“GLASS: Adapting Game content to pLayer Affective State and perSonality”

Ivo Bernardo Silva Capelo

Thesis to obtain the Master of Science Degree in Computer Engineering

Examination Commitee

Chairperson: Prof. Miguel Nuno Dias Alves Pupo Correia
Supervisor: Prof. Carlos António Roque Martinho
Member of the Committee: Prof. César Figueiredo Pimentel

November 2013
To my parents, Helena and Fernando...
Acknowledgments

This thesis was only possible due to the support and orientation of Prof. Doutor Carlos Martinho, who proved to be an invaluable mentor by guiding me along the way. His help was crucial and insightful despite having ever more responsibilities given to him.

I want to thank my uncle and aunt who gave me lodgings and who allowed me to focus on my work as well as my parents for their emotional and financial support. It was thanks to my family and their utmost dedication at raising me that this work came to be. I am forever indebted to them.

I would like to send my heartfelt appreciation to my colleagues Daniela Borges, Diogo Gil Henriques, Diogo Nunes, Daniela Fontes, António Pereira, Ana Almeida and Auguste Cunha for their constant feedback over all stages of this work, without whom I would often be lost.

I would like to thank the patience of the Squared Muffin team - Pedro Engana, Pedro Lousada and Rui Dias - who allowed me time to work on polishing this thesis where others would only ask for results.

Unconventionally I want to thank my Sattelite L500-1 TU and my 65-34-DD, who despite not being sentient, worked beyond the line of duty, making sure only to break down when the work was finalized.

Finally I am forever grateful to Sandra Lourenço, for her love, comprehension and patience over this long trial, who stood up to my workaholic tendencies and helped me find the calm to think clearly and reasonably.
Resumo

Os videogames tornaram-se excessivamente caros de produzir. Para poder agradar a diversos tipos de jogadores, os criadores dependem dos géneros com mais frequente sucesso na indústria. Conforme os mercados se enchem de jogos semelhantes, imensos jogadores permanecem desinteressados.

Trabalhos recentes no campo de modelação de entretenimento mostraram sucesso em desenvolver métodos que aumentam a imersão e tempo de sessão de videogames, permitindo que jogos cheguem a audiências mais diversificadas. Este trabalho propõe resolver o problema do reconhecimento de personalidade como primeira fase da modelação de entretenimento.

Descrevemos a nossa solução como sendo composta por um cenário baseado em tarefas e um sistema com uma rede Bayesiana. O nosso sistema solução consegue utilizar amostras oriundas do cenário para efectuar lógica probabilística na decisão da personalidade de jogadores.

Também descrevemos extensivamente a nossa metodologia para desenvolver o cenário, desde a colecção de tarefas inicial até à integração com a rede Bayesiana. Esta metodologia, baseada no modelo de personalidade de temperamentos Keirsey, foi desenvolvida para poder ser portada para diferentes jogos, géneros e modelos de jogador.

Aplicámos a nossa metodologia para construir um cenário concreto para o jogo Minecraft. Para validar a nossa solução obtivemos amostragens provindas de testes do cenário e efectuamos diversos testes de validação cruzada para obter as taxas de sucesso do nosso sistema. Conseguimos assim, com elevadas taxas de sucesso, desenvolver a primeira parte de um sistema adaptativo, um que consiga identificar personalidade e interesses do jogador para adaptar o conteúdo e aumentar o valor de entretenimento em jogos.

Palavras-chave: Modelação de Entretenimento, Videogames, Modelos de Personalidade, Aprendizagem de Máquina, Lógica Probabilística.
Abstract

Videogames have become extremely expensive to produce. In order to please all kinds of players, developers have turned to industry-proven genres. This has left the market overfilled with similar games and many potential players unengaged and uninterested. Recent work in the field of entertainment modelling has successfully developed methods that can increase engagement and play-time for different types of players, allowing games to reach a broader audience.

This work addresses the problem of recognizing player personality as a first step in entertainment modelling. We describe our solution that is comprised of a task-based scenario and a Bayesian network system that classifies personality based on sample data from the scenario.

We also extensively describe our methodology for developing the task-based scenario, from initial task collection to integration with the Bayesian network system. This methodology, based on the Keirsey Temperament Model, was developed to be ported and applied over different games, genres and player models.

To prove the usefulness of our methodology, we applied it to develop a concrete scenario for the game Minecraft and tied it to an inferring system. We then validated our concrete solution by running several cross-validation tests over our system to achieve sucess rates. We have thus successfully developed the first part of an adaptive system, one that can identify personality and interests to adapt content and increase entertainment in games.

Keywords: Entertainment Modelling, Video Games, Personality Models, Machine Learning, Probablistic Reasoning.
Contents

Acknowledgments ................................................................. iii
Resumo ................................................................................ iv
Abstract ................................................................................ v
Contents ................................................................................. vii
List of Tables ........................................................................... viii
List of Figures ........................................................................... x

1 Introduction ........................................................................... 2
  1.1 Motivation ....................................................................... 2
  1.2 Problem Description ......................................................... 3
  1.3 Document outline ............................................................. 4

2 State of the Art ...................................................................... 6
  2.1 Personality in Psychology .................................................. 6
  2.2 Personality based player models ......................................... 9
  2.3 Entertainment Modelling .................................................. 13

3 Solution ................................................................................. 20
  3.1 Design Methodology ........................................................ 21
    3.1.1 Task Identification ...................................................... 22
    3.1.2 Observation .............................................................. 24
    3.1.3 Task Clustering ........................................................ 29
    3.1.4 Scenario Design ....................................................... 31
  3.2 Solution Scenario ............................................................ 35
  3.3 Inferring System Model ..................................................... 40
  3.4 Experiment Protocol ........................................................ 42
  3.5 Solution summary ........................................................... 43

4 Data Analysis ......................................................................... 46
  4.1 Experimental results ......................................................... 46
  4.2 Result Significance ........................................................... 52

5 Conclusions .......................................................................... 54
A  Observation Results

B  MBTI Questionnaire, obtained from [1]

Bibliography
List of Tables

2.1 Correlations between TCI elements and FFM factors. ......................... 8
2.2 Relation between DGD model play styles and temperament theory skillsets. ..... 11
2.3 Unified Model relation between Keirsey, Bartle and Lazzaro models. ............... 12
2.4 Comparison of presented entertainment modelling studies. ..................... 18

3.1 Initial prospects on enjoyability of Minecraft tasks across temperaments. ........ 23
3.2 Societal distribution of Keirsey Temperaments. .................................... 25
3.3 Frequency of temperament actions over five clusters. ............................ 30
3.4 The five clusters achieved in Task Clustering and their corresponding tasks .... 30
3.5 Possible combinations of the five clusters and associated temperaments. ........ 34
3.6 Distribution of tasks across rooms of the solution scenario. ................... 36

4.1 Hit and miss rates for combinations of expected malformed and well formed rooms. 51
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Bartle’s MUD types distribution.</td>
<td>10</td>
</tr>
<tr>
<td>2.2</td>
<td>Unified Model comparison between Keirsey Temperaments and DGD types.</td>
<td>13</td>
</tr>
<tr>
<td>2.3</td>
<td>Architecture for adaptive games.</td>
<td>15</td>
</tr>
<tr>
<td>2.4</td>
<td>A model for passive Dynamic Difficulty Adjusting systems.</td>
<td>16</td>
</tr>
<tr>
<td>3.1</td>
<td>A kingdom in Minecraft.</td>
<td>21</td>
</tr>
<tr>
<td>3.2</td>
<td>Binary Tree Architecture.</td>
<td>32</td>
</tr>
<tr>
<td>3.3</td>
<td>Collapsed Rooms Architecture.</td>
<td>32</td>
</tr>
<tr>
<td>3.4</td>
<td>Solution Architecture.</td>
<td>33</td>
</tr>
<tr>
<td>3.5</td>
<td>Depiction of a checkpoint implemented in <em>Rooms</em>.</td>
<td>35</td>
</tr>
<tr>
<td>3.6</td>
<td>Generalized task room architecture of the solution scenario.</td>
<td>36</td>
</tr>
<tr>
<td>3.7</td>
<td>Implementation of room 1 in the <em>Minecraft</em> scenario.</td>
<td>37</td>
</tr>
<tr>
<td>3.8</td>
<td>Implementation of room 2 in the <em>Minecraft</em> scenario.</td>
<td>37</td>
</tr>
<tr>
<td>3.9</td>
<td>Implementation of room 3 in the <em>Minecraft</em> scenario.</td>
<td>38</td>
</tr>
<tr>
<td>3.10</td>
<td>Implementation of room 4 in the <em>Minecraft</em> scenario.</td>
<td>38</td>
</tr>
<tr>
<td>3.11</td>
<td>Implementation of room 5 in the <em>Minecraft</em> scenario.</td>
<td>38</td>
</tr>
<tr>
<td>3.12</td>
<td>Implementation of room 6 in the <em>Minecraft</em> scenario.</td>
<td>39</td>
</tr>
<tr>
<td>3.13</td>
<td>Implementation of room 7 in the <em>Minecraft</em> scenario.</td>
<td>39</td>
</tr>
<tr>
<td>3.14</td>
<td>Generalization of the solution Bayesian Network.</td>
<td>40</td>
</tr>
<tr>
<td>3.15</td>
<td>Player made paintings at room 7 of the scenario.</td>
<td>43</td>
</tr>
<tr>
<td>4.1</td>
<td>Temperament distribution across experiment sample population.</td>
<td>47</td>
</tr>
<tr>
<td>4.2</td>
<td>Comparison of temperament distributions between our test population and societal standard.</td>
<td>47</td>
</tr>
<tr>
<td>4.3</td>
<td>Hit and miss rates of initial cross-validation using sample from all rooms.</td>
<td>48</td>
</tr>
<tr>
<td>4.4</td>
<td>Average rates of cross-validation, by discarding single rooms.</td>
<td>49</td>
</tr>
<tr>
<td>4.5</td>
<td>Average rates of cross-validation for the Guardian temperament, by discarding single rooms.</td>
<td>50</td>
</tr>
<tr>
<td>4.6</td>
<td>Average rates of cross-validation for the Rational temperament, by discarding single rooms.</td>
<td>50</td>
</tr>
<tr>
<td>4.7</td>
<td>Average rates of cross-validation, by using only data from single rooms.</td>
<td>51</td>
</tr>
<tr>
<td>A.1</td>
<td>Frequency of Guardian actions over initial task list during the experiment in <em>Observation</em>.</td>
<td>58</td>
</tr>
<tr>
<td>A.2</td>
<td>Frequency of Rational actions over initial task list during the experiment in <em>Observation</em>.</td>
<td>59</td>
</tr>
</tbody>
</table>
A.3 Frequency of Artisan actions over initial task list during the experiment in *Observation.* . . 60
A.4 Frequency of Idealist actions over initial task list during the experiment in *Observation.* . . 61
Chapter 1

Introduction

1.1 Motivation

With videogames spreading across the globe and becoming more and more inclusive, the industry has reported a worth of over $74bn dollars\(^1\), far more than the entertainment giant of Hollywood\(^2\). However, it is an industry whose costs for development have grown, as assets such as graphics, animations and even scenarios require more and more manpower.

To keep profits up, publishers look for games that have solid and proven gameplay, sometimes straying new and creative ideas away from production. This led the current console generation to a mass production of similar genres (such as the first person shooter), leaving part of the industry somewhat stagnant and thousands of potential customers uninterested and unengaged.

One way to improve the success of games and reduce costs would be to adapt the games dynamically to the user who is playing, giving the ideal level of challenge and content. This would allow game creators attract a more diverse audience by presenting different kinds of content that can fulfill different necessities.

Unfortunately everyone learns at different rates and in different ways\(^2\), making player necessities very different and complex to understand. Howard Gardner\(^3\) has presented in his psychological theory, that intelligence itself is encompassed of different domains and as such different people can excel differently. While a mathematician should have greater logical-mathematical intelligence a piano player would exceed him in the musical domain.

To develop games that truly engage players, developers have turned to psychological theories such as the flow theory \([4][5][6]\). Flow is the state of complete absorption and concentration that originates from having the exact skills to tackle a challenge without it being excessively simple or complicated. Flow presents a middle ground where the capabilities of an individual allow him to engage, grow and learn to better face the same type of experiences. Different people will thus be inclined to experiences that take advantage of their aptitudes.


International Hobo, a videogame consulting group, advises[7] to move beyond the dependence of personal judgments when designing games by viewing the process of game design from a psychological perspective. Both the practical and the theoretical side of game development point to the need of player models that can categorize and explain general player features and skillsets, thus simplifying the work of game designers. Psychological theory has yet to be used to identify the player at runtime, promoting dynamically adapted content.

Previous endeavors, by the videogame industry, to identify personality have been invasive and immersion breaking. In Silent Hill: Shattered Memories 3 a mixed approach is used. The player answers several psych profiling sessions and the results are later fine-tuned according to player actions. The outcome of the sessions can easily be predicted and influenced, breaking the player immersion. The problem of the approach is that it can be fantasy breaking. It only works with Shattered Memories due to its already established sleuthing and psychological themes.

In his thesis[8], Dias has confirmed that adapting the content of videogames according to personality typologies can increase their entertainment value. As part of future work he recommends the extraction of personality before actual content adaptation. Our work will follow Dias’ recommendation and develop a “prequel” methodology to identify personality at runtime, allowing latter adaption.

### 1.2 Problem Description

As games become more complex, their development costs also increase. For costs at the order of millions the risk of failure is excessively high and companies have turned to develop industry-proven genres. With the market over flooded with similar titles some stagnation in creativity is becoming apparent, diminishing the potential of games as a medium. A potential solution to reduce risk is to adapt the content at real-time so it fits the players interests, also known as entertainment modelling[9].

Entertainment modelling in videogames can be adaptation of difficulty, the way goals and information are given to the player, or changes to overall content. Recent work has been done [9],[10] to change games according to performance data, usually discarding information regarding the characteristics, interests and habits of the player. Data regarding performance is not subjected to subjective interpretations, thus being used more often. In order to accurately depict, the mental state of players for entertainment modelling, both performance and personality data, are required.

This work will address the problem of how to infer, from data gathered at runtime, the personality preferences of players for content adaptation. It will examine the hypothesis that actions and choices reflect the preferences of individual.

Contributions from the work should include a methodology for analysing decisions and mapping them to personality typologies. This methodology can then be used and applied over different games and replicated for different personality models. Once personality has been gathered, game algorithms can perform a more precise modelling of entertainment over player preferences, increasing enjoyment value.

