a narrative context of adequate mass, with the intent of it serving as a fallback system, if the game system is not of adequate complexity. This safeguard is necessary, because players' may have different levels of expertise, making the complexity inadequate in some cases.
- For making a sufficient narrative, the ideal amount of narrative content avoids dead ends and critical mass to the extent possible.
- The prototype must strive to have understandable consequences and outcomes in response to player actions, but without making the system transparent to a degree where the player can complete it without giving thought to their actions.
- The prototype should strive to cultivate the player's suspension of disbelief.

- The prototype should consider how it can invoke agency according to the suggestions by Mallon (2008).
- The prototype should be limited to one productive interaction system for the purpose of testing, though it cannot be denied that other systems may leave room for player inherent productive interaction.
- The prototype should be considered in relation to all four types of productive interaction, though in line with the previous requirement, designing for all four is unlikely, while restricted to a single productive system. Furthermore, player inherent productive interaction cannot be designed for outside of providing the player with an open system.

5 Methods

In this chapter we will present the methodologies that will be utilized for our data gathering, and construction of the testing environment to rule out most bias for testing purposes and make the test design. For testing our Final Problem Statement, we require an implementation that follows the design requirements listed in the previous chapter and considered in the design chapter. Due to uncertainty regarding the acquisition of test participants, we utilize convenience sampling, though strive to collect data on their demographics. We test only on the delimited target group based on the requirement for game literacy established in an earlier chapter.

5.1 Test design

Since our design was made with OPIL in mind, it can seem irrelevant using OPIL to evaluate its effect: As an OPIL based design should assure productive interaction. However as argued, it only seems possible to assure system inherent productive interaction, and even then, the model has not been empirically proven. Thus, we seek to investigate both the model's use for design and for evaluation by applying it to an implementation based on our research on narrative development and dialogue systems. For discovering the presence of productive interaction and for identification of each type of productive interaction, we suggest a semi-structured interview, as it assures a higher detail level in the participants' answers. The quality of the information may also prove prudent in regard to identify missing elements of the model, making it ideal for an exploratory approach. To support and prove claims by participants, we intend to log data during the play test, such as the players' text input and to which chatbot. Data from the logs may aid us in mapping the players' approach and locate places in which they change their approach. Furthermore, the logs also allow us to record the amount of errors and fruitless inputs, allowing us to evaluate the possibility for system inherent productive interaction. The logs could also be used in redesign, as a way of removing dead ends from the dialogue system, though we are aware that all state space cannot possibly be covered, due to the combinatorial explosion. For measuring flow, we consider Moneta (2012) evaluation of measurements of flow, and the study questions he poses. As we need to determine flow after a specific activity, Moneta's suggestion seems to be the componential approach by Jackson and Eklund (2002), which he compliments on being psychometrically sound. Moneta however notes that it imposes flow on participants leading to inflated prevalence rates, and that it cannot distinguish between antecedents and facets[11] of flow and is too simple a structure to account for the complexity of these elements in relation to flow. Also noteworthy for this study, the scales by Jackson and Eklund (2002) regards sport and exercise, but it has been rephrased to include general activities by researchers Payne, Jackson, Noh, and Stine-Morrow (2011). We transferred their Activity Flow State Scale into a Google Survey, for easy collection of the data. The scale has 26 items rated between Strongly Disagree (1) to Strongly Agree (5).

5.2 Semi-Structured Interview

For the interview we consider the sub-areas that must be explored equal to the requirements proposed, creating two sub-areas: *system complexity and approach, and context, consequences and outcome.*

5.2.1 System Complexity and Approach

- Please explain how the game worked.
    - Focus on the part about gathering information.
        - What could the system understand?
        - What could it not understand?
        - How could you make the system understand your input
            - What was your approach to gathering information?
            - Did you note any changes in your own approach?
            - Why/when did you change approach?
            - What thoughts went into the change of approach?
        - What is their expected timetable for 100 percent completion?

[11] "Antecedents of flow are internal states and perceptions that precede and foster the flow state but are not themselves expressions of flow. These include, for example, clarity of goals, unambiguous feedback, and perceptions of challenge and skill in carrying out an activity. These factors are theorized to have a causal impact on flow by either increasing the likelihood that flow occurs or by augmenting the intensity of flow. Components or facets of flow are internal states and perceptions that represent expressions of flow. These include, for example, merging of action and awareness and loss of time awareness or time acceleration when carrying out an activity. These factors are theorized to be caused by flow" Moneta (2012)

      - Could they complete it?
      - Would they complete it?

5.2.2 Context, Consequences and Outcome

- What was the game about?
  - What was the narrative?
- Most interesting and least interesting parts of the narrative?
- What happened when you gave the system an input?
  - Were the system responses sensible?
  - Were the system responses satisfactory?
    - Did you note any differences between the characters?
      - Did the player perceive the stereotypes?
    - Did any of the characters stand out in a noticeable way?
  - How do you think the game ends?
    - What do you think was the point of gathering information in the game?
    - What happens after the party?
  - What happens to each of the characters?
    - Your own character.
    - The other characters.

5.3 Test Setup

Requirements per test stations:

- A laptop with the prototype
- A functional keyboard
- An undisturbed environment
- A headset

Actual Setup:

- A MX Master Mouse V1 from Logitech
- Logitech M525 Mouse Wireless Black/Red
- A MSI laptop 5 years old (runs the game on medium quality)
- A Urbanears Plattan 2 Black (USB)
- Environments: (AAU Campus in CPH: Multi-sensory lab, Seminar Room 0.93A; Frederikskaj 12, 2nd floor, room 2.18)

5.4 Test Structure

The following is the structure we expect the individual tests to follow. Any deviations from the test structure must be evaluated as bias in the discussion sessions.

1. The participant will read and sign a consent contract.
2. The participant is seated in front of the test session, which shows the controls on screen
3. The first phase of the test will begin and last for 30 minutes
   (a) 30 minutes is expected to be just long enough for the player to find a productive approach to solving the system, including extra time to learn the controls. It should not be enough to allow the player to exhaust the content.
4. The participant answers the flow questionnaire
5. The participant is interviewed according to the structure presented above.

5.5 Evaluation Methods

As the interview depends on an exploratory approach it is to be evaluated with meaning condensation, and with the intent of building models to explain the individual productive interaction of each participant. Assuming productive interaction, these models can be combined for a general analysis of the implementation. The flow questionnaire will be evaluated with a single sample right-tailed t-test, with the intent of determining if the participants experience an amount of flow significantly above from the threshold. Since the scale is a 1 to 5 scale, the threshold for whether flow is experienced must be 3 and we suggest a significance level of .05.

5.6 Target Group

We argue that it would be prudent to delimit the target group. First off, we are dealing with the game medium, and there might be great differences between non-gamers, casual gamers and the self-defined hardcore gamers. A lack of game literacy might make it impossible for non-gamers to understand the interactions, or to put higher requirements on the game design. It should also be noted that there are many more sub-groups of gamers but assuring enough participants of a particular group may prove problematic. We settle for investigating the type of gamer through self-assessed categories of casual gamer, gamer, hardcore gamer, professional gamer, and non-gamer, and rely on participant's explaining their choice to define these sub-groups. We will not restrict participants on cultural background as long as they are proficient in English, which is enforced by mainly drawing participants from the AAU campus. Preferably the target group will be in the same age group and considering the sampling method we expect a majority to participants between the age of 18 and 29.

6 Design

The following chapter discusses the design considerations based on the requirements previously established. It is subdivided in four parts; The technical limitations, the game aspects, the narrative structure, and the logging system. Following in the footsteps of Mateas and Stern (2003) we strive to use virtual agents in the form of chatbots to provide the player with open interaction delimited by the player's creativity and the system's ability to understand near infinite input.

6.1 Game Aspects

6.1.1 Game type and genre

The use of a chatbot for dialogue system delimits the possible game genres for this study. Hence, the game must focus on the dialogue and social aspects, making it a roleplaying social game. Roleplaying would imply the player taking control of a character in a virtual environment, while social games are described by McCoy et al. (2010) as "*Multi-character social interactions whose function is to modify the social state existing within and across the participants.*" This puts considerable restrictions on the mechanics, investigated below. However, the prototype could also draw on investigative elements, and put focus on the player's ability to discover information through conversation. This choice is determined by the technical limitations later in the chapter, but we presuppose that the main chatbot in consideration leaves little room for adaption without a considerable time investment, making investigative gameplay a preferable option to true social gameplay. We consider that if all participants use either first or third person view, any effects of the point of view should be coherent across all tests.

6.1.2 Mechanics

The focus on the dialogue system as a complicated system for productive interaction requires the interaction to fulfil the requirements proposed for productive interaction. The dialogue system receives an amount of complexity and a quasi-unique approach through the input it can receive. By not delimiting the player to certain responses, a part of the gameplay becomes figuring out which responses result in an answer and which do not. This puts strain on the dialogue system in that the user will pay greater attention to the system, which might challenge the perceived intelligence as discussed under the Chatbot AI subchapter.

Furthermore, the user should be restricted in non-dialogue interactions, as to not lead them onto mechanics not supported by the game. This decision restricts the prototype to a single productive system as intended. The understandable outcome and consequences emerges from the psychological relationships with the characters, and the dependence on speech and social interactions as real-life systems the users are already familiar with. To further the possibility for an unspecified approach together with an increase of complexity we consider adding multiple NPCs in the Narrative Structure subchapter, making movement and choice of conversation partner part of the complicated system. Finally, the interaction with NPCs and underlying systems should hint at the possible narrative developments as considered under the Transparency of the System subchapter.

