



Linneuniversitetet
Kalmar Vaxjö

Master Thesis Project

Procedural Narrative Generation Through Emotionally Interesting Non-Player Characters



Author: Ioseff Griffith
Supervisor: Johan Hagelbäck
Examiner: Narges Khakpour
Semester: VT 2018
Course Code: 5DV50E
Subject: Computer Science

Abstract

Procedural content generation is a technique used to produce a wide range of computer-generated content in many industries today, the video game industry in particular. This study focuses on how procedural content generation can be applied to create emotionally interesting non-player characters and through this, generate narrative snippets that can immerse and interest a reader. The main points examined are how to achieve this using a modular approach to personality and behaviour, how well readers can distinguish whether motivations and interactions are generated by a computer or written by a human, and to what degree a reader can be immersed in a computer-generated narrative. Procedural narrative could help to reduce workload on large projects or lower costs, and is an area in which there is much room for further research. To answer these problems, a literature review of existing techniques for the creation of emotionally interesting non-player characters was conducted and used to design and construct a prototype implementation for generating procedural narrative. The output of this narrative was dressed up to match the style of a human text and A/B testing was conducted utilising a survey in order to evaluate and compare responses to the two texts. Ultimately, the results showed very little difference between the perception of the human-written text and the computer-written text, with the only aspects found lacking in the computer-written text being clarity of emotion and foreshadowing.

Keywords: procedural content generation, procedural narrative, emotionally interesting non-player characters, immersion

Contents

1	Introduction	1
1.1	Background and Motivations	1
1.1.1	Procedural Content Generation	1
1.1.2	Artificial intelligence	1
1.1.3	Motivation	2
1.2	Problem Statement	2
1.3	Contributions/Limitations	3
1.4	Target group	4
1.5	Report Structure	4
2	Literature Review	5
2.1	Immersion	5
2.2	Procedural Content Generation	5
2.3	Interactive drama	5
2.3.1	Thespian	5
2.3.2	Façade	7
2.3.3	Versu	9
2.4	Models of Emotion	11
2.4.1	Emotional Elicitation Process	11
2.4.2	EMA	13
2.5	Wants and Needs	15
2.5.1	COR-E	15
2.5.2	ACORE	15
2.6	Statecharts	16
2.7	Summary	17
3	Method	19
3.1	Scientific approach	19
3.2	Literature review	19
3.3	Prototype design and implementation	19
3.4	Survey	20
3.5	Reliability and Validity	24
3.6	Ethical considerations	25
4	Implementation	26
4.1	Personality and Trait Modules	26
4.1.1	Overview	26
4.1.2	Generous	27
4.1.3	Violent	27
4.1.4	Thief	28
4.1.5	Lawful	28
4.1.6	Social	28
4.1.7	Survival	28
4.2	World State	29
4.3	Wants and Needs	29
4.4	Evaluation and Memory	30
4.5	Actions as Statecharts	33
4.6	Simulation Execution	35

5	Evaluation	36
5.1	Results	36
5.2	Analysis	44
5.3	Discussion	47
6	Conclusion	50
7	Future Work	52
	References	53
A	Appendix 1	A
A.1	Text A	A
A.2	Text B	A
A.2.1	Prototype Output Excerpt	A
A.2.2	Final Text	C
B	Appendix 2	D
B.1	Quantitative data	D
B.1.1	Text A	E
B.1.2	Text B	F
B.2	Qualitative data	G
B.2.1	Text A	G
B.2.2	Text B	I

1 Introduction

This thesis describes research in the field of procedural content generation (PCG) and artificial intelligence (AI). Its primary purpose is to explore and present modular solutions to facilitate the procedural generation of emotionally interesting non-player characters (NPCs) that can provide immersive dialogue and story lines in the form of narrative to the player within predefined bounds.

1.1 Background and Motivations

This section briefly describes the background of the subject area and the motivations behind the conception of this thesis.

1.1.1 Procedural Content Generation

Procedural content generation is often used to refer to ‘the algorithmic creation of game content with limited or indirect user input’ [1]. Such algorithms are often defined by their novelty, structure, interest, speed, and controllability [2]. In this instance, novelty refers to the unpredictability of the output, structure refers to the fact that the output should contain some semblance of structure rather than random noise, interest refers to the engagement level of the player, and controllability refers to how the content is generated based on a set of input parameters. To be considered procedural, the output must be generated at least in part by the automated algorithm with minimal user input [3].

Since the early days of the video game industry, PCG has been widely used in many ways [4]. In recent years, it has grown to be an integral part of the industry as a whole. As the industry continues to mature, it has begun to see use in almost all aspects of the creation of games, from levels and maps, to stories and characters, all of which are considered to be ‘content’ under our definition [5].

PCG is not only used within the video game industry, and not only for the purpose of creating content of the types described above; the techniques are often used for the generation of graphical artefacts and feature frequently within the film industry [6].

One of the many goals of its use within these industries is often considered to be lowering the costs of development [7] by reducing the amount of man-made content required for the release of a game (or film) in addition to providing more content for players post-release. This allows development of games with smaller teams and budgets, often by independent developers in the process of making ‘indie’ games [8].

1.1.2 Artificial intelligence

Artificial intelligence is a term that usually refers to ‘systems that exhibit the characteristics we associate with intelligence in human behaviour’, or more specifically, agents that perceive their environment in some manner and take actions to maximise their chances of achieving their goals [9]. This can involve many aspects, such as problem solving or machine learning.

Artificial intelligence is present in almost every single-player game in some format [10], whether it be a simple background AI that performs a minor role, or a more visible foreground AI in the form of adversaries or companions. This AI

usually errs on the simplistic side and often suffers from ‘artificial stupidity’ [11], leading to excessively predictable behaviours, a low level of emotional engagement from the player, and often in the case of companions, a deep sense of frustration. For many games, this is sufficient to provide the desired play experience.

Modular AI refers to the design principle in which individual components related to the interaction of an actor with the world are separated into modules which can be interchanged to provide varying behaviour [12]. This allows for great freedom of design and implementation by allowing aspects such as senses or behavioural traits to be defined on a generic level and applied in groups to modify the ultimate actions of a given actor.

1.1.3 Motivation

Immersion is often considered to be an important aspect within video games, with gamers being able to tell when a game is well constructed and reacting to the effort put in to entice them [13]. One of the key contributors to generating this immersion is the use of narrative flow [14]. NPCs and their interactions with players and each other through quests, stories, and connections, number amongst the factors that influence both the narrative flow of a game and the overall levels of immersion generated [15]. To be able to generate such aspects on the fly without need for direct human intervention would be a significant step toward to reducing development time and resources consumed.

Much research has previously been conducted in the fields of procedural generation and artificial intelligence. There are even video games that have made significant progress within the field of producing emotionally interesting NPCs, one such example being *Faade* [16].

In their paper on the architecture of *Versu* [17], Evans and Short discuss the design decisions which led to their implementation of the AI in *Versu* and how it increases the speed at which content can be created using similar systems.

Pena et al. discuss the application of an emotional model to role-playing game (RPG) characters [18] with focus on the ‘Emotional Elicitation Process’ engine. This engine processes events and environment to adapt character moods and emotions using a set of components partially grounded in psychology.

Manavalan et al. [19] also discuss an algorithm intended to procedurally generate emotionally affected behaviour through the use of *ACORE*, a system which relies on a representation of resources available to a given AI and the actions they will take in order to improve their values. This ultimately results in behaviours such as fleeing from danger or seeking items, with the actions performed presenting an illusion of emotion from the NPC.

Finally, Dragert et al. [20] detail the use of statecharts to produce modular AI through procedural generation. In particular, they focus on how this approach can be used to generate immersion through background AI characters such as squirrels or other animals.

1.2 Problem Statement

The aim of this project is to expand upon previous research in order to develop and evaluate an algorithm, or series of algorithms, that allow for the procedural generation of emotionally interesting non-player characters in video games. These NPCs should be capable of independent interaction, not only with the player, but with

other AI-controlled NPCs, forging realistic bonds and relationships based on their needs and preferences. This procedural generation should also be strongly bounded by pre-existent lore within the setting of the game, resulting in rich back-stories that give rise to individual personalities and the ability to make context-relevant requests of the player that could have the potential to advance these storylines further.

As there are many different approaches to implementing algorithms where PCG is involved, this project will be unable to explore all possible avenues. Instead, it will focus primarily on examining the results of a modular approach to personality, using statecharts to control behaviour and manage wants and needs.

Research Questions

RQ1	How can a modular approach to personality and behaviour be used to produce emotionally interesting NPCs with minimal human influence?
RQ2	To what degree can readers discern whether or not motivations and interactions between NPCs are pre-scripted or generated based on procedural algorithms?
RQ3	To what degree is it possible to immerse readers in a narrative using only procedurally generated characters and their interactions?

By answering these questions, an expanded approach to procedurally generating NPCs could be found and presented. Compared to existing research, more focus will be placed upon generating inter-personal relationships and history within a world framework, with the intent of providing character depth. Placing the product of these algorithms in front of players or readers will allow for them to be evaluated within their intended domain by their target audience, giving valuable insights on their effectiveness at achieving the goal of emotionally interesting and immersive NPCs and storylines.

The expectation is that participants will be able to discern whether or not the interactions are pre-scripted with varying degrees of difficulty. Dressed up language may help to obfuscate the differences to some degree. It is expected that participants will have less difficulty identifying emotions and motivations in pre-scripted narratives than in procedurally generated narratives.

1.3 Contributions/Limitations

This report presents an outline for the design and implementation of a system for the generation of procedural narrative and emotionally interesting non-player characters through the use of personality traits, wants and needs, and statecharts, in the context of a simulation. A thorough literature review is presented, from which several techniques are sourced for use in the design and implementation. The prototype system is also evaluated in terms of its output through the use of a survey to gauge its success in producing narratives on par with those produced by humans.

Ultimately, the first research question is answered by the design presented in Chapter 4, utilising personality and trait modules to provide behaviours, wants and needs to provide motivation, and statecharts to handle the individual execution details. In relation to the second research question, the degree to which readers could discern whether motivations and interactions were pre-scripted, no statistical significance could be found to suggest they were able to make a distinction between

pre-scripted and procedurally generated content. Finally, with regards to the degree to which readers could be immersed in a procedural narrative, they seemed to experience a similar level of immersion as with the human-written text.

As there are countless personality traits and behaviours exhibited by humans, not all of these can feasibly be implemented in a prototype system. For this reason, only a select number are included at this stage of the project which limits the number of different interactions and outcomes within the simulation.

A limitation is also presented in that the output of the system itself is purely textual. While internally, the system runs a simulation that could potentially be displayed in a visual manner utilising graphics libraries or frameworks, this would result in far less possible applications for the end result. By limiting the focus of this thesis to the output text, it can be applied to a broader subset of media in which narrative is required, which still includes visual media.

1.4 Target group

The target group for this thesis is video game developers with an interest in procedural narrative and the desire to implement such systems in their games. Developers of massively multiplayer online role-playing games with open worlds or the plethora of modern-day sandbox games may find this research of value in producing unique and interesting non-player community stories within their worlds, capable of evolving and interacting in a believable and immersive manner.

Since the prototype system outlined in this thesis is based upon the use of behavioural modules in AI, it could also potentially be of use in other forms of media which require the depiction of interactions between individuals on a medium to large scale. One example of this could be procedurally generated background crowds and bystanders in CGI films.

In addition, as the prototype implementation is designed to provide simplified textual output, with much further work and research in the computer generation of prose, this research could also potentially be used to produce simple short stories, or other similar works.

With even more expansion, there is also a possibility that the research could be of interest in the field of simulating the development and interactions of people and personalities within a restricted region.

1.5 Report Structure

Following the introduction provided in the current chapter, this report can roughly be divided into six chapters. The first of these is the Literature Review (Section 2), which briefly summarises the contents of the literature review and provides an overview of the state of the art. After this, the Method (Section 3) describes the scientific research methods used to address the research questions presented in the thesis. The Implementation chapter (Section 4) then discusses details of the prototype system's design and implementation at a high level, providing an insight into the concepts used and the overall design pattern to follow. This is then followed by the Evaluation chapter (Section 5) which provides the results of the survey, their analysis, and discussions based around this. The penultimate chapter (Section 6) provides the Conclusion to the study, including the answers to the research questions. Finally, the Future Work chapter (Section 7) discusses potential next steps and future applications of the research.

2 Literature Review

The following sections detail various theoretical concepts and frameworks related to the development of procedural narrative and emotionally interesting NPCs within the fields of video games. This includes summaries of various papers, articles, and theories deemed particularly relevant during the literature review process.

2.1 Immersion

Immersion is a term generally used to describe the deep mental involvement of an individual in a task or activity. The concept is often brought up in relation to video games and other types of entertainment media, and is frequently cited as being critical to the enjoyment a player experiences [21]. While games are diverse and fall into countless different genres with vastly different styles, all successful games share this characteristic of drawing players in and allowing them to ‘lose’ themselves in the experience [22].

In general, immersion can be created or broken by almost any aspect of the game, its gameplay elements, their cohesion, or its world. One core part of this involves the believable actions and interactions of NPCs within the game world [23], combined with the narrative they help to uphold. This means that they must act in a way that is appropriate for the context they inhabit and a manner that befits their supposed personality [24].

2.2 Procedural Content Generation

Procedural content generation, as previously described, is a concept generally used to refer to ‘the algorithmic creation of game content with limited or indirect user input’ [1]. Algorithms are designed and used to generate a variety of different types of content with minimal human interaction, providing potentially limitless content for players to experience. While PCG has historically been applied mostly to the generation of levels and maps (areas in which the game takes place), it is also frequently used to generate almost all other types of content and has even influenced higher level areas such as AI-assisted game design and NPC behaviour learning [25].

The procedural generation of narrative is a segment of PCG for which there is far less research than many of the other aspects of video game content. In a talk at the Game Developers Conference [26], Martens introduces a working definition of procedural narrative generation as ‘any automatic process that creates a narrative over time, where that narrative is not determined before the process begins’ [27].

2.3 Interactive drama

The following papers deal largely with the concept of interactive drama utilising autonomous agents designed to display emotionally interesting behaviour based on factors such as personality and motivation.

2.3.1 Thespian

Thespian [28], a framework presented by Si et al., takes the approach of training agents to perform their roles within an interactive drama. Various linear scripts are

authored, specifying the environment of the story and desired paths. With the agents constrained by these scripts and designed to act in accordance to their specific roles and personalities, they can be responsive in an open-ended manner and easily reused across multiple stories.

Each character in a Thespian-based interactive drama is controlled by its own agent, and any of these can be replaced by a human player while still allowing good interactions with the other agents. These agents possess their own subjective view of the world which they can use to reason about the environment, the state of other agents, and how their actions can affect the state. They also have goals in the form of reward functions that affect features or resources the agent might seek to increase. Actions can be selected based on reactive rules wherein they react to other interactions, a bounded optimal with a minimal degree of lookahead to future goals, and a mixture of these. Agents are comprised of a set of components. These are goal states, event states, policies, available actions, and beliefs.

Agent state in Thespian is broken into two types - goal states, and event states. The former represents the progress in achieving particular goals (which are limited to maximising or minimising related resources in Thespian), and the latter keeps track of other environmental or event related state that does not directly relate to the agent's goals.

Rather than use a direct turn-based system for agent actions and interactions, Thespian uses a more flexible variation wherein the 'arousal' levels of each agent is monitored and the agent with the highest value is given priority, allowing them to continue for several turns if necessary. Whenever an agent selects its next action, it uses a limited degree of lookahead in order to consider the responses of other agents to each action, its own action in response to those, and so on. These factors allow their actions to be driven by individual personalities.

Character personalities are modelled in terms of a profile containing goals with various weights. Actions are chosen taking these goals into account, as well as any author defined 'preferred' story paths as dictated by the scripts. Internal heuristics are also used to determine which story paths should be selected during this process.

Thespian was designed partially in order to reduce the authoring effort required when developing agents and stories by allowing reuse of story elements and using certain processes to automate the authoring of characters. Dynamic functions are used to define how agents' actions affect the world. A story is built first by creating a list of possible actions that can be performed by an agent and possible goals, then by checking to see if such a function is already defined. If so, it can be reused, otherwise a new dynamic function must be written. These steps are performed manually by the author and programmer.

When authoring a character, the goals and goal weights must be specified and adjusted and their beliefs modified. Desired paths must then be written in a similar manner to writing variations on a story, consisting of a sequence of actions that should ideally be executed. These paths are then used by an automated fitting process in order to properly configure that characters' personalities. It is possible for this fitting process to fail due to errors in the story or characters, which must then be rectified by the author. Once the fitting process has been completed, the author modifies the weights and goals accordingly and runs the fitting process again for fine-tuning.

As Thespian was designed with re-usability of characters in mind, it is simple to re-use characters in new stories, modifying them as necessary by adjusting or

dropping goals, or even combining multiple characters with similar personalities into a single, richer character. The greatest limitations of the system are ultimately related to the lack of good authoring tools to allow non-programmers to quickly and easily create agents and stories.