---

3 Silent Hill: Shattered Memories, Climax Studios, Konami Digital Entertainment, 2009
1.3 Document outline

The remainder of this document is divided into four parts.

In State-of-the-art, we report the most prevalent theories that connect the field of personality psychology to game design. Correlations between the different models will be explained in order to select an appropriate model to work with. We also present existing research in the field of entertainment modelling.

In Solution, we describe the processes of developing our solution and its description. We begin by extensively describing our methodology to design a game scenario that can be coupled to our inferring system. At each step we present our own usage of the methodology to design a Minecraft based scenario. Afterwards we explain the probabilistic reasoning algorithm for our system that can, to a certain degree, classify the temperament orientation of players. Finally we explain the procedure of the experiment to validate our solution.

In Data Analysis, we analyze the results obtained from our experiment and their significance over our work. We present a series of cross-validation tests and how we used them to identify malformed rooms. At the end of the section we debate the importance of our findings and the issues that were identified.

In Conclusions, we sum up the important aspects and decisions taken for our system as well as discuss success criteria and issues to be addressed. To finish up we discuss potential future work based on ours.
Chapter 2

State of the Art

Since our work is heavily based on previous studies done by the personality psychology community, we start by explaining some of these models and related correspondences. The compared models were chosen due to their established use inside the game development community, close relation to player behavior or claimed relevance in psychology. We also compare player models that derive from empiric study or adaptation of psychological models. Finally, we report on some studies that use entertainment modelling, a practical application of this work.

2.1 Personality in Psychology

Personality psychology is a field based on empirical observation and analysis, as such different models have appeared to categorize individuals. Study on used psychological models is required to choose one best fitted to our objectives. This section will explain the psychological theories of Meyer-Briggs typology (MBTI), Five Factor Model (FFM), Temperament and Character Inventory (TCI) and Keirsey Temperament Model (KTM). From the many personality models, we chose to analyze models that had the following criteria: models with significant widespread use within psychological or interactive entertainment fields (MBTI, FFM); models that relate behaviour to mental states (TCI, KTM).

Meyer-Briggs Type Indicator (MBTI). MBTI is a scientific method for classifying individuals according to psychological preferences. It is one of the most used psychological models in the field of interactive entertainment [11], and its simplicity allowed it to be adapted into player models such as the DGD. The theory is based on the work of psychotherapist and psychiatrist Carl Jung[12]. Jung proposed that the human conscious has four main functions: Sensation vs. Intuition for perceiving, and Thinking vs. Feeling for judging. The main functions could then be modified by two attitude types: Extraversion vs. Introversion. MBTI is built upon Jungian theory, underlining the need for a new dimension to explain interaction with the outside world - Judging vs. Perceiving. Unlike Jung’s work, that only supports eight possible personality types by identifying the most relevant function and modifier, MBTI divides the personality spectrum into sixteen possible types.
The Extraversion-Introversion dimension is concerned with world interaction. While Extroverts thrive on diverse interpersonal relationships, Introverts prefer one-on-one communication and privacy. The Sensing-Intuition dimension explains how information is acquired by individuals. While Sensing people work with clear information and instructions based on conventional consensus, Intuition people use information to build abstract theoretical models, patterns and connections. The Thinking-Feeling dimension is concerned with decision making. Thinking people follow objective reasoning and judgments while Feeling people decide subjectively as an emotional response to personal values. Finally, the Judging-Perceiving dimension considers how one manages and organizes their life. Judging individuals plan ahead and prefer routine behavior, while Perceiving individuals act upon the present necessity, using flexibility and new possibilities.

MBTI has become widespread in different areas of knowledge, including interactive entertainment. Its validity, however, as a tool has been questioned in the field of personality psychology. In [13], McCrae et al. debate the validity of the MBTI in the field of personality psychology. The first question raised is the need for mutually exclusive types. This statement can lead to loss of information when individuals that show a balance between both sides of a dichotomy are not allowed ambiguous designations. Another important question is that the added dimension for dealing with the outside world shows greater preference in identifying the extrovert individual, losing the dominant-auxiliary notions that were important in Jungian theory. They then propose a correlation between MBTI dichotomies and four of the five factors of the FFM, excluding Neuroticism.

Five Factor Model (FFM). The FFM has been a massively researched and consented model that presents a hierarchical organization of personality traits in terms of five different dimensions. Validation of FFM has been based on reports studying either natural language adjectives across different languages or theoretically based questionnaires [14]. Due to its diverse origin, the five factors can have different nomenclatures. We consider the McCrae standard: Extraversion, Agreeableness, Conscientiousness, Neuroticism and Openness to Experience. The significance of showing a high score in a specific trait is as follows:

- **Extraversion** - Shows a tendency to look for others and being outgoing.
- **Agreeableness** - Cooperative, compassionate and receptive to others.
- **Conscientiousness** - Follow strict self-discipline and plan their lives.
- **Neuroticism** - Develop easily undesirable emotions such as anxiety and anger.
- **Openness to experience** - Thrive on intellectual curiosity always seeking novelty.

Cloninger’s Temperament and Character Inventory (TCI). TCI appears as a popular model in psychiatric practice and research, to describe individual differences in psychopathologic behavior. Unlike previous models that focus on normal forms of behavior, TCI is best for describing maladaptive behaviors. Cloninger’s theory [15] introduces seven dimensions (four temperaments and three characters).
Table 2.1: Correlations between TCI elements and FFM factors. Bold elements are significant for $p < 0.01$. HA: Harm Avoidance, NS: Novelty Seeking, P: Persistence, RD: Reward dependence, SD: Self-directedness, CO: Cooperativeness, ST: Self-transcendence. Based on findings of [16].

<table>
<thead>
<tr>
<th></th>
<th>HA</th>
<th>NS</th>
<th>P</th>
<th>RD</th>
<th>SD</th>
<th>CO</th>
<th>ST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neuroticism</td>
<td>0.42</td>
<td>-0.04</td>
<td>0.15</td>
<td><strong>0.25</strong></td>
<td>-0.52</td>
<td>-0.15</td>
<td>0.16</td>
</tr>
<tr>
<td>Extraversion</td>
<td>-0.36</td>
<td><strong>0.46</strong></td>
<td>0.03</td>
<td><strong>0.50</strong></td>
<td>0.11</td>
<td><strong>0.17</strong></td>
<td>0.13</td>
</tr>
<tr>
<td>Openness</td>
<td>-0.20</td>
<td>0.09</td>
<td>0.01</td>
<td>0.16</td>
<td>0.05</td>
<td><strong>0.26</strong></td>
<td><strong>0.37</strong></td>
</tr>
<tr>
<td>Agreeableness</td>
<td>-0.03</td>
<td>0.07</td>
<td>-0.06</td>
<td><strong>0.29</strong></td>
<td>0.10</td>
<td><strong>0.60</strong></td>
<td>0.04</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>-0.04</td>
<td><strong>-0.45</strong></td>
<td><strong>0.52</strong></td>
<td>0.03</td>
<td>0.24</td>
<td>-0.08</td>
<td><strong>0.20</strong></td>
</tr>
</tbody>
</table>

The temperaments are individually heritable and can be seen in early stages of personal development. They are Novelty Seeking, Harm Avoidance, Reward-dependence and Persistence. The three characters are used to explain differences in people with similar temperaments and are connected to learning mechanisms, they are Self-directedness, Cooperativeness and Self-transcendence.

Table 2.1 shows the result of comparison between TCI and FFM [16], with each element of temperament or character correlating with at least one of the five factors. This reveals significant overlap of both models. Harm Avoidance is directly related to Neuroticism and inversely related to Extraversion and Openness. Novelty Seeking is related to Extraversion and inversely related to Conscientiousness. Persistence is directly related to Conscientiousness. Reward dependence is directly related to Neuroticism, Extraversion and Agreeableness. Self-Directedness mildly relates positively to Conscientiousness and negatively to Neuroticism. Cooperativeness relates positively with Extraversion, Openness and Agreeableness. Finally, self-transcendence relates directly with Openness and Conscientiousness.

**Keirsey Temperament Model (KTM).** Analysis of classical temperaments and MBTI allowed Keirsey [17] to present his Temperament theory. This theory, unlike previous models, focuses on behavior rather than thoughts and emotion. Keirsey identifies four archetypes from the sixteen junctions of MBTI. The SJ or Guardians, the SP or Artisans, the NT or Rationals and the NF or Idealists. According to Keirsey[18], society is roughly formed of 38% Guardians, 38% Artisans, 12% Rationals and 12% Idealists.

The Artisan Temperament must be free and unbound to do what he wishes when he wishes. Artisans live on the present and take action according to whim. They are happier when being impulsive, unbound and show strong resilience to failure.

The Guardian Temperament seeks to belong in a structured organization, such as their own society. Unlike Artisans, Guardians seek to be bound and obliged always serving their responsibilities and duties. They are seen as responsible, title-seeking and strict.

The Rational Temperament is fascinated by power over nature, by their understanding, control, prediction and explanation of it. Rationals always look for wisdom and are compelled to competence. They are seen as competent, self-critical, logical as well as curious.

The Idealist Temperament looks for self-actualization and face a constant struggle to find identity. Idealists view identity as integrity, are honest to others and invest considerable time and emotion in their relationships, not needing a similar return of devotion. They become enthralled by words and love to work with them, many becoming writers. They are seen as social, honest and sensitive.
The basis of KTM on patterns of behavior is ideal to assess the skillset [18][11], employed by players when facing a challenge. Rational people demonstrate Strategic Skills, thinking ahead and developing processes to overcome possible future contingencies and reach objectives. Idealists show Diplomatical Skills, which allow strivance for unity and conflict resolution. Artisans deploy Tactical Skills to take action and reach the best possible outcome over current events. Finally, the Guardian’s Logistical skills are mediators of resources, people and information able to optimize and standardize events.

In this section, we summarized the most relevant studies in the field of personality psychology to understand which can be used as basis for entertainment modelling. The MBTI presents a direct way to identify typologies but discards information when confronted with ambiguous data. FFM is defended as being sounder than MBTI and presents five scales of personality instead of binary dichotomies, thus holding ambiguous data for future analysis. Unfortunately, FFM and MBTI both focus excessively on the mental state of individuals and depend on self-assessment to identify personality, which would be intrusive in adaptive games. TCI presents dimensions that relate very well to game mechanics: Novelty Seeking, Harm Avoidance, Reward-dependence and Persistence. Despite this synergy, TCI is better fitted to identify maladaptive behaviors and dysfunctional individuals being inappropriate for our purposes. Finally, KTM has a similar basis as MBTI but chooses to explain behavior instead of thought and emotion. Since it can explain typologies by observing actions, KTM is the best fitted model to our goals. Additionally KTM gives information regarding motivation and skillsets of different typologies which can be helpful when developing critical tasks for games.

From the personality psychology field, KTM has appeared to be the most relevant model for our objectives. It is now required to understand how game design addresses personality. The next section will approach player models according to game design, some of these models have basis on the personality models already explained in this section, mainly MBTI.

### 2.2 Personality based player models

The diversity of players can be a problem to game developers. In order to completely engage the player in a form of flow, it is important to give them challenges that reflect their abilities and preferences. Several studies have been made with the ambition of deriving player typefication from psychological theories.

**Bartle player types.** A study conducted by Richard Bartle on multi-user dungeons (MUDs) brought one of the first, and most enduring, player model. In [19], player actions are related to their personalities. According to Bartle, four approaches to playing MUD's can be found by analysing the interaction patterns of players. The patterns can be aligned using a two dimensional graph as shown by Figure 2.1. The world-player axis explains if interests are derived from player or world interaction, while the acting-interacting axis is strictly environment related.

Socializers are players who derive enjoyment in other players and interacting with them. Killers become engaged while acting on other players, usually from bringing grief or demonstrating superiority.
Figure 2.1: Bartle’s MUD types distribution according to acting-interacting and world-player axis. Based on data and graphics of [19].

Achievers act over the world. To them the environment is an immersive experience and mastering the world is their goal. Explorers want to interact with the world; they want to be surprised and find new things and tend to rack up knowledge and experiment regarding the game’s mechanics. Unfortunately, Bartle’s work is related to a specific genre of games, the MUD. And it is extremely limited when applied to other genres.

**Demographic Game Design (DGD).** The DGD[7] is a very popular model based on the MBTI and focuses on market oriented game design. Out of the four MBTI dichotomies DGD uses only the last two, TF and PJ, as these were shown to be the most discriminative when applied to gaming preferences. In DGD, players fall into four clusters - the Conqueror (type1, TJ), the Manager (type2,TP), the Wanderer (type3,FP), the Participant (type4,FJ) - each with 2 subgroups to distinguish hardcore from casual players. The survey approached over 300 individuals and was crossed with the before mentioned Bartle MUD Model.

Bateman and Boon[11] present a comparison to match temperament skillsets to DGD. The comparison emerges with the assumption that Hardcore players have innate intuition (I) while the Casual audience has innate sensing(S). The assumption is backed by statistical tendencies found in both groups. Table 2.2 shows the relation play type to skillset: Type 1 Conquerors present strategic(H1) and logistical skillset(C1); Type2 Managers present strategic(H2) and Tactical(C2) skillsets; Type3 Wanderer presents Diplomatic(H3) and Tactical(C3) skillsets; Type 4 Participant presents Diplomatic(H4) and Logistical(C4). A division of skillsets between hardcore’s strategic-diplomatic and casual’s logistical-tactical seems evident. This comparison is a fairly crude approximation stating that all members of a set, and not the majority, are endowed of a specific skillsets.
Table 2.2: Relation between DGD model play styles and temperament theory skillsets. Adapted from [11].

**Lazzaro’s fun types.** Lazzaro’s work[20] regarding player motivation shows an important relation between play style and emotion. The work focuses on emotions emerging from gameplay and how people play games to create moment-to-moment experiences, as such the key factors that people pursue when playing are deeply connected to their play styles and personalities. From the data obtained, Lazzaro et. al. created 12 models of player experience based on four emotional keys: Hard Fun, Easy Fun, Altered States or Serious Fun and People Factor or People Fun. These four fun types explain what motivates players and what emotions can be expected during play.

- **Hard Fun:** Players seek challenge, strategy, and problem solving. Hard fun leads to emotions of Frustration, and Fiero, the emotion of triumph over adversity.

- **Easy Fun:** Players become immersed in intrigue and curiosity and prefer games that completely absorb their attention. Immersive game aspects lead to “Easy Fun” and generate emotions of Wonder, Awe, and Mystery.

