Player Expectations of Strategy Game AI

Daniel Gomme

A thesis submitted for the degree of *Doctor of Philosophy*

School of Computer Science and Electronic Engineering

University of Essex

June 2023

Acknowledgements

Here's to everyone that's helped me get this far. I won't name you all — this thesis doesn't need to double in length — but you know who you are, friends.

First, to Richard (my supervisor, not my brother): I cannot express enough thanks for the steady guidance and quiet wholesomeness. This document would be far from what it is today without you, and contain many more split infinitives. Yes, I checked that sentence for them.

To the strange coterie of creatures I'm happy to call my housemates and friends: Thank you for putting up with me while I was deep in the unsteady mineshafts of this thesis. I wasn't quite sure how it'd go, but the neverending cheerful support did no end of good for my morale.

My family (*now* this includes you, brother of mine) have been the most reassuring set of souls a Dan could ever wish to have. I'm pretty sure I've bored you mindless when I've gone on a tangent about this work, but that hasn't stopped you smiling, nodding, and wishing me luck.

Joseph Walton-Rivers, he who hath shared both office and technological wizardry, this PhD got a whole lot quieter when you went. You've always been a solid companion in these academic realms.

Last, and certainly not least, I couldn't talk of a guide in academic realms without mentioning Nathan Hughes. A staunch ally and genuinely fearsome wielder of righteous statistics, you've helped more than you could know.

Abstract

The behaviour of AI in modern strategy games is universally recognised as flawed. To compensate for this and successfully challenge humans, it must often be given significant advantages, such as luck bonuses, access to extra in-game resources or knowledge of the entire game state. Players often proclaim their dislike of these flaws, discussing nonsensical moves AI opponents have made, or the fact that the AI 'cheats' — out of necessity, as creating competent strategy game AI on consumer hardware is incredibly difficult, even with state-of-the-art techniques.

Therefore, this thesis asks: what frustrates players about the opponents — human and AI — that they play against? By asking this, we can establish the most efficient ways to improve player experience when facing AI opponents. To answer, we explore the computer science that drives AI, the psychology that drives players, and the nature of game interactivity as a whole.

Flaws in a range of popular strategy games were investigated, forming a grounded theory on how AI play typically annoys strategy game players. We find that players expect their opponents to conform to a set of expectations.

Two scenarios were crafted for an existing strategy game. A mix of qualitative and quantitative methods were used to evaluate how players' experience of one of those expectations — tension — changes under different, controlled conditions. We find that tension can be observed, and is connected to both player uncertainty and perceptions of power. In addition, analysis of player experiences allowed extraction of practical, concrete methods with which game developers can directly influence player experiences of tension in-game.

A further experiment clarifies that investment is also connected to tension, but that it is more effective to phrase it as *need* when questioning players about their investment in a given objective. It also demonstrates that too little information given to players can remove the connection between perceived powers and tension.

Finally, we connect our findings to the current literature on player experience in games, and highlight where further work needs to be done.

Contents

List of Figures

List of Tables

List of Studies

Distributed Social Multi-Agent Negotiation Framework For Incomplete Information Games (Walton-Rivers, Longford et al. 2019)

Authors: *Joseph Walton-Rivers, Edward Longford, Daniel Gomme, Richard Bartle, Michael Gardner*

Published at: CEEC 2019, University of Essex. Internal conference

A foray into how agents might meaningfully communicate in games with a great deal of communication options, by selecting from meaningful categories of possibilities and choosing within those.

Strategy Games: The Components of a Worthy Opponent (Gomme and Bartle, Richard 2020)

Authors: *Daniel Gomme, Richard Bartle*

Published at: Foundations of Digital Games 2020. Conference

From over 200 forum comments, assembles a grounded theory on expectations that strategy game players have of the opponents that they play against. Forms the basis of Chapter 2.

Tools To Adjust Tension And Suspense In Strategy Games: An Investigation (Gomme and Bartle, Richard 2022)

Authors: *Daniel Gomme, Richard Bartle*

In review at: IEEE Transactions on Games

Examines one of the expectations found in the grounded theory above: that opponents should provide tension. Puts players through a *Battle for Wesnoth* (*The Battle for Wesnoth* 2003) scenario that tweaks theorised components of tension, then measures these components afterwards. Forms the basis of Chapter 3.

Examining Player Investment as Related to Tension in Strategy Games (Daniel Gomme and Richard Bartle 2023)

Authors: *Daniel Gomme, Richard Bartle*

Not submitted

Puts players through another *Battle for Wesnoth* scenario in order to test one of the unexpected findings of the previous paper: does player investment really not matter so much when it comes to tension, or was it merely the questionnaire that was at fault? This study's questionnaire uses several wordings for its questions on investment, and compares these to the control wording and the expected findings of the previous paper. Forms the basis of Chapter 4.

Abbreviations

This thesis uses several abbreviations, and Table 0.1 enumerates some of the most common.

Abbreviation	Full Meaning
AI	Artificial Intelligence
Civ VI	*Sid Meier's Civilization VI*
DDA	Dynamic Difficulty Adjustment
EDM	Expectancy Disconfirmation Model
ES2	*Endless Space 2*
FSM	Finite State Machine
LLM	Large Language Model
MCTS	Monte Carlo Tree Search
MMO	Massively Multiplayer Online game
PP	Predictive Processing
RL	Reinforcement Learning
SDT	Self-Determination Theory
TA	Thematic Analysis

Table 0.1: Abbreviations Used In Thesis

Technical terms

The work in this thesis defines some terms to describe observed phenomena. To provide clear access to these, those definitions are listed in Table 0.2.

Term	Definition
Investment (section 2.6.2)	A player's desire for a specific future event to occur.
Tension (section 2.6.2)	Excitement derived from a player being invested in an event, and due to another entity contesting their efforts to bring the event about, being uncertain if that event will occur.

Table 0.2: Technical Terms Defined In Thesis

0.1 A Guide To The Thesis Ahead

First, we detail the ways in which strategy games present significant difficulties for both traditional and state-of-the-art game AI techniques. We explore ways those difficulties manifest and the negative impacts this has on player experience.

Having established that strategy game AI is flawed, and that with current techniques we'd need significant increases in computing power to change the situation, we turn to asking how to improve strategy AI without simply throwing more computing power at the problem. Thus, we come to the research question of this thesis:

How can player perceptions of strategy game AI be improved?

To answer this, we look at what players want from strategy-game AI. They had very little extant praise for it in their communities, so instead we look at the issues they had with the AI.

We conduct a grounded study of player comments about the issues they have with strategy game AI, finding four key areas of concern. We go on to explore one of them — tension, comprised of perceptions of power combined with players' investment in the outcome — because the specific mechanisms with which to adjust it are relatively unexplored in the games literature.

In order to test whether tension in fact exists in real gameplay, we formulate several hypotheses about its components and test these in an experiment. The results support the identified components of tension, except for one: the player's investment in the desired outcome. The results also suggest concrete, actionable guidance on what mechanics and events can alter players' perceptions of tension.

We then run another experiment investigating whether investment is not related to tension, or if it was simply measured incorrectly previously. We find that investment is indeed a factor, can be measured more effectively by referring to it as *need*, and that perceptions of player power can be decoupled from perceptions of opponent power if information is not clear enough to them.

Therefore, we reach one possible answer to the research question posed: that strategy game AI can be improved by knowing what players are motivated to seek at any given time, and manipulating perceptions of power by controlling information and resources the player has available to them.

Finally, we relate our findings to current understandings of player experience in games, highlighting other areas in which AI may be improved and the mechanisms by which this may be done.

Chapter 1

AI in Strategy Games: Its Current State

1.1 Defining Strategy Games

To adequately investigate AI in strategy games, it is first necessary to define what we mean by 'strategy game'. This is perhaps a Sisyphean task on its own. There is very little cultural consensus around the concept of strategy (Dor 2018), though it has charmingly been referred to as a "top down game in which you gather resources to build and control armies of little guys" (*The Future of the RTS?* 2008).

Happily, we can define it more usefully than nailing the whole genre down into a definition that everyone agrees on. We instead define it in terms that *make a game hard for AI to navigate*.

We define it as a game which fulfils the following:

- The game is digital.
  - As yet, the means to have an AI adequately interpret and interact with a physical game are not in wide use. Hence, we focus on the digital.
- The player plays as a whole faction or empire.
- The game requires long-term planning to defeat at least one other faction, while playing against at least one other faction.
- There are a wide array of resources to take advantage of within the game in order to assist in victory.
- A board or game state upon which players expand their presence spatially through the course of gameplay.
- Opponent factions that take the same role a human player otherwise would.
- A large number of viable routes to victory, given similar starting conditions.

- A large number of possible actions to take at any given time.

This is not a particularly rigid definition, and isn't intended to be; there may be games that people consider to be strategy games that do not fall into this definition, and vice-versa. But there is certainly a wide gamut of games generally considered to be strategy games that *do* fall into this definition: 4X games (*4X Games* 2008) such as the *Civilization* series (*Sid Meier's Civilization VI* 2016); real-time strategy such as Blizzard's *Starcraft* (*Starcraft II* 2010); grand strategy such as Paradox's *Crusader Kings* games (*Crusader Kings III* 2020); and Creative Assembly's *Total War* series (Creative Assembly 2023).

1.2 The Difficulty of Making AI for Strategy Games

Strategy games are complex beasts, and making AI for them is difficult (Schubert 2005). Players manage whole factions or empires: expanding their territory; seeking out new resources with which to build and upgrade; managing relations with other empires; commanding their units to attack, pillage and claim territory. Competently managing all this is cognitively intensive for human players. This contrasts with other types of games, which do not often have so much to keep track of. First-person shooters center around in-the-moment gunplay, rarely requiring significant thought to the future. Puzzle games require thought, but have pre-defined solutions, and do not necessarily even need AI in the first place. Strategy games need opponents, and the games' sheer depth and number of moving parts make the creation of such AI difficult.

They have information-rich maps that can span in-fiction continents (*Old World* 2022), planets (*Sid Meier's Civilization VI* 2016), or even galaxies (*Endless Space 2* 2017; *Stellaris* 2016). Players often command several units, each of which may be ordered to move to any location in range, or to engage in combat wherein *further* actions can be chosen. Players may build structures, colonise new lands or planets, and research new technologies. Fog-of-war hides information that has not yet been scouted. Some factions play incredibly differently when compared to others (*Endless Space 2* 2017; *Master of Orion II: Battle at Antares* 1996), and each automated faction has a limited time budget to calculate any actions they should take. Moreover, full game sessions can take several real-time hours to complete, making it time-consuming to test AI play over the course of a whole game, especially when involving human players. Creative Assembly partly addresses this when developing their games, by simulating playthroughs over the weekend that pit AI players against each other, then analysing the logs afterwards.