2.3.2 Façade

Façade is part of an experiment using the research and development of an architecture which allows the creation of interactive drama and is described in depth in *Façade: An Experiment in Building a Fully-Realized Interactive Drama* [16] and *Procedural Authorship: A Case-Study Of the Interactive Drama Façade* [29]. The system is described as being in part inspired by the inability of players to communicate with game systems using natural language and is compared to being present on a stage with two live actors who try to create dramatic situations. In Façade, a single story scenario is presented and the actions of the player within this scenario can drastically change the narrative flow. The game uses simple three-dimensional graphics and the player and two actors have complete freedom to move around, interact with objects, perform gestures, and speak at any time. The NPC actors are designed to react in a robust manner to almost anything the player says or does.

The papers discuss two possible approaches to creating interactive narrative experiences. The first is the hand-crafted approach in which a complex navigational graph is created by human authors which controls the flow of events based on choices made by the player. The second is the use of a procedural simulation in which a virtual world is defined with various elements, parameters, and simplistic autonomous agents. Most of the events in the second, which are experienced by the player in a similar manner to how reality is experienced, do not particularly form a strong narrative structure. Only in a few cases to emergent narratives form where a sequence of these events coincides in such a manner that something resembling narrative structure comes into play.

Façade is stated to be an attempt to find a middle ground where some degree of human authorship and structure is required, but with the procedural variation and player freedom of a simulation. Overall, Façade is a simulation foremost as defined by the virtual world and its agents. At the same time, a third agent, named the ‘drama manager’ monitors the simulation and modifies the actions of the other actors by adding or removing behaviours or discourse context. The updates themselves are referred to by the authors as ‘beats’ - a collection of behaviours that relate to a given situation or context. These individual beats need to first be annotated by the author with their preconditions and the effects they have on the overall plot such that the drama manager will be able to utilise them correctly. As this results in ordering of the beats based on these preconditions and effects, the structure very much resembles that of the hand-crafted navigational graph previously mentioned. Beats are changed extremely frequently and picked from a large pool.

One of the motivations for taking a middle ground stance between these two possible approaches is stated to be the ability for the architecture to coordinate the activity of autonomous agents. This is done through joint goals and behaviour and allows for more sophisticated and coordinated actions between otherwise autonomous agents.

Four authoring languages were developed in order to support Façade. These are: A Behaviour Language (ABL) which is a reactive planning language based

on Hap and allows for behaviours in sequence or in parallel; Natural Language Understanding (NLU) which is a template language designed to map text input by the player into discourse acts that can be used to communicate with the agents; Reaction Decider language, which is used to propose and select actions; and Beat Sequencing language which is used for the drama management.

One of the key points of interest is in how Façade achieves believable agents. One of the stated requirements in order to achieve this is that they should be able to perform several intelligent activities in parallel such as walking, talking, gesturing, and using objects at the same time. Within the ABL language used to control these agents in Façade, activities are represented as goals consisting of one to many behaviours that might accomplish the task. These behaviours are defined as a series of steps that do not necessarily need to be executed sequentially. Upon completion of these steps, the activity is terminated. It is also possible for steps to fail resulting in the need to find an alternative activity to accomplish the goal. These goals can also feature sub-goals and sub-behaviours resulting in the need for an active behaviour tree.

In order to allow story beats to be flexible and interesting, they feature a collection of behaviours which act as a set of goals. These can occur in varying orders and can be optional. Goals in this context are broken down into five types: expressing the intentions of the beat, posing a dramatic question or situation to the player, reacting to the player, waiting for the player to react to the situation, a final reaction which transitions out of the goal. These beats control both of Façade's autonomous agents simultaneously with parallel behaviours in order to achieve synchronisation between the otherwise asynchronous agents. This prevents them from performing conflicting actions in order to achieve more natural interaction.

At any given time, any beat is designed to be able to interpret and respond to player interaction and result in changes to the current beat, and potentially to future beats. These interactions can be anything from dialogue spoken by the player and gestures made, to where they move within the room. In order to facilitate this, 'contexts' are used which define how to react to any player interaction at that time. All beats trigger several contexts with varying priorities that can have both local or global consequences for the story progression. At least one unique context must be written by an author per beat. By varying responses depending on the currently active context, the narrative can evolve and move in different ways over time.

Handling player interaction through natural language processing and reacting correctly is clearly an important aspect to Façade, but also one of its limiting factors. The surface text is mapped to a large number of patterns and intermediate meanings using template rules. These are described as often being too permissive - mapping too many inputs to a given act. The context is then used in order to decide an appropriate reaction based upon these intermediate meanings. Finally, the active beat usually terminates its current behaviours in order to perform the reaction, or suspends them to be resumed afterwards if necessary. These reactions can also have effects on global values, such as affinity between actors and the player, and so on.

While the architecture offers a framework for some degree of procedural narrative, it does so within a framework of behaviours provided by a human author. The system itself cannot procedurally generate these behaviours, only mix these into sequences in clever ways. The 20 minute episode created as an experiment is stated to have taken almost two man-years to construct due to the amount of work required to author the behaviours. The complexity of the languages required for authoring

scenarios is an additional barrier to use as these two man-years are in the context of individuals who are intimately familiar with the languages having developed them. The system's natural language processing is also stated to be less than complete - only certain patterns and phrases are sought out and acted upon.

Other limitations presented by the authors include that of the player's role as an interactive observer rather than a protagonist. At best, the player's role is equal to that of the autonomous agents, but not the centre role. It is also unclear how well the architecture would scale beyond very limited and intimate experiences. The implemented scenario involves only two agents and a small area in which the player can interact. This might not scale well to the larger scope of most modern games wherein there can be hundreds of actors and large, open worlds.

2.3.3 Versu

Versu [30] is an engine for narrative video game development focused on producing emotionally interesting NPCs through dynamic interactions based on character traits. The developers describe Versu as more of an 'improvisational play' in which the actors are the driving force behind the story and interactions as opposed to the traditional approach in which the 'drama manager' (DM) dictates events [31]. This is facilitated by a system in which roles and social practices are designed and implemented to be entirely character agnostic, allowing for any character to fit into any role within a narrative. One of the big goals of the project is stated to be that Versu could be authored without knowing whether a given role was played by a human player, or an NPC. It also allows for the player to recognise patterns within simulation and build an understanding in order to anticipate the outcome of making certain choices.

The system was designed with the intent of solving several of the issues that Façade encountered, as such, it does not feature natural language processing for user input and instead relies on the use of menus and option selection by the player. In addition, all internal state and events are represented purely in text. The use of text over 3D or 2D graphics is justified as being able to better convey the thoughts and emotional reactions of the NPCs, allowing the use of adverbial modifiers to render actions 'reluctant' or 'tentative' where graphical representations may struggle to express these or require expensive interactions to do so.

At its core, the simulation is built on agents and social practices. These social practices define recurring social situations that can be used to coordinate agents. Essentially, these practices are used as a blueprint to define what each actor can do in a given situation. Multiples of these can exist at the same time and are made up of various states. These states are capable of storing arbitrary persistence data and can be used to update the state of the narrative world.

As Versu is seen as an 'improvisational play', segments with actors are split into 'episodes' or 'scenes'. These are initially defined by a number of properties - the social practices that may exist within the given scene, the initial states of the agents, and the initial state of the world itself. Once the scene begins, a scoring system is used by each agent to determine which action to execute.

Since all agents act based upon scores assigned to actions based upon their individual internal parameters, Versu restricts these to only be available during social practices in order to ensure that actions are chosen with context taken into account. In addition, agents understand what they should and should not do in any given sit-

uation and will respect these norms. They are also capable of responding to player violations of these norms. This helps to reduce the number of options available while encouraging interaction between agents that extends beyond merely heading directly for their goals.

When an agent evaluates a choice, rather than evaluating a simplified version of the consequences that reflect the intended goal of the agent (e.g. obtaining something they want), the actual chain of consequences is executed and taken into account in the evaluation. This allows them to use some simplified short term planning that takes into account consequences and advance multiple goals at once while evaluating the true consequences of actions.

Unlike many traditional simulations, Versu's core state is stored as a set of modal logic sentences in order to allow it to be more accessible for inspection and debugging. This makes it far easier for the internal state to be understood by developers as the logic behind it is intended to be very transparent. The particular type of modal logic used for the state is exclusion logic [32]. By using this type of logic, very fine-grained personalities can be specified which can result in much more expressive desires in NPCs. This also allows for roles to be evaluated using predefined role evaluations, resulting in the formation of dynamic relationships which have states of their own.

While Versu goes to great lengths to simulate emotion, relationships, and other related aspects of human behaviour for NPCs, it does not fully implement possibilities for misunderstandings. Beliefs are implemented to some degree within each agent, but as memory constraints do not allow for all agents to possess their own complete and separate representation of the world, these beliefs are far shallower than they could be.

Social practices alone would not be enough in order to adequately simulate interesting characters as it does not provide mechanisms for changing goals, desires, or emotions over time. This is partially implemented through 'character arcs', a series of details relating to the character, their backgrounds, and their possible desirable outcomes. Another limitation in the system is also present in this aspect - while the player can freely choose to set aside their desires and goals for another, agents cannot, and thus remain in a supporting role throughout.

Within the model, various internal variables are tracked relating to relationship states, beliefs, emotions, and evaluations. These are altered by character actions in order to facilitate communication between individual agents and result in repercussions and changes within the story. Emotional state changes are described as being relatively unimportant in terms of impacting the overall story in that they are transient and only affect immediate decisions rather than those with long term consequences. Instead, it is the character evaluation of other characters and their actions that causes lasting impact by altering relationship states. For this reason, relationship states and character evaluation are considered the two most significant parts of the core model.

An important design goal described in regards to the core model is that individual agent actions should result in changes to multiple qualities such as changing emotions, altering evaluations, communicating emotions or beliefs, and so on. This is intended to allow characters to appear to react both to the surface text and its subtext by communicating both the obvious qualities and those that are more concealed.

While the DM is not the central player in the drama of Versu, it does exist in

some capacity. For the most part, the manager initially sets up the scene, the actors, and assigns them social practices which they are the actors are then given free reign to act out. However, the DM is implemented as a social practice itself and, as such, does have an impact on how a scene plays out. It is described as being designed to periodically spawn new social practices and tweak individual agent goals in order to keep the story moving. The DM is not designed to plan ahead, only to move along the story in various ways at important points.

Some key limitations are identified in the system in addition to those previously mentioned. Social practices are simplified such that all agents have a shared understanding of the state of social practices. This means that two agents cannot disagree on the current state, with the example being provided that two agents cannot disagree whose turn it is in a board game. They also cannot have divergent understandings of when certain actions are appropriate or inappropriate - these are also embedded as a part of the social practice.

The system also does not implement the concept of a public mood [33] which would allow certain social practices triggered by events to take precedence over others by suppressing actions that would not be appropriate to the current situation. The example given is that, in the event of a murder, other agents should not be considering flirting or joking due to the seriousness of the situation. Versu deals with this by using an excessively broad and crude mechanism to suppress other social practices as it also suppresses those that would make sense given the context.

Future work relating to Versu involves the time and effort required for a knowledgeable user to generate their own content using the system. While it is described as faster than using Façade it could still take two to three months to author a fully fleshed out 20 minute episode. This can apparently be shortened to one week using an authoring tool and language by the name of Prompter.

2.4 Models of Emotion

The following papers deal with specific representations of emotion through the use of models with the intent of emulating personality and emotionally affected behaviour.

2.4.1 Emotional Elicitation Process

In 2011, Pena et al. proposed a lightweight emotional model designed to give players the impression that agent responses are motivated by emotion while conforming to the limited time and computational complexity allowed in the development of a video game [18]. This they refer to as the Emotional Elicitation Process (EEP). They define seven aspects that take into account the two main objectives: being able to produce coherent and expected emotional responses, and being simple enough not to cause too much overhead within the system. These seven aspects are:

1. Perception of events and consequences should be evaluated separately and both produce emotions.
2. Events should be classified as actions, consequences, and objects.
3. A robust and efficient definition of transitions between emotional states.
4. A full classification of all elements in the environment so that a correspondence can be identified between configurations and events.

5. A character profile or model that defines the initial tendencies such as personality and general mood.
6. An emotional dynamic that represents differing levels of emotion.
7. Mechanisms for propagating information in order for the emotions to affect control of the character.

The model's design is built upon psychology techniques and models. The first of these is the OCC model [34] which specifies various categories for emotion synthesis based on goal-relevant events and attractive or unattractive objects. A simplified representation of OCC is used as part of the EEP to analyse events perceived by an actor and select the emotions based on the event.

The Big Five Factors [35] are used as part of the character profile within the model in order to define parametrised aspects of their personality with a range between -1 and 1. These factors are:

1. Openness and how much appreciation they have for art, their curiosity, sense of adventure, etc.
2. Conscientiousness and self-discipline, duty, planned behaviour over spontaneous actions.
3. Extroversion and positive emotion, seeking the company of others.
4. Agreeableness and the tendency to be compassionate and cooperative.
5. Neuroticism and how easily they might experience negative emotions.

Finally, the Pleasure-Arousal-Dominance (PAD) Temperament Model [36] is used to project the OCC model emotions into tangible numerical values, translating personalities to default moods, and representing moods in a vector space. This model uses three scales to represent emotions: pleasure-displeasure which, arousal-nonarousal, and dominance-submissiveness. The space defined by the PAD model can be broken into eight parts representing different emotions and moods a character can experience based on the values of the scales: Exuberant, Disdainful, Docile, Bored, Relaxed, Dependent, Anxious, and Hostile.

Primarily, the EEP model uses the OCC and PAD models in order to quantify and manipulate numerical representations of emotions to simulate emotional responsiveness from NPCs. The architecture is made up of four components: conceptual dictionaries, character profiles, the EEP engine, and the mood vectorial space. A single cycle involves a change in the environment which is decomposed and analysed by the engine based on the character profile, the emotions generated are translated into vectors described by the PAD model and added to the mood represented in the vector space, and this space allows for a correct tag representing the mood to be pulled from the conceptual dictionary.

The conceptual dictionaries, which define all possible elements that appear in the scenarios, consist of four different types: consequences of actions, actions themselves, objects, and relationships. The character profile specifies emotional parameters that connect to the concepts described by the dictionaries, such as the desirability of consequences or the appeal of certain objects. In addition, character personality traits are defined here based on the big five factors, as are the mood space points

which represent the combination of emotions brought about by recent events. The mood vectorial space is a three-dimensional space designed to hold representations of moods. It is designed to compose and add emotions to the current mood, decay these emotions back to the base mood, and provide a mechanism for associating with a specific mood tag.

Finally, the EEP engine acts as the core of the whole system, evaluating all events as perceived by characters. This involves a multi-step process where emotions are elicited by events, quantified, and used to find a representative mood tag. When no events occur, another process is used to move the mood back towards its base levels.

While the paper draws the conclusion that the EEP model fulfils all the previously mentioned requirements, it does not discuss any potential limitations of the model. No mention is made of possible future work or improvements.

2.4.2 EMA

The EMA - or EMotion and Adaptation - process model [37] aims to appraise and model the dynamics of emotion. The authors posit that computational models of emotion must explain not only the rapid dynamics of emotional reactions, but also the gradual evolution of responses that follow. The central tenets used by the model are adopted from appraisal theories [38] of emotion, which involve interpreting an individual's relationship with their environment. Emotions are associated with certain combinations of factors. These are:

- Relevance, valence, and intensity, which relate to the detection of personally significant events, their effects, and the intensity of these effects and consequences.
- Future implications, which involve hopes and fears with reactions to violations of their expectations, likelihood, and changeability.
- Blame and responsibility, which relate to causality, agency, and actor intent, motivations, and knowledge.
- Power and coping potential, which relate to the sense of control a subject has over an event that would elicit emotions in addition to their control over their individual emotions.
- Coping strategies, which can be triggered by events and can change a person's beliefs, desires, or intentions, and are closely related to the evolution of emotion dynamics.

In addition to this, a number of assumptions are used by the model relating to appraisal. It is assumed that appraisal causes emotion, that these appraisals are part of a cycle in which events and situations are constantly reappraised, and that appraisals themselves are very quick and shallow in nature.

EMA itself is intended to be a computational model of a mental process based on the above principles. It consists of a set of processes that interprets person-environment relationships in terms of a set of appraisal values and coping strategies that can affect and manipulate these relationships. A number of representation common in cognitive architecture are used by the model in order to represent knowledge.

At any given time, a causal interpretation is made up of a record of past events, the current world state, and a set of possible future outcomes and plans.

Individual actions within the causal representation have preconditions, durations, and effects. An effect does not necessarily have to occur synchronously with the rest of the action. The states and actions within the representation include variables which represent the beliefs, desires, and intentions of an agent in the relative situation. States can be contingent on or a precondition for other states, establishing causal relations between them. Additionally, agents have preferences for states and actions have probabilities which indicate the likelihood that the agent will perform them or the probability that it can be executed at all. Actions can also update and change world state, beliefs, and other aspects of the agents.