- **Serious fun:** Players gain enjoyment from internal experiences as reaction to visceral, behavior, cognitive, and social aspects of games. Serious fun brings internal change by emotions such as Excitement or Relief from player thoughts and feelings.

- **The People Factor:** Players look for social experiences and use games as mechanisms to achieve them. These players enjoy the emotions of Amusement, Schadenfreude or pleasure gained from the misfortunes of others, and Naches or the pleasure a mentor gains from the success of the pupil. These emotions come from playing with others through competition, teamwork, as well as social bonding and personal recognition.

Despite not showing important personality-psychological conclusions, correlations to other models have been made[21]. Hard Fun describes more or less accurately the interests of Bartle’s Achievers or KTM’s Guardians. Easy Fun’s characteristic of immersion has been mentioned in Bartle’s Explorers and KTM’s Rationals. Reaction to visceral cognitive factors that leads to excitement, or Serious Fun, is the basis for Bartle’s Killers. Using games as social mechanisms for personal recognition follows the interests of KTM’s Idealists and social bonding is important for Bartle’s socializers. By considering these relations,
the model becomes interesting for understanding the entertainment values expected by players.

**Unified Model.** Comprehensive analysis has been made in order to propose a unified model [21] that can be used by the game industry, encompassing different personality and play style models. The article relates previous referred models, with the exception of FFM, as well as game design models such as Gamist/Narrativist/Simulationist[22] and Mechanics/Dynamics/Aesthetics[23] that are not considered in our work.

The first correlation is between Bartle’s MUD types and Keirsey’s KTM, defending that Temperaments are a superset of MUD types. This means that Socializers are a specific type of Idealists, Killers a type of Artisans, Explorers a type of Rationals and Achievers a type of Guardians. When examining their behaviors this relation becomes apparent for the author. Idealist’ social capabilities and their need to invest in relationships can also be a description of Socializers. Artisans’ impulsive reactions and freedom can be associated to the Killer methodology, where free will is exerted over other players regardless of consequences. Explorers’ curiosity to understand the nature and mechanics of games is reflected by the Rational temperament. A Guardian’s quest for validation of responsibilities is shared by achiever’s endeavors for higher score and titles, both being strict to duty.

By understanding the underlying motivation of each player, Lazzaro’s fun types can be associated with the player types of the Keirsey and Bartle models. Table 2.3 shows the relations between Keirsey, Bartle and Lazzaro models; it also presents their motivation, attitude towards problem-solving and overall goal of each player typologies.

<table>
<thead>
<tr>
<th>Keirsey</th>
<th>Bartle</th>
<th>Lazzaro</th>
<th>Motivation</th>
<th>Problem-solving</th>
<th>Overall Goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Artisan (tactical)</td>
<td>Killer</td>
<td>Serious fun</td>
<td>Power</td>
<td>Performance</td>
<td>Do</td>
</tr>
<tr>
<td>Guardian (logistical)</td>
<td>Achiever</td>
<td>Hard fun</td>
<td>Security</td>
<td>Persistence</td>
<td>Have</td>
</tr>
<tr>
<td>Rational (strategic)</td>
<td>Explorer</td>
<td>Easy fun</td>
<td>Knowledge</td>
<td>Perception</td>
<td>Know</td>
</tr>
<tr>
<td>Idealist (diplomatic)</td>
<td>Socializer</td>
<td>People fun</td>
<td>Identity</td>
<td>Persuasion</td>
<td>Become</td>
</tr>
</tbody>
</table>

Table 2.3: Unified Model relation between Keirsey, Bartle, Lazzaro models, and their typology characteristics. Adapted from [21].

DGD is also related, filling some gaps between Temperaments. This can be understood due to the dichotomies that each model chooses to approach, while DGD uses the T-F and P-J dimensions, Keirsey uses SJ-P and N-TF dividing the MBTI scales differently. Figure 2.2 shows the graphical comparison of
DGD and Keirsey Temperaments. To the author the most significant dimensions of human behavior are Internals (a preference for seeing possibilities and the abstract) vs. Externals (seeing the concrete and realistic), and Change (which can be thought of as freedom or opportunity) vs. Structure (which can be understood as rules or organization). The distinction between casual and hardcore skillsets is also represented, showing a non-linear division of the MBTI dichotomies. The dimensions are an interpretation by the author.

2.3 Entertainment Modelling

Our work tries to solve the problem of classifying players and increase the entertainment value of games. Entertainment modelling is a direct, practical application of our work. Modelling games can be done using different kinds of theories or empiric measures and it is crucial to understand previous work and research. In this section we present a series of studies on entertainment modelling that use different approaches, afterwards we make a simple comparison across studies.

**Rodrigo Dias’ Thesis.** Work done by Rodrigo Dias[8] validates the hypothesis that adaptive games increase their own enjoyment value. His work focuses only on how information and content should be presented once the system knows the personality traits of the player, leaving the problem of personality inferring to future work. He presents as solution a general architecture for adaptive games that is divided into two parts, offline information gathering and online adaptive game system. The offline part consists
on a knowledge base that holds information about player typology, preferences and metrics for distinguishing player types. This knowledge base is then accessed by the online part that adapts the game and iterates over the player model according to gathered data.

To build the offline part, an MBTI questionnaire was used to infer player dichotomies. The MBTI dichotomies are then used to obtain the player’s DGD personality type that holds information regarding preferences. With a solid database that relates preferences to DGD and MBTI types the online part can then adapt content.

The online part of the system is a continuous cycle divided into five modules, as shown in Figure 2.3, and as described below:

- ‘Retrieve player data and performance’: A learning module tracks and analyzes performance data.
- ‘Situate player in Experience Fluctuation Model (EFM)’: This module tries to identify player experience. This is done by inferring feelings from the intersection between given challenges and shown skills.
- ‘Re-define player personality’: Once enough data has been gathered, it can be correlated to previous knowledge built by the MBTI questionnaire. This correlation leads to a more accurate approximation of player personality.
- ‘Re-assign player type (DGD) and preferences’: Using the more accurate approximation of player personality the DGD type can be obtained. With the DGD type the game has access to a preference list that supposedly caters to the player’s interests.
- ‘Adapt game according to player type preferences and state in EFM’: The preference list is used to adapt the content of the game according to Flow theory.

In order to validate his solution, he developed a shooter game, Grim Business. The game monitors performance data and is able to adapt presentation, difficulty and control according to the four DGD player types. Validation was made using two groups. First, the DGD type of players in both groups was obtained through MBTI questionnaires. Afterwards, the first group played a version of Grim Business that adapted content to their own personality, and the second group played a version that adapted content to a different personality. Finally a survey was filled to evaluate the experience players had during gameplay.

The results of the test were consistent with assumptions of the DGD Model, such as Conquerors giving more importance to game mechanics/efficiency than visual effects, while Wanderers give higher importance to these effects. It was also able to corroborate the DGD model, despite having some wrong adaptations due to misinterpretation of the model.

Having validated the hypothesis that adaptive games can bring higher enjoyment it leaves for future work to deal with how to identify personality types dynamically. Our work intends to extend his findings by inferring personality typology for content adaptation. Dias’ research and findings not only help validate our motivation but also show a practical application for our work.
Studies by Georgio Yannakakis, et al. Georgio Yannakakis et al., have done extensive work on feature capture for entertainment modelling, using children as test users. In their endeavor[24] to better describe fun in games they followed a qualitative approach, overlapping Malone’s factors [25] for engaging game play with Csikszentmihalyi’s concepts of flow [4]. In [24], their objective was to define the minimal feature subsets that modeled children’s notions of fun. Their experiment was done on testbed using the Playware [26] playground that ran the Bug-Smasher game. Bug-Smasher is a game where bugs (colored lights) appear sequentially on a 6x6 mattress disappearing after a period of time.

During play the individual traits and performance of children were recorded and between each two sessions children were asked which was more fun. By using the gathered data as a training set for an Artificial Neural Network, an optimal subset was found that modeled the entertainment of seventy-two children with a cross-validation accuracy of 77.77% (binomial-distributed p-value = 0.0019).

A following study [9] of Yannakakis et al., extended previous work to understand the importance of physiological signals as features for entertainment modelling. Using a similar approach with the Bug-smasher game allowed the acquisition of data on Heart Rate (HR), Blood Volume Pulse (BVP) and Skin Conductance (SC) values. Their work revealed that features from HR and BVP correlated with expressed preferences and that Heart Rate Variability (HRV) constituted the best predictor of preferences.

Challenge modelling studies. A common type of entertainment modelling focuses on adapting the difficulty or challenge to the player. Different to the work of Dias, that also adapted presentation and controls, or the general modelling of entertainment by Yannakakis et al., Jenova Chen and Rani et al. explored different ways to adapt challenge exclusively.

In [27], Jenova Chen presented a unique way to adapt challenge. He presents design theory to
Develop active Dynamic Difficulty Adjusting (DDA). Active DDA attempts to immerse players in a state of flow by making them take subconscious decisions that lead to an ideal state.

In his thesis, he strongly criticizes passive DDA, shown in figure 2.4, to adapt challenge. In passive DDA data is monitored, filtered, analyzed and sequential parameterizations are made to the game. In defending his design, he points out four negative factors about passive DDA: no direct data - video games do not read what player thinks yet; performance does not mirror flow; analysis is based on assumptions; adaptation and changes are based on rigid designs.

To overcome these problems he designs and implements two games Traffic light and f10w.

Traffic light is a simple interaction game where the player has to click a button before a red light turns on, between rounds the player is asked if he wishes to go faster or slower. Testing using traffic light showed an extension of the games lifespan from 1-2 minutes to a total of 5-12.

In f10w the player controls a microscopic-like being trying to survive. The world is divided into different levels each with its own inhabitants. As the game goes on, the player can choose to confront other beings and devour them becoming stronger or simply run away. In each level special catchable creatures allow the player to change level increasing or decreasing the difficulty. An important aspect of f10w is that once damage is taken the player recedes in difficulty, making the challenge adjusted by design and player skill.

Similar to the work of Chen, Rani et al in [10] present the modelling of optimal challenge based on flow theory but enhanced by physiological feedback. The main objective was to ascertain whether an affective based adaptation could increase engagement from a performance-only approach. For this they recorded individual profiles of physiological response such as SC and HRV. After an extended period of physiological pattern gathering, testers were made to play an altered game of Pong that changed difficulty levels either according to used methodology, performance or anxiety.

An important conclusion reached was that physiological signals did indeed provide further data and increase the engagement level of users, meaning that they are powerful indicators of affective states.
This relation between affective states and physiological signals could also be used to gather data to cross with personality models. Unfortunately, due to the required profiling, the size of the population was considerably small (totaling at fifteen individuals).

**Story modelling studies.** While the work of Chen and Rani *et al.* focuses exclusively in adapting challenge to each player, other types of content can be adapted. PaSSAGE [28] is a storytelling modelling system built to complement difficulty adjustment. Developing games with extensive and meaningful story choices leads to an exponential increase in assets and expensive production. Typically, when playing games, the data regarding story preferences is ignored, presenting a limited path to the player. PaSSAGE tries to overcome this limitation by adapting storyline to the player. To do this, the player is classified according to a typology used in tabletop roleplaying games presented in [29]. Player actions are monitored by the system and weighted to identify the dominant typology, then modelling is done according to three stages: Selection, Specification and Refinement.

In Selection, decisions are made regarding the sequence of events that make up the story. PaSSAGE resorts to a library of possible events, named *encounters*, that are annotated with information concerning their relevance to each player typology. Each *encounter* can have additional *branches* with potential player actions. When choosing which event to run, PaSSAGE selects the *encounter* whose *branches* best fit the monitored typology. To ensure an adequate story structure, the events are lined according to Campbell’s monomyth[30], a recurring structure for storytelling.

In Specification, PaSSAGE delegates actor roles to the environment. It constantly monitors the potential actors of the environment using functions named *triggers* and creates suitable events using those actors. The potential actors are limited to those between the player’s starting and destination points. This method was used to provide the player with stories, appropriate to his topology, near his current location.

In Refinement, a technique called *hinting* is used to ensure that the player chooses to follow the created *branch*. *Hinting* changes the dialog and occurrences of the event to spike interest in the player. The changes are done by taking into account player preferences, explained by dominant typology, and are used subtly as part of the *branch* plot.

The system was tested with ninety university students to validate the hypothesis that the adaptive version of PaSSAGE brought greater entertainment, as well as a greater notion of agency. The results did in fact validate the hypothesis, and showed greater validation for female testers which is explained by a reduced need for control of the female audience in games.

**Comparison of Entertainment Modelling studies.** Table 2.4 shows a comparison between presented entertainment modelling studies. It is important to note that there are some similarities between the works of Dias and Yannakakis *et al.* and the works of Rani *et al.* and Chen. Despite working in a similar field, all approaches were different either in their audience or methodology to adapt content. These differences show that entertainment modelling is not a closed field of research and different ways to adapt content can lead to very different results.
Table 2.4: Comparison of presented entertainment modelling studies.

<table>
<thead>
<tr>
<th>Name</th>
<th>Modelled content</th>
<th>Theoretical background</th>
<th>Extracted features</th>
<th>Testers</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rodrigo (Grim Business)</td>
<td>Presentation, difficulty, controls</td>
<td>Flow, DGD</td>
<td>Performance</td>
<td>Male testers</td>
<td>20</td>
</tr>
<tr>
<td>Rani, et al.</td>
<td>Challenge</td>
<td>Flow</td>
<td>Performance, HRV, SC</td>
<td>Adults</td>
<td>15</td>
</tr>
<tr>
<td>Chen</td>
<td>Challenge</td>
<td>Flow</td>
<td>None</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PaSSAGE</td>
<td>Story</td>
<td>Law’s typology, Monomyth</td>
<td>Law’s typology, player decisions</td>
<td>University students</td>
<td>90</td>
</tr>
</tbody>
</table>

These studies also show that the work done in entertainment modelling is still focused on adapting challenge or difficulty. Flow appears as a recurrent theory due to its direct relation to game difficulty and balance. Affective data that can be acquired from physiological signals is also common as it is an interesting measure to complement performance.