To enumerate some of the challenges faced by AI in strategy games:

- High information density
- Large state space
- Large action space
- Hidden information

- Faction asymmetry
- Long testing periods
- Limited AI time budget

All of these present significant stumbling blocks to AI algorithms, even modern ones. How so, and to which AI techniques? We examine these in more depth, first looking at more classic forms of AI, and then moving to the contemporary.

1.2.1 Symbolic AI

Symbolic AI works with symbols: representations of logic, objects, relations between them, and combining these symbols together. These often take the form of rules, for example: “if there is an enemy soldier in a fort, don’t attack them”, or “after researching gunpowder, recruit 3 units of musketmen”. The logic created is explicit, and quite comprehensible to humans. This comes with the advantage of being able to directly and easily modify and debug their behaviour during development. However, these forms of AI must be bespoke to each specific game, using that game’s specific symbols; the AI cannot ‘learn’ new ones from experiences (Harnad 1990). Such AI can also be prone to missing specific situations or edge-cases that the original programmers did not anticipate (Roelofs 2017).

Perhaps most importantly for the sake of this discussion, symbolic AI is in *much* wider use in the games industry than other forms of AI: the postmortems contained within the Game AI Pro series (Steve Rabin 2015, 2017, 2021) are almost entirely dedicated to the logic of symbolic AIs of quite recent games[1]. So: what kinds of symbolic AI exist, and how do they fare when pitted against the challenges of strategy games?

Many of the criticisms that follow don’t hold for other types of games, wherein many of these AI techniques work just fine.

Scripts

Almost certainly the first kind of AI in games to have existed, these are a programmer’s definitions of the exact behaviour that the AI should perform. Take this snippet, for example (in pseudocode):

```
putTeabagInCup()
idle(2 minutes)
removeTeabag()
if(playerAskedForMilk):
   pourMilkInCup()
```

[1] There are some chapters on using non-symbolic AI that indicate they’re mostly used as small parts of the overall AI in a hybrid approach (N. A. Barriga, Stanescu and Buro 2017; Robbins 2019), or for automated QA or balancing purposes (Borovikov 2021; Manabe and Miyake 2021).

```
stir()
```

It's flexible, inasmuch as the programmer can make the AI do whatever they wish within the bounds of logic. They may query and process any state accessible within or without the game itself, and perform almost any action available in the game engine. However, the more complex the logic involved, the messier this gets, especially when edge cases must be covered. Look at the example once more: what happens if there's no milk left? That complicates the code. What if the player pockets the cup partway through? Why doesn't the NPC actually pour hot water into the cup? Players that witness erratic or incomplete behaviour from an edge case that hasn't yet been considered may become frustrated at the seemingly-illogical behaviour.

Each addition and consideration makes the above example more complex, and depending on the game there may be many other systems interacting with it. No: while scripting is incredibly useful, it suffers when required to produce consistent and maintainable behaviours in such large state and action spaces, where so many edge cases abound. It also takes a good AI programmer to maintain or make any changes, and the game may need to be recompiled for these changes to be made — both things to be avoided where possible, lest development be slowed. So, while scripting is an essential building block that may be combined with other AI techniques, it isn't going to provide good strategy game AI on its own.

1.2.2 Finite State Machines

Agents using finite state machines (Gill et al. 1962) exist in one of a finite set of states at any given time (eg *Idle*, *Alert*, *Fight*), and may transition between states based on conditions. Let's take a simple guard AI:

It's relatively easy to understand the flow of the guard's behaviour, and Finite State Machines can be (depending on how much power you want to give your FSM) almost as powerful as scripting. They're more digestible to non-programmers than scripts, making it far easier for a team to give feedback and collaborate on behaviour. However, in the same way that interacting systems can exponentially increase the complexity of a script, they can force FSMs to become complex and nested. Even in this example, the Fight state must specify what the guard should do; within it lies another FSM. The same is true for Alert. In this way, high-level, abstract FSMs are quite capable of becoming rabbitholes of complexity. Once more, explicitly specifying what gets done under exactly which conditions results in a large amount of manually specified behaviour. Flexible, in that almost any given behaviour can be explicitly constructed. Brittle, in that a lot of such construction is required to avoid it breaking in edge cases.

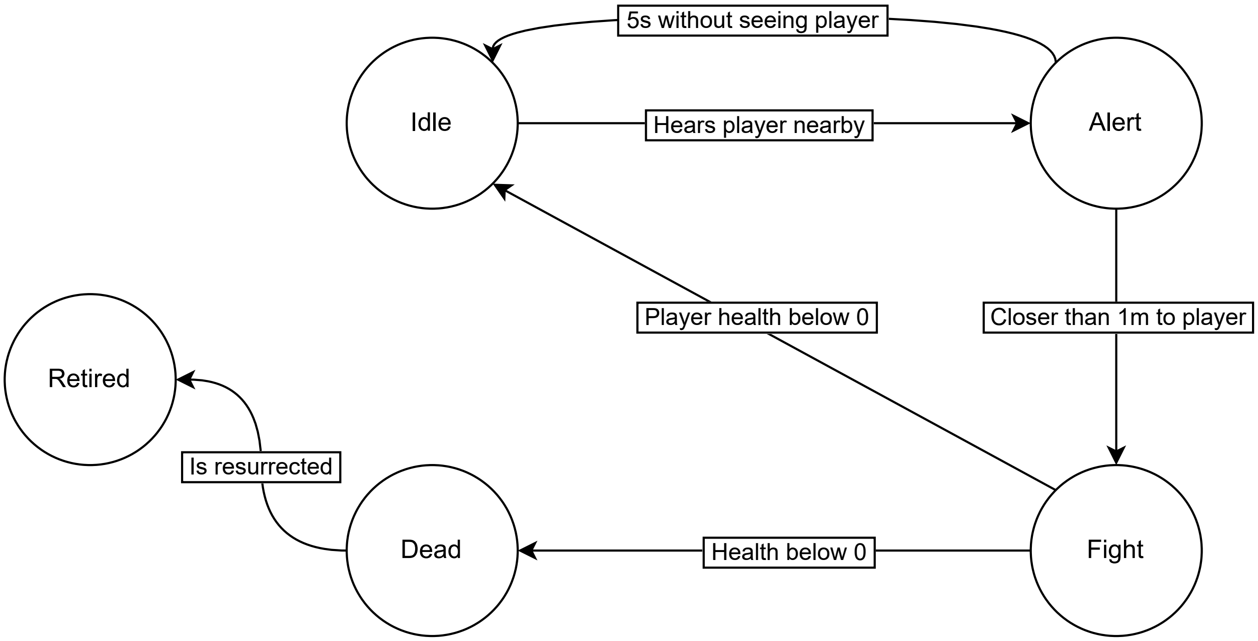


Figure 1.1: Example Finite State Machine for a Guard AI

Behaviour Trees

Originally introduced to control enemy NPCs in *Halo 2* (*GDC 2005 Proceeding* 2019; *Halo 2* 2004), Behaviour Trees describe behaviour using a tree structure (Chris Simpson 2014). The tree consists of nodes which, when evaluated, do something and return their state: either *Success*, *Failure* or *Running*. "Do something" depends on the type of node it is:

- **Composite** nodes have children which will be evaluated in some order. For example, a *Sequence* node would attempt to evaluate each of its children in order, only eventually returning *Success* if all its children do so also. A *Selector* node would evaluate its children in order, until one of them returns *Success*. At this point, it will stop executing and return *Success*. If none of its children succeed, it'll return *Failure*.
- **Decorators** can only have one child. It controls under what conditions its child can run.
- **Leaf** nodes have no children, and when executed they do something within the game world, or ask a success/fail query of that world. For example: *Cast fireball*, or *Flail madly with staff*, or *Am I really a wizard?*.

Put together, behaviour trees are suprisingly effective. The leaf nodes dictate what actions *can* be done by the AI, and composite and decorator nodes dictate when to execute these. Nodes returning their state allows for robust fallback behaviour: imagine a composite node that attempts to execute each of its children in turn until one succeeds. Don't have a key to open the chest? Bash it open. That didn't work? Pray for divine intervention. Your god didn't listen to you? Go home. That behaviour tree might look something like Figure 1.2:

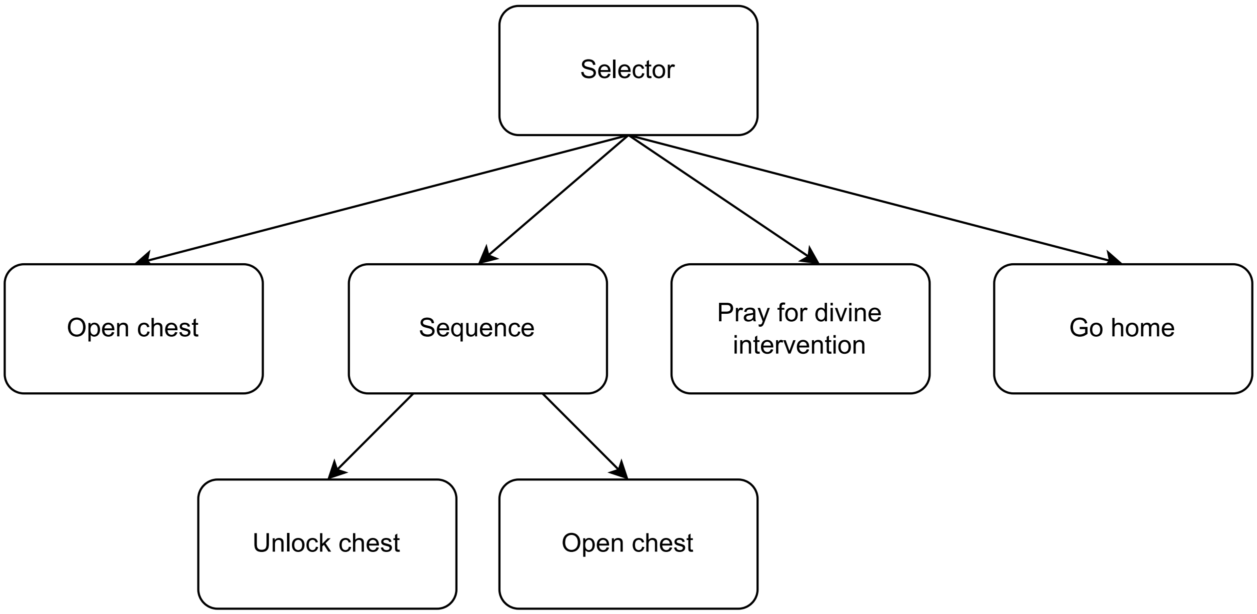


Figure 1.2: Example Behaviour Tree for trying to open a chest

As they quite easily handle failure of any given 'branch', Behaviour Trees scale better than Finite State Machines. Perhaps more importantly, they're *accessible*; given a drag-and-drop editor and a small library of leaf nodes, it's possible for non-programmers to put together an AI, and as the tree itself is a data structure, it can be edited without recompiling the whole game — a grand improvement. These qualities have given the technique staying power in the industry: it's still a prominent AI component in *Unreal Engine 5* (*Behavior Trees* 2023).