6.1.3 Aesthetics

The field of aesthetics is however as wide, or perhaps wider, than that of narrative, likely as heatedly discussed, and we must admit the majority of the academic discussion on the matter to be outside the scope of this project. Our choice of aesthetics therefore relies on what we are capable of creating in the timeframe of this thesis, and the information we have gathered in our analysis. We do not believe it relevant for testing OPIL to go to quite the same lengths as Mateas and Stern (2003), who sought to create an environment with truly believable characters, covering animation, facial expressions, and an extensive level of responsiveness from the system. However, for cultivating suspension of disbelief and providing contextual narrative content, the environment and characters must be defined. To help us implement visuals, we choose a 3D virtual environment, which was to resemble a virtual reality world, such as basic shapes, humanoid robots, and abstract landscape. This fits the theme of a social event in virtual reality, while seeking to avoid problems such as the uncanny valley, and other high graphic issues. For simplicity of implementation we strive to create a uniform appearance of the virtual environment, relying on the setting to support the choices of visuals, and asset packs for Unity[12]. Mallon (2008) argues that the drawing quality of the user's character can enhance or diminish agency, but the gain is considered too small for such a system to be implemented. The aesthetics are also used to aid the player to identify the various NPCs, which is done by color grading and positioning in the scene. This helps construct the narrative context, and support the personalities of each character, as well as making the characters stand out, lessening the burden on memory.

[12]Unity Technologies (2005) `http://unity3d.com/`

6.1.4 Agency

Regarding the suggestions for invoking agency by Mallon (2008), we consider which ones we will fulfil, and which ones we will not, as a consideration of the player's possibility for meaningful interaction. To provide the player with a sense of being the central protagonist, none of the NPCs will act outside of receiving input from the player. This is also to not overcomplicate the time it will take the player to learn the position of the different characters. While agency is expected to be higher when the player has greater precision of the tools, the study requires us to only provide precision at the first-order choices, while leaving second and third order vague to allow for productive interaction. The aesthetics of animated characters and sound feedback in the form of speech is intended to provide the sensory response to support agency. The difference of players makes it hard if not impossible to design a perfectly fitted challenge curve, so as part of the design we test the system with possible test participants to strive as close as possible to a fitting skill level. Depending on the player's understanding of the system, they may use different points of the evaluation of interaction by Mallon (2008). Due to the simplicity of the system the interaction is likely to be described as that the user has manual control of the tools, but with a possibility to get better. The interaction does not require high motor-skill or provide possibility for the combination of multiple skills (outside of the cognitive process that we do not intend to narrow down). While the system's outputs are designed to be proportionate with the responses required to trigger the outputs, this is one of the areas in which chatbots generally struggle, because it depends on their ability to seem intelligent. As formerly presupposed the chatbot chosen will not be capable of real-time learning, so having it remember previous conversations or provide satisfactory responses to intelligent, moral or attitudinal choices is unlikely. While it was argued that the player getting a chance to customize their character could be a source of agency, the majority of such possibilities lie beyond the scope of this project as the systems in play do not depend on any numeric values the player can modify, and we can spare no time to add visual customization to the player character. Players will however get the chance to define their in-game character through choice of words and approach, though limited by the narrative context provided.

6.1.5 Transparency of the system

The prototype faces the problem that it on one hand seeks to divert the user's attention away from the limited narrative outcomes, but on the other hand requires a complex system that the user can learn to understand and use. For this reason, we must consider the application of tools to help the user understand their progression, while holding it against the chance that it reveals too much of the system to the user. The systems we consider must therefore find a balance between making the player aware of their possibilities, and not giving away too much of the system. Below we discuss and choose the helping systems we will include:

Highlight Keywords: Highlighting specific words said by the NPCs to hint at other words that may provoke responses with the particular NPC. Our main concern with this approach is that highlighting specific keywords may become selective in the fashion that if every keyword unlocks another keyword, the player's conversation is controlled by the revealed keywords.

Save Keywords: Another option is to make keywords for knowledge the player has of the ongoing narrative and adding them to a log that it may aid the player's ability to remember the various plotlines. This too suffers from making a too coherent information database that we expect will leave the player with the impression that they have unlocked a certain part of the narrative, rather than that they have created a narrative of their own.

Character Traits/Public Information: Alternatively, instead of saving or showing the player the material they have unlocked, we could implement a UI system that shows information regarding the NPC the player is currently engaging with, as a way of hinting of various character relations and character traits. Showing the character's personality traits might help alleviate stress off of the chatbot, as prior research has suggested that chatbots of limited intelligence can more easily pretend to be personify some traits, like whimsical. The same concern as in the prior ideas persist, in that giving players choice from a limited amount of options, even if they are free to go beyond the bounds of those options, might railroad the player along a path defined by those words. If the number of traits become large enough that the player cannot simply try all of them in turn, the interaction may pass from selective to productive.

Hidden Jigsaw: Seeking to avoid the problem of prior suggestions, we consider giving the player a hint when they uncover information, but without relaying what information was important. Thus, the player is not made aware of what they specifically learned, making it more difficult to get an overview, but showing that progress has been made. This could be carried out like in Life is Strange[13], where an icon will appear and a sound is played when the player unlocks an alternative dialogue option. As the amount of information grows it might be necessary to help the player remember what exactly was said, and thus this system can be implemented with an index of the full sentences that revealed something important,

[13]Episodic game developed by Dontnod Entertainment and published by Square Enix, 2015

as seen in Figure 10. Note that this system only works if the sentences carry multiple pieces of information or leaves information out, because a simplistic sentence such as ‘I enjoy the sunset’ which triggers the feedback cue, can only relate to the NPC’s enjoyment of the sunset. The alternative being ‘I’m married to Juliet’ which offers opportunities to investigate both the new character ‘Juliet’ and the opinions regarding the ‘marriage’.

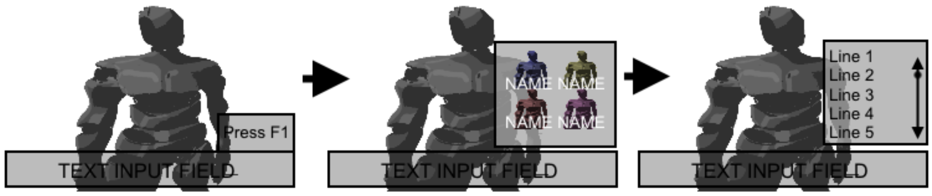

Figure 10: A sketch of the interface. This interface allows the user to access earlier findings.

6.1.6 Chatbot AI

To choose a chatbot, we first consider whether the chatbot is going to be in an open or closed domain, and the level of intellect the chatbot has to possess in order to fulfil its purpose. Open domain is the designation for chatbots designed to be able to talk about anything, without restriction as to where the conversation can develop. Closed domain is the designation for chatbots made to work with a specific topic and often focuses more on giving detailed information regarding that topic, but without the capacity for conversation outside of the focus field. Open domain may leave too many dead ends and make it hard to achieve critical mass for emergence, but the closed domain could generally be expected to provide too little or too limited interaction. Therefore, we are taking elements from both, and making the plot related content closed domain, so that the correct questions or words will lead to the correct progression of the game but make small talk and general topics appear as separate non-progression related system inputs. Due to our requirement to control the system’s output, we are also forced to rely on retrieval-based models, meaning that all of the chatbots’ answers are generally scripted. It would be possible to fill more state space by using generative content, but as J. O. Ryan et al. (2015) point out, generative content lowers the possibility for expressive content, which in our case lowers the narrative context. While it would be preferable to use a chatbot of high intelligence, the time restraints and desire for authorial control forces us to choose something simpler.

6.2 Technical Limitations

When it comes to choosing a chatbot approach, there are several varieties that must be considered before choosing a chatbot to settle on simply by what is required of it. Therefore, we here review different types and their strengths and drawbacks.

Template based chatbots (AIML) This type of chatbot uses a pre-defined template often structure with XML. The bot will quickly go through all the templates and find matches. These kinds of bots have enormous databases of AIML templates, mostly hand-crafted. While they work very well, they fall short on acting smart when communicating with users if not enough state-space has been covered.

Crowdsource Chatbot (Cleverscript) This type of chatbot uses a huge database, and learns through human conversations, like when it encounters a question the chatbot does not understand. It will then mimic the response to another human and learn from the results. This makes the crowdsource chatbots seem very humanlike. However, the huge drawback is inconsistency, as the chatbot can one moment pretend to be a 18 year old male, and the next moment it claims to be a 40 year old female, as it cannot maintain the context of the conversation.

Markov Chain Chatbot (MegaHAL) This type of chatbot store words in Markov chains, by taking a sentence and splitting it into pieces learning to phrase new sentences from it. The biggest strength of this type is that it is easier for it to appear knowledgeable. The drawback is that it does not understand the environment and context it is in.

Roy Van Rijn's attempt at creating a Natural Language Processor (NLP) using Wordnet with Java Wordnet Library (JWNL)[14] This approach takes the previous methods into account and tries to fix the problems they have by creating a chatbot that learns through conversation and reading text, understands relations and concepts and have different scopes, global knowledge, and conversation scope. The drawback of this method is the huge database, and machine learning making it difficult to personalize for individual NPC's in a game environment, without considerable time and effort put into training it first.

Machine learning Chatbots (Google Chatbot, DialogFlow[15], Motion.AI[16], WIT.AI[17],

[14]Source: `http://royvanrijn.com/blog/2014/04/creating-a-chatterbot/`
[15]Source: `https://dialogflow.com/docs/getting-started/basics`
[16]Source: `https://www.hubspot.com/bots`
[17]Source: `https://github.com/wit-ai/pywit`

SiriKit[18]**)** Sequence to Sequence learning, NN, with Conversational Model with two LSTM layers (First layer: Thought vector, Second Layer: Response) Flexible as it focuses on Keywords and understands adjectives for Intent of the sentence. Vinyals and Le (2015) The more data the better it works. A drawback is that takes a long time to train (3-4 days), plus it is more difficult to personalize, but possible nonetheless. (Cournoyer, 2016)[19]. The chatbot used the Cornell Movie–Dialogs Corpus Dataset[20] for initial training. If we are to use this we must provide additional data for the chatbot to become personalized.

6.3 Choice of Chatbot

We proceed with DialogFlow due to the hybrid format between using a simple search-based machine learning algorithm while its approach remains template-based. Additionally, DialogFlow offers an easy-to-use graphical user interface to setup their agents and has the capability to support additional platforms including Unity.