For our work it is important to note that all presented studies suggest that extracting player personality can play an important part in increasing entertainment value. Dias’ thesis revalidates the need for detecting personality traits previous to content adaptation. The work of Yannakakis et al., despite showing positive signals still fails 32% of the time in modelling children; his conclusions indicate that potential data could be acquired from personality to increase its success rate. Chen comments that at runtime games still do not understand what players are thinking, something that can explained by personality theory. Rani et al. have shown how profiling is important in feature extraction as well as physiological response to map affective states. PaSSAGE used a rudimentary typology associated with tabletop roleplaying games to model storytelling.
Chapter 3

Solution

In the previous sections we presented the divergences across personality and player models common to the game industry. Due to the empiric nature of psychological studies no model has completely described the scope of the human psyche and ad hoc compromises must be met when choosing a model. For our work used the Keirsey Four Temperaments theory due to it being a behavior oriented model. Additionally due to its connection to MBTI dichotomies it can be associated to DGD and Bartle MUD types giving additional insight to expected temperament preferences in game.

Our main objectives are the development of a framework that allows the design of tasks to characterize player typology; and the development of a system that can, with a considerable degree of certainty, identify player personality using extracted data from task-based play sessions. We hypothesized that the choice of actions in a task based scenario is sufficient to extrapolate personality. For this, it was critical to use a model that not only described and related player actions to behavior, but simplified the design of coherent player tasks.

A more complete and extensive list of critical objectives is as follows:

- Identify psychological and player models that can be adapted into game mechanics and scenarios.
- Adapt the Keirsey Temperament model as a task framework that can be incorporated into games.
- Build a system with learning capabilities that can infer personality based on tasks chosen by players.
- Use the task framework to build a non-invasive scenario that can gather data and increase the efficiency of the classification system.
- Validate the system with extensive user testing using the built scenario.

In sum, a deep analysis into what tasks can be used to identify player personality will be followed by scenario testing using a learning system.

To reach our objectives, our solution is comprised of two modules: the inferring system and the game scenario. The inferring system keeps hold of a knowledge base of previous player actions and can perform probabilistic reasoning to infer temperaments. The scenario module is coupled with the game code
and keeps track of player decisions. To evaluate our solution the modules were developed and coupled with the game *Minecraft*.

*Minecraft.* *Minecraft* is an indie sandbox game, made by *Mojang* and released in 2011. Its open attitude towards modding, complex world-based mechanics, and iconographic aesthetics have made it a widely popularized success. The game has branched from the computer space to consoles such as the *Microsoft XBOX-360*. It is also being used for teaching collaboration and discussion at classrooms around the world [31].

The main game mechanics revolve around the breaking and placing of blocks in the world, as well as the acquisition of materials. The materials can then be used to create tools, weapons, armor or aesthetic objects. Populating the world are monsters, villagers and animals that enhance the player interaction. Since it has no concrete goals it allows players a large amount of freedom, from building a kingdom, as shown in figure 3.1, exploring dungeons, to creating new inventions with circuits.

In the remainder of this section we extensively describe our methodology for developing a *Minecraft* task-based scenario, from identifying tasks to implementing them. We proceed to explain the scenario architecture and its contents as well as the technological decisions made to couple it to *Minecraft*. Finally we describe the protocol used when testing, the environment where tests occurred, and the nature of the testers procured.

### 3.1 Design Methodology

We have mentioned that our solution is comprised of an inferring system model and a scenario environment. However, these comprise only a part of our research. As stated in our goals, we intend to develop a design methodology that could be replicated for different games and personality models, besides the final personality inferring system and environment.

This section will describe our methodology regarding task and scenario design so that the process
can be replicated. We also exemplify how the methodology was used to create *Rooms*, a scenario that coupled with *Minecraft* was used to evaluate our work. The methodology is comprised of four steps: *Task Identification; Observation; Task Clustering; Scenario Design.*

In *Task Identification*, we describe how we created the initial tasks that would later be refined, as well as the adversities of using these tasks for the final scenario.

In *Observation*, we explain how we set-up a small experiment to reduce subjectivity when pairing tasks with temperaments.

In *Task Clustering*, we exemplify how we reached task clusters using data resulting from the previous experiment, as well as the final relation between tasks and temperaments.

Finally in *Scenario Design*, we unfold how we designed, compared and chose from different possible scenario architectures, based on emergent patterns.

### 3.1.1 Task Identification

For the success of the inferring system, different choices need to be pleasing to different personalities. Therefore we must begin by establishing meaningful tasks that can be instantiated using the game engine. In order to establish the initial set of tasks we followed an iterative process of collecting tasks.

*Task Identification* can be understood as the initial process of developing tasks. In it, we must cross game mechanics and interactions with known information regarding personality and game model preferences. From this merger of information we collect a set of potential tasks for our scenario.

Inputs for this stage are all researched personality theories, both from psychology and game design. We process the inputs by identifying game mechanics, known *a priori*, that can be tied to the already known personality preferences. The mechanics are then translated as tasks that can be performed by the player.

At the end of *Task Identification* we have as output a task list of in game implementable tasks and their relation to the different spectrums of personality. This task list has to be extensive enough to cover most of the possible game mechanics and depends greatly on previous knowledge over the game.

*Minecraft* based tasks. For our solution, we began by taking advantage of the relation between Keirsey temperaments and other personality models. Special attention was given to Bartle MUD types and DGD, as well as their relation to the MBTI and Keirsey models. Initial project research indicated that Keirsey temperaments were the best personality model for our goals. Being a behavior oriented model, we could understand player personality from their in game choices.

We proceeded to look at the information regarding each temperament, and created a profile definition for their preferences. Guardians are goal oriented individuals who try to validate their in game duty by acquiring rare and powerful materials or achievements. Artisans are impulsive and enjoy combat and killing. Rationals seek to understand and explore both the world and game mechanics. Idealists enjoy artistic expression and are pulled away from straining dangerous situations such as combat.

Using these profile definitions we looked at each temperament preferences and *Minecraft* mechanics.
Table 3.1: Initial prospects on enjoyability of Minecraft tasks across temperaments. Tasks were based on game mechanics and temperament game preferences known by their link to Bartle types and DGD. This initial list is still crude and requires further refinement of its tasks before being used in the final scenario.

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Guardians</th>
<th>Rationals</th>
<th>Idealists</th>
<th>Artisans</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acquire all visible materials</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Avoid combat with weapon</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Avoid combat without weapon</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Build according to social constraints</td>
<td>Yes</td>
<td>Unknown</td>
<td>No</td>
<td>Unknown</td>
</tr>
<tr>
<td>Build complex structures</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Build original creations</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Compulsive organization</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Eat</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
</tr>
<tr>
<td>Explore</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Follow goals</td>
<td>Yes</td>
<td>No</td>
<td>Unknown</td>
<td>Unknown</td>
</tr>
<tr>
<td>Ignore NPCS</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Improve armor</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Improve tools</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Interact with animals</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Kill animals for fun</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Kill animals for Need</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Kill enemies with weapon for fun</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Kill enemies with weapon for need</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Make utilities</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Play with physics</td>
<td>No</td>
<td>Yes</td>
<td>Unkown</td>
<td>No</td>
</tr>
<tr>
<td>Play with recipes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Play with water</td>
<td>No</td>
<td>Yes</td>
<td>Unkown</td>
<td>No</td>
</tr>
<tr>
<td>Search for materials</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Use aesthetic materials</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 3.1 shows the initial list of tasks as well as the expected enjoyment of each task by each temperament. Theory does not possess full information regarding some combinations of tasks and temperaments, but the purpose of reaching this initial list is to have working tasks to refine in a further process that we define as Task Clustering.

Once again it is crucial to state that this is not the final list of tasks nor, does it represent the real preference of temperaments. This initial process of identifying tasks is still crude and based on subjective interpretation. It can therefore contain wrong assumptions and needs to be validated by experimental observation.

Another important concern regarding our task collection is that the link between Keirsey temperaments and DGD or Bartle types appeared to facilitate the designing of games. This means that despite being well documented, the unified theory was built for ad-hoc practice rather than full academic investigation.

Either due to subjective interpretation or forced correlations across theories, our initial task list requires validation. The following sections, Observation and Task Clustering, will explain how we set up...
an experiment to validate our task list and how it was further refined and simplified for its final version.

3.1.2 **Observation**

Our initial task list was collected by crossing knowledge regarding player preference and game mechanics. Each task should incite an expected enjoyment, either positive or negative, on different personalities. The enjoyments reflected in this initial list are a starting point and need to be polished and supported by empirical data. Two issues can be raised over the soundness of this list.

The first is the crudeness and subjectivity inherent to the initial collecting process. Insight of the researcher is required to do the crossing between theory and game mechanics, which can lead to false assumptions over player enjoyment.

The second is that for our tasks we are bringing theory from different game models that can be associated with Keirsey Temperaments. These associations were not made for extensive academic investigation but for the practice of making videogames.

It became necessary to develop an experiment to corroborate our assumptions as well as to reaffirm the usefulness of Keirsey Temperaments as a behavioral based model that has connections to game mechanics. The experiment consisted on observing the different personality temperaments while playing the game that would be used for the scenario. Player’s personality was measured using a personality sorter, found in annex B, and their actions were registered for further analysis.

**Experimental logistics.** The experiment required the most recent version of *Minecraft* that was supplied to the observant. The nature of the experiment allowed both observations in person as well as through the internet using the *Skype* platform.

To avoid influencing the observer by giving them a goal, the experiment had to be unstructured, allowing complete freedom within the game world. Participation by means of conversation was mandatory, this directive was necessary to understand the underlying motivations behind player actions, giving some insights to the non-behavioral aspects in play.

**Experimental procedure.** The experiment lasted between thirty to forty minutes, and could be divided into three phases: **Explanation**, **Questionnaire** and **Play**.

**Explanation** served as a preliminary phase to explain the layout of the experiment to the observant. This phase took no longer than three minutes as it was merely informative and attempted to satisfy any questions the observant might have had.

During the explanation, the observant was informed that the experiment would take between thirty to forty minutes, during which time they would fill out a personality survey and be allowed to play *Minecraft* freely. They were also informed that questions were allowed and encouraged to help to obtain data.

At this stage it was required to understand whether the observant was already familiar with the game. In the case of the observant being unfamiliar with the game, a brief introduction to the game was required in the **Play** stage.
Once the experiment was explained, in Questionnaire, players were asked to fill in a personality questionnaire. A seventy question MBTI personality survey was used, since it is possible to extract Keirsey Temperaments directly from the MBTI typology. The reason as to why an MBTI survey was used lies solely in the ease of acquiring a free and already validated survey, presented in annex B.

Once the observant filled all questions, the data was input to a previously arranged spreadsheet that returned the corresponding Temperament. The observant was then informed of his or her corresponding personality without being informed of its meaning in terms of behaviors. This restraint of information was intentional to avoid influencing behavior and player decisions.

At the beginning of Play a brand new world of Minecraft was generated. This world was different across all players so as to ensure that they began with a blank slate.

If the individual was a new player, a simple introduction to the key bindings of the game, as well as the recipe engine and general mechanics, was required. These players would spend at least five minutes adjusting themselves to the game and its mechanics. Players were then asked to play in their world. For the whole duration of the play session, player actions were registered according to our initial task list.

Constant conversation helped align actual decisions with the expected actions from the list. Whenever the motivation behind an action was not understood, further questioning of the player was required.

Often, during early tests, players performed actions that were not on the list. These new actions could either be a subset of already existing tasks or brand new ones. In the case of new tasks, these were added to the list and considered for the following experiments. These initial samples were not considered for the experimental results section.

**Experimental results.** The initial observation experiment approached a total of 10 individuals, from both sexes. Most of the individuals were part of different degrees in the field of Computer Engineering. Temperament distribution of testers consisted of 60% Guardians, 20% Rationals, 10% Idealists and 10% Artisans. Table 3.2, however, shows the standard societal distribution of temperaments, as considered in [18]. The surplus of the Guardian and Rational Temperaments as well as the lack of Artisan Temperament can be explained by having collected the data of individuals from a particular field, Computer Science.

To understand the impact of chosen tasks on each temperament we explain the most relevant experimental results, divided by temperaments.

**Guardians.** The Guardian temperament was overly represented in our experiment. Of all individuals, 60% were Guardians. Both experienced players as well as new players were observed. Experienced players accounted for 33% of the temperament’s population, while new players accounted for 66%.
Guardians have a preference for following goals, acquiring better weaponry and armor and collecting points. Expected most chosen actions included ‘Follow goals’, ‘Improve tools’ and ‘Improve armor’.

Figure A.1 in Annex A shows the incidence of chosen actions for Guardians, both expert and new. The most frequent actions performed where ‘Follow goals’, ‘Improve tools’ and ‘Avoid combat without weapon’. Out of the expected actions only ‘Improve armor’ was never done.

Expert and new players present very different actions. ‘Follow goals’ and ‘Avoid combat without weapon’ seem to be the only common actions for this temperament. These can be a strong indicative of motivators for the Guardian.

Regarding combat, the Guardian seems to enjoy it as long as he possesses weaponry. Since ‘Kill x for fun’ actions were performed more than the ‘Kill x for need’, we believe the Guardian has intrinsic motivation for fighting. This temperament appears to be evasive when not holding a weapon but enjoys combat otherwise.

By looking at actions that refer to building, we can understand that this temperament has no aversion to construction but neither does it have a preference. We should expect that this temperament will only choose building constructions after exploration and combat with weaponry.

**Rationals.** 20% of players belonged to the Rational temperament. Both experienced players, as well as new players, were observed. Expert and new players were balanced, each being 50% of observed Rationals.

Rationals have enjoyment for discovery of both the world and game mechanics. The temperament was expected to enter combat only when necessary and to enjoy obtaining materials. Expected chosen actions included ‘Explore’, ‘Play with recipes’ and ‘Play with water’ or ‘Play with physics’.

Figure A.2 in Annex A shows, in percentages, the incidence of chosen actions for Rationals. Preferences for this temperament are ‘Search for materials’, ‘Make utilities’, and ‘Play with recipes’. Unexpectedly, this temperament did little interaction with water and physics, but due to the initial nature of the setting, players lacked proper tools that allowed that interaction.

As expected, this temperament has preference for exploration, ‘Search for materials’ and ‘Explore’ appear high as selected actions. Also expected was the exploration of the recipe engine which had a greater preference by new players.

‘Make utilities’ ranked very high on the list of preferences. When observed, players seemed to understand that developing utilities led to more resources and consequentially more game content.

Regarding combat, this temperament avoids it even when holding a weapon. New players still found fun in killing animals but expert players only killed animals for need. We understand that this difference between new and expert players is related to new players still requiring to learn the killing mechanics of the game.

Building actions appear spread across the middle of the graphic with ‘Build according to social constraints’ higher in the list. We can understand that this Temperament, similar to the Guardian, has neither an aversion nor preference to building.
Artisans. Contrary to Guardians, Artisans were under represented in our experiment. Gathered testers came from a strong IT background, where this temperament seems to be lacking. Only one observant, 10% of total, was an artisan. The artisan in question was also an expert at the game.