Once more though, this technique has a limit on the complexity of behaviour it can reasonably govern before becoming a sprawling tangle. Such behaviour trees can also take more time to evaluate, and many strategy games are complex enough to strain their limits, and behaviour trees would almost certainly not be able to approach human levels of play, resulting in players that don't feel challenged — unacceptable in strategy games. There has been some success in using genetic evolution to improve behaviour trees for some strategy games (Hoff and Christensen 2016; Lim, Baumgarten and Colton 2010), but these successes are in games that are quite small compared to the industry's current larger games (*Endless Space 2* 2017; *Sid Meier's Civilization VI* 2016; *Stellaris* 2016).

Utility

Utility agents use a deceptively powerful idea: that each possible action they could perform gets a calculated *utility score* (Fishburn 1970). The action to perform next is simply the action with the highest utility each tick of decision-making. For example, take an agent in a survival simulator. If the *Sleep* action is given 10 points, and the *Eat* action 5, you will have an agent that sleeps all the time, eventually perishing from hunger. But if the value of *Eat* goes up over time... that agent will eat periodically.

The power of this approach is that agents can flexibly take their environment into account, with much less need to specify edge cases. You could have actions and utility values related to combat, tactics, positioning... but if the agent gets hungry enough, they'll still take out their sandwich and munch it down mid-battle. Most importantly, no connections between these two distinct areas of behaviour need be written out to allow this kind of emergent behaviour.

A *weakness* of this approach in the context of strategy games is that game designers must still create the utility functions the AI uses, and then debug them. With a lot of utility functions running, and a whole lot of possible actions, it can be difficult to ensure that the agent is still doing exactly what you intend it to be doing. Tweaking the utility functions to provide exacting behaviour in complex games over a long period of time can be a realm of difficulty on its own. There's a lot of emergent behaviour that wouldn't make sense in a strategy game — for example, performing a utility action that improves the agent's *present* situation, but hampers its situation severely later on.

Another weakness is that their behaviour can be prone to oscillation if the utility of one course of action hovers quite close to that of another ("An Introduction to Utility Theory" 2014). A unit might acquire a target and start walking towards it, but soon *another* target might become a juicier target, so it starts walking towards that one... Seeing enemies repeating unfruitful behaviour ends up damaging the character's believability, and being quite exploitable.

1.2.3 Tree Search: Looking Into The Future

This is a type of artificial intelligence that looks into the possible futures that could occur from the present (N. A. Barriga, Stanescu and Buro 2017). Starting at a given game state, you may take any one of several actions. These actions lead to other game states from which you can perform other actions in, and so the branching tree of possibilities grows into the future. By exploring this tree, we can find which sets of actions result in desirable states, and which do not.

What is a "desirable" state? It depends on the game, and it depends on the behaviour you intend to produce. An RPG character who does not wish to die might judge a state's desirability as "the number of hit

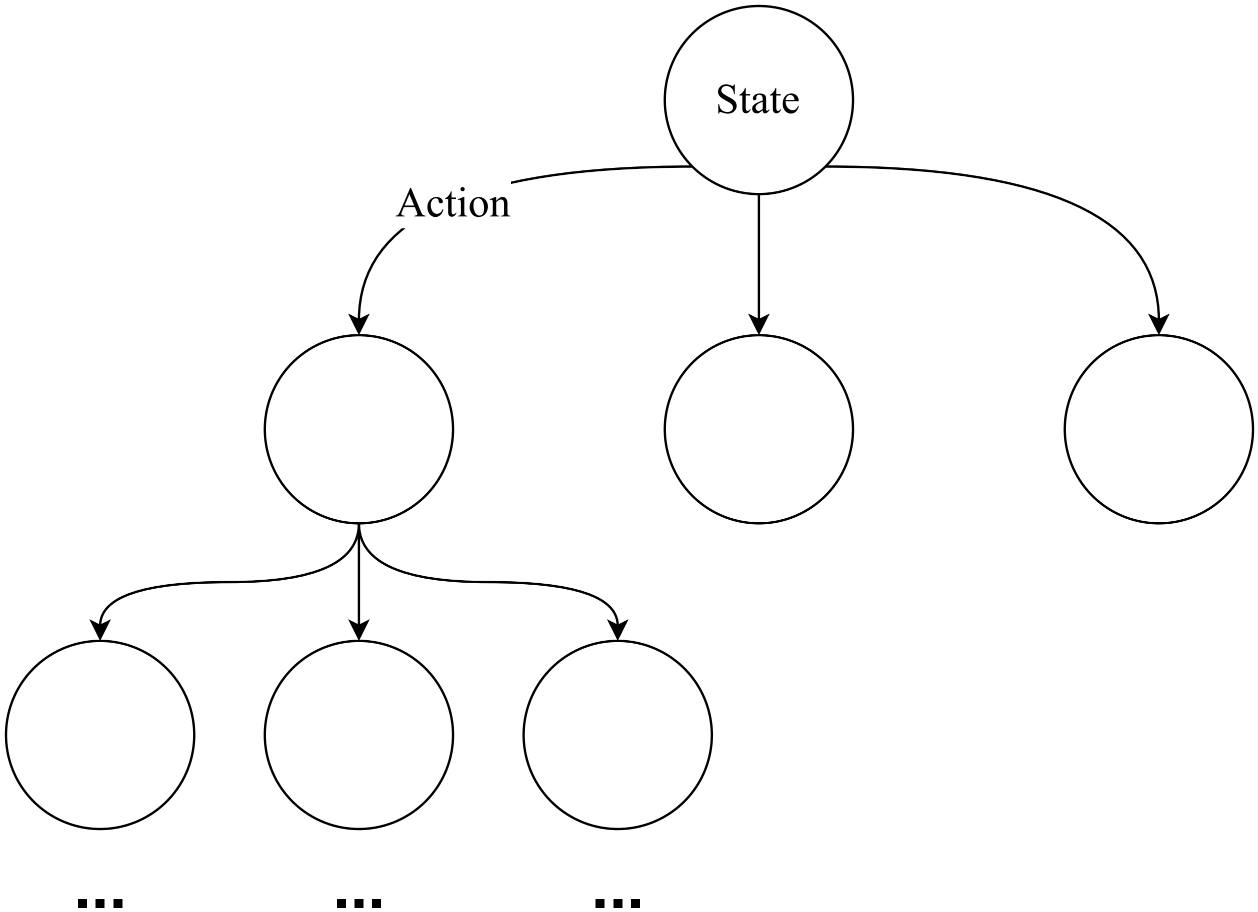


Figure 1.3: Example of a state tree

points I have". An agent in a racing game might judge by "how fast I'm travelling towards the finish line". This function that takes a state and returns how desirable it is is called the *heuristic*. Heuristics may take into account a great many factors, but it's often difficult to create one that reflects the true importance of a state for a given outcome.

To summarise, in order to grow, explore, and evaluate a tree of game states, tree search algorithms share the following requirements:

- A function returning the actions that can be performed from a given state.
- A function that takes a state, and an action, and returns the resulting state.[1] This is called the *forward model*.
- The aforementioned heuristic.

[1]This gets messy when it comes to actions with uncertain outcomes, but we'll abstract that away for now.

These might seem like stringent requirements, but to have a working game, one must already simulate what happens when an input is given, so the bulk of the work is already done; as methods of *exploring* the tree are largely game-agnostic, they're extremely flexible: you can take the same tree-search algorithm and easily apply it to several games.

However, they also present some disadvantages.

- **Not quite as effective in real-time games** Compared to games with large and discrete timesteps, actions in real-time games occur with much smaller lengths of time between them, often delaying rewards far into the action future. This also means that exploration has to be done before those small timesteps complete, lest the agent lag behind the actual gamestate.

- **Vulnerable to large action spaces** The amount of states to explore and evaluate grows exponentially with the number of actions that may be taken. A game in which you can take 3 possible actions a turn multiplies possibilities by 3 each progressive action: 3, 9, 27, etc. Increasing the possible actions to 10 makes for 1000 states[1] by the time you get to the third turn.

- **Inscrutable behaviour** It is often hard to discern *why* the action a tree-search algorith performs is considered better than others. This can make it difficult to debug, and difficult to use for design purposes.

Some variants of tree search can alleviate these disadvantages, and research in this area is ongoing (Świechowski et al. 2023).

Vanilla Tree Search

The most naïve version of tree search, vanilla considers a state's score to be the best score of all the futures that can be accessed from it. Once a future state is evaluated with the heuristic, its score is passed up through all the states that led to it, until reaching a state with a higher score. Each state ends up having the highest score of the states explored in its future. By choosing actions that lead to the next highest-scoring state from the present, the agent moves towards the best future it has seen.

Minimax

Minimax is used for games where multiple players take turns opposing each other, such as *Chess*, *Battle Chess* 1988, Zubrin 1972, or *Sid Meier's Civilization VI* 2016. It modifies vanilla tree search by assuming an agent's

[1] Or fewer, because same may be duplicates

opponents will do the absolute worst thing possible for the agent's goals. Each future state where an opponent chooses an action, instead of passing the *best* score for the agent, the *worst* score is passed up.

Thus, in a two-player game, the algorithm passes up the minimum score, then the maximum score, then minimum... hence the name minimax.

This isn't incredibly useful for many NPCs, as directly and absolutely opposing the player isn't often the goal, especially if the NPC is to have their own agendas to complete.

Monte Carlo Tree Search

Monte Carlo Tree Search (MCTS) (Chaslot, Bakkes et al. 2008; Chaslot, Saito et al. 2006) acknowledges the fact that there are often immense amount of gamestates to search through, but limited time before an agent must choose an action. So it doesn't evaluate all future gamestates; it instead evaluates a portion of them, and builds up a statistical average of how good a state's future might be.

It consists of four stages:

- **Selection** Starting at the root of the tree, one of the current state's children is chosen. The choice is returned by a *tree policy*, which usually balances *exploitation* (choosing a node which has not yet been explored much) and *exploitation* (choosing one of the more promising explored nodes).