Each proposition or possible outcome relating to events is appraised by the EMA model based on a set of variables. These are:

- Relevance, which is determined based on the relative utility of a proposition or its causal impacts to the agent, an example being that injury would have negative utility and thus be relevant, whereas something with zero utility would not.
- Perspective, which defines the viewpoint from which the appraisal takes place, which is usually that of the agent themselves.
- Desirability, or the value (positive or negative) a proposition has to the agent.
- Likelihood that an event will occur.
- Expectedness, or the extent to which the outcome of an action is known and observed. In the EMA model this is always assumed to be high.
- Causal attribution, which involves the credit or blame for actions and who it falls upon.
- Controllability, or the degree to which an agent can alter outcomes based on their actions.
- Changeability, which relates to whether another agent can alter the outcome.

Within EMA, emotion is broken into two aspects - appraisal and mood. Multiple appraisal frames are used to represent specific emotions and their intensities which will decide an agent's coping response based on a combination of these factors. These are also aggregated into a mood which can change slowly over time based on different appraisals. This mood also has an intensity in the same way as the appraisals. The coping strategies based on these moods and emotions determine the action the agent takes. Many of these relate to the agent's relation with the environment, such as encouraging them to seek information, to their beliefs, changing how they view others or their wishful thinking, managing their desire to achieve a goal, or adjusting whether they carry out certain actions, and so on.

There are several limitations of the EMA model. After using the model to analyse a real interaction with a bird, it was observed that the predicted responses do not always correspond with the actual behaviours. These primarily related to emotions of fear, and actions taken in these situations that do not necessarily involve typical fear responses such as running away. This could be rectified with some further

adjustments to the model. Additionally, EMA lacks a detailed model of the time course of events - this is considered by the authors to be excessively shallow and under-constrained..

2.5 Wants and Needs

These papers deal with the use of resources and their desirability as a trigger for motivation, and through that an emulation of personality and emotion via believable and understandable behaviours.

2.5.1 COR-E

Campano et al., in a 2013 paper, present the COR-E model [39] for producing believable emotional behaviours in NPC agents. The model itself is based on a principle wherein agents attempt to acquire and protect desirable resources. Resources are associated with two types of behaviour - protective and acquisitive - and the resource values are weighted differently in terms of desirability for different NPCs. Based on these weights, values are computed for individual behaviours and behaviours are selected based partially on this value. The more desirable a resource, the higher its weight, and thus the value of the behaviour that will increase it.

Behaviours are defined with preconditions which must be true in order for the NPC to execute them. Each behaviour will adjust resources in different ways, allowing them to be acquired, protected, threatened, or lost by any given action. As such there are essentially two subsets of behaviour - acquisitive and protective. Behaviours are selected based on which resources an NPC desires at a given time, taking into account any loss and gain of resources that might result from this. If the action of another NPC will threaten one of its resources and the threatened resource is more desirable than those that can be obtained by another action, the NPC will take protective action instead. Otherwise, an acquisitive behaviour will be triggered.

In order to evaluate the model, the authors conducted an experiment in which observers were shown pre-recorded video clips of NPCs simulated by the model. The observers evaluated whether the NPCs displayed believable emotional behaviour and filled in a questionnaire. The results of the questionnaire suggested that an extension to the model involving dynamic threat level would be a good addition for future versions. This dynamic threat level would cause agents' responses to threats to its resources to become more severe as it loses more of this resource. Ultimately, the results showed that COR-E fulfilled all the proposed hypotheses and produced emotionally believable actions in NPCs, though part of this is attributed to the phrasing used to display the emotions on-screen to the player. Despite this, the most believable video clip is stated to have been one in which the model was not used.

2.5.2 ACORE

In a 2015 paper, Manavalan et al. discuss an implementation of a lightweight algorithm for procedurally generating emotionally affected behaviour and the appearance of these emotions [19] based upon ACORE [40] as discussed in a previous paper by the author.

ACORE, or Appraisal for Conservation Of Resources Engine, is an extension of the COR-E model [39] with the addition of the ability to compute emotions. The

model is focused mainly on resources wherein each NPC is modelled with weighted resources based on their importance to the NPC. These weights essentially define the personality of a given NPC. At any time, an NPC will have a vector of resource values which shows how much of each resource is in their possession at that time. It is possible for weights to change over time.

Each of an NPC's possible actions or behaviours results in a change in their resource vectors. Actions and behaviours are selected by the NPC in order to increase their resources based on the weights provided. These changes are then appraised in order to elicit emotions. Only four emotions are defined for ACORE: hope, joy, fear, and distress. Hope refers to the possibility that a resource may increase based on an action but is not guaranteed to do so, joy occurs when an action is guaranteed to increase a resource, and conversely, fear and distress act as the negative equivalents of the former two. Intensity of emotion is based upon the desirability of the resource change based upon values and weights.

An experiment was conducted in which a scenario was modelled where six people stand in line waiting to purchase a video game. Each NPC had three resources - health, reputation, and rank - and could either pass the person in front of them or keep waiting at each step. Every step, the person at the front of the line would purchase the video game and leave. A series of one line phrases were used to represent the different emotions felt by the NPCs.

Three groups were formed for the experiment: the experiment group which would use ACORE and display emotions through the one line phrases; the control group using COR-E which would only perform the actions and not display these emotions; and the random group, which would use the same parameters as the experiment group but with randomly selected facial expressions and statements. 94 individuals were recruited to observe instances of this experiment and rate whether virtual characters showed believable emotions on a scale of 1 to 7. Ultimately, it was determined that the emotions displayed by the experiment group were significantly more believable than the two other groups.

The primary limitations discussed in this implementation of ACORE relate to the use of hand-coded values within the weights and resource changes of the NPCs which may have contaminated the results of the experiments. In addition, the lack of procedurally animated facial expressions meant that emotions were displayed with static images and phrases.

2.6 Statecharts

Dragert et al. discuss the need for designing AI for minor characters in a populated game world in order to facilitate player immersion. They propose using high-level statecharts representations to allow for modular generation of AI with minimal designer input [20][41].

Statecharts are based on discrete states and the transitions between them. They take a timed sequence of discrete events as an input and return another sequence as output. State transitions can happen based on both internal and external events. When an event occurs, the first matching transition is taken.

The modelling formalism used here divides AI into components and each component is modelled as a statechart. There are sensor components which learn about game state, and actuators which can alter the game state. These form the interface between the game world and the AI and are distinct in terms of functionality. A

hierarchy of components exists between these which process the events and pass them onwards. The hierarchy consists of:

- Analysers which analyse the events as received by the sensor and can create new events based on these
- Memorisers which keep track of the state that has been observed
- Strategic deciders which decide strategy for the actor based on memory and state
- Tactical deciders which decide how to pursue the chosen strategy
- Executors which translate tactical decisions into specific commands to be executed
- Coordinators which understand how the actuators relate and can adjust the commands further before they are executed, such as ensuring that the actor moves in a correct or natural way

In addition to presenting the use of statecharts to model minor character AI, Dragert et al. also discuss approaches to generating variations of these based on a designed AI.

The first and most basic of these is accomplished by varying parameter values such as movement speed, energy, health, and so on. However, in order to prevent incorrect behaviour, predefined value ranges are necessary to prevent invalid values from being chosen. Parameters can also have dependencies on other parameters which may necessitate the generation of dynamic ranges which take this into account.

The second method involves varying component configurations, such as removing, replacing, or adding components to the AI. Removing components could involve removing some sensory components, actuators or others, which could result in things such as extra characters who are blind, mute, and so on. Replacing components can be used to completely change reactions to certain events by dropping in a new component with different transitions and states. In the same way, adding new components can result in new perceptions and possible actions.

Finally, it is possible to vary component models to generate variations in AI. These arbitrary modifications to the structure of the AI would bring about the most drastic variations. These variations are modelled in the form of transformation rules.

2.7 Summary

Ultimately, there are many approaches to the design and implementation of systems for adaptive and interactive NPCs in video games, intended to mimic human behaviour and provide ‘realistic’ interactions. Of particular interest for this thesis, perhaps, are those which utilise psychological factors such as wants and needs, and the sections dealing with modular AI and statecharts.

AI behavioural modules act as a way to define personality traits and the actions associated with them. Versu’s ‘social practices’ provide an elegant solution to defining and delimiting possible interactions in any given state. Façade’s ‘beats’ act as an avenue to provide more author control over a given situation by allowing the

interjection of solid, planned plot elements. Needs and wants systems provide an NPC with reasons and motivations to act in ways that will eventually benefit them. Statecharts provide a modular solution by which NPCs can interact differently with their world.

Each of these aspects provide a groundwork for the development of procedurally generated narrative by way of simulating an environment inhabited by NPCs designed to interact organically with others in their surroundings. Their combination could provide interesting narratives by way of emergent behaviour and designed interactions.

3 Method

This chapter describes the scientific methods used to address the research questions presented in this thesis. It includes the purpose of these methods, their execution, their reliability and validity as a whole, and any ethical considerations that were required.

3.1 Scientific approach

In order to ascertain good strategies and approaches for implementing the algorithms, a systematic literature review was first conducted, exploring related work to a greater degree and evaluating pre-existing options. This was used in combination with the design and presentation of the algorithms themselves in order to answer the first research question.

Following this, an implementation of the algorithms was utilised to generate a series of logs detailing what individual NPCs in a small village did in the space of a single day, formatted as a narrative. This was distributed to individuals interested in video games, along with a qualitative survey containing a series of questions intended to address research questions two and three. A/B testing was used so that some individuals received the generated narrative, while others received a narrative tailored by a human writer. Simplistic language was used in both cases in order to ensure that no overt discrepancies in language existed between the two. The survey asked questions relating to how believable, immersive, and interesting the narrative they received was.

3.2 Literature review

The literature review included as a part of this thesis is primarily intended to serve as a primer to the state of the art in the procedural generation of NPC AI and interactions.

Compared to many other fields in procedural content generation, research into procedural narrative and AI interactions is not as widely researched or documented. Many papers also specifically discount AI and behaviour as a valid form of procedurally generated content [1], despite the fact that research has been conducted in this field.

Papers were selected based on relevance to the topic of the thesis using specific keywords, and focused on frameworks and methods used to generate interesting NPC interactions and believable emotion. Each paper is summarised in enough depth that its contents and implementation could be used as a facet in the design and implementation of the primary system that was produced as part of this thesis. The ultimate aim of the literature review was to discover and summarise methods that could be used as part of the implementation in order to generate interesting procedural narrative.

3.3 Prototype design and implementation

Following the completion of the literature review, individual aspects of preceding research were selected and integrated into a design with the intent of constructing a simple prototype system. The intent of this system was to provide an outlet within which to test the design constructed from theories and methods presented by the

papers summarised in the literature review and to provide output which could be used as part of the survey to judge their effectiveness.

The prototype was built to utilise wants and needs from within a selection of resources, coupled with a modular behaviour system, and actions based around the concept of statecharts. Each module was designed to represent an aspect of a personality, such as ‘timid’ or ‘aggressive’ and thus affect both the interpretation and resulting execution of an NPC based on input. The system was built to allow the specification of initial NPCs and a time period for which to simulate, resulting in interactions between the NPCs based on their individual personalities.

As it was not feasible to provide true graphical output of these interactions and the primary objective is the generation of procedural narrative, the system was designed to provide a basic textual output in the form of a series of actions and interactions by and between individual NPCs within the system.

3.4 Survey

In order to evaluate the effectiveness of the algorithms implemented in the prototype, a survey was distributed along with procedurally generated narratives and similarly formatted narratives constructed by a human writer.

Producing the procedural narratives involved defining a small group of NPCs in a constricted area with initially specified personalities. These were then simulated over a period of many years and the resulting actions over a brief period following this were catalogued and formatted into a simplistic narrative. In order to facilitate better immersion in the texts, a basic level of prose was applied to both the computer generated narrative and the pre-scripted narrative.

A/B testing was used in order to gauge the differing reactions of respondents based on whether the narrative they evaluated was written by a human, or generated using procedural algorithms. The survey was distributed amongst several video game developer communities, video game discussion communities, writing forums, and reading forums. These were selected as they most closely relate to the subject matter of the thesis, i.e. the procedural generation of narrative. Writers in particular were selected as they would be the most likely to notice flaws in the use of narrative techniques and might provide unique insights from their experiences. The survey was made publicly available for a period of three weeks during which time anyone could respond, and an A/B testing system was used in order to ensure that the url to the survey would evenly distribute the participants across both version of the survey. Both versions were identical in all aspects except the narrative presented. The duration for which the survey was available was set in order to ensure enough time to complete the prototype development and analyse the final results.

Participants were first presented with a short explanation of the survey and its purpose, followed by the narrative that they were asked to evaluate. After this, a series of questions was presented, most of which were designed to obtain quantitative data, and a smaller number of which were qualitative in nature. The contents and nature of each of these questions is explained in this subsection. All quantitative questions were marked as mandatory as this was the primary data required for the study, whereas qualitative answers were optional for respondents who might have further thoughts or opinions that could benefit or direct future studies.

Question one, as seen in Figure 3.1, is a quantitative question using a Likert scale. It was designed to assess whether or not the actions and emotions of the

To what degree would you say that the actions and emotions of the actors were in agreement within the context of the narrative?

*

For instance, were the actions of each of the actors believable given their apparent emotions or circumstances? Where such circumstances and emotions were not portrayed or justified, did they act in a believable way?

	1	2	3	4	5	
Not very believable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very believable

Figure 3.1: Survey Question 1

actors within the narrative were considered by participants to be in agreement with the context of the narrative as a whole, i.e. how believable they were. From this, it would be possible to gauge how cohesive and coherent the narrative was to a participant and whether or not actors acted in an appropriate manner.

Did you find the interactions between actors within the narrative to be relevant to the narrative given the context? *

Specifically, did actors behave in an appropriate manner in their interactions with other actors? Or did individual actors act in ways that reduced the quality of the narrative?

- ☐ Character interactions were appropriate
- ☐ Character interactions were random/irrelevant and reduced the quality of the narrative
- ☐ I'm not sure

Figure 3.2: Survey Question 2

In question two, seen in Figure 3.2, quantitative data was obtained using a multiple choice question. Its intent was to gauge whether or not actor interactions within the context of the narrative were relevant, i.e. whether they acted in ways that contributed to the story, or if they were nonsensical and unnecessary. This data was intended to discern the levels of coherence and cohesion within the narrative as perceived by the participant.

Did you find the actor personalities to be detrimental or beneficial to the narrative? *

Were any of the personalities inappropriate for the narrative given its context? Did the personalities make the narrative less interesting?

- ☐ Beneficial
- ☐ Detrimental
- ☐ I'm not sure

Figure 3.3: Survey Question 3

The third question, shown in Figure 3.3, is also quantitative. It was designed to evaluate whether or not personalities depicted in the narrative were beneficial to the narrative's purpose and flow. Actors with unusual or nonsensical personalities could potentially result in a break in immersion.

Did you find the developments to be feasible given the context of the narrative, its actors, and their actions? *

Would you say that the narrative made sense? Did developments in the narrative happen naturally given their context, or was there no logical thread?

☐ Feasible

☐ Unfeasible

☐ I'm not sure

Figure 3.4: Survey Question 4

Question 4, shown in Figure 3.4, is a quantitative question designed to evaluate the perceived narrative flow of the text. An infeasible development would imply that there is no logical thread in the narrative, essentially rendering it more a description of random events than a well-flowing story.

To what degree could you determine the motivations of the actors? *

	1	2	3	4	5	
Could not determine	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Could determine with ease

To what degree could you determine the emotions of the actors? *

	1	2	3	4	5	
Could not determine	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Could determine with ease

Figure 3.5: Survey Questions 5 & 6

Questions 5 and 6, as seen in Figure 3.5, are intended to determine the clarity of the narrative. They provide quantitative data through multiple choices. These questions were designed to provide an insight into the differences in perception of clarity between the two separate texts.

The penultimate quantitative question, as seen in Figure 3.6, was intended to gauge general participant interest in the narrative itself. Hooking a reader is often considered to be important in fiction and narrative writing, therefore it was of interest to assess how successful both texts were at doing so.

Question 8, the final quantitative question as seen in Figure 3.7, asks the participant to rate the likelihood that this text was written by a human. This question

How would you rate your interest in the expansion of this narrative? *

Would you continue reading if the narrative was expanded and extended e.g. as a full-length book or video game? Would you be interested in learning more about the town in question or its inhabitants?

	1	2	3	4	5	
Not interested at all	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very interested

Figure 3.6: Survey Question 7

How likely is this text to have been written by a human? *

	1	2	3	4	5	
Very unlikely	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very likely

Figure 3.7: Survey Question 8

exists to attempt to determine whether there is any statistically significant difference in the perception of these texts by the participants.

Was there a particular part of the text that was strange? If so, why?

Your answer

Was there a particular part of the text that was compelling? If so, why?

Your answer

Do you have any further comments?