6.4 Narrative Structure

We must consider the narrative context because it serves as a fallback system if the complexity of the system is inadequate. A requirement of the interaction is that it must result in understandable outcomes, which requires consideration towards the inclusion of structure in the narrative. Also considering the decision to not use generative content, a certain level of coherence must be scripted into the narrative. Initially we consider estimating each possible player outcome and dividing them in larger categories, making it more generic but ensuring an overall structure. We also consider that some of the players actions will impact the progression directly, while others will not, and lend the term of decision point to describe points in the narrative in which the player makes a choice that matters (akin to what M.-L. Ryan (2001) uses, and the game mechanic found in the tabletop roleplaying game Exalted 3ed[21]). We must consider how each of these decision points can be resolved, and consider a variety of approaches: **Scene-based Threshold:** Each scene has an amount of parameters,

[18] Siri a virtual personal assistant made by Apple. `https://developer.apple.com/sirikit/`

[19] Marc-Andre Cournoyer (2016) Neural Conversational Model in Torch `https://github.com/macournoyer/neuralconvo`

[20] An open-source database filled with movie dialog. `http://www.cs.cornell.edu/~cristian//Cornell_Movie-Dialogs_Corpus.html`

[21] The game mechanic in Exalted describes the point in a social engagement in which a character must choose to allow themselves to be persuaded, or to draw on the quantification of their opinions (intimacies) to deny the persuasion.

which have thresholds that specify when the scene is resolved and the next scene starts. **Sequenced Pattern:** Progress based on whether the correct sequence of words/lines have been triggered. Spelling mistakes are a problem, because the room for error is rather limited. **Sequenceless Info Provided:** Continuation happens when all information of the scene has been relayed. **Hybrid:** Some information counts up to a threshold for information provided, while other pieces must be relayed in correct order. **Intimacy system:** Each NPC has a series of intimacies like Exalted(see previous footnote). Each intimacy can be affected by using other intimacies. Thus, the player can influence the characters, and they provide natural thresholds for advancing the scene. However, it is complex to the extent where some chatbots cannot support this form of advanced behaviour without considerable modification. Particularly, the chosen chatbot cannot encompass this without serious modifications, or an extensive library of alternate intimacies.

A problem with the above suggestions, are that they extend the problem of the combinatorial explosion to scene development. Left side of Figure 11 shows a snippet of a potential branching scene structure with multiple characters. As the focus of this thesis is not to tackle the combinatorial explosion, we change our design to aim for an emergent narrative. By doing so, we can argue for assembling all the NPCs in one room as seen on the right side of Figure 11, and give the player the freedom to choose who they wish to talk to. This will make it unlikely that the players proceed through the game along the same path, which supports the requirement of quasi-unique progression.

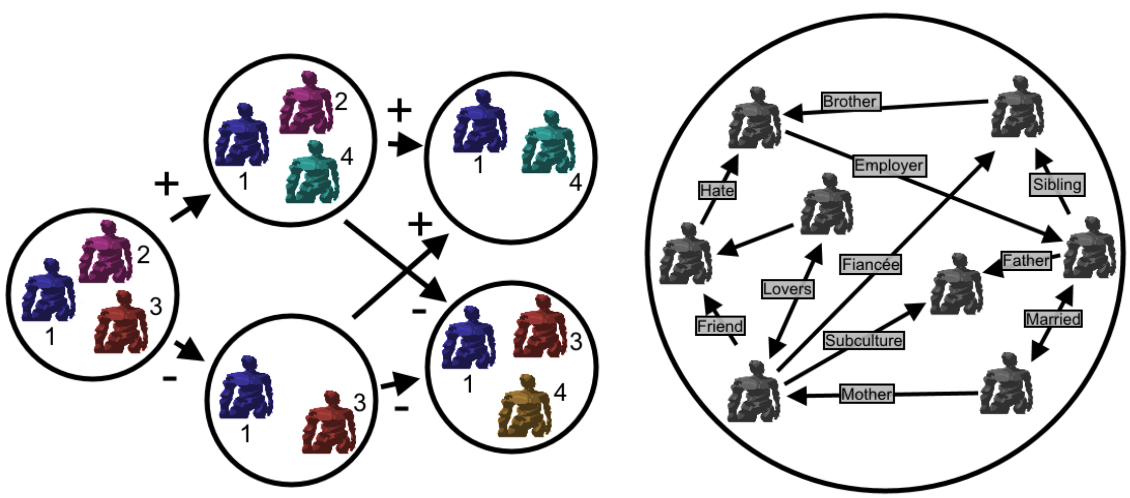

Figure 11: Two design possibilities: The left present a series of individual scenes in a branching structure. Available characters are defined by what scene the player is in. For simplicity, the model to the left only branch to a positive and a negative outcome. Right model shows an example of character relations, for characters gathered in one scene.

The combined scene faces the risk of having too much narrative material, so that the player becomes overwhelmed (refer to critical mass by Louchart et al. (2008)). However, leaving a character in the scene after the plot has been fulfilled might be considered a dead end, as the character no longer serves a purpose. On the other hand, having them leave the scene when their plotline was fulfilled would limit the player's possibilities in an obvious fashion, not to mention leaving the scene depopulated as it nears its conclusion. This might be preventable by narrowing the space and ending the game before the player has reached the conclusion of all plotlines, as for the player to perceive the interaction as productive they should not have the concept of finishing plotlines, rather they should have the impression of having made an impact on the scene as a whole.

6.5 Logging

The prototype will use a logging system that will log the conversations with the different chatbots, the path the player took throughout the narrative, and the NPC relations towards the player. This data will be used for the observational part of the evaluation as it provides useful data regarding how the player communicated with the chatbot, and whether the chatbot had fallbacks throughout the conversation. This means that we can correlate the questionnaire responses from the player afterwards with the observational data and potentially find a connection as to why the player responded in a particular way.

7 Implementation

The virtual space and characters' visual aspects were implemented in Unity. The chatbot(s) were implemented through integrating DialogFlow with Unity's SDK, but were modified using the DialogFlow UI provided at the homepage.

7.1 Player Controller

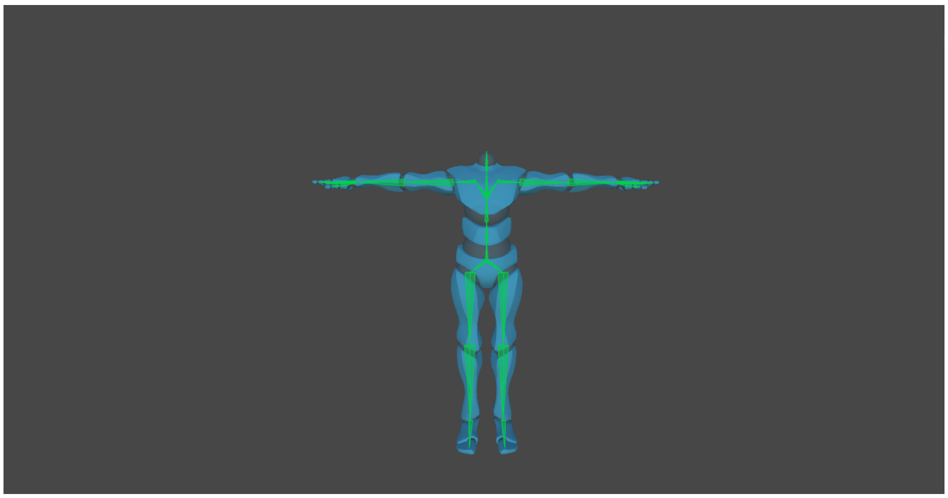

Figure 12: The edited ybot from Mixamo applied to the rigged player controller

The player controller is based on 'Invector's third person controller - basic locomotion'[22] package while rigged together using a custom ybot from Mixamo with animations. This means the player character is fully animated, rigged, and capable of moving and jumping with precision.

7.2 Dialogue System

DialogFlow uses a structure of intents, entities, context and so on to help the search algorithm to perform. An intent covers the input and output of a topic the chatbot can communicate. It takes one or more training phrases as input and can randomly select an output in response to a training phrase covers the input and output of a topic the chatbot can communicate. Entities are based on certain items conveyed through a conversation such as a name, colour, drinkables, and other objects that can vary based on the conversation. If a certain intent

[22]Unity asset by Invector, http://www.invector.xyz/project

were communicated with a certain entity the chatbot would understand that specific context and respond accordingly. For our system we mainly utilize the concept of intents up to a maximum of 2 layers, to keep the communication accessible. The problem with more layers of intents is if a player leaves an intent branch, the only way to resume it is to start over from the first layer. Another problem is that if a player was to randomly input words it may accidently result in information that they were not supposed to have.

7.3 Personality and Plot Creation

To create a narrative as previously described, we made ten characters. To ease the creation of the characters, they were created based on characters from a tabletop role-playing session both authors had attended. This was mainly done because it assured the work of character to character relation had already been carried out. These character relations were expanded upon and collected in a Table 13 (full table in Appendix A).