- **Expansion** An action that has not yet occurred from the chosen state is performed. The resulting game state is added to the game tree, as the child of the previously-chosen node.

- **Simulation/Rollout** From the new node, actions are chosen by a *default policy* until a terminal state or maximum depth is reached. The final state is given a score via a heuristic.

- **Backpropagation** This score is backed up through the chain of nodes that led to it, updating each node's simulated score.

These steps are continuously performed, either until a time budget is reached or there has been a certain number of rollouts. Given enough iterations, MCTS is able to gather a statistical aggregate of how 'good' actions are from the current state.

A significant advantage of MCTS is that in some applications, it can use wins or losses reached from a state in place of a crafted heuristic. However, this becomes unfeasible if the game tree is particularly deep or wide; it becomes impossible to actually reach those terminal states enough to gather useful statistics. Unfortunately, strategy game trees can be very deep and wide indeed (Roelofs 2017), so this advantage does not apply in their context. This means that a heuristic must be crafted, and once more this becomes prone to edge cases and complexity. A game state might score highly for one faction, but that score might be inappropriate for another faction that plays differently. When issues in behaviour arise, it can be difficult to change a heuristic without damaging other parts of the AI's behaviour.

There are a significant number of alterations to vanilla Monte Carlo Tree Search, though these are often applied to simplified versions of strategy games (N. Barriga, Stanescu and Buro 2017; Browne et al. 2012; Yang and Ontañón 2019).

1.2.4 Reinforcement Learning

Reinforcement learners receive a reward signal for each action they perform, and their purpose is to maximise the amount of reward received (Sutton and Barto 2018). They contain an internal *policy*, a mapping of which action to perform given a state. By exploring actions and their given rewards, the policy can be updated and an agent's received reward signal increased. Good reinforcement agents can have impressively nuanced and skilled behaviour, the most famous being *AlphaGo* (Silver et al. 2016) and *AlphaStar* (Vinyals, Babuschkin, Chung et al. 2019), though both had astronomical training costs. While reinforcement agents that use a table of states and actions soon grind to a halt with the large state spaces as large as those of strategy games, the field has other agents better suited to the task: those that approximate the state-action function. These manage to somewhat address the large state-space problem, but are confounded by the delayed rewards that strategy games present (Krishnan et al. 2016). For example, training a unit early on in the game and therefore securing territory may drastically increase chances far, far into the future of the game, with the benefits compounding over time. This is quite difficult to get around.

Reinforcement agents are also a black box; it's difficult to take their internal policy and convert it to a logical interpretation that humans would understand. This can make it more difficult to reproduce and debug bad behaviour, and any changes to given rewards must be carefully applied in order to maintain the integrity of good behaviours. The same concept applies when the agent's environment changes; change the environment, and behaviour may be broken until the policy can once again match up to the environment. As a game changes rapidly during development, it would be difficult to keep agents up to pace with these changes while also debugging them.

1.2.5 Markov Decision Processes

A Markov Decision Process (MDP) (Puterman 1990) codifies what happens when you perform multiple sequential actions in an environment, including what rewards you get while making these actions. It consists of the following:

- A set of states

- An initial state
- A set of possible actions that may be taken from each state
- A transition function $P(s'|s, a)$. That is, from a state, the probabilities of reaching other states with a given action
- A reward function $R(s, s')$: the immediate or expected reward after transitioning from one state (s) to another (s')

Markov Decision Processes are very useful frameworks for reinforcement learning agents. They can be used for *passive learning* (evaluating how accurate its policy is) or *active learning*, whereing the agent attempts to modify its policy to maximise reward. Passive learning can be used to evaluate how balanced an AI's behaviour is, while active learning is useful for an agent's initial training, and possibly in-game[1].

However, MDPs have some constraints. The environment must be fully observable, and Markovian. That is, a state is self-contained; transitioning from that state does not depend on any states visited previous to this one. Strictly speaking, strategy games are *not* fully observable, as information is hidden in fog-of-war. It might therefore be sensible to use Partially Observable MDPs (Shani, Pineau and Kaplow 2013; Spaan 2012) , which instead of using the actual state uses its partial observations to construct possible actual states and act accordingly.

As this presents more difficulty, both here and in other methods of creating AI agents, many strategy games simply cheat, and make the entire game state fully observable to agents. Many players, as seen later in this thesis, do not appreciate this.

1.2.6 Neural Networks

It was earlier established that storing expected rewards of actions from states in a table is untenable for the large state and action spaces of strategy games, and that instead it's often more suitable to approximate the utility function.

One such function approximator is a *neural network*. These are formed from a connected collection of smaller, less flexible function approximators called *neurons*.

Figure 1.4 displays a neuron:

Each *input* has a matching *weight*, and the neuron itself has a *bias*. The weighted inputs, plus the bias, are summed and passed to the (usually simple) function, which produces the neuron's ouput (Siddharth Sharma, Simone Sharma and Athaiya 2020).

[1] This may not be necessary; most games do not change their rules during play, and those that do could be considered to have the 'new' rules be simply permutations of the initial rules

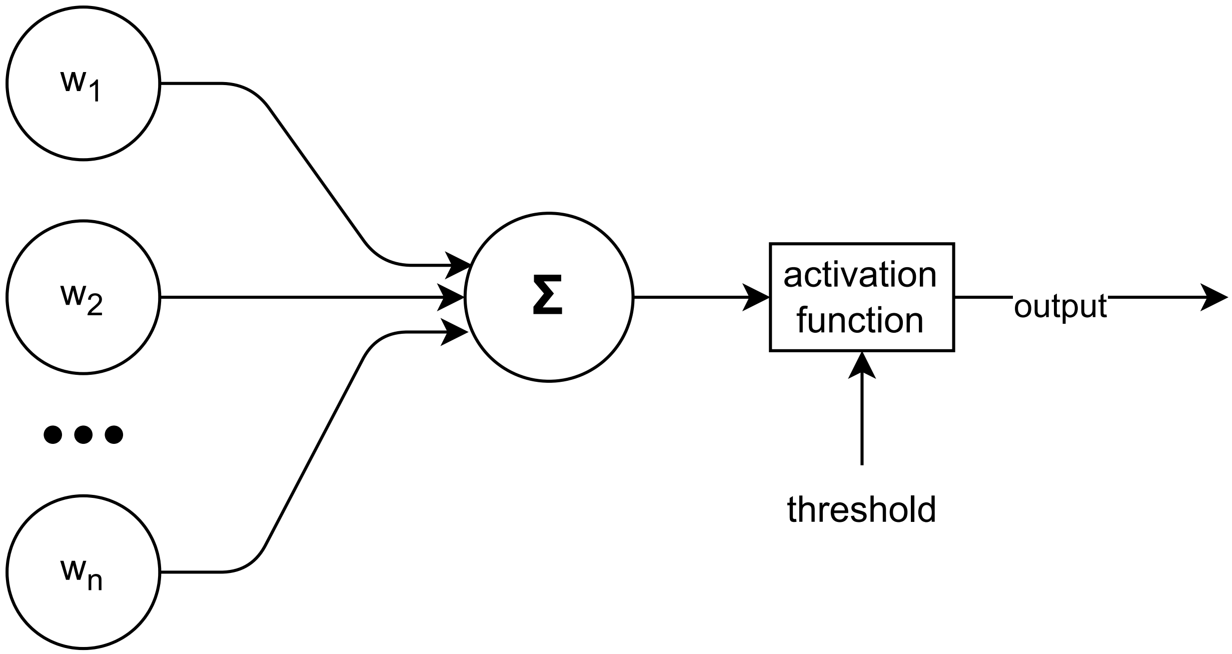


Figure 1.4: A neuron

$$\begin{aligned} \text{Input count} &= n \\ \text{Inputs} &= x = x_1, x_2, x_3 ... x_n \\ \text{Weights} &= w = w_1, w_2, w_3 .. w_n \\ \text{Bias} &= b \\ \text{Neuron Function} &= f \\ \text{Output} &= f(w \cdot x + b) \end{aligned} \tag{1.1}$$

By adjusting the weights, bias, and neuron activation function, the imitated function can change. For example, different neurons can effectively perform several boolean operations.

While a single neuron may only approximate simpler functions, more complex functions may be approximated by connecting the inputs and outputs of different neurons together. In a neural network, the outputs of one layer are the inputs to neurons in the next layer, until the final output layer of the network is reached. This allows for progressively more complex functions to be approximated.

This is interesting, but it becomes *useful* when you consider that a neuron's approximated function may be changed by changing its weights and bias. Therefore, a neural network's approximated function may be changed by altering the weights and biases of its consituent neurons. This allows for *adjustment* of the function towards desired outcomes.

1.2.7 Genetic Algorithms

Originally conceived by Holland (Holland 1992) , genetic algorithms move towards better solutions for a problem by evolving a population of programs over time. Each program is executed, and its *fitness* is measured — how good was it at solving the problem?[1] Some programs are then selected from the population, and are altered by genetic operations, resulting in changed 'child' programs that join the population. By periodically culling the least fit from the population, the average fitness of its members increases over time.

The changes that are made to programs are dependent on the fitness function, how individuals are selected from the population to 'breed', and the genetic operations themselves, which often incorporate a level of randomness to the changes they make.

Program Representation Programs are often stored as a tree, similar to Abstract Syntax Trees used in the compilation of programming languages. These trees consist of operations[2] and their operands. To evaluate a given tree, its subtrees are first evaluated, and then the final operation performed.[3]

Genetic Operations Genetic operations change individuals in a population. Apply them to parent program(s), and you get a child program that is different. Two such operators are *crossover* and *mutation*. Crossover produces a child sharing some nodes of the tree with their parent, and the rest with the other. Mutation needs only one 'parent', and applies a small amount of random change to some nodes in the parent tree. For example, one operator might change to another, or an operand's value shifted slightly.

Selection To change individuals in a population, you must first choose *which* individuals. Therefore, it's reasonable to assume the fittest in the population would be the best foundation for change, but this isn't always the case; focusing solely on these might overlook a completely different kind of solution that is a great deal better. So instead, an individual's fitness merely affects the probability that it will be selected for change. For example, one method is tournament selection, where a random sample of individuals is taken from the population, and the best of that sample selected.

An advantage that genetic algorithms hold over neural networks is that the contents of its population can also give some clues as to why its fittest individual is effective; as it contains mostly programs that have been working well so far, common elements and shared behaviours can hint at keys to their success.