Your answer

Figure 3.8: Survey Questions 9-11

The remaining questions relating to the narrative are qualitative in nature and are shown in Figure 3.8. These provided the participants with an opportunity to share any other thoughts on the two separate narratives to see if there were any common threads that would appear frequently in the responses.

At the end of the survey, basic information on demographics was requested (as seen in Figure 3.9), with the primary focus on gathering data on which forms of narrative media participants consumed.

How often do you spend time doing the following? *

	Every day	A few times a week	A few times a month	Irregularly	Very rarely	Never
Reading fiction	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Reading nonfiction	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Playing video games	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Watching films or TV series	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

What types of entertainment (not listed above) do you regularly enjoy?

Your answer

Figure 3.9: Survey Questions on Demographics

3.5 Reliability and Validity

In order to ensure the reliability and validity of this thesis, the survey was designed around both the research questions and the target audience for the narratives produced. To gather as wide a range of respondents as possible, the survey was distributed to multiple different groups, as described in Section 3.4 on different days across a lengthy period of time, and care was made to ensure that an equal number of respondents participated in both the A and B groups. Additionally, the survey was reviewed by peers before distribution to ensure its relevance.

As the prototype system itself is built on procedural generation, with a given input and defined seeded random source, the same output will be produced. This ensures that the prototype implementation itself is reliable in that it will always produce the same narratives under the same conditions to allow for repeatable production.

To ensure construct validity, all terms are carefully defined, both in this report and in the survey to ensure that no misunderstandings or misinterpretations can arise. Where possible, simple and clear terms were used in place of more complex alternatives to reduce this risk further.

Pairwise comparison was considered as an alternative to using A/B testing for evaluating the narrative texts, however this was ultimately rejected as it would require a much greater time investment from the respondent, potentially resulting in less reasoned selections. It would also provide far more room for the respondents to compare and contrast the texts in order to deduce which one is actually computer generated rather than responding in a natural way based on the individual texts themselves. In addition, to ensure that there was no bias in the selection of the randomly generated text, the first generated text from the completed prototype was selected.

The decision was made to dress up the computer generated text as opposed to formatting the human-written text as a simple list of actions since this would allow

respondents to focus more on the narrative content in a context that they would be familiar with and might be able to immerse them, as opposed to reflecting on a simple list of instructions. This, however, could potentially introduce a small amount of bias wherein the narrative text might be coloured by the individual performing the enhancement of the prose. This issue is mitigated by ensuring that both texts were written in a similar style of prose by the same author.

3.6 Ethical considerations

As this project makes use of surveys, one ethical consideration that must be taken into account is the confidentiality of user information. In order to ensure that no confidential information is published, the surveys will be anonymised and no personal information beyond basic demographics will be collected.

4 Implementation

The following sections detail the design and implementation of the prototype system that was developed to generate simple series of actions and interactions for transformation into a narrative for the survey. This system can roughly be split up into personality definition via modules, world state and representation, wants and needs, statechart evaluation, and the execution of the simulation itself. The prototype implementation was developed from the ground up using C# and .NET Core, and follows a modular architecture in which each aspect of the system can be expanded with further modules. Modules include the core simulation module, the core traits module, the core resources module, and the world state module.

4.1 Personality and Trait Modules

This section discusses the design and implementation of actor personalities through the use of modular traits. An overview of the system as a whole is provided, including a few examples of some key traits.

4.1.1 Overview

In order for interesting interactions to occur between actors in the simulation, personalities need to be defined [42]. These personalities are constructed from individual ‘trait’ modules (as seen in Figure 4.1) which act as the controllers for defining and deciding upon behaviour within any given situation. Each of these modules is purely behavioural, and judgements are made based on world state, actor memory, needs and wants, and a few other factors which are evaluated within the module to assign a score reflecting the value of executing the given behaviour. This results in actors that act differently depending on context, memory, and personality traits within the context of the current simulation.

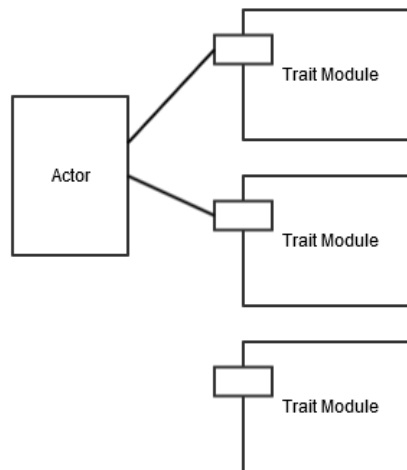


Figure 4.1: Representation of the modular nature of traits.

Individual traits can interact with each other organically (through indirect means such as two people who happen to participate in the same activity), or through combination by way of a system inspired by Versu’s ‘social practices’. Each trait defines a set of situations in which it might come into play, such as responding to

another actor's request for a resource, or defending oneself from violence, and so on. This is coupled with a combination principle which selects appropriate responses based on the action being performed and the actor's traits. Traits that have a social practice dictating a response in the context of the request are selected and their individual generated actions are weighted and evaluated based on factors such as the probability of success, outcome, wants and needs, actor memory, and so on. This allows for reasonable actions and reactions to be defined to some degree (preventing complete chaos), while still leaving room for the emergence of unexpected and interesting behaviour. An example equation for computing the weight of an action could be

$$weight = \frac{\max(0, (d - r))}{d} \cdot w_d \cdot w_t \cdot p$$

where r is the amount of a given resource currently held and d is the desired amount, multiplied by the predefined weighting (importance) for that particular resource w_d the predefined weighting for the trait itself w_t , and a probability value p which indicates the probability of success based on expected outcome and calculated using actor memory and world state.

For the sake of simplicity, these traits are also used to define an actor's job and role within the simulated society. By using traits such as 'farmer', an actor can be designed to perform critical roles within the simulation, producing resources or distributing them. This allows for another avenue of interaction, wherein there can be a bottleneck on acquisition.

The following subsections provide examples of key traits, along with their definitions, functionality, and role within the system.

4.1.2 Generous

The 'Generous' trait facilitates interactions between actors that usually involve the distribution of resources without benefit to one or more other actors. For some actors, this can fulfil a want or need through adjusting their resource values, including resources such as friendship, companionship, or karma.

This trait is usually activated through the use of social practices, such as if another actor approaches to petition the actor for a resource. Depending on the weighting of the trait and other factors of their simulated personality, a generous action can frequently result in an undesirable unbalance of the actor's resources, in some extreme cases, even its complete depletion.

4.1.3 Violent

Conflict is the backbone of any and all narrative writing or storytelling [43]. For this reason, many of the traits represented within the prototype implementation are those designed to give rise to conflict between individual actors.

As the 'Violent' trait is rather complex and multi-layered, the representation within the prototype implementation is simplified to be triggered only through the use of social practices, i.e. in response to actions taken by another actor towards the individual in question. This usually involves reacting to an action by another actor that has a negative effect on the balance of the individual's desired or needed resources. This could be reacting to a refused petition, interacting with a disliked actor, or defending one's self from another form of aggression.

4.1.4 Thief

Unlike the other key traits mentioned, the ‘Thief’ trait is similar to other job-defining traits in that it results in a role of sorts within the simulation. Depending on the weighting of the thief trait, actors might seek to pro-actively steal from others within the simulation, or may do so only in times of extreme resource imbalance. It can also be triggered as a social practice in the event of a failed petition.

When evaluating the trait for execution, memories of relationship are taken into account, along with likelihood of success based on known traits of the target, previous attempts, and other related factors. Desire and desperation (large imbalances in resources) also play a role in weighting the likelihood of the behaviour being run.

4.1.5 Lawful

The ‘Lawful’ trait is designed to act as a counterpart to traits such as the Thief trait, preventing them from wreaking too much havoc upon the simulation. When an individual commits a crime, this is recorded in the world state and, in the event of being noted by another actor, in that actor’s memory. When a crime occurs and has witnesses, this trait can evaluate options for solving said crime, including talking to witnesses, interrogating, and ultimately, arresting a criminal. The trait itself combines with many others through social practices depending on complementing traits both in the actor and any the actor interacts with. Potentially unique and interesting situations can emerge in instances where an actor is given both the Lawful trait and one of its opposites.

4.1.6 Social

Acting as the main avenue by which relationships can be improved or worsened, the ‘Social’ trait acts as the drive for actors to socialise on a basis which does not involve the exchange of material resources. Personal desires such as companionship and friendship naturally degrade with the progression of time and socialising with others helps to rebalance them.

An actor’s opinion of another can have wider effects on their behaviour as a whole. It can decide the likelihood of granting petitions, of protecting another during an interrogation, of standing up for another individual, or even of treating them poorly and making them a primary target. The ‘Social’ trait is not the only trait that involves modification of these opinion and reputation vectors, but it is one of the primary traits through which opinions are improved and friendships are formed.

4.1.7 Survival

The ‘Survival’ trait is not a trait in the same vein as the others. It exists as a representation of the base level needs of a human to survive, and as a fall-back in the event that no other actions can be performed to improve the actor’s resource balance in a satisfactory manner. Most other traits will invariably take precedence, but in dire circumstances of extreme imbalance with no other avenues through other traits, the Survival trait provides a way for actors to petition others for resources or seek out other potential avenues for survival.

Most actors will rarely evaluate this trait favourably when selecting a behaviour to execute. Traits such as the Thief trait, or other job traits that can produce or

acquire resources will most often take priority. In many cases, the actions of other actors may also result in the activation of this trait being unnecessary, such as a Generous individual providing the required missing resource.

4.2 World State

In order to facilitate interactions between actors, a world state is defined in order to keep track of all actions and interactions, the actors themselves, and the current state of the world. The prototype implementation of the world state is designed with simplicity in mind, where it could be far more comprehensive, such as in the case of Versu, where modal logic is used as its representation [17].

Modified world states to represent an individual's view of the world are not directly included as part of the prototype implementation, but are somewhat represented in each individual's memories of and opinions on various aspects. The world state itself also does not represent relationships, as these are relegated to the individual actors themselves and are accessible through the world state.

What the world state represents is the more fundamental actions and interactions that have taken place, in addition to the structures, the current date, and initial state present in the simulation. This includes aspects such as workplaces and who they belong to, what actions and interactions have been performed by actors, which crimes are publicly known and who solved them or apprehended the criminal, and what tasks are delegated to which individuals.

Each update of the system results in an update of the world state by way of actions, which are discussed in a later section. The world state itself is treated as a singleton by the system and is referred to frequently in order for actors to make decisions.

4.3 Wants and Needs

The wants and needs system forms the core backbone of an actor's behavioural decisions within the prototype implementation. Each individual actor is defined with a few different wants and needs. Each of these is defined as a vector consisting of a weight and a signed value indicating the amount of the abstract resource the actor wants or needs. These resources can be defined as anything, from friendship and companionship to more material desires such as wealth, food, or individual items such as clothing and tools. Some resources exist in a finite quantity and merely change hands, whereas some can be created or decay with time and interaction.

Any need or desire can be defined in one of two ways. Either the actor seeks to amass as much of the resource as possible with a minimum threshold that is more critical to obtain, or they can seek to balance it. In cases where the actor seeks to balance a given resource, a surplus can result in the actor choosing an action to pass on the resource to another, allow it to decay, or discard it entirely.

The main purpose of these wants and needs and their associated resources is to drive actor behaviour. Actors are designed to make choices in order to balance or maximise their given wants and needs as necessary. This is accomplished by having each action an actor can take have an effect on one or more of their current resource values. An actor is naturally more likely to take an action that will be beneficial for their resources, though this may not always be the case if methods to obtain different resources are conflicting.

Some of these wants and needs are universal. Food, for example, is both a decaying and produced resource as each individual must consume it on a daily basis while other individuals must farm in order to create and trade it. If an actor finds themselves completely without food, they will starve and ultimately perish. No actors are exempt from such universal resources and rules, leading to the need for some core behaviours to accomplish them as described in the previous section.

Examples of some of the resources present in the prototype system include food (as previously mentioned); money, a finite resource which only passes between actors; and friendship/companionship, a more abstract decaying and produced resource that results from casual interactions with other actors with a positive result and decays on a negative result or lack of interactions. It is simple to add additional resource types to the system for expansion.

4.4 Evaluation and Memory

In order to understand the method by which the prototype system evaluates and selects behaviour, the general structure of the actor's internal flow first needs to be discussed. The paradigm utilised is heavily based on that described by Dragert et al. in their paper on statecharts [20]. The entire flow can roughly be broken down into the following parts and is represented in Figure 4.2.

- The sensors, a group of modules designed to obtain data from the world around the actor. For the purpose of the prototype simulation, these are simplified to an understanding of the world data as a whole, whereas in other implementations they might include more detailed sensory data such as sight, hearing, or smell. The analyser is also folded into this grouping to work with the simplified sensory data.
- The memoriser, which keeps track of the state seen and the outcome of any actions.
- The strategist, which takes all this information and uses it to evaluate all trait module behaviours and make a final decision on which action would be most beneficial to the actor at this stage.
- The executor, which executes the final action generated.

In this simplified prototype implementation, the analyser for the sensory data is largely omitted as the world state is very straightforward and requires very little interpretation. The actuators are also omitted as individual actions act as state machines which control the required behaviours for their execution and no complex aspects (such as moving individual limbs, etc) are required for the purpose of this simulation.

Of these individual modules within the actor, the memoriser plays one of the more important roles in influencing the decisions made. For every action executed, 'memories' of this action are added in the form of changing opinions, knowledge of events, and so on, both for the actor initiating the action and any actors involved in the interaction. When an actor tries to evaluate a behaviour, it does so based on its knowledge and memories relating to the behaviour being evaluated. For example, an actor who has repeatedly been denied resources by another actor will be significantly less likely to attempt to approach them again unless other circumstances change.

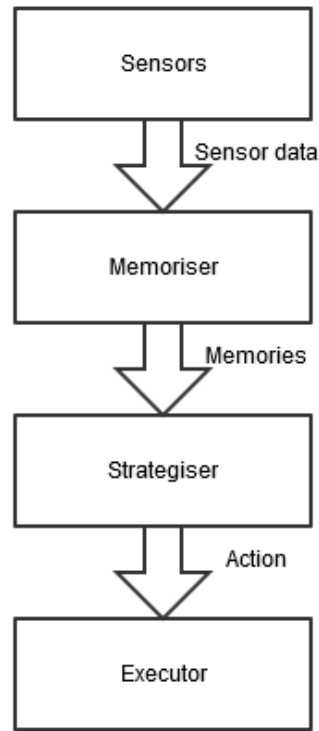


Figure 4.2: Representation of the actor's control flow.

The evaluation of each individual personality trait and its associated behaviour is a multi-part process. Each trait is evaluated in a different manner depending on its own internal logic and a final weighted score is generated that represents the perceived value of executing the behaviour to the actor at this point in time. This process is shown in Algorithm 1. In general, this score takes into account the weight of the trait overall in the personality specification, the effect it will have on the actors resources, the likelihood of success based on world state and memory, and the possible combinations this could have with other complimenting traits by way of social practices as previously described.

Result: An action to execute

let *actions* be a list of tuples, where each has a weighted score and an action attached;

foreach *Trait trait attached to the actor* **do**

 evaluate *trait* based on state and memory and produce a tuple of *score* and *action*;

 add the tuple of *score* and *action* to *actions*;

end

if *actions is empty or all scores are 0* **then**

 return an action the equates to doing nothing;

else

 select and return the highest weighted option from *actions*;

end

Algorithm 1: Pseudocode evaluation of traits and action selection.

Once each trait has been evaluated, the actor selects that with the highest weight, i.e. the action that has been determined to be the most beneficial at this point in time. In some circumstances, there are no possible behaviours the actor can execute resulting in the actor merely idling until a possibility presents itself.

As an example of the evaluation of a single trait, consider the Thief trait. First, the actor's resource values are evaluated and compared to any wants and needs. If no imbalances are found, the weighted score is significantly reduced and a resource is chosen at random from amongst those with the highest weights. Next, the world state and memories of the actor are used to determine targets that may possess the chosen resource. If none exist, the process repeats with another of the wants or needs. Each target is evaluated based on the actor's opinion of them and any prior experiences. The latter affects what the actor considers to be the likelihood that they will successfully steal from the individual and is factored into the weighted score. In some circumstances, no suitable targets or resources are found, in which case, the behaviour is scored with a zero and no associated action is produced. A pseudocode representation of this process is depicted in Algorithm 2.

Result: A weighted score for the trait

```
while Resources remain to be evaluated and no resource has been selected
do
  foreach Resource resource possessed by the actor do
    if resource is imbalanced and not evaluated in this execution then
      select resource;
      break out of the loop;
    end
  end
  if no imbalances then
    select a want or need that has not been evaluated and reduce the
    final weighted score;
  end
  set chosenLikelihood to 0;
  set chosenActor to null;
  foreach Actor actor other than this actor do
    if actor is known to possess the resource then
      if actor is liked and the imbalance is not too severe then
        continue to next loop;
      end
      evaluate the likelihood of success;
      if likelihood > chosenLikelihood then
        set chosenLikelihood to likelihood;
        set chosenActor to actor;
      end
    end
  end
  if chosenActor is not null then
    generate an action;
    return a tuple of weighted score based on chosenLikelihood and
    the action;
  end
  clear the resource selection and mark it as evaluated;
end
```

Algorithm 2: Pseudocode evaluation of the Thief trait.