	Dorian Lavigne / Damarion	Collene Cayne / Seraph	Eryn Bancroft / Pretas	Shenna Grippen / Stenson	Keno Yorin / Yui	Tanya Simril / Aldersville	Keira Lien / Nyx	Denna Albach / Juliet	Veronika Reade / Atwood	Gavin Madigan/ Afeld
Dorian Lavigne / Damarion	Hyperelite and director of OBXILUS. Ambitious, arrogant, but with a soft spot for culture.	Work together.	Considers lower class	Works together. Blackmailed her.	Dislikes him due to squandered ambitions and failure as a hyperelite. Why would one serve in the army, when one has money?	Works together.	Dislikes due to interest in Denna, and to how he humiliated her.	Likes due to appearance and intelligence.	"Just a low class police officer"	Dislikes him being a lowly police officer.
Collene Cayne / Seraph	Works together	Hyperelite and director of OBXILUS. Vain, petty, tries to appear friendly and kind. Wants a good image.	Is interested in the newest Lhosts, and Pretas is wearing one	She is so far stuck up her own ass, she doesn't know how to enjoy life	Eccentric and paranoid, doesn't know how to enjoy life.	Competitive rivalry in the firm. Likes her personality, but dislikes her erratic nature.	Uninterested	Suspecious of what she wants with Damarion on behalf of the firm, and considering her prior involvement with leakage of information	She works for a better world, though not always efficiently	He works for a better world, but lacks amibtion. Who cares about small scale theft?
Eryn Bancroft / Pretas	Suspects him of doing illegal business. Is here to investigate. Damarion is suspected of manipulative advertisement, bordering on brainwashing.	Suspects her of doing illegal business. Is here to investigate. Seraph funnels money out of the firm.	ALEPH agent here to root out suspecious activity in OBXILUS. Patronising, 'older sister', currently investigating so asks a lot of questions.	Suspects her of doing illegal business. Is here to investigate. Cue: Stenson has private meetings with someone out of town.	Fellow agent. Must be observed for erratic behaviour. He may be infested with something.	Suspects her of doing illegal business. Cue: Has been hiring people for non-firm related business.	Fellow agent.	Fellow agent.	Seeks to avoid the involvement of the police captain. Likes law and order however, as long as it is not in the way.	Seeks to avoid the involvement of the police. Otherwise, he is below notice.
Shenna Grippen / Stenson	He has secrets, that might endager the company. Otherwise, I would push him out.	Artificial to her very core, has nothing but masks to cover her empty center. Oh and a massive bitch.	Something seems odd with the way this woman tries to get information out of every meeting. Almost like an interrogation.	Hyperelite and director of OBXILUS. Ambitious, intelligent, modern business woman. She has a weird fetish for taboes, dislikes artificialness.	He is odd and exotic, two things she hates and loves.	She is a sensible person, driven but a slave to her imperfections. Likes her for her personality and richness.	Doesn't know who this is, but they provided a generous donation to the fundraiser.	Suspecious of what she wants with Damarion on behalf of the firm, and considering her prior involvement with leakage of information	Is it important to stay on good terms with the police.	It is important to stay on good terms with the police, as long as they keep their nose out of OBXILUS business.
Keno Yorin / Yui	Suspects him of doing illegal business. Is here to investigate. Damarion is suspected of using illegal technology.	Suspects her of doing illegal business. But she's friendly with the police officer, so that couldn't possibly be.	Considers her a friend and reliable backup.	Suspects her of doing illegal business. Is here to investigate. Cue: Stenson is rumoured to make strange requests of her clients.	Paranoid hyperelite. Served in the military, and is mainly here because the other ALEPH agents need his hyperelite status.	She is a different as class allows it, and Yui suspects her of doing illegal business. Cue: She has Atwood, the police captain, under her complete control.	Fellow agent, not quite trustworthy yet.	Fellow agent, not quite trustworthy. Has a suspicion that she as a hacker is a greater danger to his privacy.	Avoid the police at all costs.	Avoid the police at all costs. Especially Afeld since he almost revealed him as an agent.
Tanya Simril / Aldersville	Works together. Finds him boring and arrogant. She has already expended the interesting aspects of his character.	Predictable, vain and boring. Jokingly accepts the concept of their rivalry, but loathes the repetitiveness of it	Pretty and interesting. Is suspicious of the fake ID ..	Of tedious people, this is the worst. Embodies boring in every aspect of her life, but knows a valuable thing or two	Too crude for her liking. And the paranoid ticks frustrate her endlessly.	Frivolous hyperelite. Dislikes the uncouth, and the boring. Near perfectionist. Is generally well-liked.	Hardly a person worth talking to	Has heard of Juliet, and is anxious to discuss a potential contract working for OBXILUS	Good friend, who's trust she uses to hide OBXILUS secrets.	A lowly officer
Keira Lien / Nyx	Dislikes and is jealous of him. He might be doing illegal activities, but Nyx is more interested in stopping him from approaching Juliet.	Suspects her of doing illegal business. Is here to investigate. Seraph arranges excessive feasts to draw attention away from something.	Fellow agent. Knows that she is not actually a human.	Suspects her of doing illegal business. Cue: Stenson has several journalists on her payroll, to move their focus away from OBXILUS	Dislikes his hyperelite status, and would rather have gone in without him. Or would rather not have gone in at all.	Suspects her of doing illegal business. Cue: She's been aiding Damarion in a funding his research.	Corsair turned mercenary. She's here because of the job and Juliet. Introvert, about to panic, about to blow her cover.	Fellow agent and in love with.	Facist pig, using her authority to hurt the lower classes, and protect the high and mighty.	Pig. Being a cop makes him a horrid person. His kind is the reason the lower class is opressed.
Denna Albach / Juliet	Is charmed by his appearance and language. Suspects him of doing illegal business, and was initially trying to seduce him to press him for information.	Her desire to appear 'good' is clearly a coverup.	Fellow agent. Sure there is something she's not telling, and curious about what it is.	Hacked her personal network, and found that she'd been billed for a drug purchase. Other than that and a bit of tax fraud, she is squeeky clean.	Fellow agent. He's too technologically impaired and paranoid to be cery useful outside of a combat.	Hacked her personal network and found correspondences between her and Atwood. It appears the two are childhood friends.	Fellow agent and old friend brought together by coincidence.	Skilled hacker and agent. Came here for a job, but got smitten by Damarion's charm. Arrogant, genius, lower class than she behaves. Technophile.	Hacked her personal network. Found correspondances with Aldersville, but nothing too crazy. Her level of security was too tough though, and Juliet is scared of what she might be hiding.	Met him once during a mexican standoff. He wanted help to find out things about OBXILUS. She knows he's on the the agents' side, but was informed to keep him out of it by employer.
Veronika Reade / Atwood	Knows Damarion to be arrogant, but knows from the police records that he had a tough upbrining, his family's possessions taken by the very firm he now serves.	Knows that Seraph's desire to appear good is stemming from the knowledge that she is very sick, and wants to leave a good image behind. Atwood	Is suspicious of who this is, but figures she's a corporate spy. Intends to have her arrested.	Looks the other way when Stenson makes an illegal purchase and is caught. Though she keep it is cases she'll ever need it.	Knows he's outside of reach, but also recognizes the appearance of a warrior. Is wary of him.	Childhood friend, which is is very loyal towards.	Suspicious individual, who would never be capable of attending such a party without powerful backup.	She knows Juliet is a hacker, as her face was plastered all over the internet some years ago, when she scammed another hyper elite.	Stern, serious, and in charge of security. Is aware of OBXILUS illegal activities, and paid to keep mouth shut.	Underling with a penchant for persistance and going against orders. Is careful with secrets around him.
Gavin Madigan/ Afeld	He had the hacker skullduggy killed, and will someday answer for his crimes.	Her erratic actions are causing commotion for the people she 'tries' to help, moving them back and forth, showing them off like prized piglets. And she doesn't see it.	Knows Atwood desires her arrested, but is keeping the two apart, as he knows Pretas is working to uncover the truth about OBXILUS.	Knows that she takes drugs, but one of the drugs she's purchased 'Elathiol' is a poison, and he fears who she'll use it against.	Not sure who this stranger is, and doesn't care too much. He does seem familiar though..	Knows that she is untouchable due to her relation with Atwood. Suspects the relation may be blackmail.	Not sure who the introvert is, but she behaved quite poorly upon their meeting, so he expects she holds something against him.	Knows she's an agent seeking to uncover the truth about OBXILUS. Is determined not to blow her cover, as long as she doesn't blow his. Expects that she knows more than she lets on.	Superior. Was a good cop, is a good captain, but fears she may have been compromised. If so, he would hesitate to expose her, trying to talk her over to the right side first.	Detective on the case. Has followed leads on OBXILUS crime trail. Is also here as security detail. Tries to keep calm surronded by people who could destroy his life.

Figure 13: The Character Relations table. Black boxes are descriptions of the characters, grey are plot relevant relations, and white are non-plot related relations. The red names are the original names of the characters from the tabletop session (full table in Appendix A).

General data regarding the characters can be found in Appendix B, such as original names, new names, genders, and default small talk templates. Default small talk is a series of small talk sentences a user might present the chatbot with, that upon being filled out in the DialogFlow interface can cover general non-specific state space.

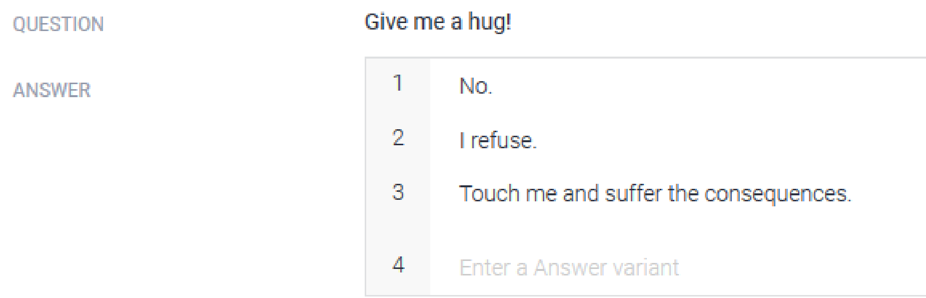

Figure 14: Example of a small talk question and answer for the chatbot Dorian, who is categorized as aggressive and bored. Taken from DialogFlow interface.

Because the small talk template is extensive, we opted for creating stereotypes and sharing them across characters. Afterwards, each template was individually fitted to the character, but it does mean that multiple of the characters have access to the same responses regarding non-plot related user input. The templates created were: Aggressive, Kind, and Inquisitive. These templates matched the general personality traits of the characters and made it simpler to make minor changes while keeping some content similar across all characters of the same template.

After looking at patterns in the character relations, 9 subplots were conceived, and some character relations were fitted to the new subplots see Figure 15 (See Appendix C for full).

Love Triangle	**Dorian's Crimes**	**Collene's End**	**Rivalry in OBXILUS**	**Eryn's Cover**	**Shenna's Addiction**	**Keno's Cover**	**Tanya's Fraud**	**Rogue Investigation**
B8 Keira's jealousi	B4 Manipulative brainwashing	C4 Funnels money out	B5 Secrets	D5 Eryn is interrogating	E4 Private meetings	F4 Infested behaviour	G4 Abusing business privileges	K5 Keep the police out of OBXILUS business
B9 Denna is charmed	B6 Illegal technology	C5 Empty husk	C7 Grudging accept of rivalry rumours	D7 Eryn fake ID	E6 Strange requests of clients	F9 Useful in combat	G6 Controls the police	K6 Keno encountered Gavin before, wants to hide
H2 Dorian humiliated Keira	B10 Tough upbringing	C6 Police immunity	G3 Competitive rivalry	D8 Not human	E8 Journalist distraction	F10 Appearance of a warrior	G8 Funding Dorian's crime	J8 Keira believes the police to serve only the hyperelite
H9 Keira and Denna old friends	B11 Skullduggy murder	C8 Distraction feasts	E2 Blackmail Shenna	D10 Corporate spy	E9 Drug purchase	F11 Familiar face?	G9 Just friends with the police	K8 Keira believes the police oppresses the lower class
I2 Dorian likes Denna's intelligence	E2 Blackmailed	C9 Distraction	Eryn knows that OBXILUS	D11 Keeping		F2 Served in	G11 Suspect blackmail of	K9 Denna and Gavin know each other's

Figure 15: A snippet of the plotlines table. The designated titles are marked in bold. Each of the underneath entries includes keywords and a reference to the Character Relation table (Appendix 3).