Genetic algorithms have been successfully used to create AI for *Backgammmon* and *Core War* (*Core War the Ultimate Programming Game* 1984). They are also able to develop behaviour trees, such genetic

[1]This equates to the *heuristics* used in other kinds of AI
[2]These are not the *genetic operations* that alter populations' individuals.
[3]This does not result in infinite evaluation, as subtrees are finite, and leaf nodes resolve to a value.

algorithms being used to moderate success in a Mario AI competition (Togelius et al. 2013). The resulting agent was particularly effective at reactive behaviour.

However, these successes have been limited to smaller games; in the field of strategy games they've thus far only been used to augment smaller facets of the AI (Tavares et al. 2017; Watson et al. 2008).

1.2.8 AlphaGo and AlphaStar

While it has been mentioned that many reinforcement-learning approaches have difficulty with particularly complex and long-form games, *AlphaGo* and *AlphaStar* are notable exceptions (Silver et al. 2016; Vinyals, Babuschkin, Chung et al. 2019; Vinyals, Babuschkin, Wojciech M. Czarnecki et al. 2019).

To summarise, *AlphaGo* reached superhuman levels of play in *Go*, and *AlphaStar* reached grandmaster levels of play in *Starcraft II* (Vinyals, Babuschkin, Wojciech M. Czarnecki et al. 2019), but the sheer scale and expense of the training involved for each of them makes it impossible for game studios to replicate.

There are no studios currently that can command this level of capital, and certainly not for one aspect of a single game. So while both *AlphaGo* and *AlphaStar* remain sterling technical achievements that are incredible examples of competitive AI, the resources that are used for training would have to be reduced significantly to become usable in industry. The resources used to *run* them are much smaller, apparently running on a "single desktop GPU" (Vinyals, Babuschkin, Chung et al. 2019). However, information is scarce on how much of the GPU's resources were used; just running the agent may still push the total hardware requirements to well beyond what most consumers have available to them.

1.2.9 Overview

We've now seen that while there are many methods with which to create AI in today's world, none seem to fit the bill of creating good strategy game AI; all have tradeoffs or downsides. Some simply cannot provide an adequate challenge to the players it's pitted against, unless they are provided boons the player does not have. Others produce erratic behaviour and damage characters' believability.

1.3 What Makes For Good AI?

To ask whether the methods we have are enough, we must first evaluate what is needed by an AI opponent, and the literature offers some options. First is the concept of *flow*.

1.3.1 Flow, and Challenge

The concept of flow was first introduced by Csikszentmihalyi (Csikszentmihalyi 1990); it is a state of mind in which nothing else seems to matter, and the subject is completely and wholly immersed in the activity producing flow. Importantly, it's also very enjoyable (Csikszentmihalyi 1990; Sweetser and Wyeth 2005).

For example, many readers will be completely consumed while they read an excellent book. They will not notice the world around them, and when they eventually stop reading they are often surprised at how much time has passed.

Similarly, other circumstances can bring about flow: rock climbing, playing games, creating art, or athletics (Csikszentmihalyi 2014).

Csikszentmihalyi proposed that several things come together to form a flow state:

1. A balance between a challenge and one's skill to complete it
2. Clear goals
3. Direct, immediate feedback
4. Concentration on the task at hand

Resulting in:

1. A merging of action and awareness
2. A sense of control
3. An altered sense of time

When these are combined, flow becomes possible. Most relevant to AI in this is *a challenge requiring skill to address* — in games, AI opponents are often the means by which the player is challenged. More specifically, flow requires the challenge to be roughly proportional to the subject's skill at navigating that challenge. Too difficult for them and they will be unable to continue. Too easy, and they may well become bored. The sweet spot between the two is what some call the ”flow zone”, and can be diagrammed as in Figure 1.5

As flow is so desirable in games, it's surely critical to manage the amount of challenge that is presented to the player (Jenova Chen 2007). Dynamic Difficulty Adjustment aims to achieve this.

Dynamic Difficulty Adjustment

Dynamic **D**ifficulty **A**djustment is the practice of adjusting a game's difficulty in real-time to match the skills of the player more closely (Hunicke 2005). For example, a player who returns to an earlier area of a game

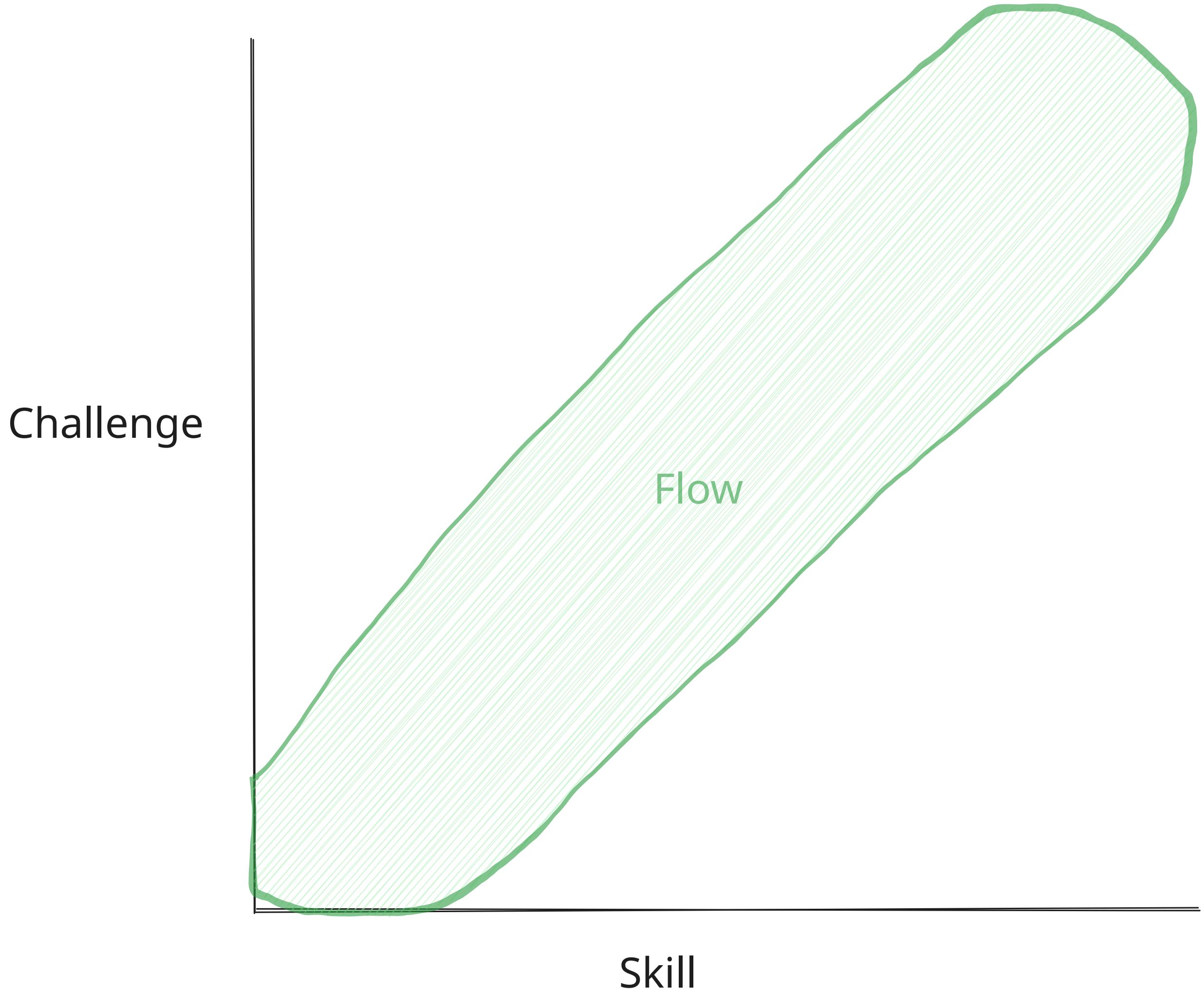


Figure 1.5: Diagram of Flow

might find the low-level monsters there still a challenge to defeat in combat; in addition to supporting flow, the idea is to ensure the player does not become bored and that parts of the game don't become an easy routine. Other games might slightly tone down monsters that would otherwise be extremely difficult at that stage, though this is rarer. DDA has been implemented in a great many games, from *The Elder Scrolls V: Skyrim* to *Left 4 Dead 2* (Bethesda Game Studios 2011; Jordan Baranowski 2018; *Left 4 Dead 2* 2009), and certainly has its uses.

Unfortunately, while DDA can successfully adjust the amount of challenge presented to the player, it often becomes detrimental to player experience when players become aware of it (Alex Vu 2018; Zohaib 2018), and can even be exploited.

Imagine you are throwing basketball hoops in an arcade while blindfolded. You can hear the machine ding when your ball falls through the hoop, and you reckon you're doing pretty well... this is difficult stuff, but you're managing it. Then your friend takes off your blindfold mid-game. Oh, the *hoop has been moving to catch your throws* (Stuff Made Here 2020). Suddenly, your sense of achievement has withered, and you realise that you can just... lackadaisically toss the ball in the vague direction of the hoop.

With the coverage of the internet and games communities, it becomes possible to find out the game is dynamically adjusting difficulty even if you yourself haven't discovered it; the façade is much easier to break. It's already a fragile mask, as DDA can stretch the game's fiction to breaking point.

However, without DDA, game balance must carefully be considered in order to not alienate players of different skill levels[1].

1.3.2 Believability: As Character

Providing challenges for the player is not the only role of the NPC; they are also part of the fabric of the game's fiction. They are explicitly a character — not necessarily a human one, take The Luggage in *Discworld* as an example (*Discworld* 1995) —- and must be believable as such. There are different types of NPC as identified first by Bartle then iterated upon by Warpefelt et al (Bartle 1996; Warpefelt and Verhagen 2015), and each has its own mechanisms for augmenting believability, depending on the mechanics they interact with and their environments: see Table 1.1.

1.3.3 Believability: As Human

A separate but related concept is the idea of making it look like the actions of an autonomous agent are not in fact being driven by a machine, but a human.

It can sometimes be desirable to make players believe they are playing against or with a human. Their perception of the agent changes, and their overall experience as a result. For example, Killers from Bartle's original player types (Bartle 1996) prefer to inflict emotional damage, and thus derive no joy from crushing what they believe to be a robot. There are other players that are aware that their opponents aren't human, but are willing to suspend their disbelief. But if the opponent breaks enough from this human-fiction, that disbelief will break, along with their sense of immersion. Players rather enjoy their immersion (E. Brown and Cairns 2004), so breaking it is likely to incur their frustration.

What it means to 'play like a human' is somewhat difficult to define, and seems rather subjective; players will use the same logic to conclude an entity is an AI as they do for concluding it's a human (Pacheco, Tokarchuk and Pérez-Liébana 2018).