4.5 Actions as Statecharts

Once the actor has decided on a course of action based on their personality traits, memory, and the evaluation of utility and benefit, they must then execute this action in some manner in order to effect the world state and interact with other actors. In the prototype implementation, this is handled by using ‘actions’, where each action defines a series of instructions and potential outcomes for the actor to follow during their period of activity. Many such actions are defined for a variety of different tasks, such as simply moving from one place to another, conversing with another actor, purchasing resources, or even attacking or injuring their target.

These actions can range in complexity from very simplistic (moving from one point to another), to moderately complex (engaging in conflict). As previously mentioned, conflict is an important aspect of any narrative since it is a primary factor in driving interest and immersing the consumer. For this reason, actions were designed in such a way that they could be chained to create a series of more complex actions based on the responses and reactions of actors other than the one performing the execution. At its simplest level, this chaining can involve linking the actions ‘go to target’ and ‘converse’, both of which are created and changed immediately upon the evaluation of the personality trait triggering them. More complex interactions can arise with conflict, an example of this being an individual who is about to be arrested resisting arrest and injuring their would-be captor. The reaction of the actor being captured is a chained action produced by way of the previously discussed ‘social practices’ and can, in turn, be chained even further.

Each action is represented and executed as a simplistic statechart. While the statecharts themselves could potentially be very complex and contain many states, transitions, and outcomes, since the resulting narratives in the prototype implementation are purely textual, simplicity was preferred since most states and transitions would not be directly observable. Transitions are run based primarily on actor interaction and the responses received, in a largely turn-based manner where each actor executes their current node in sequence or as required.

An example of one of these statecharts can be seen in Figure 4.3. This depicts a chain of actions that could result from the Thief behaviour. The actor first travels to the target’s position. If the target is not present, the actor idles until they appear or a timeout occurs. If the target is present, the actor will attempt to steal. At this point, the actor is either not caught and leaves, terminating the action, or the actor is spotted and will either flee or trigger a new action in the chain.

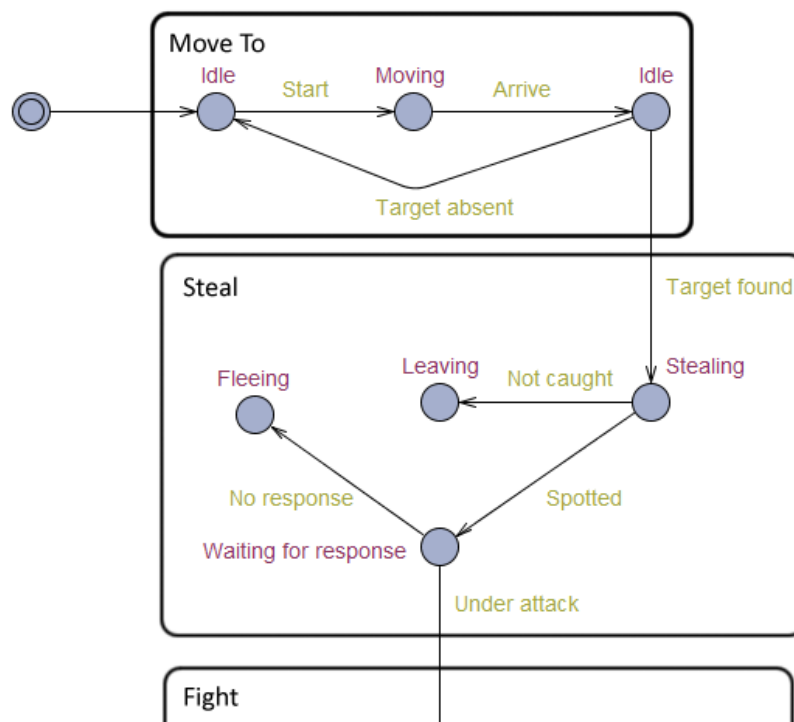


Figure 4.3: A chain of actions resulting from the Thief behaviour.

4.6 Simulation Execution

The final aspect of the prototype implementation involves the execution of the simulation as a whole. This is perhaps the simplest part of the entire system in terms of functionality as it mostly consists of a game loop designed to update logic at regular intervals and run the simulation over a predefined period of time.

As the current prototype implementation is designed to provide purely textual output, the updating of actors has been simplified to some degree into a turn-based system, wherein each actor takes an allotted turn to plan and execute their actions. The ordering of turns is based on an ‘intuition’ stat, decided at the same time as personality traits, wants, and needs. For this reason, actors often act in a fixed order, though the interactions can adjust this to some degree by involving other actors in actions outside their normal periods of activity.

The simulation itself is run for a simulated period of time as defined at initialisation time. This is measured in terms of in-simulation days, and can be specified for days, weeks, or years. In the even that all actors within the simulation perish, it can also terminate early as there is no longer any potential for interaction. Once the simulation has completed, a full list of all events that took place over the course of the run are recorded to a textual logfile which can be interpreted, either by another system, or by the initiator.

To run the simulation, an initial set of actors must be defined. These can either be hand-crafted by the individual running the simulation, or they can be generated entirely at random using the pool of traits and resources to assign varying personalities, wants, and needs.

5 Evaluation

This chapter contains the results of the A/B survey distributed and the analysis of these results. A total of 48 surveys were completed, with an exactly even distribution across both the A and B versions of the text.

5.1 Results

The following section shows the collected results of the surveys with average values shown for each Likert scale question. A total of 48 responses were collected, 24 on the A version of the text, and 24 on the B version of the text. The survey was distributed to several communities interested in video games and reading or writing, and was made publicly available for a period of three weeks. In order to ensure that an even distribution of responses was obtained across the two versions of the survey, a web-based system was used wherein clicking on the link would direct the participant to each survey with the same frequency based on the total completed number of surveys. As such, an even balance was struck across both versions. As the survey was posted publicly, it is not possible to deduce an exact number of individuals to whom it was presented.

Within the two versions of the survey, Text A consisted of a narrative written by a human, whereas Text B was a dressed up version of a narrative created through the execution of the procedural algorithms and prototype simulation. Both used similar themes and prose, and are included in Appendix A.

Table 5.1 shows the average values selected by participants for the degree to which the actions and emotions were in agreement with the context of the narrative in terms of believability. The question used a 1-5 Likert scale, with 1 being not very believable, and 5 being very believable. These values are also represented in Figure 5.1.

Actions and emotions in agreement with context	
Text A	3.250
Text B	3.208

Table 5.1: Average values for the degree to which participants found actions and emotions to be in agreement with the context of the narrative.

Tables 5.2 and 5.3 show the average values selected by participants for the degree to which they could discern actor motivations and emotions. The questions used a 1-5 Likert scale with 1 being unable to determine and 5 being easy to determine. Graph of these values are also shown in Figure 5.2 and Figure 5.3.

Clarity of actor motivations	
Text A	3.125
Text B	3.208

Table 5.2: Average values for the degree to which participants could determine actor motivations.

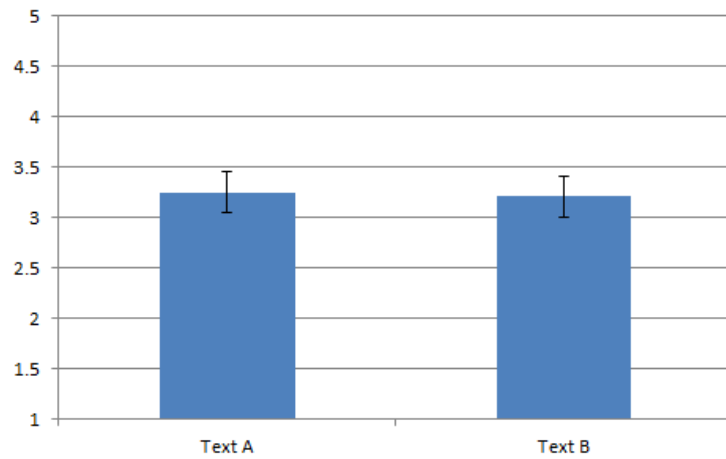


Figure 5.1: Average values for degree to which actions and emotions were in agreement with context, depicted with confidence intervals.

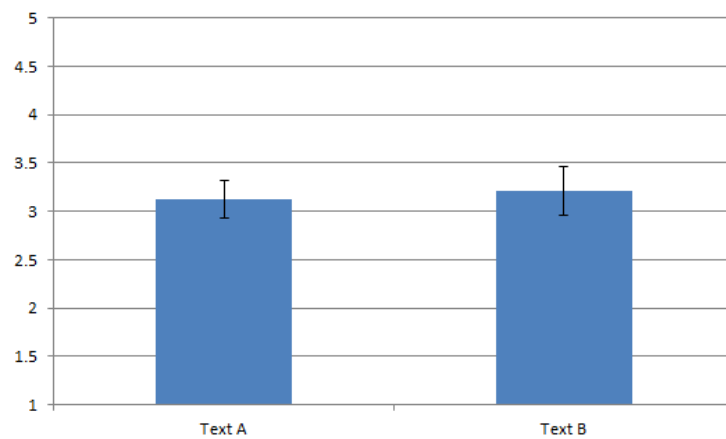


Figure 5.2: Average values for clarity of actor motivation, depicted with confidence intervals.

Table 5.4 shows the average values for respondents' interest in seeing the narrative expanded into a larger work, such as a longer story. This question used a 1-5 Likert scale where 1 was not interested at all, and 5 was very interested. These results are also depicted in graph form in Figure 5.4.

The results of the last of the Likert scale questions are shown in Table 5.5. This 1-5 Likert scale question set 1 as very unlikely, and 5 as very likely. A graph version of the results is shown in Figure 5.5.

Respondents were also asked to evaluate whether the interactions between actors were relevant given the context of the narrative, whether their personalities were beneficial to the narrative, and whether the developments in the narrative seemed feasible given the context. These questions used simple multiple choice questions, where respondents could answer positively, negatively, or state that they were unsure.

Clarity of actor emotions	
Text A	3.583
Text B	3.042

Table 5.3: Average values for the degree to which participants could determine actor emotions.

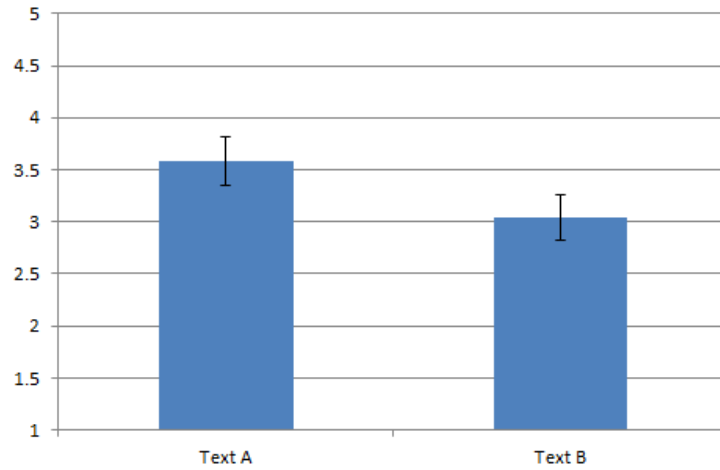


Figure 5.3: Average values for clarity of actor emotions, depicted with confidence intervals.

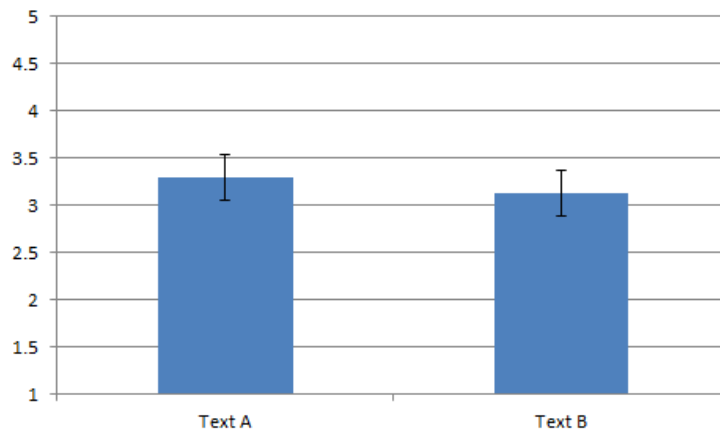


Figure 5.4: Average values for interest in narrative expansion, depicted with confidence intervals.

Table 5.6 shows the responses to whether or not the interactions between actors were relevant given the context of the narrative. These values are also depicted in Figure 5.6. Table 5.7 and Figure 5.7 show the responses to whether the personalities of actors were seen as beneficial. Table 5.8 shows the responses as to how feasible the developments within the narrative were and these values are also depicted in Figure 5.8.

Interest in narrative expansion	
Text A	3.292
Text B	3.125

Table 5.4: Average values for interest in the expansion of the narrative.

Likelihood of being written by a human	
Text A	2.750
Text B	3.333

Table 5.5: Average values for the likelihood the text was written by a human.

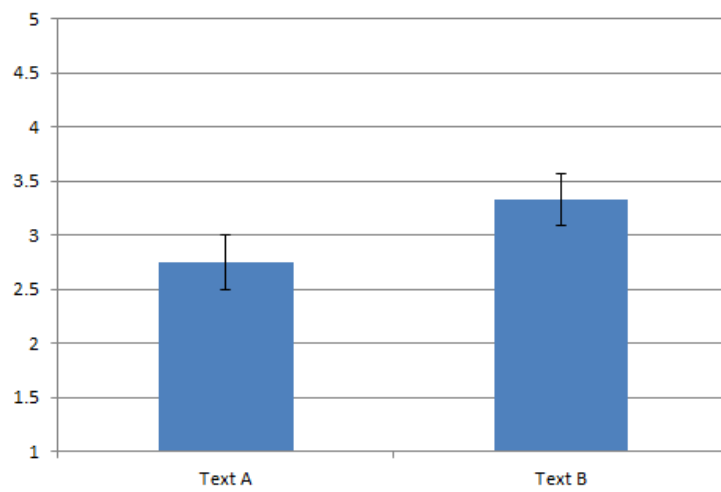


Figure 5.5: Average values for the likelihood the text was written by a human, depicted with confidence intervals.

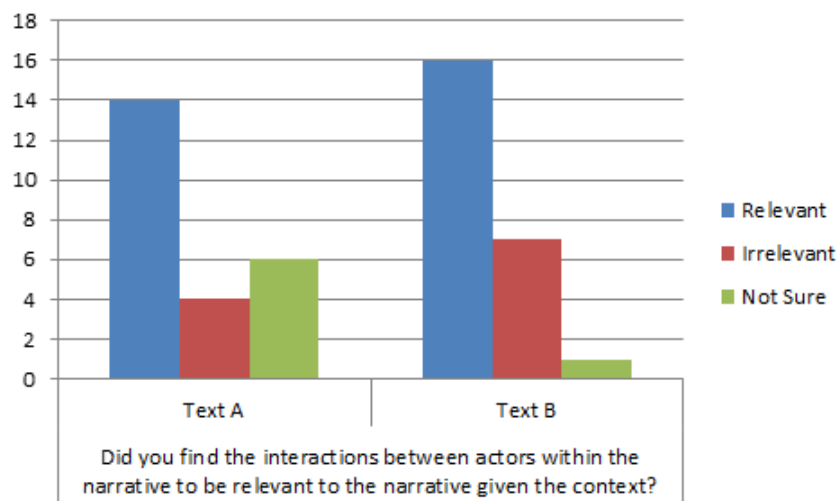


Figure 5.6: The relevance of actor interactions given narrative context.

	Text A	Text B
Relevant	14	16
Irrelevant	4	7
Not sure	6	1

Table 5.6: Responses to the relevance of actor interactions given narrative context.

	Text A	Text B
Beneficial	13	11
Detrimental	1	2
Not sure	10	11

Table 5.7: Responses to whether or not actor personalities were beneficial or detrimental to the narrative.

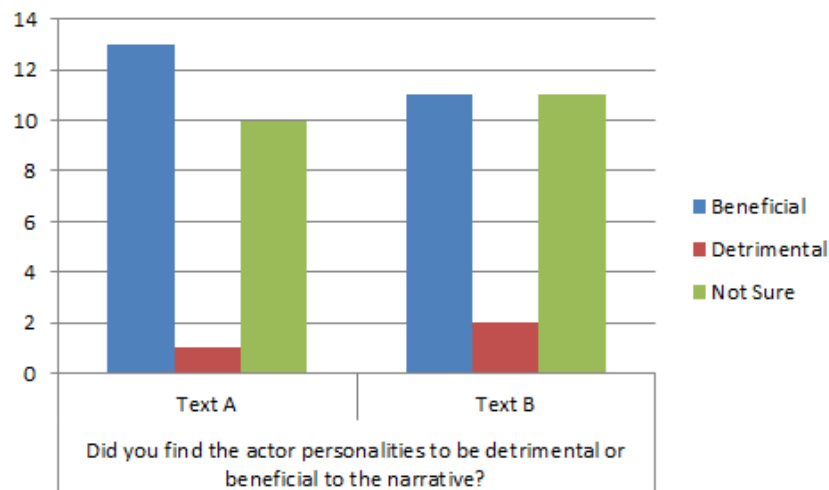


Figure 5.7: How beneficial actor personalities were to the narrative.