Plots with too few entries were written out, as they were considered too short, and they might therefore appear as dead ends to the player. Each of the plots were divided into a series of keywords, and whether an NPC would respond to the given keyword. This was made to ensure that only the characters relevant to a specific plot had something to say about it and allowed us to fill in loosely how the character would respond to the keyword.

Titles and keywords					
	Dorian Lavigne /	Collene Cayne /	Eryn Bancroft / F	Shenna Grippen	Keno Yo
Love Triangle					
Love_attraction	Dorian loves Denna				
Humiliate_degrade	Dorian humiliated Keira				
OldFriends_friends	Dorian knows that Denna and Keira are old friends				Keno kn
Intelligence_charm	Dorian likes Denna's intelligence and charm				
distrust_trust		Collene is suspicious of Denna's n		Shenna is suspic	Keno doe
Deceit_trickery		Collene knows Denna tricked anot		Shenna knows D	Keno kn
Dorian's Crimes					
Brainwashing_research_science	Dorian's research involves behavio		Eryn knows that	Shenna knows th	Keno kn
IllegalTechnology	Dorian's research involves state of the art tech			Shenna knows th	Keno kn
Blackmail	Dorian admits it to be necessary.		Eryn knows She	Shenna was blac	Keno kn
Childhood_upbringing_Dorian's past	Dorian doesn't want to talk about it		Eryn knows that	Dorian's past is a	Keno thi
Murder_Judith	Dorian believes the death of Judith to be an unfortunate consequence				
Collene's End					
Money_laundering_finances		Collene is using	Eryn knows that	Shenna believes	Keno is
Distraction_party_coverUp					

Figure 16: A snippet of the Plot Dialogue by Plot and by Character table. The green squares show that the character has one or more sentences related to the plot given on the left side. The grey boxes underneath include the general keywords that will trigger the responses. Ex. Love and attraction are variations of the two are all keywords that trigger the same response (Appendix 4).

From this we created a list of all the dialogue responses and implemented them as separate intents in DialogFlow's interface. Each character was also given an intent for every other character, using their name as the training phrase. This allows the chatbot to relay their relation to the other characters present in the scene. They were also given additional small talk responses for their opinion of the party, in which the game is placed, and OBIXLUS, a fictional organization mentioned by NPCs in the game world. All characters were also given personalized variations of the fallback intent. The fallback intent is a series of responses the chatbot will use when it does not understand the user's input. While the defaults usually ask the user to repeat their input, it was decided that asking the user to repeat would be fruitless, because the chatbot would likely not understand if it did not understand it the first time. Instead the fallback answers were changed to fit the personalities. This also serves to obscure the system's fallacies, making the system less transparent. It is however not possible writing fallback intents that give the appearance of a sensical answer, due to the amount of possible situations in which it must be used. Thus, each chatbot has around 12 fallback answers, most striving to agree with the user or to excuse why the chatbot could not answer in the context of the scene. Sometimes, it is possible for the fallback intent to appear as a sensical answer, while at other times it will not fit at all. This was intended to strengthen the players' perception of the chatbot as being capable of answering.

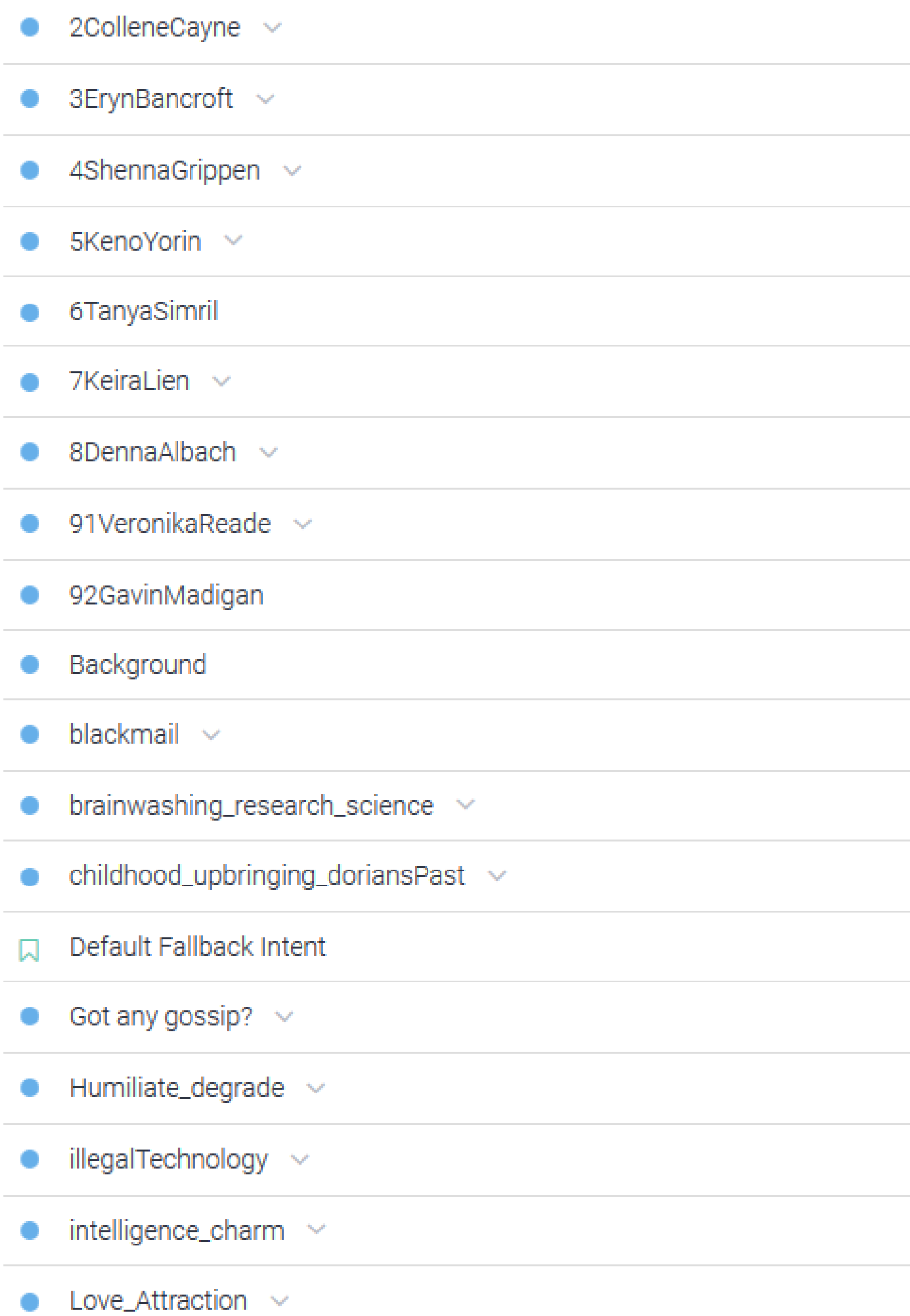

Figure 17: Dorian's list of intents

7.4 Conversation Handler

A script that handles interaction between chatbot and player

Tutorial elements

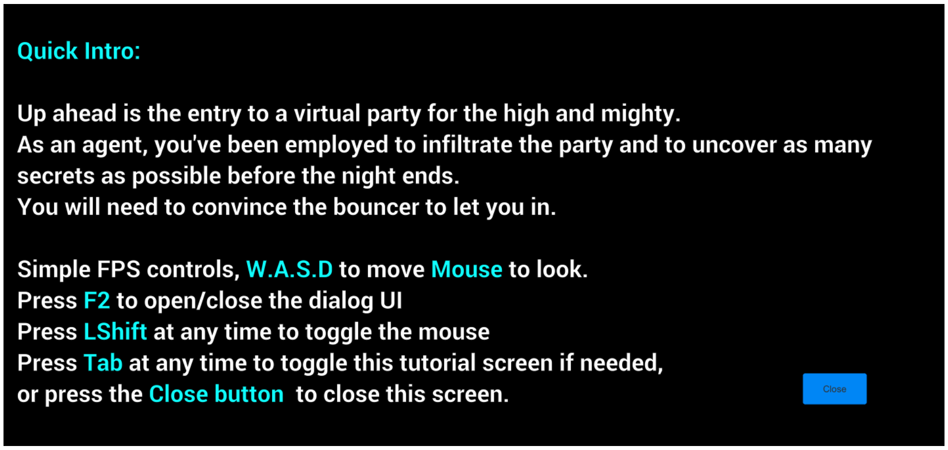

Figure 18: The tutorial canvas at the start of the game

These elements were added simply to demonstrate the controls and purpose of why the player is here and how to play the prototype.

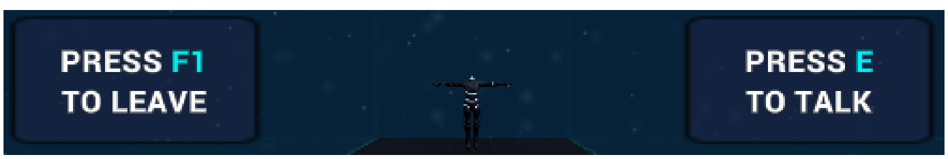

Figure 19: Pop up interaction help messages for entering/leaving a conversation

UI and Roaming mode. We decided to create a simple UI and roaming mode to ensure keyboard availability for conversations so that you could simply switch between the two without having too much trouble with interaction overlaps on the key-bindings.

7.5 Log System

Here we handle the information that is saved during game play by simply writing to a text file, the player text input and the name and response of the NPC they are talking to. This data tells us which chatbot the player is talking to and what they are talking about. Then at the end of the game-testing period we save the time played and a percentile measure of the player's progression though the narratives.