Artisans are whimsical and act according to will, they enjoy combat and unexpected novel situations. Actions that involved ‘Kill x for fun’ were expected to be the most chosen.

Figure A.3 in Annex A shows the incidence of chosen actions for Artisans. Preferences for this temperament were ‘Acquire all visible materials’. ‘Kill enemies with weapon for need’ and ‘Kill enemies with weapon for fun’.

As expected, Combat is part of Artisans main choices. It is interesting to note that ‘Kill animals for fun’ did not rank as high, during play. The player in question always killed animals on sight but mostly to acquire food. ‘Avoid combat without weapon’ ranks high while ‘Avoid combat with weapon’ was never done, this indicates that when holding weapons the artisan enjoyed combat and procured it.

‘Acquire all materials’ ranked top which was highly unexpected. Information on more artisans would be required to fully understand whether or not this plays an important part on their motivation.

Regarding building, this temperament seems to have less of an interest on such actions than the previous Guardian and Rational temperaments. We can assume that actions referring to building are not part of the Artisans interests.

Idealists. Similarly to Artisans, Idealists were hardly represented in our experiment but unlike Artisans, 10% is closer to their societal distribution. The single Idealist in question was a new player.

Idealists have a preference for self-discovery by expression and an aversion to conflict. This means that actions such as ‘Build original creations’, ‘Use aesthetic materials’ and ‘Interact with x’ would be the most frequent, while combat the most avoided.

Figure A.4 in Annex A shows the incidence of chosen actions for Idealists. Preferences for this temperament include ‘Search for materials’ and ‘Build original creations’. Unexpectedly, no ‘Interact with x’ actions occurred, more data would be required to confirm if Idealists enjoy interaction with Non Playable Characters (NPC), being they creatures or humanoids.

As expected this temperament focuses a lot on building and chose very little of other actions. ‘Search for materials’ was done in order to acquire building materials reinforcing this idea. ‘Use aesthetic materials’ did not rank as high as expected, but higher than in any other temperaments.

This temperament shows a clear aversion towards combat. It constantly acted ‘Avoid combat without weapon’ and never felt the necessity to develop weapons nor attack other creatures for fun or need.

Exploration does not seem to be of much interest to Idealists as ‘Search for materials’ was done to acquire building blocks. Another fact that supports this theory is that Idealists did not worry about playing with water, physics or recipes.

Observation Discussion. All gathered data was helpful in profiling the temperaments according to Minecraft mechanics. In the case of the Artisan and Idealist temperament, additional information would help resolve some questions that were raised. It is important to remember that the profiles below
are not definitive conclusions but part of exploratory research, especially so with the less represented temperaments.

Guardians have a greater preference for following goals and improving their tools. They can enjoy combat when given fighting gear and avoid it otherwise. Regarding exploration they show some interest in it, especially with game mechanics. Having rooms with clear goals should attract Guardians easily.

Rationals seem to enjoy exploring both the world and its mechanics. Combat and building seem to be secondary to the temperament. This indicates that for the scenario, they should be given the choice of a room with materials and mechanics to explore.

Artisans clearly enjoy combat regardless of being armed or not, building seems secondary. The scenario should cater to their visceral needs by providing clear combat choices.

Idealists prefer building original creations. Combat is detested by this temperament which distinguishes it from the others. Exploration is done only to acquire new materials and the scenario should provide rooms filled with aesthetic materials to lure in Idealists.

These general profiles can be used to refine tasks when developing the scenario. They also give us some rules of thumb to separate temperaments. The rules are as follows:

- Separating Guardians from Artisans could be done by providing combat choices that do not provide weapons. Results from this experiment show that Artisans will continue to crave the fight while Guardians might choose something else.

- Separating Guardians and Rationals should be done through goals. Providing stale rooms that fulfill some long term goal without allowing interesting explorations, and rooms with the opposite properties, should appeal differently to both temperaments.

- Separating Guardians from Idealists should be done by providing rooms with conflict and goals versus rooms that serve no long term goal but provide experimentation with building and self-expression.

- Separating Artisans from Rationals and Idealists should be done by use of combat/conflict rooms. While Artisans adore combat, Rationals are only interested in it for need while Idealists seem to be completely turned away by it.

- Separating Rationals from Idealists can be difficult since Rationals can enjoy building to some extent. Providing rooms with stimulating but dangerous mechanics versus rooms that allow self-expression is a viable option. It should be done by providing rooms with exploration of violent mechanics, such as lava or explosions; ideally Rationals will enjoy playing with mechanics while Idealists can feel some aversion to the conflict and danger.

**Observation Conclusions.** The process of Observation gathered different individuals, classified their personalities according to the Keirsey Temperament model and observed their actions while playing Minecraft. These actions helped gain some understanding of the preferences of each temperament.
From the outcome we were able to establish a set of rules of thumb for separating personalities using *Minecraft* elements. We expect these to work better with the Guardian and Rational temperament as they were more represented in the tested population.

Despite all findings, some doubts remain for the Artisan and Idealist Temperaments, who were underrepresented in the tested population. These doubts are why Artisans focused on gathering materials? And why Idealists did not concern themselves more with aesthetic materials compared to other chosen actions? More data might be required to understand these questions and further experimentation is advisable.

At this point we possess a list of tasks and rules to separate the temperaments. The list is still extensive for our purposes and needs to be decreased. For this we resort to clustering, a process that takes advantage of acquired data from the experiment and compacts the list into shorter clusters, that have a stronger relation to each temperament.

### 3.1.3 Task Clustering

The initial set of tasks is excessive for the purpose of classifying the typologies. Using the data collected in *Observation* we can resort to clustering algorithms to obtain clusters. These clusters agglomerate tasks according to personality preferences making them smaller and more manageable information that is easier to use and analyze.

Ideally, we wish to achieve four clusters, each with tasks that were performed more often by a single temperament. We use the K-Means cluster algorithm to successively calculate different cluster combinations until we reach a $K$ value that satisfies our needs, $K$ being the number of clusters the algorithm wants to achieve. We begin with $K = 4$ since it is the optimal required number of clusters to distinguish the four temperaments.

We used SPSS Statistics 2.0 software from IBM to calculate the K-Means clustering algorithm for the several values of $K$. To do this we input data from our previous experiment using our list of tasks and corresponding frequency. The software then calculates $K$ clusters by grouping actions according to their proximity within four categories – the temperaments.

It was required to fix a percentage value that reflected the dominant preferences of a temperament. We believe that 25% can be used to identify the dominant clusters as it can create ambiguity. This ambiguity in preference can be used as self-assessment tool over the usefulness of collected data. If there are less temperaments preferring the same dominant cluster then the data is more useful for our purposes.

Results indicated that the four clusters obtained were not enough to fulfill our needs. The first cluster was the only dominant choice for the Artisan temperament but was also the dominant choice for the other three. This meant that we did not have a unique dominant cluster for each temperament.

We increased the value of $K$ to five and repeated the process to obtain the results in table 3.3. Five clusters were enough to distinguish the temperaments. Rationals preferred clusters 2 and 4, while Idealists clusters 3 and 5. Guardians and Artisans both preferred the same clusters, 1 and 4, however
Table 3.3: Frequency of temperament actions over the five clusters obtained from K-Means Clustering, k=5. Frequencies superior to 25% are shown in bold and are indicative of temperament preference. These clusters were obtained using IBM SPSS 2.0. These five clusters are enough to distinguish temperaments; Rationals and Idealists have unique preferences, while Artisans and Guardians have different highest cluster preference.

Table 3.4: The five clusters achieved in Task Clustering and their corresponding tasks. Obtained from SPSS2.0 by performing K-Means Clustering using our Observation results as input, Temperaments as proximity categories, and k=5. Each cluster agglomerates tasks closer to similar temperament preferences.

<table>
<thead>
<tr>
<th>Cluster ID</th>
<th>Guardians</th>
<th>Rationals</th>
<th>Idealists</th>
<th>Artisans</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25.90%</td>
<td>5.19%</td>
<td>9.09%</td>
<td>50.00%</td>
</tr>
<tr>
<td>2</td>
<td>14.46%</td>
<td>37.66%</td>
<td>0.00%</td>
<td>5.88%</td>
</tr>
<tr>
<td>3</td>
<td>5.42%</td>
<td>12.99%</td>
<td>36.36%</td>
<td>2.94%</td>
</tr>
<tr>
<td>4</td>
<td>40.36%</td>
<td>35.06%</td>
<td>18.18%</td>
<td>38.24%</td>
</tr>
<tr>
<td>5</td>
<td>13.86%</td>
<td>9.09%</td>
<td>36.36%</td>
<td>2.94%</td>
</tr>
</tbody>
</table>

their highest preferred cluster was not the same. Taking into account these preferences we would be able to instantiate the tasks of the scenario by using these five clusters.

These five clusters are the final result of our clustering process. Each is tied to a sublist of our initial tasks as presented in Table 3.4. These clusters also possess empirical information over the preference of the different temperaments. We should point out that the clusters were achieved after observing only 10 individuals. For future repetitions of the process we recommend achieving a higher number of individuals during the observation process to achieve better clusters.

<table>
<thead>
<tr>
<th>Cluster ID</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Eat</td>
</tr>
<tr>
<td></td>
<td>Kill Enemies with weapon for fun</td>
</tr>
<tr>
<td></td>
<td>Kill Enemies with weapon for need</td>
</tr>
<tr>
<td></td>
<td>Avoid Combat without weapon</td>
</tr>
<tr>
<td></td>
<td>Acquire all visible Materials</td>
</tr>
<tr>
<td>2</td>
<td>Explore</td>
</tr>
<tr>
<td></td>
<td>Make Utilities</td>
</tr>
<tr>
<td></td>
<td>Avoid Combat with weapon</td>
</tr>
<tr>
<td></td>
<td>Play with recipes</td>
</tr>
<tr>
<td>3</td>
<td>Search for Materials</td>
</tr>
<tr>
<td>4</td>
<td>Build Complex Structures</td>
</tr>
<tr>
<td></td>
<td>Build According to Social constraints</td>
</tr>
<tr>
<td></td>
<td>Follow Goals</td>
</tr>
<tr>
<td></td>
<td>Improve Tools</td>
</tr>
<tr>
<td></td>
<td>Kill Animals for fun</td>
</tr>
<tr>
<td></td>
<td>Kill Animals for Need</td>
</tr>
<tr>
<td>5</td>
<td>Build Original creations</td>
</tr>
<tr>
<td></td>
<td>Compulsive organization</td>
</tr>
<tr>
<td></td>
<td>Improve Armor</td>
</tr>
<tr>
<td></td>
<td>Interact with Animals</td>
</tr>
<tr>
<td></td>
<td>Ignore NPCs</td>
</tr>
<tr>
<td></td>
<td>Play with Physics</td>
</tr>
<tr>
<td></td>
<td>Play with water</td>
</tr>
<tr>
<td></td>
<td>Use Aesthetic Materials</td>
</tr>
</tbody>
</table>
separate temperaments and a set of clusters that can be assigned to the different rooms. All that is left, is reaching a functional scenario architecture, populate it with the necessary clusters and choose specific tasks from the clusters when implementing the scenario.

3.1.4  **Scenario Design**

Different players enjoy doing different tasks and our work hypothesizes that task preference can be an indicator of personality. We have explained our process to develop tasks, from the initial observational experiment to task clustering. Now it was required to design a scenario that allocates task clusters and presents them efficiently to the player. This section describes our scenario design process, as well as the different architectures considered for our final experiment.

The goal of developing a scenario is to allow player choice between several clusters. At each stage the player is presented with tasks and he is free to choose which he enjoys best. This follows our hypothesis that decisions can indicate personality and that data acquired from player decisions can be used as evidence for our network to infer personality.

Our scenario design process is comprised of two stages - Architecture Design and Task Assignment. In Architecture Design the architecture of the scenario is fleshed out to meet the experiment goals by analysing various options available; In Task Assignment the scenario is populated with clusters by looking at emergent patterns from expected selections.

**Architecture Design.** As previously stated, each scenario must present tasks as choices to the player. As the player moves through the scenario choices can appear at different moments. These moments can be the choice of the next room, or choices pertaining the room the player is in now. These possible interactions need to be defined and analyzed before the scenario is built and they can be understood as designing an architecture for the scenario.

Architecture Design of a scenario is a stage for deciding where the player choices are located and how the player is transposed from choice to choice. This stage does not require task clusters, but they need to be assigned to the architecture to achieve the final design of the scenario.

We began by considering that each room would present the players with several tasks. The number of tasks in a room can vary according to the experiment, for ours we established two tasks per room. This decision was taken to simplify the parsing of data from each room.

It becomes important to analyze how the player traverses the scenario and how choices are placed. For this, two questions must be posed:

- Should the player choose which set of tasks to do next?
- If the player decides which tasks to do, should he also choose which tasks he preferred?

The first question leads us to consider how players move from room to room. Our first design was similar to a binary decision tree as depicted in figure 3.2 where after finishing a room the player would be asked to choose the next one from a set of two.
Figure 3.2: Binary Tree Architecture. This initial architecture has an excessive number of rooms, $2^{h+1} - 1$ rooms and $2^h - 1$ decisions for a height of $h$. By allowing players to choose the next room the system becomes too dependent on the personal interpretation of each description.

Figure 3.3: Collapsed Rooms Architecture. This architecture collapses the rooms from the Binary Tree Architecture to achieve $2h + 1$ rooms, and $h$ decisions, for a tree height of $h$. Similar to the Binary Tree Architecture, in figure 3.2, the system is too dependent on the personal interpretation of each description.

This binary tree architecture has the disadvantage of having an excessive number of rooms to implement, where a height of $h$ would lead to $2^{h+1} - 1$ rooms and $2^h - 1$ decisions. Another issue arrives with describing the following rooms. This description becomes subjected to personal interpretation and can influence his decisions in unexpected ways.

Another possible choice would be to collapse decisions, rejoining the player at a common decision hub after each room, as depicted in figure 3.3. This architecture reduces the number of necessary rooms to $2h + 1$ rooms and $h$ decisions for $h$ tree height. However, the architecture does not fix the problem of subjectivity in interpreting descriptions.

Looking at the second question can give insight over this recurring problem. Instead of focusing on room choices, we should focus on task choices and player preference after he has interacted with each task.

Figure 3.4 depicts an architecture comprised of a sequence of rooms that do not require choice to enter but force choice on exit over the task the player preferred. This helps reduce the issue of creating appropriate descriptions but raises other issues.