	Text A	Text B
Feasible	14	17
Infeasible	9	5
Not sure	1	2

Table 5.8: Responses to whether or not the narrative developments were feasible.

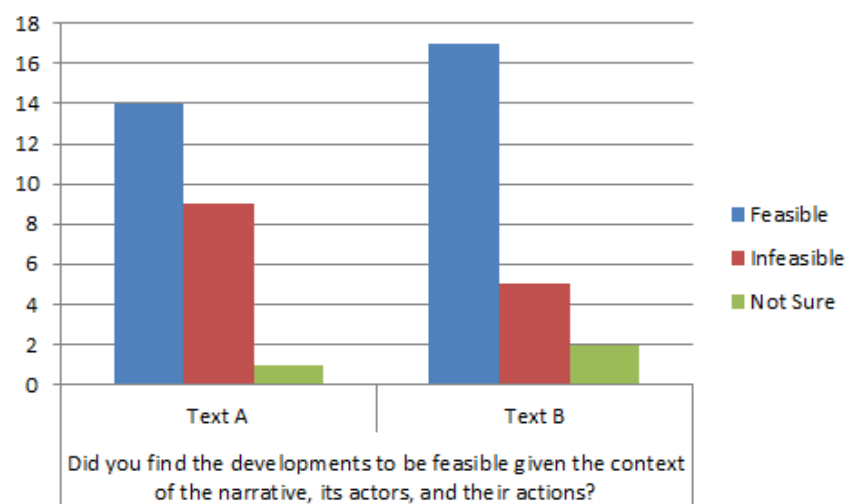


Figure 5.8: How feasible the narrative developments were.

The final section of the survey was used to collect basic demographic data about the respondents' consumption of narrative media. Tables 5.9 and 5.10 show how often respondents read fiction and non-fiction, play video games, and watch films or TV series. This data is also shown in Graphs 5.9, 5.10, 5.11, and 5.12.

	Reading fiction		Reading non-fiction	
	Text A	Text B	Text A	Text B
Every day	8	3	9	7
A few times a week	11	11	5	9
A few times a month	2	2	6	4
Irregularly	3	3	3	2
Very rarely	0	4	1	0
Never	0	1	0	2

Table 5.9: How often respondents read fiction or non-fiction.

	Playing video games		Watching films/TV series	
	Text A	Text B	Text A	Text B
Every day	8	7	13	8
A few times a week	5	7	6	12
A few times a month	2	3	3	3
Irregularly	4	2	2	1
Very rarely	2	3	0	0
Never	3	2	0	0

Table 5.10: How often respondents play video games or watch films/TV series.

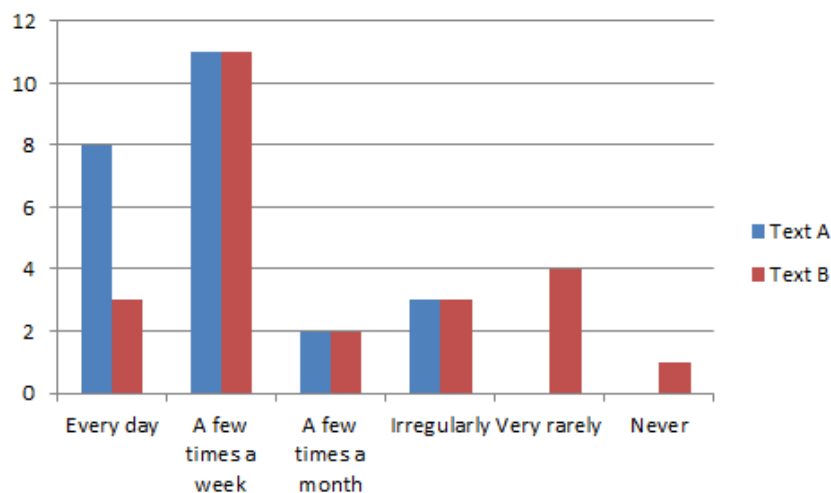


Figure 5.9: How often respondents read fiction.

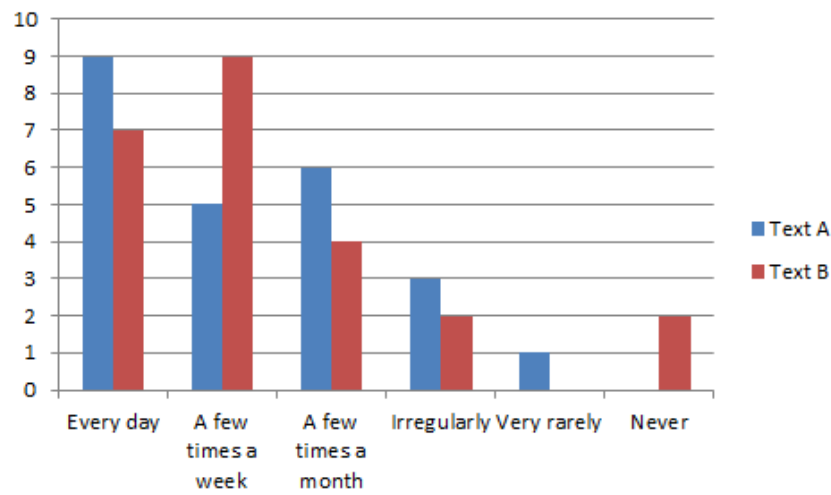


Figure 5.10: How often respondents read non-fiction.

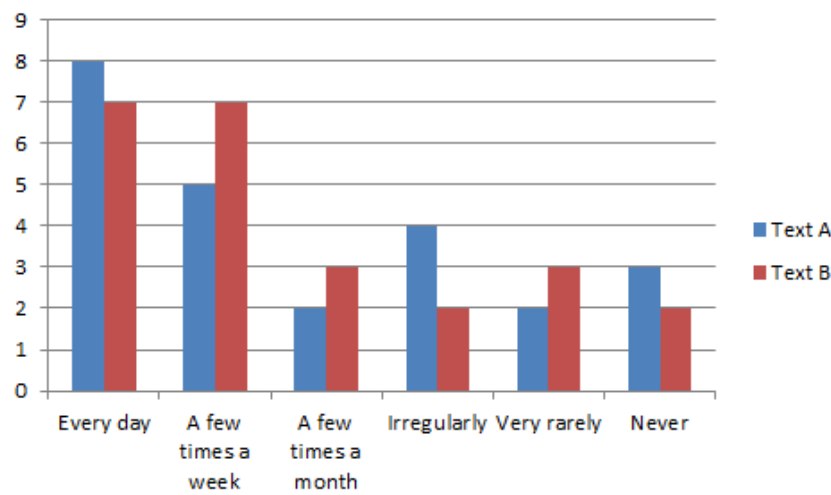


Figure 5.11: How often respondents play video games.

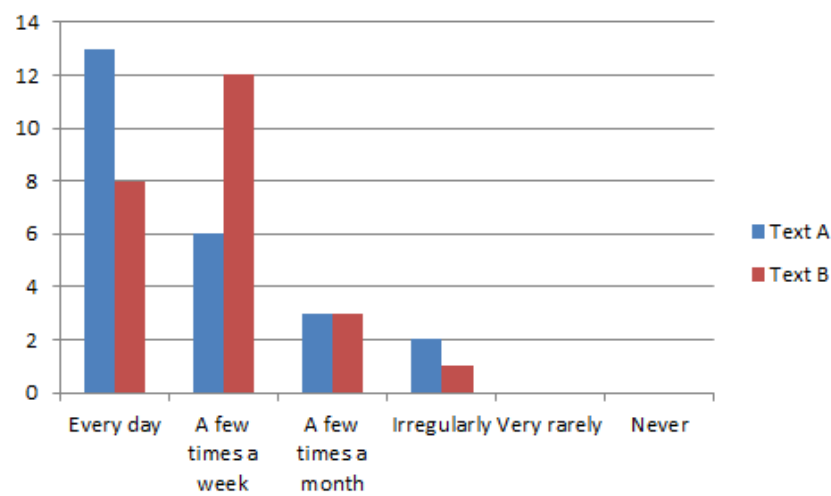


Figure 5.12: How often respondents watch films or TV series.

5.2 Analysis

This chapter presents an analysis of the results gathered from the survey. The data is used to find insights as to how the two separate texts (included in Appendix A) were evaluated by potential consumers, and how these reactions differ between a text written by a human, and a text generated via the procedural methods outlined in this report. Particular focus is placed on how these texts are perceived in terms of clarity of emotions and motivations, relevance of interactions, feasibility of narrative flow, and fit of actor personalities. Interest and thoughts on whether or not the text was produced by a human are also examined.

Statistical testing is used where appropriate in order to try and locate statistically significant differences between reactions to the two texts, thus suggesting whether or not the procedural algorithms were successful in producing something that could be considered close to what a human might produce. The Wilcoxon Rank-Sum test was used to evaluate statistical significance where Likert scales were used, as this data is not usually normally distributed and the Wilcoxon Rank-Sum test is nonparametric in nature. This allows for testing if there is a significant difference between the two sets of data, using the assumption that the alternative hypothesis is that the distributions are not equal. Each of the Wilcoxon Rank-Sum tests were performed as one-tailed tests with a cut-off of 0.05, since the intent was to discover if either of the texts outperformed the other in any given situation.

Dataset	N	Mean	Stdev
Text A	24	3.12	0.947
Text B	24	3.21	1.250
P-value		0.394	

Table 5.11: Results of the Wilcoxon Rank-Sum test on the degree to which actor motivations could be determined.

From both the results and analysis of the degree to which actor motivations could be determined by the respondent, it can be seen that there is no significant difference between the two texts with a P-value of 0.394, as shown in Table 5.11. This suggests that in both cases, respondents found the motivations of actors to be relatively clear, but not very clear, as the averages are roughly around the middle of the scale.

Dataset	N	Mean	Stdev
Text A	24	3.58	1.139
Text B	24	3.04	1.083
P-value		0.039	

Table 5.12: Results of the Wilcoxon Rank-Sum test on the degree to which actor emotions could be determined.

The results and analysis of the degree to which actor emotions could be determined, however, do show a significant difference between the two datasets, with a P-value of 0.039 as shown in Table 5.12. This means that the respondents who read

the human-written text found actor emotions to be significantly clearer and more understandable than those who read the computer generated text.

Based on the analysis of statistical significance shown in Table 5.13, there was also no significant difference in respondents' interest in seeing the narrative expanded into a longer story. In both cases, respondents were moderately interested in seeing an expansion, falling roughly into the middle of the Likert scale.

Dataset	N	Mean	Stdev
Text A	24	3.29	1.197
Text B	24	3.12	1.191
P-value		0.344	

Table 5.13: Results of the Wilcoxon Rank-Sum test on interest in the expansion of the narrative.

Dataset	N	Mean	Stdev
Text A	24	2.75	1.294
Text B	24	3.33	1.204
P-value		0.067	

Table 5.14: Results of the Wilcoxon Rank-Sum test on how likely it is that the narrative was written by a human.

There was also no statistically significant difference in the evaluation of the likelihood that the text was written by a human. The P-value, as shown in Table 5.14, was close to the cut-off at 0.067, with respondents slightly more likely to believe that the computer generate text was written by a human than the text that was written by a human, though not to a statistically significant degree.

Pearson's Chi-squared test was used in order to look for statistically significant differences in the relevance of actor actions, the benefit of actor personalities to the narrative, and the feasibility of the narrative developments. This test was chosen as the data in question is categorical and the observations are independent. Specifically, the Chi-squared test of independence was used, with the null hypothesis being that no differences will be observed between the human-written and computer generated text, and the alternative hypothesis being that significant differences will be observed between the two texts. In each of the following cases, the calculation was made with two degrees of freedom.

In Table 5.15, it can be seen that no statistically significant difference was found between the individual texts and the relevance of actor interactions given narrative context, and so the null hypothesis cannot be rejected. In both cases, the majority of respondents indicated that the actor interactions were relevant.

Table 5.16 similarly does not allow the rejection of the null hypothesis, suggesting that respondents viewed the benefit of individual actor personalities in a similar manner in both texts. The majority of respondents receiving the human-written text found the actor personalities to be beneficial, though a similar number were unsure. 11 respondents to the computer generated narrative were unsure about the benefit of the personalities and 11 considered them to be beneficial, resulting in an equal proportion.

Observed	Expected	Difference squared	Chi-squared value
14	15	1	0.07
16	15	1	0.07
4	5.5	2.25	0.41
7	5.5	2.25	0.41
6	3.5	6.25	1.79
1	3.5	6.25	1.79
Chi-squared value			4.5229
P-value			0.104

Table 5.15: Results of the Pearson Chi-squared test of independence on the relevance of actor interactions given the context of the narrative.

Observed	Expected	Difference squared	Chi-squared value
13	12	1	0.08
11	12	1	0.08
1	1.5	0.25	0.17
2	1.5	0.25	0.17
10	10.5	0.25	0.02
11	10.5	0.25	0.02
Chi-squared value			0.5476
P-value			0.76

Table 5.16: Results of the Pearson Chi-squared test of independence on whether or not actor personalities were beneficial or detrimental to the narrative.

From Table 5.17, it is apparent that there is also no significance between the choice of texts and respondent perception of whether or not the narrative developments were feasible and the null hypothesis still holds. In both instances, the majority of respondents found the developments to be feasible.

Observed	Expected	Difference squared	Chi-squared value
14	15.5	2.25	0.15
17	15.5	2.25	0.15
9	7	4	0.57
5	8	4	0.57
1	1.5	0.25	0.17
2	1.5	0.25	0.17
Chi-squared value			1.7665
P-value			0.413

Table 5.17: Results of the Pearson Chi-squared test of independence on whether or not the developments in the narrative were feasible.

Table 5.18 shows that there was also no statistically significant difference in consumption of narrative media between the groups two whom the separate A and B tests were distributed. In each case, the calculations were performed with five degrees of freedom, except when calculating the significance of how frequently

respondents watched films or TV series. In this instance, no respondents selected very rarely or never, and thus those two options were discounted in the calculation.

Type	Chi-squared value	df	P-value
Fiction	7.2727	5	0.201
Non-fiction	4.9929	5	0.417
Video games	1.6667	5	0.893
Films/TV series	3.5238	3	0.318

Table 5.18: Results of the Pearson Chi-squared test of independence on the frequency with which respondents consumed certain types of narrative media.

5.3 Discussion

Based on the results of the survey and their analysis using statistical testing, it is difficult to be sure just how differently the two texts were perceived by survey respondents. Since no statistically significant difference was found between the two test groups in terms of media consumption and the null hypothesis was shown to hold, it could potentially be said that a similar distribution of respondents was achieved in both cases, meaning that no one test group was skewed to a particular type of media, or its specific style of narrative.

On the whole, most of the results of the Likert scale questions tended to hover around the mid-range, with the averages being slightly above or slightly below the middle mark. Other than where clarity of character emotion was involved, there were no significant differences between respondent perception of motivation, interest in the expansion of the narrative, and whether or not the narrative was written by a human or computer generated. This does not provide a great deal of clarity, but seems to suggest that the respondents feel similarly about both texts, indicating that beyond clarity of emotion, the prototype implementation is able to generate feasibly human narratives.

It should not be discounted, however, that the computer generated text was dressed up. This could potentially be a factor in the similarity in scores for how likely the text was to have been written by a human, though the averages were not high in either case. It is possible that using simple bullet point narratives instead, while less likely to immerse the reader, would have a similar outcome.

Despite the original hypothesis stating that respondents would more easily identify emotions and motivations, respondents seem to have found motivation similarly clear in both texts, suggesting that the prototype implementation is capable of generating stories with motivations that are as clear as a human-written narrative. This is reasonable given that the actors in the prototype implementation essentially act with motivation based on internal resources as their primary driving force. Emotions, however, were significantly clearer in the human-written text as opposed to the computer generated text, suggesting that there is still a way to go for the prototype implementation to be able to mimic emotionally feasible actions and interactions between actors, as originally expected.

One reason why the computer generated narrative may have lacked clarity of emotion and sees much uncertainty in terms of actor personalities could be related to the hardcoded values for each of these traits, as discussed in relation to the ACORE

system [39]. The use of values and weights that do not dynamically adapt to the situation at hand, or suppress actions that may not be appropriate in the given situation, as mentioned by Evans in discussions on Versu [31] does not allow for the concept of public mood [33]. This leads to actors not always acting in a believable manner given the circumstances as they will still prioritise their own personal wants and needs in a vacuum, with only a limited amount of effect from the general world state and history. While they will react to the events and it may have a bearing on the likelihood that they will choose a given action, it will not outright forbid events or actions that would be seen as severely inappropriate in reality.