[1]Exceptions are made for games whose difficulty or lack thereof are an integral part of the game: see Soulsborne games and cookie clickers (Deterding et al. 2022)

NPC Type	How they augment believability
Buy/Sell/Make Stuff	Via the details of their trading environs, and the wares they sell to the player.
Provide Services	These characters are often visibly matched to the kind of service they provide. A chef *looks* like a chef.
Provide Combat Challenges	The more challenging this is, the more badass they should look.
Provide Mechanical Challenges	NPCs, similarly to humans in this world, can present problems without you having to fight them: in *The Secret World* (*The Secret World* 2012), NPCs on the Tokyo underground act as the walls of a maze.
Provide Loot	Are they wearing cool gear? They should drop loot. Are they an animal? They may drop fur and hides. They may even be designed to drop whole hoards of loot when killed, such as treasure goblins in *Diablo III* (*Treasure Goblin* 2019)
Give or advance quests	Their quests will often be related to their narrative; a druid might ask you to bring a great sacrifice to the strange prehensile tree in the swamp (*Roadwarden* 2022)
Provide narrative exposition	Has likely been heavily involved in the events they're detailing
Assist the player	This category is quite vague, but there are often narrative (or tutorial) reasons a given NPC is assisting the player, and this assistance and that narrative lend weight to each other.
Act as an ally in combat	Same as above. Their combat style is character-dependent, though most of the difficulty here lies with not getting in the way of the player.
Accompany the player	Also vague, but are likely to spend a lot of time around the player, so they need extra attention to detail, aesthetic, and player interactions.
Make the place look busy	The base NPC, the foundation of civilization, these just need to look appropriate to their environment, and possibly be going about their business. One of the most well-done examples of this are the fauna of *Subnautica* (*Subnautica* 2018)

Table 1.1: Methods NPCs use to enhance believability

However, that there are some genres in which believable-as-human is much harder to achieve than others. It would take a long time of consistent behaviour to fool someone in an MMO, and until recently with the advent of believable language models (*ChatGPT* 2022), natural-language communication with non-human agents could not be achieved, which almost immediately 'outs' an agent.

1.3.4 Immersion

AI believability feeds into player *immersion*. This is essentially the degree of a player's involvement with a game (Bartle 2001; E. Brown and Cairns 2004), and progresses through three tiers, four if 'no immersion' counts as one.

1. **Engagement:** The player has invested time, effort, and attention to the game. For this, the player needs both a willingness to play the game, and for the controls to allow that player to master them.

For example, an avid lover of shooter games would likely be willing to play another game in that genre. However, fewer players would engage with that same game if they found themselves unable to aim effectively enough.[1]

2. **Engrossment:** Players have become more involved, and their emotions are directly affected by the game. Game 'construction' contributes to this state of affairs: visuals (whether directly rendered or text-based), interesting tasks, narrative. All increase the player's willingness to dedicate attention to the game. At this point, players can be emotionally affected by events: they might be upset that their character died.

3. **Total Immersion:** This adds presence (Bartle 2007; Lombard and Ditton 1997), in which the mediation between the player and the virtual environment disappears: the player sees themselves *as* the entity in the virtual environment. Moreover, the 'real' world essentially disappears too; the game is the only thing on the player's mind. They're in the game.

These levels have also been referred to as *Avatar*, *Character*, and *Persona* respectively (Bartle 2001).

AI can contribute to immersion by driving the behaviour of Non-Player Characters (NPCs), as those NPCs are part of a game's fiction. If the player is able to suspend their disbelief regarding NPC behaviour, the NPC can form part of the game world the player immerses themselves into. The driving AI does not need to be complex at all to do this: just as a city in *World of Warcraft* does not need to be photorealistic for players to immerse themselves in it, an NPC shopkeeper can serve the fiction perfectly well simply by selling goods to the player (Warpefelt and Verhagen 2015). That is, an NPC does not have to fool players into thinking it's human, just that it's believable as the character it's presenting.

There are claims that AI can damage immersion with egregious or obviously patterned behaviour (Warpefelt and Strååt 2013). While this is yet to be empirically tested, and many players' suspension of disbelief may cover the 'gap' in behaviour (D. W. Brown 2012), some guidelines are presented to at least improve the player's experience (Warpefelt and Strååt 2013): NPCs' awareness of the world should occupy something of a sweet spot. NPCs should not be omniscient, but neither should they be unaware of fictionally relevant events. Their reactions to events were also important. Should an NPC not react at all to what they know, or act in a way that makes their present situation worse, that too can break immersion. .

Balancing the contradictions within these guidelines is something of an art form, and exceptions can occur. Shopkeepers in *The Elder Scrolls V: Skyrim* were vulnerable to the *Buckethead Exploit*(McMullen 2021). If a player placed a wooden bucket upon a shopkeeper's head, they could rob the shopkeeper blind without them

[1] This is likely why many console games have 'auto-aim', which snaps the player's aim onto entities when their aim is near enough to that entity.

knowing they were being robbed at all[1]. This seems to violate many of the above principles, but it became a part of *Skyrim*'s culture.

It should be noted that not all games need NPCs, and therefore AI agents, to immerse their players. *PowerWash Simulator* does it perfectly well (Vuorre et al. 2023) without using them. However, there are relatively few game fictions that support the player being the only entity in the game, so AI certainly remains relevant.

1.4 So What Do Agents Need?

In short, what agents need to do is dependent on their role. For example, bosses in games exist to give the player a huge challenge initially, and for that challenge to be mastered by the player over time. To that effect, they're traditionally predictable. They exhibit fixed patterns of behaviour that are initially very difficult to get past, but with more experience fighting the boss, players find it easier and easier. A boss, as a rule of thumb, should not deviate from their patterns of behaviour; exhibiting human believability by changing their behaviour would in most cases be disastrous.

However, agents specifically designed to draw the attention of Killers in MMOs so that human players did not suffer the detrimental effects of their murderous tendencies? Those would need to be *very* believable as humans.

Veteran undead guards of a long-forgotten tomb? These must be able to provide the player with a challenge in combat, but would also benefit greatly from being walking glimpses of the traditions of a dead empire.

It is clear that different NPCs must fulfil different criteria, but some of these are more important to some players than others.

1.4.1 Types of Player

No two humans are quite the same, and that remains true when considering why people play games. Some like to network with others. Others like ganking[2] unsuspecting players that made the mistake of taking a shortcut through a dark alley. Others still prefer venturing to the bounds of the world to see all it has to offer.

In short, there are different types of players that want different things from the games they play. By covering literature on what players want from gameplay, we can highlight that the current state of industry AI is unable to fulfil these completely.

[1] This was an emergent interaction between *Skyrim*'s physics system and NPCs making active use of vision. The physics engine allowed the bucket to be placed on the NPC's head, and the bucket actually blocked the NPC's vision. While the behaviour of the NPC was anomalous, it was also *understandable*.

[2] Murdering with substantial and possibly gleeful violence, ideally non-consensually

Bartle's Player Types

One of the first models of why players play was created by Bartle (Bartle 1996), derived from extensive forum discussion between many of the most experienced players of *MUD2* (Richard Bartle and Roy Trubshaw 1985). While it has been iterated on since, it's useful to start here so that evolutions may be better understood.

It posited 4 types of **M**ulti **U**ser **D**ungeons, the precursors to **M**assively **M**ultiplayer **O**nline **R**ole**P**laying **G**ames. Those player types are distinguished by whether they *act* or *interact*, and whether they are *world-oriented* or *player-oriented*.

- **Killers** Act on *players*. They demonstrate their superiority over others by doing things to them, often without consent. Interestingly, some claim "there's no fun unless it can affect a real person instead of an emotionless, computerised entity" (Bartle 1996). This is one case where ticking 'believable as human' might be useful, though difficult to do.

- **Achievers** *Act* on the *world*. They focus on mastering the game, manipulating its mechanisms to make it do their bidding. For example, players that create fully-functioning computers out of redstone in Minecraft. (Torb 2023)

- **Socialisers** *Interact* with *players*. The people and characters are far more interesting to them than the world itself. They gather friendships, contacts, and influence.

- **Explorers** *Interact* with the *world*. They enjoy the surprises, novelties and beauty that the world offers, and seek out a sense of wonder through exploring.

Figure 1.6 depicts these axes and where players lie on them.

However, players change types over time, and there are subtypes of the above 4 player types. So the typology was expanded to 8 player types with another axis: *implicit* <–> *explicit*. *explicit* subtypes act with plans in mind, goals towards which they are working. *Implicit* subtypes act more spontaneously, with less forethought and more reaction.

For example, *Griefers* are *implicit Killers*: they grief[1] other players opportunistically, without a scheme in mind. But *Politicians* are *explicit*; they plan and manipulate others towards their own goals.

Is what we have enough? Assuming this lens of player types holds for strategy games — Bartle only defends it for virtual worlds — it's clear to see that strategy games can cater for Explorers; it's quite possible to craft a beautiful world and to represent its denizens in artistic glory. However, they cannot cater for Socialisers — AI is currently capable of little natural-language communication, with almost all communcation

[1] Grief: Harass other players by sabotaging their gameplay.