This last architecture defers decisions to the moment when tasks have been experienced and is thusly dependent on the implementation of each task as mechanics. If tasks are overly attractive or repulsive, different temperaments can all end up taking the same choice. When this happens we consider
that we have a malformed room, a room whose implementation does not add useful information to our network by not achieving its design goal.

For our scenario we decided to use the last architecture. We expect the player to cross a sequence of rooms, and decide after experimenting with each task in a room. We deal with the possibility of malformed rooms with statistical analysis over play session samples. The exact method is explained in the Data Analysis chapter.

Having the architecture and task clusters we near the final stage of our scenario development methodology. All that is left is to identify and allocate the combinations of clusters that reveal the most information about the temperaments.

**Task Assignment.** Having both the scenario architecture and usable task clusters, we can build the final scenario model. This was done by assigning the clusters obtained from *Task Clustering* to the rooms of the architecture obtained from *Scenario Design*.

In our final scenario, each room will contain exactly two tasks, so we begin by combining task clusters in pairs. If we had kept all tasks we would reach 276 possible combinations. Since we only have five clusters this leads to $5 \choose 2 = 10$ possible combinations. To understand whether these combinations are interesting, we need to remember the cluster preferences of each temperaments and compare them when paired. The comparisons are depicted in table 3.5.

**Emergent Patterns.** The table also shows how the different temperaments are expected to behave when choosing between different cluster pairings. From these temperament preferences some patterns emerged:

1. Separate Guardians and Artisans from Rationals;
2. Separate Guardians and Artisans from Idealists;
3. Separate Rationals from Idealists;
Temperaments

Cluster A: that prefer A
1) Guardian, Artisan
2) Rational
3) Idealist
4) Guardian, Artisan, Rational
5) Idealist

Cluster B: that prefer B
1) Rational
2) Guardian, Artisan, Rational
3) Idealist
4) Guardian, Artisan, Rational
5) Idealist

Emergent patterns
1) Separate Guardians and Artisans from Rationals
2) Separate Guardians and Artisans from Idealists
3) Separate Guardians and Artisans from Idealists
4) Separate Rationalists and Idealists
5) Separate Idealists

Table 3.5: Possible combinations of the five clusters and associated temperaments. From the different combinations of temperament preferences, five patterns emerged, these will be assigned to the architecture obtained from Scenario Design.

4. Isolate Idealists.

Each of these patterns brings an interesting pairing that can be used for a room in the final model and define the goal of each room. Each pattern has two possible cluster combinations so we chose the combination that ensured that each cluster is used at least once with as little repetitions as possible.

Task selection. Once the cluster pairings of each pattern were chosen, we proceeded to choose a specific task from each cluster. This process can be done by choosing tasks that align with the goal of the room. Each cluster attracts specific temperaments so tasks that do the same are the appropriate choice. We resort to the information shown in table 3.1 and retrieve tasks that appeal to the same temperaments as their cluster. Ideally the chosen task should also be uninteresting to temperaments that are meant to choose the other side of the room.

This leaves us with the first four rooms of our solution. There is, however, an issue regarding the emergent patterns: Guardians and Artisans appear coupled. It is important to remember that these clusters came from an experiment that had little representation of the Artisan temperament. We took this into consideration and to the four rooms that came from emergent patterns we developed additional rooms to try and separate Artisans from the other temperaments.

To separate the Artisans from other temperaments we went back to our tasks list. Then we tried combining tasks that attracted Artisans and repulsed a specific Temperament, with tasks that did the opposite. We chose a total of six tasks that would occupy three additional rooms leaving our scenario with the total of seven rooms.

Having all tasks chosen we were then able to implement the scenario on Minecraft. The implementation process and description of each room are presented in the next section.
3.2 Solution Scenario

Data acquisition for the inferring system required a multi-choice scenario that we entitled *Rooms*. By following the process described in the previous sections we were able to reach a total of seven task rooms. Additionally the final scenario contains an entrance and an exit. *Rooms* is set so that the player begins at the entrance and must cross all task rooms before being allowed into the exit.

The entrance contains the starting point and is filled with information referring to the different elements that compose the scenario and how the player should act. It also contains checkpoints; these are special pads that allow the player to return to the previous room in case of avatar death, shown in figure 3.5. During his first stay at the entrance the player is asked to leave by a teleportation pad, these will always allow progress to the next task room or exit.

As seen in figure 3.6, each task room can be divided into an entrance, two tasks, a task hallway, two decision hallways and an exit. On entry, the player can choose which task to do first. At any point he or she can use the task hallway to change tasks. Once both tasks have been completed or abandoned the player is asked to make a decision regarding the most enjoyed task. Once the decision has been made, the player exits by the corresponding decision hallway and can proceed through the exit. Finally he can use the teleportation pad on the exit to move to the next room.

**Room Tasks.** Special attention was taken when selecting room tasks. The choice of preferred task aligned with player personality is used to feed sample data to our Bayesian network. To allow effective inferring, the fourteen tasks divided across seven rooms were paired so as to filter different personalities.

Table 3.6 shows how tasks are distributed across rooms. Each room intends to separate different temperaments to increase the precision of the inferring system. This was the final scenario attained from the Design Methodology chapter.

Room 1, attempts to isolate Idealists from the other three temperaments. The tasks it contains are ‘Kill Animals for need’ and ‘Use Aesthetic Materials’. The first task should create aversion to the Idealists mindset whilst the second pleasure. From this room we should gather higher Idealist preference for the
Figure 3.6: Generalized task room architecture of the solution scenario. This architecture is used to present two simultaneous tasks to the player, as well as to force a decision over preferred task on exit. Players start at the entrance and choose a task. At any point they can move from one task to the other by the task hallway. Finally they choose their preferred task by going to the exit by the adjacent decision hallway.

<table>
<thead>
<tr>
<th>Room</th>
<th>First Task</th>
<th>Second Task</th>
<th>First Temperaments</th>
<th>Second Temperaments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Use aesthetic materials</td>
<td>Kill animals for need</td>
<td>Idealist</td>
<td>Remaining Temperaments</td>
</tr>
<tr>
<td>2</td>
<td>Play with recipes</td>
<td>Build original creations</td>
<td>Rational</td>
<td>Idealist</td>
</tr>
<tr>
<td>3</td>
<td>Kill enemies with weapon for fun</td>
<td>Explore</td>
<td>Artisan, Guardian</td>
<td>Rational</td>
</tr>
<tr>
<td>4</td>
<td>Kill enemies with weapon for need</td>
<td>Play with water and lava</td>
<td>Artisan, Guardian</td>
<td>Idealist</td>
</tr>
<tr>
<td>5</td>
<td>Kill enemies with weapon for fun</td>
<td>Search for materials</td>
<td>Artisan</td>
<td>Guardian</td>
</tr>
<tr>
<td>6</td>
<td>Make utilities</td>
<td>Kill animals for fun</td>
<td>Rational</td>
<td>Artisan</td>
</tr>
<tr>
<td>7</td>
<td>Improve tools</td>
<td>Build original creations</td>
<td>Artisan</td>
<td>Idealist</td>
</tr>
</tbody>
</table>

Table 3.6: Distribution of tasks across rooms of the solution scenario. Each room contains two specific tasks that intend to separate different sets of temperaments. The first and second tasks are preferred by the first and second set of temperaments respectively.
second task. Remaining temperaments are expected to choose arbitrarily but with a higher inclination for the second task, due to its more violent nature. Its implementation is shown in figure 3.7.

![Image](image1.png)  
(a) ‘Kill Animals for need’  
(b) ‘Use Aesthetic Materials’

Figure 3.7: Implementation of room 1 in the Minecraft scenario.

Room 2 attempts to separate Rationals from Idealists. The tasks chosen are ‘Play with recipes’ and ‘Build original creations’. The first and second task should appeal more to Rationals and Idealists respectively. Guardians and Artisans are expected to choose arbitrarily. Its implementation is shown in figure 3.8.

![Image](image2.png)  
(a) ‘Play with recipes’  
(b) ‘Build original creations’

Figure 3.8: Implementation of room 2 in the Minecraft scenario.

Room 3 attempts to separate Artisans and Guardians from Rationals. The tasks chosen are ‘Kill enemies with weapon for fun’ and ‘Explore’. Artisans and Guardians should enjoy the visceral first task while the curiosity of Rationals should make them enjoy the second task. Idealists are expected to prefer the second task due to its non-violent nature. Its implementation is shown in figure 3.9.

Room 4 attempts to separate Artisans and Guardians from Idealists. The tasks chosen are ‘Kill enemies with weapon for need’ and ‘Play with water and lava’. Similar to room 3, Artisans and Guardians are expected to prefer the more visceral but also goal oriented first task while Idealists are expected to choose the second task due to its less violent nature. Rationals are expected to choose arbitrarily. Its implementation is shown in figure 3.10.

Room 5 attempts to separate Artisans from Guardians. The tasks chosen are ‘Kill enemies with weapon for fun’ and ‘Search for Materials’. Artisans are expected to prefer the more visceral first task while Guardians should prefer the goal oriented second task over the goalless first task. Rationals...
and Idealists are expected to prefer the second task, due to its non-violent exploratory nature. Its implementation is shown in figure 3.11.

Room 6 attempts to separate Rationals from Artisans. The tasks chosen are ‘Make utilities’ and ‘Kill animals for fun’. Rationals are expected to prefer the first task due to it requiring discovering new utilities. Artisans are expected to prefer the second due to its violent and schadenfreude nature. Guardians are expected to choose arbitrarily while Idealists should opt for the first task due to its non-violent nature. Its implementation is shown in figure 3.12.

Room 7 attempts to separate Artisans from Idealists. The tasks chosen were ‘Improve Tools’ and ‘Build original creations’. Artisans should choose the first task due to its potential to create harm (since
tools refers to weapons and armor). Idealists should prefer the second task due to its creative nature. Guardians are expected to prefer the first task, since it is goal-driven while Rationals are expected to choose arbitrarily. Its implementation is shown in figure 3.13.

Once all rooms have been crossed and enough data on temperaments has been gathered, it is expected for decision patterns to emerge. These form the knowledge base of our system that can, with new sample data, determine player personality.

**Implementation.** Building tasks that can relate to every temperament requires mechanics and variety that can be found more easily in the sandbox genre. *Minecraft* fits well into our work since it allows the crafting of a game world without the need for implementation of visual content or gameplay. Its content caters to different play styles and the world can be populated according to the necessity of the tasks.

The vanilla version was edited using the *Minecraft Coder Pack* (MCP) tools that allow decompilation and recompilation of the source code into Java format. Using the *Eclipse IDE*, a series of teleportation mechanisms were implemented to allow the decoupling of rooms and the logging system for the experiment sessions.

The creation and furnishing of each room was accelerated by resorting to *MCEdit*, an open source world editor for *Minecraft*. This tool allowed the saving and copying of template structures, as well as the mass substitution of blocks. Each room started from a previously made template and was augmented
3.3 Inferring System Model

Personality inference can be understood as an uncertainty problem. The goal of our system is to ascertain and decide what temperament better fits the player. We hypothesized that task preference in a game could be an indicative of personality. Player preference however does not always follow a logical route. To overcome subjective interpretations the system was built around probabilistic reasoning. This section will describe this system, its functionality and general conceptual model.

Probabilistic reasoning[32] is an area of investigation in the field of artificial intelligence that deals with uncertainty in data by using probabilistic analysis. It allows machines to bypass the bipolar states of true and false by analysing sets of data and inferring the maximum optimal output.

In our system, probabilistic reasoning is modeled with a Bayesian network, a probabilistic graphical model that maintains the state of random variables and their conditional dependencies by use of an acyclic graph.

**Network Model.** To design the network we looked at the task rooms in our final scenario and for each one created a node with its two outcomes. Figure 3.14 shows a generalization of our network. The parent node, Temperaments, has four possible outcomes, one for each Keirsey temperament. The child nodes represent choices, each with two outcomes that correspond to a scenario room.

For our network the \( N \) value, number of decisions, was fixed at 7, one for each room. The value was chosen after a long process of scenario design, extensively described in the Design Methodology section.

Before finishing, the player is forced to make a choice seven times. As the player progresses through the scenario the decision sample data is registered in a text log. Once finished the text log can be used
to infer personality according to the current probability values in the knowledge base. With no sample data, our system assumes that all temperaments have the same probability of leading players to make a choice. This starting stage is uninteresting for classifying players but its probabilities can be increased by machine learning.

**Learning.** To increase the success rate of our solution, machine learning is required. By coupling decision sample data and player temperament it is possible to increase the knowledge base of the network. Each coupling, or training data, helps refine probabilities and leads to a better system. The more data fed to our system the more accurate its predictions become.

Initially the network contains all values and conditional probabilities for our world space equidistributed across temperaments. Once sample data is fed to the system the probabilities are updated to reflect actual player preferences.

The only requirement for learning in our system is the offline acquisition of actual personality. This can be done using a personality sorter, a psychology questionnaire that indicates personality once it is filled.

**Classifying temperaments.** The network serves as a means for storing conditional values obtained from training data. The classification is then done by calculating the maximum optimal cause, or temperament, based on these conditional values and subject data from current session.

Equation 3.1 shows Bayes’ rule for conditional probability, it allows us to derive $P(Cause \mid Effect)$, or in our model $P(Temperament \mid Choice)$. By using subject data our system calculates the conditional probability for each temperament knowing what task was chosen at each Choice.

$$
P(Cause \mid Effect) = \frac{P(Effect \mid Cause)P(Cause)}{P(Effect)}. \quad (3.1)
$$

However, instead of a unique choice we have several choices so we need an answer for $P(Choice_1, ..., Choice_N \mid Temperament)$; $N=7$ being the total number of choices in the scenario. Note that choices in our model are independent from one another; this was done to take advantage of the mathematical property of conditional independence, as depicted in Equation 3.2. From conditional independence we can obtain Equation 3.3.

$$
P(X, Y \mid Z) = P(X \mid Z)P(Y \mid Z) \quad (3.2)
$$

$$
P(A_1, A_2, ..., A_n \mid temperament) = P(A_1 \mid temperament) \ast P(A_2 \mid temperament) \ast ... \ast P(A_n \mid temperament) \quad (3.3)
$$

Conditional independence allows us to simplify the problem of computing the Bayes classifier and allows simple calculations using the values already stored in our network. This process allows us to obtain the probabilities of the player belonging to any of the possible outcomes of the Temperaments Node.
The outcome with highest probability, or the maximum optimal cause, is best indicative of personality and result of our inferring system.