7.6 UI-Puzzle and Log Book.

Puzzle Piece Checker

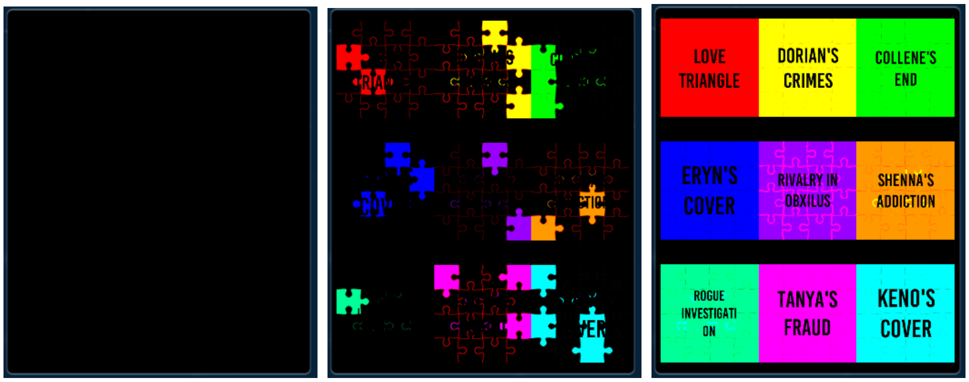

Figure 20: The puzzle canvas from empty (left) to full (right)

Each of the pieces in the 9 puzzles holds a script that checks for a certain trigger sentence in the narrator'sA gameObject in the Unity scene that holds information regarding scripting and other scene wide information. data-log and if found enable the specific puzzle piece(s). For each puzzle piece found the completion percentage would increase up to a total of 100 percent if all 9 puzzles had been found. The 9 puzzles hold in total 144 pieces which is 16 for each puzzle, though some sentences unlocked multiple pieces.

Pop up Canvas script with UI controller

Figure 21: Pop up feedback graphic when a puzzle piece(s) has been found

This small feature was created using a UI controller that simply ensures easily manageable animations with UI canvas elements, as we felt the importance of showing the player progression throughout the prototype and to notify them to check the Puzzle UI for the new-found puzzle piece(s).

Log book

Figure 22: The UI Logbook demonstrated

The log-book was created so that the player had a chance to investigate some of the important sentences the player would find during the game, so they would not lose track of their findings for further conversation with the NPCs. This was done by creating a small UI with a profile picture and title for easy identification arranged in a grid, that the player can click to open a sub-menu with a conversation log of only important sentences.

7.7 Text-To-Speech VoiceManager (Based on Microsoft's Speech API library)

VoiceManager This script checks the computer for installed windows narrator voices installed on the computer and fetches them in an array for easy access. It also creates compatibility with C-Sharp and the DLL library SAPI by creating external function calls to the SAPI library. Due to the way the VoiceManager is setup means it only has support for windows at the moment but can be made available to all platforms if a standalone voice library with voice samples that are not dependant on the SAPI library is made ready instead.

Test Voice A simple script, which were attached to each chatbot in the unity environment which translated the DialogFlow response to voice feedback.

7.8 The setting and scene

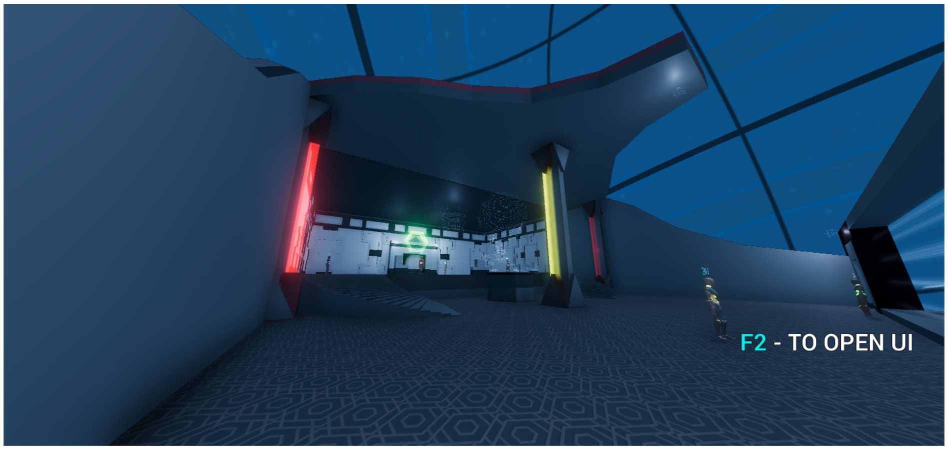

Figure 23: The Gala Room

The gala room was designed after the theme of a virtual reality party with a sci-fi setting. The open space leading out to a balcony serves as a visual pointer for the player to navigate by.

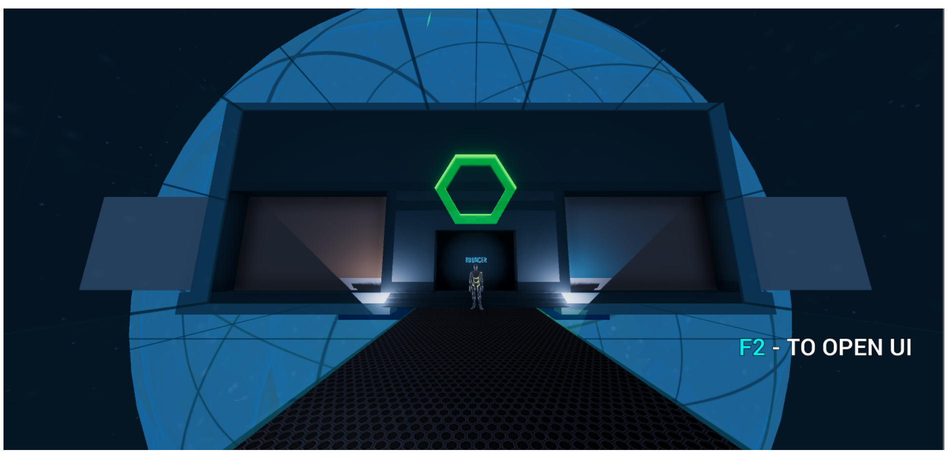

Figure 24: The Gala Entrance

The gala entrance was designed to appear as a gateway. The area was purposefully left empty except for the bouncer NPC. The bouncer served as a tutorial for the player to get a grasp of the initial mechanics or writing responses and provided clear feedback in the form of the opening gateway.

8 Evaluation

In this chapter we go through the results from our tests. In the following, we refer to participants by numbers, so P1 being participant 1, etc.

8.1 Demographics

Two participants were gathered from volunteering on a Facebook group for AAU students looking for people to test their project. The rest were contacted by the researchers either through general posts on Facebook or directly. This was an undesired method to gather participants but was deemed necessary when 6 other participants who had signed up failed to show up for their tests. All the twelve final test participants were males, and most were in the age group 24-26 (a single participant was in the age group 21-23, and a single participant in the age group 27-29). Regarding game literacy; Three categorized themselves as casual gamers, and defined it as only playing games rarely, noting that it was a less prevalent hobby of theirs. Six participants categorized themselves as gamers, which they defined as playing games for multiple hours daily. Two categorized themselves as hardcore gamers, defined by striving for the competitive scene of professional gamers, without having reached it. One categorized himself as being a professional gamer but failed to elaborate. Their level of English proficiency were split between the four option: Untrained(0), Novice(0), Intermediate(3) and Expert(9). P7 who chose intermediate identified himself as having dyslexia. While we cannot claim to be knowledgeable regarding dyslexia, we noted that his test results did not stand out particularly when it came to errors of writing.

8.2 General Bias

Due to complications with gathering test participants, multiple different testing environments were used to raise the amount of potential test participants. The majority of tests were performed in the basement of the main building on the AAU CPH campus, in a closed off lecture room. P6's test was carried out in the Multi-Sensory Lab at AAU, with the consequence that students had access and were present. P4's test was performed at his home, where he was interrupted by a phone call. The three last tests took place in an unused copy room on the AAU CPH Campus. P1, P2, P10, P11, and P12 used a different brand PC mouse for the test than the rest. The instructions to leave conversations with NPCs failed to appear for some players and was accidently left out of the guide available to players.

This lead to some frustration and waste of time. Due to convenience sampling, some of the participants knew the researchers personally. This is not expected to have influenced the test significantly, as a player's possibility to solve and interact with the system is not believed to be related. The program crashed approximately 10 minutes into the test for participants 5 and 9. The game was restarted and a new timer set, to ensure that they received approximately 30 minutes of gameplay, but this removed data regarding their completion percentage. This data was later restored by reiterating their data, checking for sentences that would provide progression.

8.3 Flow Questionnaire

Across Participants	
Mean of means	3.5
STD	0.60024
Single Sample T-test(Right tailed, 3)	H = 1

Table 5: Shows the results of a right tailed single sample T-test to see whether the mean of means (across all participants) is higher than 3 on a 5 percent significance level. In other words, it tells us whether the participants were generally in flow during the test, as 3 is the centre of the 1 to 5 scale. As the mean of means is 3.5 it shows some presence of flow across the sample, but with a standard deviation telling us that some values might border on close to the threshold of 3. As H = 1, the null hypothesis that the mean of means is equal to the mean of 3 is rejected on a five percent significance level.

T-test sample for each participant												
Participant	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
Mean	3.44	4.44	3.64	3.92	3.88	3.24	4.32	3.00	2.92	2.56	3.8	2.84
STD	1.08	0.92	1.04	1.12	0.67	1.20	0.80	0.81	1.52	0.82	1.00	1.28
T-test	1	1	1	1	1	0	1	0	0	0	1	0

Table 6: We perform a right tailed single sample T-test for each of the participants, bearing in mind that the repeated use of a test statistic increases the chance of random sampling error. However, looking at each participant independently should still be possible, as here we note that P6, P8, P9, P10 and P12 fails to be significantly higher than the threshold of 3. If we group their logging and interview data and does the same to the rest of the participants, we may see key differences between those who experience flow and those who do not. We do this in the discussion.