It is also possible that actor personalities were not adequately discernible in either of the text due to the limitations on length of the narrative piece. ‘Show, don’t tell’ is an extremely common teaching within the field of narrative design [44], which stipulates that a reader or viewer should be shown the facets of an actor (or any other part of the narrative), rather than told outright. As these narratives were very short and the number of actors introduced was large, there was little room in which to show, or expand upon, particular actor personalities, potentially resulting in personalities that were difficult to discern, or did not come through in the text itself.

In each case, actor interactions were considered largely relevant to both texts, and the narrative flow was considered to be feasible in both cases. As there were no significant differences between the responses to the two texts, this could suggest that the prototype implementation is capable of creating immersive and well-structured narratives that are similar to those written by a human, if only through logical flow of events and interactions.

While most of the qualitative data gathered indicates areas of confusion for respondents in both cases, perhaps due to the limited length of the narratives in question and a lack of space in which to resolve all questions, one important aspect of narrative is discussed only in the comments of the human-written narrative and is not mentioned in responses to the computer generated narrative. Foreshadowing is the act by which a reader’s inference of the story is manipulated by the writer [45] and significant events hint at things that are yet to come. Several respondents to the survey indicated that the foreshadowing in the human-written text was a particularly compelling part of the narrative. This is an aspect that is not represented at all in the prototype implementation, as a simulation of interactions between actors merely follows a logical thread towards the future and lacks many of the capabilities required to create these kinds of literary devices.

The prototype implementation itself shares many similarities with the related work discussed in the literature review, but also many differences. Social practices are implemented in a manner similar to Versu [31] to ensure that actors behave appropriately given the situation. While the current system is not designed to involve a player in quite the same way as Versu, it provides more concrete motivations in the form of resource vectors and wants and needs. The use of statecharts for action execution also facilitates more interaction with the world itself. Though the prototype implementation does not quite match the levels of emotional depth presented in the EEP model [18], it does provide motivations through the use of wants and needs and is more capable of enabling interaction between actors. While CORE [39] and ACORE [19] involve the use of wants and needs with resources in a similar manner to the prototype implementation, they lack depth in interaction and properly defined personalities. Statecharts are used in a similar manner to the work

presented by Dragert et al. [41], but with the further addition of more configurable personalities and resource vectors to act as motivations for the statechart generation. Furthermore, none of the aforementioned works make use of procedural generation or produce narratives, which is the core focus of this study.

6 Conclusion

In this thesis, a detailed literature review was conducted in order to summarise all relevant literature on the implementation of emotionally interesting non-player characters and how this could be applied to creating believable interactions between them. A prototype implementation was developed using several of the specific techniques described in order to simulate these actors within the confines of a village, and the results of this simulation were formatted into a simple text-based narrative with dressed up prose. An A/B test was conducted wherein participants in a survey were asked to answer questions based on two separate texts, one which was generated by this prototype implementation, and one written by a human. Respondents were largely uncertain as to whether or not their given text was written by a human, resulting in similar results for both the A and the B groups.

RQ1: How can a modular approach to personality and behaviour be used to produce emotionally interesting NPCs with minimal human influence?

This is discussed in detail in Chapter 4, where the implementation of the prototype simulation is detailed in full. Personality and trait modules were utilised to provide behaviours based upon individual personality traits, wants and needs were used to provide a motivation to execute individual behaviours, world state and memory provided a means by which to evaluate the feasibility of these actions in optimising the actor resources, and statecharts handled the execution of individual aspects of each action and their effects on the world.

RQ2: To what degree can readers discern whether or not motivations and interactions between NPCs are pre-scripted or generated based on procedural algorithms?

From the results and their analysis, no statistical significance could be found to suggest that participants in the survey were able to discern which text was generated by a computer and which was not. In both instances, participants rated the likelihood of the text being written by a human roughly the same, though the computer generated text was considered slightly more likely to have been written by a human. In both instances, the motivations and interactions were considered to be beneficial and believable with roughly the same frequency. Only in terms of clarity of emotion was there any significant difference, where respondents found the emotion to be clearer in the human written text. Ultimately, readers were largely unable to discern a difference between the narrative generated by the prototype implementation, and the narrative written by a human, except where emotion was involved.

RQ3: To what degree is it possible to immerse readers in a narrative using only procedurally generated characters and their interactions?

Based on the results and analysis presented, readers seemed to experience a similar level of immersion in the procedurally generated text as they did in the human-written text. As previously mentioned, clarity of emotion was the only significant deviation where the human-written text was seen as superior. On the whole, respondents were slightly more interested in seeing an expansion of the human-written text, but not to a statistically significant degree, and on the whole, the narrative flow and feasibility of both texts was rated roughly the same. This suggests that it is definitely possible to immerse readers in a narrative using only procedurally generated characters and their interactions.

Some of these results could most likely be applied to the wider industry, not only where procedural narrative generation is involved, but also for the development of

emotionally interesting non-player characters in other respects. In particular, the results point a potential flaw that could be further addressed and is likely to affect other similar systems. When actors are generated and interact purely on motivation and resources, it does not always guarantee the illusion of emotion. More work may be needed to clarify and facilitate such interactions. The use of modular personality and traits is something that could be beneficial to any system in which the simulation of interesting non-player characters is required.

Though the initial results have been promising, it is possible that they could have turned out differently based on a number of factors. The distribution of the survey could have been handled differently, such as only distributed to players of certain types of video games, or only to those with detailed knowledge in analysing narrative texts. This would likely have resulted in more clear selections and slightly less vague results in terms of the value of the two narrative texts. It could also have been beneficial to distribute to a more 'casual' audience that might provide a more neutral ground for responses. A larger number of respondents would help to clarify the data and ensure more statistically significant results to better justify the evaluations and conclusions, as the number surveyed in this study was only high enough to produce one statistically significant result.

Additionally, the way in which the texts were presented could have been handled differently and potentially resulted in a different perception by the participants. If the text was instead presented as a bullet-point list of actions, readers may have been less inclined to consider the texts as being written by human and thus rated them purely on the narrative content, but also may have found themselves less immersed or interested in them in general.

Providing a visual representation of the actions might also have provided clearer results from a video game perspective, as respondents would be able to experience the interactions from the perspective of its inclusion in a video game. This would also allow for more robust representation of personalities (an aspect many found to be unclear) through body language and motion. However, as this research was meant specifically for the procedural generation of narrative as a whole, this would also have been severely limiting as to its potential in that regard and restrict it only to relevance within video games.

7 Future Work

There are many ways in which this project could be expanded further with future work, in order to create a more robust and powerful system for the procedural generation of narrative and for the development of emotionally interesting non-player characters for inclusion in video games. One of the major questions raised during the implementation phase of the prototype is how best to evaluate segments of the system's output in order to find the most compelling sections to comprise the narrative. Most of the interactions occurring in the system are relatively benign and do little more than advance the daily lives of the actors. Some form of evaluation system designed to search out and find segments with particularly compelling conflict would help to ensure that produced narratives are all interesting and sufficiently concise.

It would also be pertinent to conduct similar surveys with a larger and more varied audience in order to more accurately evaluate the performance of the system. This could also be conducted using different methods of representation in order to address other potential ways in which the results could have been affected. An implementation within a video game would be a useful method to further evaluate the performance of the prototype, perhaps as a simulation of a small village within a larger open world massively-multiplayer online role playing game. This would allow individuals to experience the simulation and evaluate its capability for immersion directly.

Other ways in which the system could be expanded would be to include more traits, wants, and needs, to allow for an even more diverse pool from which to generate the actors themselves, resulting in more intricate and complicated behaviours. The addition of dynamic weighting to individual resources rather than hardcoded initial variables per actor might also result in more natural behaviours as actor personalities and goals evolve over time.

Divergent world views using multiple copies of the world state could also be an avenue to pursue to provide actors with more noticeable beliefs and emotions based on how they perceive the world as opposed to how it might truly exist. A system similar to *Faade's* 'beats' [16] could also be used to allow for more author control over the narrative as a whole, by providing individual scripted sections towards which the procedural segments advance.

Finally, it would likely be beneficial to the procedurally generated narratives as a whole to develop a method to inject foreshadowing into the narrative, as this is one of the areas in which it was most lacking. In its current state, the system can only produce foreshadowing through coincidence, despite the importance of the literary device to many types of narrative storytelling.

References

- [1] J. Togelius, E. Kastbjerg, D. Schedl, and G. N. Yannakakis, “What is procedural content generation?: Mario on the borderline,” in *Proceedings of the 2Nd International Workshop on Procedural Content Generation in Games*, ser. PCGames ’11. New York, NY, USA: ACM, 2011, pp. 3:1–3:6.
- [2] J. Doran and I. Parberry, “Controlled procedural terrain generation using software agents,” *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 2, no. 2, pp. 111–119, June 2010.
- [3] N. Shaker, J. Togelius, and M. J. Nelson, *Procedural content generation in games*. Springer, 2016.
- [4] G. Smith, “An analog history of procedural content generation,” in *Proceedings of the 2015 Conference on the Foundations of Digital Games (FDG 2015)*. Monterey, CA, June.
- [5] N. Shaker, J. Togelius, and M. J. Nelson, *Procedural Content Generation in Games: A Textbook and an Overview of Current Research*. Springer, 2015.
- [6] D. Sheerin, “Procedural tentacle bundles in “Edge of Tomorrow”,” in *ACM SIGGRAPH 2014 Talks*, ser. SIGGRAPH ’14. New York, NY, USA: ACM, 2014, pp. 33:1–33:1.
- [7] L. Cardamone, D. Loiacono, and P. L. Lanzi, “Interactive evolution for the procedural generation of tracks in a high-end racing game,” in *Proceedings of the 13th annual conference on Genetic and evolutionary computation*. ACM, 2011, pp. 395–402.
- [8] M. B. Garda and P. Grabarczyk, “Is every indie game independent? towards the concept of independent game,” *Game Studies*, vol. 16, no. 1, 2016.
- [9] G. Tecuci, “Artificial intelligence,” *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 4, no. 2, pp. 168–180, March 2012.
- [10] M. Treanor, A. Zook, M. P. Eladhari, J. Togelius, G. Smith, M. Cook, T. Thompson, B. Magerko, J. Levine, and A. Smith, “Ai-based game design patterns,” in *Proceedings of the 10 International Conference on Foundations of Digital Games, FDG*, 2015.
- [11] L. Lidén, “Artificial stupidity: The art of intentional mistakes,” *AI game programming wisdom*, vol. 2, pp. 41–48, 2003.
- [12] K. Dill, C. Dragert, and T. Humphreys. (2016) Nuts and bolts: Modular ai from the ground up. [Online]. Available: <https://www.gdcvault.com/play/1023439/Nuts-and-Bolts-Modular-AI>
- [13] E. Brown and P. Cairns, “A grounded investigation of game immersion,” in *CHI ’04 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA ’04. New York, NY, USA: ACM, 2004, pp. 1297–1300.

- [14] Y. Douglas and A. Hargadon, “The pleasure principle: immersion, engagement, flow,” in *Proceedings of the eleventh ACM on Hypertext and hypermedia*. ACM, 2000, pp. 153–160.
- [15] M. Johansson, “Do non player characters dream of electric sheep?: A thesis about players, npcs, immersion and believability,” 2013.
- [16] M. Mateas and A. Stern, “Façade: An experiment in building a fully-realized interactive drama,” in *Game developers conference*, vol. 2, 2003.
- [17] R. Evans and E. Short, “The ai architecture of versu,” in *Imperial College, London*, Unknown Year.
- [18] L. Pena, S. Ossowski, J. M. Pena, and J. A. Sánchez, “Eep—a lightweight emotional model: application to rpg video game characters,” in *2011 IEEE Conference on Computational Intelligence and Games (CIG’11)*. IEEE, 2011, pp. 142–149.
- [19] Y. Manavalan, V. Bulitko, and M. Spetch, “A lightweight algorithm for procedural generation of emotionally affected behavior and appearance,” in *Eleventh Artificial Intelligence and Interactive Digital Entertainment Conference*, 2015.
- [20] C. Dragert, J. Kienzle, H. Vangheluwe, and C. Verbrugge, “Generating extras: Procedural ai with statecharts,” Technical Report SOCS-TR-2011.1, Tech. Rep., 2011.
- [21] C. Jennett, A. L. Cox, P. Cairns, S. Dhoparee, A. Epps, T. Tijs, and A. Walton, “Measuring and defining the experience of immersion in games,” *International Journal of Human - Computer Studies*, vol. 66, no. 9, pp. 641–661, 2008.
- [22] G. Christou, “The interplay between immersion and appeal in video games,” *Computers in Human Behavior*, vol. 32, pp. 92–100, March 2014.
- [23] P. Lankoski, “Character design fundamentals for role-playing games,” *Beyond Role and Play*, pp. 139–148, 2004.
- [24] H. Warpefelt, M. Johansson, and H. Verhagen, “Analyzing the believability of game character behavior using the game agent matrix,” 2014, pp. 1–11.
- [25] R. Lara-Cabrera, M. Nogueira-Collazo, C. Cotta, and A. J. Fernandez-Leiva, “Procedural content generation for real-time strategy games.(special issue on digital economy)(report),” vol. 3, no. 2, March 2015.
- [26] (2017) Game developers conference. [Online]. Available: <http://www.gdconf.com/>
- [27] C. Martens and R. Cardona-Rivera. (2017) Procedural narrative generation. [Online]. Available: <https://www.gdcvault.com/play/1024143/Procedural-Narrative>

- [28] M. Si, S. C. Marsella, and D. V. Pynadath, “Thespian: Using multi-agent fitting to craft interactive drama,” in *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*. ACM, 2005, pp. 21–28.
- [29] M. Mateas and A. Stern, “Procedural authorship: A case-study of the interactive drama façade,” *Digital Arts and Culture (DAC)*, 2005.
- [30] (2012) Versu. [Online]. Available: <https://versu.com/>
- [31] R. Evans and E. Short, “Versu—a simulationist storytelling system,” *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 6, no. 2, pp. 113–130, 2014.
- [32] R. Evans, “Introducing exclusion logic as a deontic logic,” in *International Conference on Deontic Logic in Computer Science*. Springer, 2010, pp. 179–195.
- [33] H. Dreyfus, *Being-in-the-world: A Commentary on Heidegger’s Being and Time, División I*. MIT Press, 1991.
- [34] A. Ortony, G. L. Clore, and A. Collins, *The cognitive structure of emotions*. Cambridge university press, 1990.
- [35] J. S. Wiggins, *The five-factor model of personality: Theoretical perspectives*. Guilford Press, 1996.
- [36] A. Mehrabian, “Analysis of the big-five personality factors in terms of the pad temperament model,” *Australian Journal of Psychology*, vol. 48, no. 2, pp. 86–92, 1996.
- [37] S. C. Marsella and J. Gratch, “Ema: A process model of appraisal dynamics,” *Cognitive Systems Research*, vol. 10, no. 1, pp. 70–90, 2009.
- [38] R. Reisenzein, “Appraisal processes conceptualized from a schema-theoretic perspective: Contributions to a process analysis of emotions.” 2001.
- [39] S. Campano, N. Sabouret, E. De Sevin, and V. Corruble, “An evaluation of the cor-e computational model for affective behaviors,” in *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*. International Foundation for Autonomous Agents and Multiagent Systems, 2013, pp. 745–752.
- [40] Y. B. Manavalan and V. Bulitko, “Appraisal of emotions from resources,” in *International Conference on Interactive Digital Storytelling*. Springer, 2014, pp. 224–227.
- [41] C. Dragert, J. Kienzle, and C. Verbrugge, “Toward high-level reuse of statechart-based ai in computer games,” in *Proceedings of the 1st International Workshop on Games and Software Engineering*. ACM, 2011, pp. 25–28.
- [42] D. Gundogdu, A. N. Finnerty, J. Staiano, S. Teso, A. Passerini, F. Pianesi, and B. Lepri, “Investigating the association between social interactions and personality states dynamics,” *Royal Society Open Science*, vol. 4, no. 9, 2017.

- [43] D. Glover, *Attack of the Copula Spiders: Essays on Writing*. Biblioasis, 2012.
- [44] R. M. Burns, “Creative writing 101: Show vs. tell,” *Colorado Springs Fiction Writer’s Group*, 2003.
- [45] B.-C. Bae, Y.-G. Cheong, and D. Vella, “Modeling foreshadowing in narrative comprehension for sentimental readers,” in *International Conference on Interactive Digital Storytelling*. Springer, 2013, pp. 1–12.

A Appendix 1

This Appendix contains the two texts presented in the separate surveys as part of A/B testing.

A.1 Text A

Tom's cat has been missing for nearly a week. He's spoken to almost everyone in the town and had no luck.

Today, Tom goes to see Erica. They talk for a while, but Erica is nervous and evasive. Tom doesn't want to suspect that she had something to do with the missing cat, but doesn't like the way she acts towards him.

Aaron, a local miscreant, breaks into Tom's house while he's out. While searching for valuables, he comes across traces of a murder and flees in terror. Someone sees him leave and reports the robbery.