**Implementation.** To implement the network we used the Decision Systems Laboratory SMILE (Structural Modelling, Inference, and Learning Engine) software, made in the University of Pittsburgh \(^1\). SMILE is a fully portable Bayesian inference engine written in C++ and is released as a dynamic link library (DLL). It has several wrappers available such as SMILE.NET (.NET interface), SMILEX (Active X) and jSMILE (Java interface).

Due to its portability across languages, SMILE can be used in conjunction with different games and scenarios, including ours. Since the language used for the scenario is Java we use the jSMILE wrapper of SMILE in our inferring system.

### 3.4 Experiment Protocol

Having our scenario and Bayesian network set up to work during playtime, it became necessary to acquire sample data to build the knowledge base of the system. To do this we resorted to several sessions of player testing. Sessions were comprised of four stages: Scenario Introduction, Questionnaire, Play and Appreciation.

**Scenario Introduction.** At the start of the experiment players were thanked for participating and were directed to a seat where testing would take place. We explained the goal of our project as being an experiment to analyze different game mechanics. This hoax was used to avoid influencing players with our real goal and created a testing environment where they could feel comfortable exploring each room.

**Questionnaire.** Building the knowledge base required not only information regarding player preference, but also their temperaments. This meant that having players answer a personality survey was required during the test.

As was shown in previous sections, Keirsey temperaments can be extrapolated from MBTI dichotomies. This allows us to use MBTI surveys to obtain player personality. The advantage of using MBTI surveys is their low or non-existing cost, as well as having been used on previous studies of entertainment modelling. The sorter used can be found in Annex B.

To facilitate this stage, the questionnaire was given as a Google form. This allowed instant acquisition of answers at the end of the survey. It also allowed us to minimize errors, such as omitted answers, when filling the document.

Once the questionnaire was filled, the answers were passed to a previously prepared spreadsheet that returned the temperament of the player. Having obtained the player’s personality, it was crucial to keep the details of personality behaviour hidden to avoid influencing behavior.

---

\(^1\)SMILE software acquired at http://genie.sis.pitt.edu/
**Play.** At the beginning of this stage, players were told the expected test length and how the scenario was composed. They were also informed of their goals and the general architecture of each room. One important aspect that players needed to know was that despite being free to abandon any task, they were required to visit both tasks in a room.

The explanation ended once the player understood the different mechanics of our scenario such as choosing a task, moving between tasks and teleporting across rooms.

If players were new with Minecraft, a brief tutorial over its mechanics was required. This tutorial explained movement, acquisition of items, usage of recipes and general interaction with the world.

After the explanations, players were allowed to begin the scenario. The initial state of the game was the same for all players. To ensure independency between choices, the order of rooms was chosen randomly, at runtime, and differed from player to player. Each player had their own unique experience with the scenario and it allowed them freedom of choice and of expression, as evidenced by figure 3.15.

![Painting of the Portuguese flag](image1.png) ![Painting of a person](image2.png)

Figure 3.15: Player made paintings at room 7 of the scenario.

**Appreciation.** When players finished the scenario they were greatly thanked for their help and given a coffee coupon to use at a near-by coffee shop. We felt that this was necessary to show our gratitude to players who helped us.

Once coupons were given, the players were escorted out. At this point acquired session data and temperaments were stored and the experiment was reset for the next test session.

### 3.5 Solution summary

This chapter described our final solution as being comprised of a task-based scenario and an inferring system. These two modules intended to classify personality as part of entertainment modelling. This required a design methodology that could also be applied to different games.

Our proposed methodology was extensively described in order to be replicated by future work. It began by identifying possible tasks from the game and creating an initial task list. Tasks in this list were then associated to temperament preferences and clustered to simplify analysis. Pairs of task clusters were then assigned to a scenario architecture with the goal of separating temperaments. The result of applying this methodology to the game Minecraft was our scenario entitled *Rooms.*
We also described the architecture of the system used to infer personality based on player preference. This system was composed of a Bayesian network that updated its knowledge base with data obtained by Rooms.

Finally we described an experiment that used both the inferring system and Rooms to acquire player data from game sessions and classify personality. Obtained data and results from the experiment, as well as further analysis that helps improve our methodology is depicted in the next chapter.
Chapter 4

Data Analysis

In previous sections we described our experimental process to acquire user data by allowing players interaction with Rooms. Data analysis and validation of our system resorted to the acquired user samples. These samples were then fed to our system and several cross-validation runs of the system were performed with different permutations of rooms.

In this chapter we present the data acquired and the sequential analysis that were made in order to validate and improve our scenario and inferring system.

Sample Population. Our experiment collected data from thirty two individuals, mainly from different universities of Lisbon. Most of these belonged to courses related to Computer Science.

The temperament of testers was acquired using the MBTI sorter in Annex B. From these thirty-two individuals, three have ambiguous dichotomies making it impossible to accurately identify their Keirsey Temperament. These three special cases were not considered for the evaluation of the system.

The frequency of the distribution of temperaments of our sample population can be seen in Figure 4.1. Rational and Guardian are the most frequent temperaments. Despite the considerable number of samples acquired, the Idealist and Artisan temperaments are underrepresented.

Ideally, we would have preferred our test population to have a similar distribution as that belonging to the normal society distribution of temperaments.

Figure 4.2 shows the comparison between the temperaments of our sample population and the expected values. While Guardians and Idealists show similar distributions, Rationals are overrepresented while Artisans are underrepresented. We believe that these discrepancies in distribution are related to the field of study of our samples, Computer Science.

4.1 Experimental results

Having conducted our experiment, we used the data collected to evaluate the success rate of our solution. We feed the different player choices as samples to our system and test its hit and miss rates. Due to the flexibility of Bayesian networks, it becomes possible to omit specific room data when increasing
Figure 4.1: Temperament distribution across experiment sample population. The ??? category represents individuals whose ambiguous dichotomies did not allow for proper temperament classification.

Figure 4.2: Comparison of temperament distributions between our test population and societal standard. The Guardian and Idealist distribution appears similar to societal expectations. The Rational temperament is overrepresented in our experiment while the Artisan temperament is underrepresented. These discrepancies in distribution might be related to the field of study of most of our samples, Computer Science.
Figure 4.3: Hit and Hiss rates of initial cross-validation using sample from all rooms. The system shows positive results in identifying the Guardian and Rational temperaments, the most represented in our samples. The possibility of malformed rooms was not considered for this run.

The system shows positive results in identifying the Guardian and Rational temperaments, the most represented in our samples. The possibility of malformed rooms was not considered for this run.

We began by resorting to k-fold cross-validation and selectively dividing our samples into training sets and evidence sets. For k-fold cross validation we select a $k$ value that divides samples into $k$ sets of $n/k$ elements. For our validation we used $k = 10$, this left us with 9 sets of 3 samples and one set with 2 samples.

At each step of cross-validation a new set from the $k$ sets was selected as the evidence set. Non-selected sets were used, as training sets, to train the knowledge base of our network. The network then proceeded to predict the temperaments of the samples in the evidence set. To ensure that each set was selected at least once, the process was repeated $k$ times. At the end we acquired the hit and miss rates depicted in figure 4.3.

Considering the data from all rooms for our initial k-fold cross-validation led our system to have an average hit rate of only 51.72%. The best predicted temperament was the Guardian with 72.73% hit rate. These results show that the system was unable to predict any of the less represented temperaments, Artisan or Idealist.

These results stem from using all rooms when doing cross-validation. We should also consider that some rooms can be influencing the system negatively by being malformed. Malformed rooms can be considered rooms that do not fullfill their design goal once implemented and that negatively affect

**Discarding malformed rooms.** We have previously explained that such scenarios can contain malformed rooms, rooms that do not fulfill their design goal once implemented and that negatively affect
Figure 4.4: Average hit and miss rates by discarding single rooms during cross-validation. Removing rooms 1, 5, 6 and 7 led to a reduction of hit rate which can be indicative of them being well formed. Removing rooms 2, 3 and 4 led to an increase of hit rate which can be indicative of these rooms being malformed.

the results of the inferring system. Identifying these rooms can be made by analysing the individual influence of each room, and eventually removing malformed rooms to achieve higher success rates.

Alternatively we could run a brute force analysis that calculated all possible combinations of rooms and their associated hit rate. Although simpler, from an implementation point of view, this method does not allow for a deeper understanding of the effects of malformed rooms. Since this work intends to be a research study we wish to acquire as much information as possible over malformed rooms and techniques to identify them. As such we decided to analyze the individual influence of each room and try to pinpoint malformed ones, against running a brute force algorithm to achieve the maximum hit rate.

Data regarding individual room influence can be acquired by repeating the process of k-fold cross-validation but removing the sample data from each room. A total of seven k-fold cross-validations were run, discarding a different room at each run. Figure x shows the average results from removing each room for all temperaments.

Removing rooms 1, 5, 6 and 7 led to a reduction of hit rate when compared to having all rooms. This seems to indicate that these rooms are well formed and contribute positively to identifying temperaments. A look at their individual efficiency should give more information over their influence.

Removing rooms 2, 3 and 4 led to an increase of hit rate compared to having all rooms. This can indicate that these rooms were not designed or implemented properly, and contribute negatively to identifying temperaments. Once again, information regarding their individual efficiency can help validate this indication.

The rooms influence over the Guardian Temperament can be seen in figure 4.5 and over the Rational Temperament in figure 4.6. In both cases removing rooms 1, 5, 6 and 7 either maintains the hit rate or reduces it. While removal of rooms 2, 3 and 4 also increases the hit rate of predicting these two temperaments. These results are aligned with the average results before mentioned.
Figure 4.5: Average hit and miss rates of cross-validation for the Guardian temperament, obtained by discarding single rooms. Removing rooms 1, 5, 6 and 7 led to a reduction of hit rate which can be indicative of them being well formed. Removing rooms 2, 3 and 4 led to an increase of hit rate which can be indicative of these rooms being malformed.

<table>
<thead>
<tr>
<th>Discard Room 7</th>
<th>Guardian Hit %</th>
<th>Guardian Miss %</th>
</tr>
</thead>
<tbody>
<tr>
<td>63.64%</td>
<td>36.36%</td>
<td></td>
</tr>
<tr>
<td>Discard Room 6</td>
<td>36.36%</td>
<td>63.64%</td>
</tr>
<tr>
<td>Discard Room 5</td>
<td>27.27%</td>
<td>72.73%</td>
</tr>
<tr>
<td>Discard Room 4</td>
<td>81.82%</td>
<td>18.18%</td>
</tr>
<tr>
<td>Discard Room 3</td>
<td>81.82%</td>
<td>18.18%</td>
</tr>
<tr>
<td>Discard Room 2</td>
<td>72.73%</td>
<td>27.27%</td>
</tr>
<tr>
<td>Discard Room 1</td>
<td>63.64%</td>
<td>36.36%</td>
</tr>
<tr>
<td>All Rooms</td>
<td>72.73%</td>
<td>27.27%</td>
</tr>
</tbody>
</table>

Figure 4.6: Average hit and miss rates of cross-validation for the Rational temperament, obtained by discarding single rooms. Removing rooms 1, 5, 6 and 7 led to a reduction of hit rate which can be indicative of them being well formed. Removing rooms 2, 3 and 4 led to an increase of hit rate which can be indicative of these rooms being malformed.

<table>
<thead>
<tr>
<th>Discard Room 7</th>
<th>Rational Hit %</th>
<th>Rational Miss %</th>
</tr>
</thead>
<tbody>
<tr>
<td>53.85%</td>
<td>46.15%</td>
<td></td>
</tr>
<tr>
<td>Discard Room 6</td>
<td>53.85%</td>
<td>46.15%</td>
</tr>
<tr>
<td>Discard Room 5</td>
<td>46.15%</td>
<td>53.85%</td>
</tr>
<tr>
<td>Discard Room 4</td>
<td>76.92%</td>
<td>23.08%</td>
</tr>
<tr>
<td>Discard Room 3</td>
<td>61.54%</td>
<td>38.46%</td>
</tr>
<tr>
<td>Discard Room 2</td>
<td>61.54%</td>
<td>38.46%</td>
</tr>
<tr>
<td>Discard Room 1</td>
<td>53.85%</td>
<td>46.15%</td>
</tr>
<tr>
<td>All Rooms</td>
<td>53.85%</td>
<td>46.15%</td>
</tr>
</tbody>
</table>
Figure 4.7: Average hit and miss rates of cross-validation, obtained by only using data from single rooms. Rooms 1, 5 and 6 continue to seem well formed and rooms 2, 3 and 4 malformed. Room 7 in isolation appears with worse hit rate than the average of all rooms, a possible indication of it contributing negatively to the average hit rate.

![Table 4.1: Hit and miss rates for combinations of expected malformed and well formed rooms. Combination 
{1,5,6} is seen as the most successful and efficient, as it takes less time to achieve the highest hit rate so far. It is interesting to note that malformed rooms {2,3,4} were able to predict an Idealist entry, that so far always missed.](image)

<table>
<thead>
<tr>
<th>Rooms 2,3,4</th>
<th>Rooms 1,5,6,7</th>
<th>Rooms 2,3,4,7</th>
<th>Rooms 1,5,6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hit%</td>
<td>31.03%</td>
<td>68.97%</td>
<td>24.14%</td>
</tr>
<tr>
<td>Miss%</td>
<td>68.97%</td>
<td>31.03%</td>
<td>75.86%</td>
</tr>
<tr>
<td>Guardian</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hit%</td>
<td>18.18%</td>
<td>81.82%</td>
<td>18.18%</td>
</tr>
<tr>
<td>Miss%</td>
<td>81.82%</td>
<td>18.18%</td>
<td>81.82%</td>
</tr>
<tr>
<td>Idealist</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hit%</td>
<td>25.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Miss%</td>
<td>75.00%</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
<tr>
<td>Rational</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hit%</td>
<td>46.15%</td>
<td>84.62%</td>
<td>38.46%</td>
</tr>
<tr>
<td>Miss%</td>
<td>53.85%</td>
<td>15.38%</td>
<td>61.54%</td>
</tr>
</tbody>
</table>

Regarding the Idealist and Artisan temperaments, the removal of any room did not influence the hit rate that remains at 0%. Once again, the number of samples for these temperaments is too small for the system to be able to predict them.

To reinforce previous information regarding influence of rooms, we can look at the influence in isolation by repeating the k-fold cross-validation using only sample data from the room we want. By repeating the cross-validation process seven times we obtain figure 4.7 that shows the average influence of each room across all temperaments.