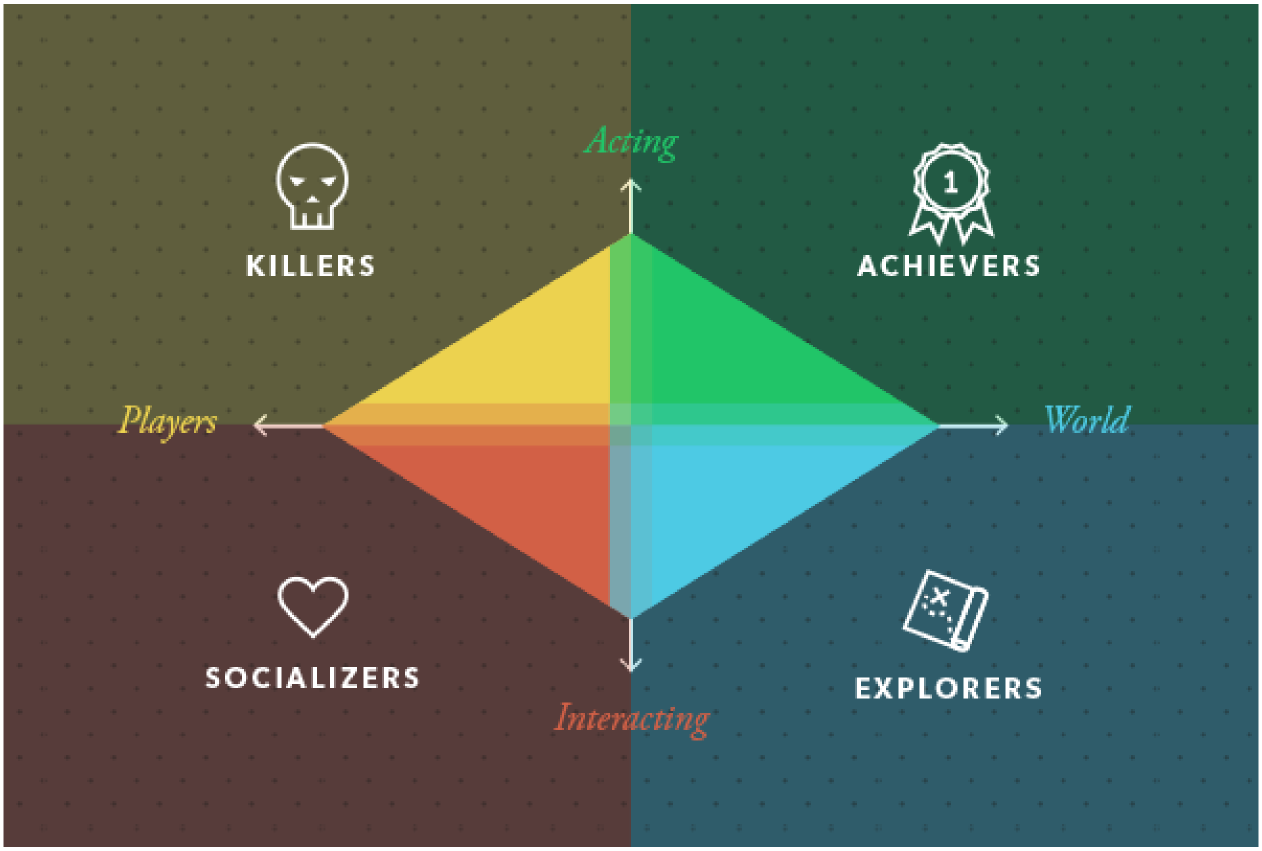


Figure 1.6: Graph of where player types lie on world/player and act/interact axes

being done via pre-written text conditionally presented to the player. Killers find no joy in killing robots, and Achievers in strategy games…well, they're particularly potent strategists, and current games have the AI cheat to provide them a challenge significant enough to master.

Yee's Motivational Model

Yee presents a related model of player motivation (Yee 2005), using factor analysis to validate and expand the model iteratively. He takes possible motivations from literature and open-ended surveys, and converts these into survey questions. Factor analysis is then performed on the data to separate statements into clusters of correlated statements. The aim was to produce motivations that had been statistically verified as being separate concepts.

Their results revealed three main components: *Achievement*, *Social*, and *Immersion*. Each component has subcomponents emphasising what each component is made up of. Table 1.2 is a reconstruction of the original table from Yee's paper (Yee 2005) that enumerates the components of his model:

These findings allow player motivations to be constructed from granular components that aren't necessarily mutually exclusive. It's also somewhat statistically grounded, as the three main components are mostly

Achievement	Social	Immersion
Advancement Progress, Power, Accumulation, Status	*Socialising* Casual Chat, Helping Others, Making Friends	*Discovery* Exploration, Lore, Finding Hidden Things
Mechanics Numbers, Optimisation, Templating, Analysis	*Relationship* Personal, Self-Disclosure, Find & Give Support	*Role-Playing* Storyline, Character History, Roles, Fantasy
Competition Challenging Others, Provocation, Domination	*Teamwork* Collaboration, Groups, Group Achievements	*Customisation* Appearances, Accessories, Style, Colour Schemes
		Escapism Relax, Escape from RL, Avoid RL Problems

Table 1.2: Yee's motivational model

uncorrelated with each other ($r = .10$).[1]

These motivations seem to be what drives players in virtual worlds. While these are useful perspectives to fulfil those players' desires, and are seemingly used across the videogame industry (Salmond 2019), strategy games are not virtual worlds, and it would be presumptuous to assume they share the same desires. Therefore, we look to other theories of motivation. There don't seem to be any for strategy games in particular, so we look towards theories that address human behaviour, and seek common ground there.

1.4.2 Self-Determination Theory and PENS

Self-**D**etermination **T**heory (Deci and Richard M Ryan 2012) attempts to explain what factors affect human motivation, whether intrinsic or extrinsic. It's been tested extensively, and puts forward that humans have several psychological needs, the fulfilment of which fosters motivation and well-being:

- **Competence** The need to feel *effective*, and to grow in mastery.
- **Relatedness** The need to connect with others, to have a sense of belonging.
- **Autonomy** The need to have *choice* in one's actions.

The satisfaction of these needs by media has been shown to directly affect the enjoyment of consumers (Tamborini et al. 2011). In games, this has been explored in more detail (A. K. Przybylski, Rigby and Richard M. Ryan 2010), and in many genres the fulfilment of these needs related highly to players' reports of fun and immersion (Scott Rigby and Richard Ryan 2007). This suggests that there are parallels between game-specific constructs and these more general needs: challenging the player appropriately bolsters players' feelings of

[1] Pearson's r is a measure of linear correlation between -1 (perfect negative correlation) and 1 (perfect correlation).

competence while giving them room to grow their skills; giving players agency (Murray 1997; Wardrip-Fruin et al. 2009) fills players' needs for *autonomy*, and social interaction with other players fulfils the need for *relatedness* (Scott Rigby and Richard Ryan 2007).

There are also similarities with the research on player motivations in virtual worlds. Deci and Ryan's *Competence* and *Relatedness* appear to cover similar semantic ground to Yee's *Achievement* and *Social*, but there are statistically significant differences when both models are measured simultaneously (Richard M. Ryan, Rigby and A. Przybylski 2006) even within virtual worlds. This suggests that while motivations of players in virtual worlds are a good starting point, we should be aware that there will almost certainly be differences when it comes to strategy games.

It's also worth noting that while SDT provides a good framework for gameplay that nourishes our psyche, it doesn't take into account *why* a particular player is playing a particular type of game. It may not be to fulfil such needs. Those playing *Cookie Clicker* might not be seeking to challenge their skills, but still wish to see the satisfaction of numbers rising due to their input. Another day, they might feel up for more of a challenge, and play *Candy Crush*. Therefore, while it suggests guidelines for wholesome gameplay, what strategy game players seek might not quite follow those guidelines.

1.5 Continuing

It's clear that strategy game AI does not currently meet all players' needs. Providing challenge is difficult as-is. Fulfilling social needs, at the time of writing, is beyond the capabilities of consumer hardware. Improving the situation may be difficult; we may not be able to fill all the gaps in such a challenging environment without significant leaps forward in computing technology.

Therefore, it makes sense to be economical: find what ways to make better use of the technology we do have to improve player experience. We can theorise ad infinitum about what needs to be improved, but those theories hold far less worth until they are tested, especially as player preferences in strategy games have not yet been explored specifically.

So, with the aim of playing to the strengths of strategy game AI, what do players like about it, that we can make it even better?

Chapter 2

Finding Out What Strategy Game Players Want

2.0.1 Reading Positive Player Comments

The first attempted course of action was to search strategy game forums for positive experiences that players had with the forums' respective games. This turned out to be difficult; posts containing such content were few and far between.

Perhaps I'd been searching with the wrong keywords for positive experiences, so I looked for experiences with AI in general: simply searching for the keyword 'AI' in order to gain a better picture of which experiences players *did* post about.

It turned out the vast majority of such posts were *complaints* about the AI.

2.0.2 There Weren't Positive Comments

Players didn't praise AI much on the forums, but they *did* damn it, with varying levels of intensity and topic. So I switched context from highlighting strengths of AI opponents to highlighting their *weaknesses*.

To explore AI weaknesses from extant player comments, qualitative research has some well-established tools: Thematic Analysis and Grounded Theory. Let's explore both of these to evaluate which would be the most appropriate.

2.1 Thematic Analysis

Thematic **A**nalysis is a method to discern and organise patterns within a dataset, and then to provide insight into those patterns (Braun and Clarke 2012). It consists of several steps.

- **Immersing oneself in the data** Whatever form your data takes, review it. Take notes as you go through it, highlight interesting areas, *examine* it and read between the lines. What's being assumed? What's left unsaid? How does the participant interpret the situation? The idea is to get *familiar* with your data, noting down insights as you go, to prepare you for the next stages.

- **Initial coding of data** Here begins *systematic* analysis. To code data, you extract the essence and concepts of what is present in portions of the data. For example, *"Toodle around doing whatever you like… until you only see a dull slog ahead of you leading to your inevitable victory"* might be coded with *Foregone conclusion takes the fun out of the game* and perhaps *Lack of difficulty is fine until you realise you've already won*. Prefer being inclusive; if you think something *might* be relevant to your research question, code it. When this step is done, each data item should be thoroughly coded. Some segments may have multiple codes, and many will likely share codes. This phase ends when your data are fully coded, and each code has had the data it's attached to collated.

- **Search for themes** Themes are "something important about the data in relation to the research question, and represents some level of *patterned* response or meaning within the data set" (Braun and Clarke 2006, p.82). To find these, examine your codes and their data: Are there any topics that several orbit around or fall under the umbrella of? These codes can be clustered into the *theme* they fall under. There may be some overlap, and sometimes themes may need to be split into multiple themes if they begin to address differing points. Again, this phase should end with collating the data relevant to each theme.

- **Review potential themes** In short, quality checking. Check your themes against the data they're connected to (as well as the big-picture view of the dataset, which is why it's so important to immerse yourself in the data first). Does it still make sense completely? If not, it may need to be tweaked: codes discarded, themes merged or split, or even discarded. If your themes manage to successfully capture the most important parts of the data when it comes to your reseach question, this phase ends. Otherwise, keep refining.

- **Define and Name Themes** Themes should each be distinct and concrete, though they may be related. Importantly, they should impact your research question. Here is where you write them out, describing each *theme's* essence — it often involves taking a portion of the data that particularly exemplifies the theme, and building the story of the theme around that anchor. Show what's interesting about that extract, the story behind it that matches the theme. With your theme defined, you must name it. It should capture the theme, but not dryly: "Players lose motivation when they know they'll win" is a lot less memorable than "When victory is inevitable, fun breathes its last".

- **Write the Report** Tell the story of your data and the themes that carry through it. Connect your themes logically in this narrative.

This process ends up with a holistic and comprehensive view of the themes underlying the data. However, we want to know not just *what* annoys players about strategy game AI , but *why*. Can the themes of those annoyances be connected in a logical, grounded way to put forward a theory on what lies at the heart of these frustrations?

That's where grounded theory comes in.