Around midday, Tom is out to buy groceries when he thinks he spots a familiar tail disappearing into an alleyway. He decides to follow it.

Elsewhere, Erica groups up with another of the town's law-keepers and arrests Aaron for robbery. They toss him in a cell. Before they leave, Aaron confesses that he broke into Tom's home and tells them what he saw.

Eventually, Tom arrives back at a familiar house, but there's no sign of the cat anywhere. Somehow, Tom knows that his cat is dead. Somehow, he knows that the owner of this house is responsible. Consumed by his rage, Tom sets fire to the house and bars the door, trapping the owner inside.

Hours later, a large group, gathered by Erica, comes to question Tom.

They find his house burnt to the ground.

A.2 Text B

This subsection includes both the original bullet-point output of the prototype system and the dressed up version of Text B.

A.2.1 Prototype Output Excerpt

- Everyone is starving.
- Emma goes to see Anna at Home
- Emma steals 2 food from Anna (2nd theft)
- Anna spots the action...
- Anna starts to dislike Emma
- Rob has nothing to do
- Anna goes to see Henry
- Anna steals 6 food from Henry (11th theft)
- Alex tends to the farm.
- Harry tends to the farm.

- Alex produces 15 food from farming
- Jamie goes to see Anna
- Jamie tries to arrest Anna but is resisted
- Anna kills Jamie
- Matt goes to see Anna
- Matt tries to arrest Anna but is resisted
- Anna kills Matt
- Henry goes to the Bakery
- Henry steals 6 food from Frank (1st theft)
- Chris has nothing to do
- Francesca goes to see Anna
- Francesca tries to arrest Anna but is resisted
- Anna beats Francesca
- Francesca dies from starvation
- Tim goes to see Anna
- Tim steals 4 food from Anna (1st theft)
- Farmer tends to the farm.
- Graham tends to the farm.
- Harry goes to see Anna
- Harry tries to arrest Anna but is resisted
- Anna kills Harry
- Emma goes to see Alex
- Emma steals 9 food from Alex (3rd theft)
- Robert goes to see Alex
- Robert steals 4 food from Alex (2nd theft)
- Anna goes to see Henry
- Anna steals 4 food from Henry (12th theft)
- Alex plants a seed at the farm.
- Henry Sr goes to see Alex
- Alex sells Henry Sr 4 food for 8 gold

- Henry goes to see Henry Sr
- Henry Sr gives Henry 6 food
- Henry starts to like Henry Sr more
- Aria7 produces 5 food from farming
- Frank goes to see Alex
- Alex sells Frank 4 food for 8 gold
- ...
- Anna tries to arrest Anna but is resisted
- Anna starts to dislike Anna
- Anna kills Anna

A.2.2 Final Text

The town is starving.

A chain of thefts begins when Emma sneaks into Anna's home and steals some food. Anna spots her in the act and immediately starts to dislike her.

Now without food, Anna finds herself equally as desperate. With no other recourse remaining, she, in turn, steals from another of the town's residents—a simple man named Henry.

But this was not Anna's first crime. She has been robbing the town blind for weeks. Jamie and Matt, two of the town's more fervent believers in the law, decide to confront her and bring her to task for her actions. As they drag her off to the town's jail, Anna flies into a fit of rage and kills them both.

Chaos descends upon the town.

Taking advantage of the confusion, Henry gives in and steals from the local bakery while he thinks no one will notice. Emma joins him in the act and they leave the building almost empty.

Francesca, another believer in the law, attempts to avenge Anna's victims, but is severely weakened by hunger. Anna knocks her down and leaves her to starve. No longer entirely sane, Anna kills the next man who attempts to apprehend her and mugs Henry of his stolen bread.

Amidst the bloodbath, some few residents remain rational. Henry Senior seeks out a local farmer to purchase food, which he gives to Henry, his only son. Their bond strengthens. The baker, now left with nothing to his name, visits the very same farm and purchases just enough food to survive the week.

Five days later, consumed by self-loathing and following her own warped sense of justice, Anna takes her own life.

B Appendix 2

This Appendix includes the raw data from the survey.

B.1 Quantitative data

The questions in this section are defined as follows:

- **Q1:** To what degree would you say that the actions and emotions of the actors were in agreement within the context of the narrative?
- **Q2:** Did you find the interactions between actors within the narrative to be relevant to the narrative given the context?
- **Q3:** Did you find the actor personalities to be detrimental or beneficial to the narrative?
- **Q4:** Did you find the developments to be feasible given the context of the narrative, its actors, and their actions?
- **Q5:** To what degree could you determine the motivations of the actors?
- **Q6:** To what degree could you determine the emotions of the actors?
- **Q7:** How would you rate your interest in the expansion of this narrative?
- **Q8:** How likely is this text to have been written by a human?

B.1.1 Text A

Respondent	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
1	3	Random/irrelevant	Beneficial	Not sure	2	4	3	1
2	4	Appropriate	Not sure	Feasible	4	4	2	3
3	2	Not sure	Beneficial	Feasible	3	2	2	3
4	4	Appropriate	Beneficial	Unfeasible	1	4	5	3
5	3	Appropriate	Beneficial	Unfeasible	4	3	3	2
6	2	Not sure	Not sure	Unfeasible	2	2	4	1
7	4	Appropriate	Not sure	Feasible	3	5	5	4
8	4	Not sure	Beneficial	Feasible	3	4	2	3
9	2	Not sure	Not sure	Unfeasible	3	4	3	1
10	5	Appropriate	Beneficial	Feasible	5	5	5	5
11	4	Appropriate	Not sure	Feasible	2	4	3	5
12	4	Not sure	Not sure	Feasible	4	4	2	3
13	2	Random/irrelevant	Not sure	Unfeasible	3	4	2	2
14	3	Appropriate	Beneficial	Feasible	2	2	3	3
15	2	Random/irrelevant	Not sure	Unfeasible	3	1	2	1
16	3	Appropriate	Beneficial	Unfeasible	3	2	4	2
17	2	Random/irrelevant	Not sure	Unfeasible	2	4	5	2
18	4	Appropriate	Beneficial	Feasible	3	4	4	4
19	4	Appropriate	Not sure	Feasible	4	4	4	5
20	5	Appropriate	Beneficial	Feasible	4	5	3	4
21	4	Not sure	Beneficial	Feasible	4	4	3	3
22	4	Appropriate	Detrimental	Feasible	3	4	5	2
23	2	Appropriate	Beneficial	Unfeasible	4	5	1	1
24	2	Appropriate	Beneficial	Feasible	4	2	4	3

B.1.2 Text B

Respondent	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
25	2	Appropriate	Detrimental	Unfeasible	3	2	1	3
26	4	Appropriate	Not sure	Feasible	2	1	4	5
27	3	Appropriate	Not sure	Feasible	2	1	3	4
28	4	Random/irrelevant	Beneficial	Unfeasible	4	3	3	4
29	4	Random/irrelevant	Not sure	Feasible	2	2	3	4
30	3	Random/irrelevant	Not sure	Unfeasible	2	3	2	1
31	4	Appropriate	Not sure	Feasible	4	4	3	3
32	4	Appropriate	Beneficial	Feasible	4	3	2	2
33	4	Appropriate	Beneficial	Feasible	5	5	4	4
34	4	Appropriate	Detrimental	Feasible	5	4	2	2
35	2	Not sure	Beneficial	Unfeasible	2	4	4	4
36	4	Appropriate	Not sure	Feasible	4	3	3	3
37	4	Appropriate	Beneficial	Feasible	2	2	5	5
38	2	Random/irrelevant	Not sure	Not sure	3	3	2	2
39	2	Random/irrelevant	Not sure	Feasible	1	3	2	2
40	5	Appropriate	Beneficial	Feasible	3	4	5	5
41	4	Appropriate	Beneficial	Feasible	4	5	5	2
42	3	Appropriate	Beneficial	Feasible	4	3	4	5
43	1	Random/irrelevant	Not sure	Not sure	1	2	1	3
44	2	Appropriate	Not sure	Feasible	3	3	2	3
45	4	Appropriate	Beneficial	Unfeasible	5	4	3	2
46	3	Appropriate	Beneficial	Feasible	4	2	4	4
47	3	Appropriate	Beneficial	Feasible	5	4	4	5
48	2	Random/irrelevant	Not sure	Feasible	3	3	4	3

B.2 Qualitative data

The questions in this section are defined as follows:

- **Q9:** Was there a particular part of the text that was strange? If so, why?
- **Q10:** Was there a particular part of the text that was compelling? If so, why?
- **Q11:** Do you have any further comments?

B.2.1 Text A

Respondent 1

- **A9:** Tom doesn't seem confused at all until right at the end, when you realize that he killed his own cat. Tom is a psychopath. Erica being suspicious of him is a bit of a non sequitur, even considering the ending.
- **A10:** The ending, when you realize that tom killed his own cat and couldn't remember.

Respondent 2

- **A9:** I can't see why would Erica try to arrest Aaron. The rest made sense.
- **A10:** Aaron finding traces of murder. The break-in is suspenseful in itself, but twisting it even more takes it to the top. It was also nice how Aaron told the others about the murder afterwards.

Respondent 4

- **A9:** Tom burning his own house down, thinking it was the house of another, was strange, but my brain was able to concoct psychological reasons for his strange behavior.
- **A10:** The dead body was compelling, and Tom trapping himself in the house was too. Murder often seems compelling, and even moreso, people acting against their own interests are often compelling.

Respondent 6

- **A9:** How Tom didn't know he was burning his house.
- **A10:** Why Erica was nervous.
- **A11:** It would've been easier to rate if a little more detail was added.

Respondent 8

- **A9:** Burning down a house and murdering a person over a cat. Doesn't seem believable to me.
- **A10:** The thief was interesting

Respondent 9

- **A10:** Tom seems to give up on actively looking for his cat very abruptly. Also, "buying groceries" doesn't sound quite right for the medieval setting. Something like "visiting the market" or "buying bread" would probably fit the feel better. It also reads like Tom burnt his own house down, but leaving the homeowner (someone else) trapped contradicts this.
- **A11:** It's a decent plot outline, but it jumps around a bit too much. It feels like there are a lot of unanswered questions (which of course may be answered in a larger work but aren't concluded in this piece). Why is Erica with the law-keepers? How does Tom know this person killed his cat? What's the deal with Aaron finding evidence of a murder?

Respondent 13

- **A9:** Grammar, sequence of events. Context - background to events and personalities

Respondent 14

- **A9:** not really. I just found it to be very fast paced and non-descriptive.
- **A10:** The twist at the end. It was foreshadowed and it peaked my interest making me want to read more.

Respondent 15

- **A9:** Somehow, Tom knows that his cat is dead + is willing to kill someone-not logical at all and very strange, all together there is no logical development of the story

Respondent 16

- **A9:** why was the house of Tom suddenly burned down when he was the one that did set the other house on fire. rage?
- **A10:** What is the role of Erica in the story, she is somehow linked to all
- **A11:** it was oddly captivating

Respondent 17

- **A9:** Is Erica's house burnt?
- **A10:** No.
- **A11:** It is hard to connect them.

Respondent 21

- **A9:** It was strange how Erica acted towards Tom in their interaction, since that wasn't expanded upon at all later.

- **A10:** The ending was compelling because you had to rethink briefly what exactly had happened.

Respondent 22

- **A10:** The part where Tom burnt down his house was very compelling. It works like a cliffhanger.

Respondent 23

- **A9:** It seems like he himself murdered his cat and then knew somehow and killed himself by burning down his own house. But the cat has been gone for a long time thus the evidence for murder should have been there for a while as well. The only logical explanation for me given the supplied narrative is mental illness and that his brain is unable to accept that he killed the cat himself. But for him to accept it in the end and burn the house down while killing himself there has to be a reason or a trigger. Without that it becomes a lot less believable. I mean why does he suddenly know? And why then of all times? Also if this theory is true grief or regret from his end would have made it more believable.
- **A10:** Not really any specific part. Most parts are generic and often used in narratives (doesn't necessary make them bad).
- **A11:** Only given the above narrative the "red thread" is a little bit hard to follow and a lot of own assumptions has to be made for it to even make sense.

B.2.2 Text B

Respondent 25

- **A9:** The escalation of theft is illogical given the context of the story. Anna catches Emma in the act and starts to dislike her is not an appropriate response when it results in her having to steal from others. More likely, she would defend her stash and hate Emma, resulting in escalation against Emma, not others.
- **A10:** The opening line carries heavy impact due to the lack of secondary description.
- **A11:** While this is a step above the chatbots that were around ten years ago, the general feel is the same.

Respondent 27

- **A9:** abruptly introducing characters
- **A10:** no

Respondent 28

- **A9:** The next to last paragraph seemed odd, irrelevant, and came out of nowhere

- **A10:** Most people who were stolen from end up stealing themselves
- **A11:** No

Respondent 29

- **A9:** Everyone seemingly just "did" something as a reaction. There wasn't much interaction or real story. A did X so B did Y causing C to do Z.
- **A10:** Anna, presumably the main character, went crazy and that was the most realistic part.
- **A11:** If there's any way to make the story follow a particular character, it would allow for better storytelling as opposed to an omniscient "telling" or a series of events. It was definitely not "shown."

Respondent 30

- **A9:** I'm not sure how could Anna steal so much food to starve the whole town, what she did with all that food, or why she committed suicide instead of just stopping. She sounds really evil, but without any real motivation. She clearly want to bring destruction, not just eat.

Respondent 31

- **A9:** Anna having been robbing the town for weeks—makes more narrative sense if it her first offence after Emma steals from her (since it is said to be due to desperation)

Respondent 32

- **A9:** It seems a bit weird, because in the text there's quite a lot of focus on emotions, motivations and characteristics of the actors, but very little on their actions, so it is kind of unequal. The time perspective is weird as well because at first it seems like everything is happening within a short time, then people start getting hungry and go shopping so suddenly it seems to be going on for days.
- **A10:** No particular part, but the whole thing was pretty funny and could only been written by wither a child or AI.
- **A11:** I wish questions 2-5 had Likert-scaled answers because it's difficult to be certain. You might end up with more "not sure" than necessary. Interesting topic, good luck!

Respondent 34

- **A9:** ""A chain of thefts begins with [Emma's act]"" , when it is said 2 lines later that Anna had stolen multiples times already. Most of the emotions provided did not bring anything to the story, or would even contradict it, e.g. Henry being described as a ""simple man"" but being the only ""clever"" one in his acts by figuring out he could steal more easily in the confusion. Some names used do not sound relevant in a medieval context. But maybe you rank that within the ""language context"" non-relevant to this survey. Characters resorting, in the end, to buying food from a farm to solve their problems, sounds... non-logical."

- **A10:** "The description of the chain of thefts is fluid, so is the global story-telling, considering it is not supposed to be a ""complete story"". The next to last paragraph, ""Amongst the bloodbath [...]"", sticks out as we forget Anna for a while and get to know secondary characters. Switching subjects in that way sounds very ""human-like""."
- **A11:** I never heard about procedural narrative generation, and it never came to my mind that stories could be computer generated. I probably wouldn't have even considered that this story could be computer generated (whether it actually is or not) without the context of the survey, I think it would have passed as a regular human -yet flawed- text without a second thought.

Respondent 35

- **A9:** The text begins with "a chain of thefts begins with...", but apparently Anna has been "robbing the town blind for weeks"

Respondent 36

- **A9:** That Emma and Henry suddenly becomes partners in crime, after she has stole from him.
- **A10:** When Anna is brought to jail. Action and justice and I'm curious on Annas career as a thief.
- **A11:** The suicide is both a logical and inlogical end.

Respondent 38

- **A9:** More in general, how quickly it went from "stealing food" to "killing everyone"
- **A11:** While in general it is believable that starvation leads to irrational behaviour, the events in this story just seem exaggerated.

Respondent 39

- **A9:** why she can live on her own life at last
- **A10:** When Anna change her mind
- **A11:** A sad story

Respondent 41

- **A9:** too much focused on action of the characters less description about environment and feelings?
- **A11:** this was interesting.

Respondent 42

- **A9:** The suicide in the end is very sudden.

- **A11:** Everyone was stealing, it felt a bit too much when even Henry was doing it. It's probably that this is just an exerpt, but since we jump right into a theft and then there's crimes and whatnot, there is no normalcy to put these things into context, this is just what people do in this particular village. Some people are buying bread in the end, yes, but then that is framed somewhat as an anomaly.

Respondent 44

- **A9:** Random new characters appearing out of nowhere was a tad off.
- **A10:** The one sentence explainers about what's happening, they broke up the pace in a good way and pulled together all the random bits of information, would be nice to have another at the end o.o
- **A11:** Nope, aside from good luck this seems a very interesting project.

Respondent 45

- **A9:** A character is knocked down and starves.

Respondent 46

- **A9:** Their bond strengthens' is strange when it's a father-son relation.

Respondent 47

- **A11:** Some of the actions seem a bit too harsh, like the killing and suicide. But the rest was believable.