Looking at rooms in isolation shows that rooms 1, 5 and 6 have greater isolated hit rate than the average of all rooms together. Rooms 2, 3 and 4 have a worse hit rate than the average of all rooms. Room 7, that before seemed to be well formed now appears, in isolation with worse hit rate than the average of all rooms, albeit it superior to the reduction of rooms 2, 3 and 4.

At this point, we understand that rooms 1, 5 and 6 have positive influence over the success rate of our system, rooms 2, 3 and 4 have negative influence and room 7 seems to be a special case. To try to maximize the hit rate and reduce scenario time, we proceeded to group the well formed rooms and malformed rooms and tested different combinations with the special case.
Table 4.1 shows the percentages regarding hit and miss rates for the different combinations chosen. Combinations of malformed rooms did in fact show a low hit percentage, 31% for \(\{2,3,4\}\) and 24% for \(\{2,3,4,7\}\). Combinations of well-formed rooms have the highest hit percentage; so far with 68.97%.

It is interesting to note that one of these combinations was able to have a positive prediction over the Idealist temperament, that up until this point had 100% miss rate. Despite this, the overall hit rate of the combination is too low to be considered a good result for the system.

Despite having two combinations with the highest average hit rate, we believe that combination \(\{1,5,6\}\) is better than \(\{1,5,6,7\}\), due to having one less room to traverse. With fewer rooms the scenario is more efficient, as it takes less time to predict temperaments.

With the most efficient combination achieved, the well represented temperaments, Rational and Guardian, hit percentages of over 80%. The other temperaments, Artisan and Idealist, remained with no positive predictions.

### 4.2 Result Significance

Evaluative analysis of our system was able to show its success in identifying two of the four temperaments over 80% of the times. These two temperaments were also the most represented by our sample. When predicting all temperaments our system is correct 68.97% of the times when using a *Minecraft* based scenario that contains only three rooms.

Our system was unable to predict the two lesser represented temperaments. To ensure that future systems increase their success rate it is required and recommended that the experiment be repeated with a sample of larger size, preferably with the same distribution of temperaments as the one present in society.

The positive results achieved confirm the usefulness of our design methodology for the game chosen. More study must be done over whether effectiveness is maintained over other games, besides *Minecraft*, and other genres, besides the sandbox genre.

The results also demonstrate that our scenario design methodology, despite iterative still generates malformed rooms that can and should be removed from final models. Future work based on ours should take this into consideration and schedule time to do cross validation to pinpoint malformed rooms. If the goal of future work is merely to achieve the highest success rate possible we recommend running a brute force algorithm that explores all possible combinations of rooms and their success rate.
Chapter 5

Conclusions

Playing a game is a personal experience and each individual will enjoy it differently, this makes the design and creation of games a hard task. We began our work by looking at the field of entertainment modelling and what it could do to increase the enjoyment of videogames to different players. One problem, unresolved in the work of Dias[8], is how to identify player preference before adapting content? Player preferences have been associated with personality and knowing a player’s personality can give insights into what content to adapt. Our work tried to solve this problem of classifying the personality of players while they are playing.

We began by hypothesizing that player actions and choices could be an indicative of personality. We fixed Keirsey Temperament Model as our personality model due to its behavioristic nature. Using temperaments we developed a methodology to design scenarios that allowed the collection of player data for a particular game, as well as an inferring system that classified players taking into account the scenario data. To validate and exemplify our methodology, we created Rooms, a task based scenario for the game Minecraft.

The methodology process began with collecting tasks by crossing game theory with specific game mechanics. These tasks were crude and numerous and their relation with temperaments highly subjective. We validated some tasks from our list with a simple experiment that observed testers freely playing Minecraft. The preferences of the testers were then tied to their personality, thus complementing and validating some tasks.

Due to the large number of tasks, it became necessary to cluster them using data from the experiment. The clustering process made use of the IBM SPSS2.0 software to perform K-means clustering with different values of $K$, $K$ being the number of clusters the algorithm wants to achieve. Ideally we wanted to reach four clusters but the variety of information provided by five clusters was more useful.

Finally, we analyzed different scenario architectures and chose the one that had less subjectivity tied to the decision process. Having our architecture chosen we assigned clusters and then tasks to each room finally reaching a final model.

This methodology process was a crucial part of our solution. The resulting scenario was then tied to a temperament inferring system implemented as a Bayesian network. The data acquired from the
scenario was fed to the inferring system that could then increase its knowledge base or classify player temperaments.

To validate our solution, we implemented the scenario Rooms for the game Minecraft. Afterwards we set up an experiment that collected sample data from thirty-two players. Three individuals were identified as ambiguous temperaments and their data was discarded.

The scenario used in the experiment still contained malformed rooms. Malformed rooms can be considered as rooms that when implemented do not achieve their design goal, being a negative influence over the inferring system. Through cross-validation analysis we were able to identify malformed rooms and increase the success rate of our system. Future work based on ours more worried with achieving the highest success rate can try to bypass cross-validation by using a brute force approach.

After removing malformed rooms, the system was able to predict temperaments with an average success rate of 68.97%. The two best represented temperaments showed success rates of over 80%, while the less represented temperaments were never successfully identified.

We believe that these discrepancies can be explained by the lack of Artisan and Idealist temperaments from our sample. The sample was gathered mostly from individuals closely related to the field of Computer Science, that appears to have a temperament distribution different from the standard societal distribution. Future work should be done to confirm our beliefs and researchers should take some time to gather even larger samples than ours.

**Future work.** As our work built upon Dias’ thesis[8] we notify of the possibility of joining both researches in a larger study. This study would first recognize personality using our solution and then adapt content according to player preferences in a seamless and smooth play session.

For future work, we recommend using our methodology to create systems to infer personality for other games besides Minecraft, and other genres of games besides the sandbox genre. We also leave the possibility of adapting our design methodology and inferring system to use other personality models besides the Keirsey temperaments.
Appendix A

Observation Results
Figure A.1: Frequency of Guardian actions over our initial task list during the experiment in Observation. Expert players accounted for 33.33% of the temperament and new players 66.66%. Many differences can be seen between new and expert players, with common liking to the tasks ‘Follow Goals’ and ‘Avoid Combat without weapon’ which can be indicative of them being strong motivators for the temperament. Temperament shows no particular preference or aversion to building.
Figure A.2: Frequency of Rational actions over our initial task list during the experiment in Observation. Exploratory actions such as ‘Search for materials’ and ‘Explore’ can be seen as important for this temperament. New and expert players were balanced, each representing 50% of the temperament population. Contrast arises in the lack of expert players enjoying combat situations, while new players still enjoy killing for fun. Temperament shows no particular preference or aversion to building.
The results stem from a single expert Artisan. Results confirm that combat are part of the Artisans preferences especially when holding a weapon. This temperament seems to have less interest in Building than others.
Figure A.4: Frequency of Idealist actions over our initial task list during the experiment in Observation. The results stem from a single Idealist, unexperienced with the game. Results confirm that building are the preferred set of actions for the temperament, with ‘Search for Materials’ done to acquire building materials. The results confirm that Idealists have aversion for combat. The temperament shows no particular preference for exploration.
Appendix B

MBTI Questionnaire, obtained from [1]
## MBTI Personality Type Test

1. **At a party do you:**
   a. Interact with many, including strangers
   b. Interact with a few, known to you

2. **Are you more:**
   a. Realistic than speculative
   b. Speculative than realistic

3. **Is it worse to:**
   a. Have your “head in the clouds”
   b. Be “in a rut”

4. **Are you more impressed by:**
   a. Principles
   b. Emotions

5. **Are you more drawn toward the:**
   a. Convincing
   b. Touching

6. **Do you prefer to work:**
   a. To deadlines
   b. Just “whenever”

7. **Do you tend to choose:**
   a. Rather carefully
   b. Somewhat impulsively

8. **At parties do you:**
   a. Stay late, with increasing energy
   b. Leave early with decreased energy

9. **Are you more attracted to:**
   a. Sensible people
   b. Imaginative people

10. **Are you more interested in:**
    a. What is actual
    b. What is possible

11. **In judging others are you more swayed by:**
    a. Laws than circumstances
    b. Circumstances than laws

12. **In approaching others is your inclination to be somewhat:**
    a. Objective
    b. Personal

13. **Are you more:**
    a. Punctual
    b. Leisurely

14. **Does it bother you more having things:**
    a. Incomplete
    b. Completed

15. **In your social groups do you:**
    a. Keep abreast of other’s happenings
    b. Get behind on the news

16. **In doing ordinary things are you more likely to:**
    a. Do it the usual way
    b. Do it your own way

17. **Writers should:**
    a. “Say what they mean and mean what they say”
    b. Express things more by use of analogy

18. **Which appeals to you more:**
    a. Consistency of thought
    b. Harmonious human relationships

19. **Are you more comfortable in making:**
    a. Logical judgments
    b. Value judgments

20. **Do you want things:**
    a. Settled and decided
    b. Unsettled and undecided

21. **Would you say you are more:**
    a. Serious and determined
    b. Easy-going

22. **In phoning do you:**
    a. Rarely question that it will all be said
    b. Rehearse what you’ll say

23. **Facts:**
    a. “Speak for themselves”
    b. Illustrate principles

24. **Are visionaries:**
    a. somewhat annoying
    b. rather fascinating

25. **Are you more often:**
    a. a cool-headed person
    b. a warm-hearted person

26. **Is it worse to be:**
    a. unjust
    b. merciless
27. Should one usually let events occur:
   a. by careful selection and choice
   b. randomly and by chance

28. Do you feel better about:
   a. having purchased
   b. having the option to buy

29. In company do you:
   a. initiate conversation
   b. wait to be approached

30. Common sense is:
   a. rarely questionable
   b. frequently questionable

31. Children often do not:
   a. make themselves useful enough
   b. exercise their fantasy enough

32. In making decisions do you feel more comfortable with:
   a. standards
   b. feelings

33. Are you more:
   a. firm than gentle
   b. gentle than firm

34. Which is more admirable:
   a. the ability to organize and be methodical
   b. the ability to adapt and make do

35. Do you put more value on:
   a. infinite
   b. open-minded

36. Does new and non-routine interaction with others:
   a. stimulate and energize you
   b. tax your reserves

37. Are you more frequently:
   a. a practical sort of person
   b. a fanciful sort of person

38. Are you more likely to:
   a. see how others are useful
   b. see how others see

39. Which is more satisfying:
   a. to discuss an issue thoroughly
   b. to arrive at agreement on an issue

40. Which rules you more:
   a. your head
   b. your heart

41. Are you more comfortable with work that is:
   a. contracted
   b. done on a casual basis

42. Do you tend to look for:
   a. the orderly
   b. whatever turns up

43. Do you prefer:
   a. many friends with brief contact
   b. a few friends with more lengthy contact

44. Do you go more by:
   a. facts
   b. principles

45. Are you more interested in:
   a. production and distribution
   b. design and research

46. Which is more of a compliment:
   a. “There is a very logical person.”
   b. “There is a very sentimental person.”

47. Do you value in yourself more that you are:
   a. unwavering
   b. devoted

48. Do you more often prefer the
   a. final and unalterable statement
   b. tentative and preliminary statement

49. Are you more comfortable:
   a. after a decision
   b. before a decision

50. Do you:
   a. speak easily and at length with strangers
   b. find little to say to strangers

51. Are you more likely to trust your:
   a. experience
   b. hunch

52. Do you feel:
   a. more practical than ingenious
   b. more ingenious than practical

53. Which person is more to be complimented – one of:
   a. clear reason
   b. strong feeling
54. Are you inclined more to be:
   a. fair-minded
   b. sympathetic

55. Is it preferable mostly to:
   a. make sure things are arranged
   b. just let things happen

56. In relationships should most things be:
   a. re-negotiable
   b. random and circumstantial

57. When the phone rings do you:
   a. hasten to get to it first
   b. hope someone else will answer

58. Do you prize more in yourself:
   a. a strong sense of reality
   b. a vivid imagination

59. Are you drawn more to:
   a. fundamentals
   b. overtones

60. Which seems the greater error:
   a. to be too passionate
   b. to be too objective

61. Do you see yourself as basically:
   a. hard-headed
   b. soft-hearted

62. Which situation appeals to you more:
   a. the structured and scheduled
   b. the unstructured and unscheduled

63. Are you a person that is more:
   a. routinized than whimsical
   b. whimsical than routinized

64. Are you more inclined to be:
   a. easy to approach
   b. somewhat reserved

65. In writings do you prefer:
   a. the more literal
   b. the more figurative

66. Is it harder for you to:
   a. identify with others
   b. utilize others

67. Which do you wish more for yourself:
   a. clarity of reason
   b. strength of compassion

68. Which is the greater fault:
   a. being indiscriminate
   b. being critical

69. Do you prefer the:
   a. planned event
   b. unplanned event

70. Do you tend to be more:
   a. deliberate than spontaneous
   b. spontaneous than deliberate
# Scoring

<table>
<thead>
<tr>
<th></th>
<th>Col 1</th>
<th>Col 2</th>
<th>Col 3</th>
<th>Col 4</th>
<th>Col 5</th>
<th>Col 6</th>
<th>Col 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td>27</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>30</td>
<td>31</td>
<td>32</td>
<td>33</td>
<td>34</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>37</td>
<td>38</td>
<td>39</td>
<td>40</td>
<td>41</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>44</td>
<td>45</td>
<td>46</td>
<td>47</td>
<td>48</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>51</td>
<td>52</td>
<td>53</td>
<td>54</td>
<td>55</td>
<td>56</td>
<td></td>
</tr>
<tr>
<td>57</td>
<td>58</td>
<td>59</td>
<td>60</td>
<td>61</td>
<td>62</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>65</td>
<td>66</td>
<td>67</td>
<td>68</td>
<td>69</td>
<td>70</td>
<td></td>
</tr>
</tbody>
</table>

Copy to ➔ Copy to ➔ Copy to ➔

<table>
<thead>
<tr>
<th></th>
<th>E</th>
<th>I</th>
<th>S</th>
<th>N</th>
<th>T</th>
<th>F</th>
<th></th>
<th>J</th>
<th>P</th>
</tr>
</thead>
</table>

1. Copy your answers to this answer key carefully.
2. Count the number of checks in each of the A and B columns, and total at the bottom.
3. Copy the totals for Column 2 to the spaces below the totals for Column 3. Do the same for Columns 4 and 6.
4. Add totals downwards to calculate your totals.
5. Circle the letter with this highest score. This is your type.
Bibliography