8.4 Flow Bias

The following is a list and discussion of bias we consider having impacted the results of the flow questionnaire. An error in the flow questionnaire forced us to leave out the item "I felt that I had everything under control" from all participants. This may compromise the scale and especially the sense of control element, to no longer have to accuracy as intended by Payne et al. (2011). Due to the nature of flow and the possibility for distractions, we noted down intended interruptions of the testing environment. Unexpectedly the AAU CPH campus proved a poor testing ground, as passersby frequently thought to open the door, and in some instances even entered the room regardless of it being occupied. One participant received a phone call during the test, which was a sure breach of flow. During the tests of participant 5 and 9 the program crashed. This is bound to have broken any present flow state, and restarting may have been frustrating. It should also be noted that multiple participants noted during the interview that they felt time pressured, which was likely a response to having been informed that the test would run for slightly less than an hour and having been told that the prototype would stop itself. We consider this a bias in regard to the flow questionnaire, as the relation between being aware of time and being in flow are mutually exclusive. Some participants were gathered on a Facebook group for AAU students seeking other student to test their projects, which meant some participants had academic background, some even from the same education. P2 noted that he recognized the questionnaire structure. It would be unfortunate if his or others' meta-understanding of the test affected their responses and would suggest participants from other educations for future studies. It should also be noted that the used scale is generally considered to inflate the prevalence of flow (Moneta, 2012).

8.5 Interview

The majority of participants realized that the system was looking for keywords (aka. 'trigger words', 'buzzwords'). P5 and P12 alone noticeably do not mention keywords. There seems to be a self-reported tendency to start out with a narrative context dependant approach where players would assume the role given and try to stay in character. This includes writing in full sentences and trying to form a continued conversation with the chatbot. When this failed, everyone reported that they adapted by shortening sentences or resorting to only keywords. Participants 2, 4, 5, 6, 7, 8, 10 and 12 reports resorting to asking about names as keywords. There seems to be five general reasons for why participants shorted sentences: They grew frustrated with the system or input system (P8), they wanted to be more efficient (P5; P6;

P7; P9; P10; P11), it was necessary to progress (P2; P3; P11), they felt time pressured (P2; P5; P6; P7), or it was considered easier (P6). The categories of efficiency and time pressure overlap, in that participants strive for finding as much as possible in the time given. The necessity category is opposed the category of ease, as the distinction may speak of the degree to which the users understand the system. Participant 4, 8, and 12 commented on initially believing that they could be discovered and expelled from the party, but upon realizing that this was not the case, they changed their approaches to pay more attention to solving the system, and less attention to being discovered. P6 and P11 suggested a notebook be added to the game, because they found it difficult to keep track of the many keywords and pieces of information. P4 and P5 believed question marks to be important for the system, while P6 considered them, but could not make a conclusion. P9 noted that question marks did nothing. All participants could repeat the narrative context presented at the onset of the game. The participants generally report the system output to be sensible when it does not use the fallback responses, which were mostly considered nonsensical. Participants 2, 3 and 11 thought they would continue playing the game outside of a test setting due to continued curiosity. Participant 7 was undecided. P1 reported that the game was not the kind of game he normally enjoyed, and he and P10 thought that it was too difficult. P4, P6, and P12 noted that the game was not developed enough and with too many bugs for them to play it outside of testing. P5 imagines that he would try to play the game outside of testing but imagines he would be distracted before completing it. P8 found the interaction frustrating and did thus not desire to play it again. P9 did not think the game interesting enough to be played outside of the first impression.

8.6 Interview Bias

Due to unexpected circumstances only, a single researcher was available for tests with P3 through P12. This means the researcher had to perform both the interview and take notes, which may have lead to a decrease in the quality of taken notes, and less mental capacity to go into depth with each topic. While not a bias as such, the participants could answer in Danish or English as they pleased and were also presented with the opportunity of having the interview questions presented in either language.

8.7 Logging Results

To determine the relationship between players and system, we had three points of interest regarding the player input and system output. The first measure we call Fallback. It is counted every time a participant triggers a default fallback response from the NPCs, and thus measures how often the system fails to understand the user, or when the user provides inadequate input. It is also likely to trigger if users make too many spelling errors. The second measure is Nonsensical Response, which is counted when the system responds with something that is completely unrelated, but not a fallback response, likely as a machine learning error. The Approach Thresholds are the line numbers in which a user turns to using mainly keywords for the rest of the session.

Fallback Example (P1, line. 3): Participant Input: *"What brings you here?"* DialogFlow response from Keira chatbot: *"Was I meant to respond to that?"*.

Nonsensical Response Example (P1, line. 30): Participant Input: *"Hello Keno"*. DialogFlow response from Keno chatbot: *"Whatever Dorian is doing, it is dangerous. It's gonna come crashing down if he isn't stopped."* This is the response belonging to the brainwashing research science intent in Keno, which uses the trigger phrases: brainwashing, experiment, science, and research. It should have used the smalltalk template for being greeted.

Logging Results												
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
Fallbacks	36	20	32	34	25	59	65	35	64	45	32	51
Nonsensical	10	3	1	4	5	1	5	4	1	1	0	5
Threshold	-	L84	-	-	L66	L86	L87	L87	-	L93	L25	-
Total Lines	148	111	92	91	154	164	211	108	153	151	121	136
Completion %	12.5	4.9	14.6	4.9	18.8	20.1	7.6	8.3	0	27.7	18.1	5.6

Table 7: Logging results of Fallbacks, Nonsensical Responses and Approach Thresholds (given in line number). The total lines measure is the amount of inputs the player provided, which is equal to the amount of system outputs. Note that P5 and P9 suffered crashes, and that the numbers are the combined values from both pre- and post-crash gameplay time.

8.8 Logging Problems

An error occurred, in which the participants would send their movement input as a text input to NPCs they were not currently engaged in a conversation with. This resulted in the original logs having some instances of ex. *"wwwwwwwwwaaawwwdd"* as the W, A, S and D keys were used for movement, as per the common first-person shooter control scheme for PC. These

Logging Results Totals			
	Mean	Std	Median
Fallbacks	41.5	15.1	35.5
Nonsensical	3.3	2.8	3.5
Threshold[1]	75.4	23.8	86
Total Lines	136.7	34.4	142
Completion Percentage	11.9	8.1	10.4

[1] Note that the thresholds in Table 8 are made only from the participants in which a threshold was observed: P2, P5, P6, P7, P8, P10, and P11.

Table 8: Calculated means, STDs and medians of the totals of the logged results.

errors have been removed from the logs. From P5 and P9 the game crashed approximately 10 minutes into the test. For them, the prototype was restarted, and a new timer was set for an estimate of 5 minutes, though problems with P5 lead him to have 5 more minutes of play. We went through their first logs by hand and counted sentences that would have resulted in progression. P5 had unlocked two additional puzzle pieces, but P9 had not unlocked anything pieces in either of his sessions.

9 Discussion

First, it should be noted that due to the sampling, the empirical observations and conclusions should only apply to the target group of males in the age group 21-29, who have at least some experience with videogames. Furthermore, we note that the sample is rather small due to unexpected complications, and for future studies would suggest a larger sample. Additionally, the OPIL model does not consider the player's preferences regarding productive interaction systems, such as genre and gameplay elements. This originates from the model's other purpose as an evaluative theory usable on an interactive product, without the involvement of players. As we do not adequately investigate these differences in players, we cannot know how their personal preferences may have affected their interaction with the system other than the participant's responses as to whether they would play the game outside of the testing environment. Only participants 2, 3, and 11 thought they wanted to play the game more outside of testing environment, while participant 7 was unsure, and many of the rest provided reasons for being uninterested such as game type, genre, or the underdeveloped state of the game. The flow scores should provide some measure of the efficiency of the system to catch a player's interest and attention, though it may be affected by the novelty of the experience, and hence not be relevant as to their desire to play the game further. Due to our convenience sampling on the AAU campus, the participants could have had knowledge of the study, test method, or a misguided desire to provide the best possible results, which may have influenced the flow questionnaire results. Some tests also suffered interruptions which likely influenced the flow questionnaire results, and it was not possible to ensure homogeneous environments. The inclusion of flow scores cannot prove the presence of productive interaction in a system in and of itself, but it can point to participants who do not per OPIL's definition experience productive interaction. Afterwards, a researcher must rely on alternative test methods such as those used in this study, to explore and define why a system fails to provoke productive interaction. The mean of means and single sample T-tests show a general presence of flow across the sample, but with multiple participants falling below the required threshold. Adding the H values from the single sample T-test for each participant, and dividing the values in Table 7 those above and those below the median values show us two general patterns: Participants who triggered fewer fallbacks have a higher flow score, and participants who wrote more lines generally had a higher rate of fallback responses. The former observation stems from only participant 1, 7, and 8 having high fallback rates and significant flow level, or low fallback rates and not significant flow score. As per the interview results, we know that the participants generally considered the fallback responses to be nonsensical, which points to them not providing an understandable consequence of the player's interaction, and thus not fulfilling the requirement for productive interaction. The latter observation, that participants who write more lines generally receive more fallback responses, seems logical. The two observations suggest that the fallback response system does not work as a suitable alternative for a lack of critical mass for emergence, and that it may even remind the players that they do not understand the system. The interview data regarding narrative context tells us that most participants could imagine a variety of different endings and developments of the narrative, which hints at a sufficient level of contextual material with an understandable though undefined outcome. However, some participants noted that they changed approach when they realized the narrative did not have the expected consequences initially being of the understanding that the chatbots could blow their cover. By using the same thresholding as above, there could also be a vague correlation between nonsensical responses and completion percentage, which could be argued to be due to the nonsensical responses providing players with more information making it easier to progress or randomly triggering completion sentences. There are three interesting player profiles that we would like to expand on: Participant 8 could not elaborate on the narrative other than what was provided on the starting screen and ended with a low flow score. This suggests that participant 8 did not perceive adequate narrative content or that he did not find the particular narrative engaging. Participant 7 suffers the most fallbacks and writes the most lines but has a high flow score. Considering the amount of total lines, his threshold comes quite early, though compared to the other participants it does not. Similarly, participant 11 understands the system very early (l. 25) and gets a good amount of progression in a short amount of lines, ending with a positive flow score. Both cases go against the background reasoning regarding the motivation for productive interaction as breaking the system means that any further interactions should not be adequately complicated. We however consider that it could be an example of meta dependant productive interaction, but we lack data for further explanation. An interesting tendency described in the interviews and visible in the logs, is how participants engaged the system with interactions guided by the narrative context, but for different reasons they abandoned the narrative context in favour of fast and efficient interactions. These interactions often take the form of simplifying input and repetition, which seems to go against what the OPIL model suggests. However, the first layer of the OPIL model has certain requirements for any productive interaction system, and this degeneration of player interaction may reveal that the system does not have the correct level of complexity. Alternatively, we consider that the prototype adds an unintended productive system through the use of natural language, as the process of constructing a sentence in itself could be described as productive. This unintended system is not necessary for solving the intended productive system, which may explain the observation above: The players are not required to form sensical sentences if they only seek to complete the intended system. Considering the prototype as two separate systems of productive interaction, (1) that of gathering information, and (2) that of forming sentences, which support the narrative context, we can test each against the requirements of OPIL. The first system may lack an adequate level of complexity, as the players decrease their interaction to repetition of keywords, but it does have a quasi-unique progression and leads to understandable and defined outcomes of uncovering information. The unique progression is who the player talks to and what information they investigate, and the consequences are unlocking or not unlocking pieces of a puzzle. The second system has the complexity and quasi-uniqueness of natural language combined with the narrative context, but is limited by the intelligence of the chatbot, which results in fallback responses that were not understandable. The first system seeks to provide system inherent productive interaction, where the second system provides context based productive interaction for a limited time. Alternatively, seeing as the second system carries no weight in game, and was not designed with the intent of having any, it can be considered to be player inherent productive interaction. The player inherent productive interaction can be seen by the creative sentences participants initially use. Ideally OPIL serves to lay bare all productive interaction subsystems, but as described in the model, the player inherent interaction is specifically that which is not designed for, which suggests it could not have been predicted. As only a minority of participants pointed to this influencing their approach, which may be proof of the interaction being inherent to those particular players. The above discussion and earlier design provides examples of how to design and evaluate using the OPIL model. Though designing by use of the model is something that requires iteration and a fitting target group, which this thesis did not deal with adequately. These requirements are concerns much similar to designing an engaging experience, or experiences meant to invoke flow in a participant: Dependant on the user. The model seems capable of aiding discussion and mapping of the various elements that make out productive interaction in systems and furthering a designer's understanding of the underlying elements to make a positive experience for the user.

10 Conclusion

We start by considering narrative theory and interaction design and use it for the creation of the OPIL model. The OPIL model uses a series of requirements from our definition of productive interaction, possibility for two different levels of detail, and a measure of flow, for identifying and categorizing productive interaction in a digital application. We then designed a prototype based on the desire to invoke productive interaction and tested the prototype on the criteria. The discussion showed the potential for evaluation that the OPIL model provides, while also revealing the shortcomings. The primary shortcoming of OPIL is that in its current state does not take into account the player and context but presents productive interaction as a universal state of mind that can be provoked in anyone with any design. The flow score used may reveal participants who are outside the intended target group, but this approach is not viable when using OPIL on theoretical systems. Another potential shortcoming of the model is the generousness in which a researcher divides a system into productive subsystems. The discussion show that the prototype tested appears to be not one, but two productive systems. Doubt persists in whether the unintentional secondary system makes context based or player inherent productive interaction. Testing OPIL with a measure of flow on the prototype revealed that the fallback responses were not understandable nor did they lead to a defined outcome, which appears to have a correlation to the observed presence of flow in participants. OPIL however shows potential as to the identification and classification of productive interaction, combined with an encompassing model for describing systems in detail, or narrowing down focus to particular pain points.

11 Future Works

Here we seek to consider the problems of OPIL, and which developments or considerations are necessary for its future uses. As the primary problem is the lack of respect to player and context, OPIL should not be used without target group and demographic considerations. Rather, we would see future versions of OPIL being specified to particular groups, be they divided by player types, genre preferences, game literacy or similar. Perhaps then, a general model for defining players could support OPIL in more varied applications. Second, the prototype in question did not allow us to investigate OPIL used with Aylett's levels, as all interaction was on the concrete level. The player's approach is however built into the concept of productive interaction, as we present it, though it might be interesting to see whether it is possible to split the two. It is also with regret that we did not approach OPIL with more studies of the cognitive sciences and consider this as an important stepping stone for further development.

12 References

Andreen, M. (2017). *Choice in digital games: A taxonomy of choice types applied to player agency and identity* (Unpublished doctoral dissertation).

Aylett, R. (2000). Emergent narrative, social immersion and "storification". In *Proceedings of the 1st international workshop on narrative and interactive learning environments* (pp. 35–44).

Aylett, R., Louchart, S., Dias, J., Paiva, A., Vala, M., Woods, S., & Hall, L. (2006). Unscripted narrative for affectively driven characters. *IEEE Computer Graphics and Applications*, *26*(3), 42–52.

Bevensee, S. H., & Schoenau-Fog, H. (2013). Conceptualizing productive interactivity in emergent narratives. In *International conference on interactive digital storytelling* (pp. 61–64).

Bruni, L. E., & Baceviciute, S. (2013). Narrative intelligibility and closure in interactive systems. In *International conference on interactive digital storytelling* (pp. 13–24).

Caillois, R. (1961). *Man, play, and games.* University of Illinois Press.

Cavazza, M., & Charles, F. (2005). Dialogue generation in character-based interactive storytelling. In *Aiide* (pp. 21–26).

Czisikszentmihalyi, M. (n.d.). *Flow-the psychology of optimal experience, 1990.* Harper & Row.

Dishonored developer Harvey Smith details "Chaos" system and Dunwall. kernel description. (n.d.). https://attackofthefanboy.com/e3/dishonored-developer-harvey-smith-talks-chaos-system-immersive-world-dunwall/. (Accessed: 2018-02-20)

Eco, U. (n.d.). *The role of the reader, 1981.*

Emirbayer, M., & Mische, A. (1998). What is agency? *American journal of sociology*, *103*(4), 962–1023.

Flowers, A., Magerko, B., & Mishra, P. (2006). Gamemasters and interactive story: A categorization of storytelling techniques in live roleplaying. *Futureplay, London, Ontario*.

Gee, J. P. (2003). What video games have to teach us about learning and literacy. *Computers in Entertainment (CIE)*, *1*(1), 20–20.

Hill, J., Ford, W. R., & Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. *Computers in Human Behavior*, *49*, 245–250.

Jackson, S. A., & Eklund, R. C. (2002). Assessing flow in physical activity: The flow state scale–2 and dispositional flow scale–2. *Journal of Sport and Exercise Psychology*, *24*(2), 133–150.

Jenkins, H. (2004). Game design as narrative. *Computer*, *44*, 53.

Laurel, B. (1991). Computers as theatre addison-wesley. *Reading, MA*.

Lebowitz, M. (1983). Creating a story-telling universe. In *Ijcai* (pp. 63–65).

Louchart, S., & Aylett, R. (2003). Solving the narrative paradox in ves–lessons from rpgs. In *International workshop on intelligent virtual agents* (pp. 244–248).

Louchart, S., Swartjes, I., Kriegel, M., & Aylett, R. (2008). Purposeful authoring for emergent narrative. In *Joint international conference on interactive digital storytelling* (pp. 273–284).

Mallon, B. (2008). Towards a taxonomy of perceived agency in narrative game-play. *Computers in Entertainment (CIE)*, *5*(4), 4.

Mateas, M., & Stern, A. (2003). Façade: An experiment in building a fully-realized interactive drama. In *Game developers conference* (Vol. 2, pp. 4–8).

McCoy, J., Treanor, M., Samuel, B., Tearse, B., Mateas, M., & Wardrip-Fruin, N. (2010). Comme il faut 2: A fully realized model for socially-oriented gameplay. In *Proceedings of the intelligent narrative technologies iii workshop* (p. 10).

Meehan, J. R. (1977). Tale-spin, an interactive program that writes stories. In *Ijcai* (Vol. 77, pp. 91–98).

Moneta, G. B. (2012). On the measurement and conceptualization of flow. In *Advances in flow research* (pp. 23–50). Springer.

Murray, J. (2000). Hamlet on the holodeck. 1997. *Murray presents views to computer-based and interactive narratives and discusses notions of procedural authoring, immersion and agency*.

Nakamura, J., & Csikszentmihalyi, M. (2009). Flow theory and research. *Handbook of positive psychology*, 195–206.

Payne, B. R., Jackson, J. J., Noh, S. R., & Stine-Morrow, E. A. (2011). In the zone: Flow state and cognition in older adults. *Psychology and aging*, *26*(3), 738.

Ryan, J. O., Mateas, M., & Wardrip-Fruin, N. (2015). Open design challenges for interactive emergent narrative. In *International conference on interactive digital storytelling* (pp. 14–26).

Ryan, M.-L. (2001). *Narrative as virtual reality: Immersion and interactivity in literature and electronic media.* Johns Hopkins University Press.

Ryan, M.-L. (2006). *Avatars of story.* U of Minnesota Press.

Ryan, M.-L. (2015). *Narrative as virtual reality 2: Revisiting immersion and interactivity in literature and electronic media* (Vol. 2). JHU Press.

Salen, K., & Zimmerman, E. (2004). *Rules of play: Game design fundamentals.* MIT press.

Stern, A. (2008). Embracing the combinatorial explosion: A brief prescription for interactive story r&d. In *Joint international conference on interactive digital storytelling* (pp. 1–5).

Vinyals, O., & Le, Q. V. (2015). A neural conversational model. *Computation and Language, arXiv, Cornell University*.

Weizenbaum, J. (1983). Eliza—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *26*(1), 23–28.

Zubairm, P., Bhat, H., & Lone, T. (2017). Cortana-intelligent personal digital assistant: A review. *International Journal of Advanced Research in Computer Science*, *8*(7).

13 Appendix

In order to view the different appendices. Appendix will be uploaded to the digital eksamen

Appendix by occurrence in report:

- Appendix A: Character Relations
- Appendix B: Name Transfer Sheet
- Appendix C: Plotlines
- Appendix D: Plothooks by character and trigger