2.2 Grounded Theory

Grounded Theory (Charmaz 2006) is another qualitative research method. It too begins with data, but instead of ending with a collection of focused themes, it iterates over themes until it produces a theory that attempts to explain the themes and phenomena that arise in the data.

- **Initial Coding** This is much the same as in thematic analysis (TA).
- **Focused Coding** This is akin to reviewing potential themes in TA. The most useful initial codes are tested against more and more data in order to refine them and to elaborate them.
- **Memo-writing** After coding, and perhaps during, it is essential to write *memos*. Memos are essentially mini documents in which you record your thoughts, explore connections between concepts, and sometimes just asking questions of yourself. They're informal, from you to yourself, and can take whatever format best works for you. They also serve as tools for consistency; when coding, it's possible to refer to memos of a code to check whether the segment of data you're coding truly matches the code you're aiming to apply[1].
- **Theoretical Sampling** When current categories are reviewed and are found to be thin, nebulous or ill-defined, it is time to seek more data on them. Specifically, any data that elaborates, refines, or challenges these nascent categories. This informed data collection is where iteration comes into play; gather data, refine categories, and repeat.
- **Saturation** When data collected no longer adds new theoretical concepts, or challenges current ones, saturation is reached.

In this case, Grounded Theory was chosen over Thematic Analysis, in order to attempt creation of a theory that explains why players become frustrated with a game.

[1] This becomes especially useful when working in teams (MacQueen et al. 1998).

2.3 What Kind of Data?

In order for the data and resulting conclusions to be directly relevant to current industry titles, it's necessary to gather opinions *about* such industry titles. This also sidesteps issues of laboratory conditions being different to the conditions that players would usually be in, and preserves ecological validity. Moreover, the difficulties of creating strategy game AI have already been covered. It would take much work and testing to produce an AI for a researcher-produced game that didn't bring its own flaws to the player experience.

Where to find player opinions on strategy game titles? One option would be to interview several players about what annoys them regarding AI in strategy games. This would allow for a relatively small sample size to provide adequately rich data, but players would need to be carefully selected to cover a range of experience levels, their recollections of events may be fuzzy if it's been quite some time since they played, and continuously gathering more data (as is required in Grounded Theory (Charmaz 2006)) would be a laborious matter.

Instead, we can take advantage of pre-existing discussions of game AI between players: community forums. By using discussion threads on AI in these forums, we can gather data from a greater range of players, covering topics that perhaps might not have come to mind through interviews.

2.4 Which Games? Which Communities?

The games to be chosen, and the extant player discussions attached to them, have some prerequisites:

- **Reflect the state of the art.** [1] To accurately gather players' opinions of the state of the art, the games shouldn't be old. It would be both unfair and inaccurate to claim X or Y about player issues if those issues were from the original *Heroes of Might and Magic* (*Heroes of Might and Magic: A Strategic Quest* 1995). It would also be unfair to do so with indie games that might not have had the resources to develop their game's AI significantly.[2]

- **Are strategy games.** As it is strategy game AI we're investigating, it'd be out of scope to investigate other genres. That's not to say that such investigation wouldn't be fruitful or useful; it's just not the author's focus at this time.

- **Are turn-based.** This study aims to find where AI has been pushed to its limits and players have noticed. Real-time strategy games offer frantic moment-to-moment gameplay that may distract the player from

[1] As of this study's writing, 2020.

[2] Indie games are often the bleeding edge of game design. However, their players may not have plumbed the mechanical depths of their design space yet, and therefore might tolerate an AI's mistakes more than veterans of an established genre would. Indie games also generally have smaller (albeit enthusiastic) communities.

any objectionable behaviour presented by the AI.

- **Have significant English-speaking communities.** If we are to investigate extant player opinions, the communities and discussions they contain must be in English, as the author is not fluent in any other languages, as well as rich and diverse. This somewhat conflicts with 'reflect the state of the art', as the *newest* games will likely have smaller communities immediately after release.

- **Are not the same game.** By choosing multiple games within the strategy genre, we can avoid idiosyncracies particular to one specific game — for example, Gandhi's propensity for nuclear warfare in *Sid Meier's Civilization VI* (*Sid Meier's Civilization VI* 2016).

Games That Didn't Make It There are several games that fit the above criteria; some were considered and eliminated from consideration for other, more individual reasons.

- ***Starcraft II* 2010** Huge community, became free-to-play in 2017. However, it was felt that the real-time nature would not give players as much time to dwell on opponent flaws as turn-based would; think blitz chess vs a regular game of chess.

- ***MicroRTS* 2023** Is also real-time. Its interface is also, frankly, headache-inducing, which may make players rather disinclined to play; it's better for testing reinforcement agents than playing as a human.

- ***Stellaris* 2016** Ticks every condition above, as well as being configurably slow; this makes its real-time play less of an issue. Unfortunately, searching the *Stellaris* community for threads mentioning 'AI' doesn't work. There is a whole branch of technology related to in-fiction AI. In-fiction AI propels one of the game's crises. It's therefore *much* more difficult to harvest player opinions for.

- ***Sid Meier's Civilization V* 2010** While it's had more time for community development than *Sid Meier's Civilization VI*, the latter still has a large community (with 43,558 average concurrent players in April 2020 (*Sid Meier's Civilization VI - Steam Charts* 2023), compared to *V*'s 29,266 (*Sid Meier's Civilization V - Steam Charts* 2023)), and better reflects the state of the art.

A couple candidates remained, however: *Sid Meier's Civilization VI* (*Sid Meier's Civilization VI* 2016) (henceforth *Civ VI*) and *Endless Space 2* (*Endless Space 2* 2017) (henceforth *ES2*). These were used for the study. At the time of the study (2020), they were relatively recent, but both have large, rich communities around them: their Steam forums (*Steam Community :: ENDLESS™ Space 2* 2023; *Steam Community :: Sid Meier's Civilization VI* 2023) had over 19.8k and 10.4k discussion threads respectively, with any given thread typically made up of a handful to several dozen comments. They are both 4X games (*What Is a 4X Game?*

2021) in which players control an empire and "eXplore, eXpand, eXploit and eXterminate": revealing hidden areas; expanding territory into those areas to secure resources; exploiting those resources and researching new technologies; and destroying other empires, typically through deployment of military units. They both allow for multiple types of victory, such as military, science or diplomacy.

With all these similarities in mind, they are certainly not the same game; they merely fall within the same genre. *Civ VI* is set on a planet's surface, while *ES2* operates at a galactic scale. *ES2*'s factions are asymmetrical, some incorporating entire game mechanics unavailable to other factions (*Factions* 2023), where *Civ VI*'s factions are largely similar with key changes to give individual nations thematic and mechanical flavour.

Figure 2.1: A screenshot of *Civ VI*

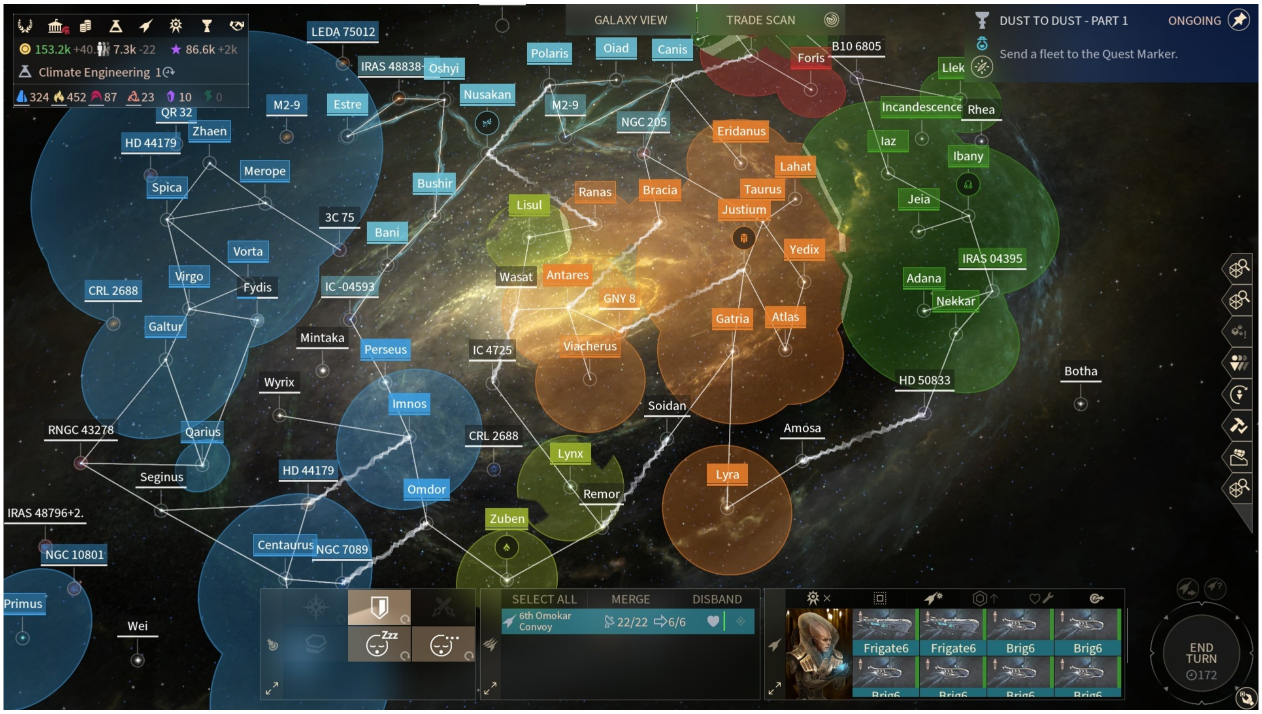


Figure 2.2: A screenshot of *ES2*

2.5 Methods

2.5.1 Overview

Comments on *ES2*'s AI were gathered from both its Steam Community forum (*Steam Community :: ENDLESS™ Space 2* 2023) and the */r/endlessspace* subreddit (*The Subreddit for Endless Space.* 2019).

These comments were iterated over several times using grounded theory (Charmaz 2006) , until coherent themes emerged on what players do not enjoy about opponents they play against. These themes were then tested for generalisability by iterating over comments from *Civ VI*'s Steam Community forum; for concepts to be general to strategy games as a whole, they should be able to apply to both games. Discrepancies were resolved, and further iterations of analysis were undertaken on an ever-growing set of comments until a clear, logical theoretical framework was reached. This theory describes what pitfalls opponents should avoid, applying to both games without needing to make concessions for either.

2.5.2 Data Gathering

Player forums are ripe sources of discussion between players on everything from what factions are the most overpowered, to memes, and how players think the game should be improved or what frustrates them about its current state. It is this last category we are interested in.

Each discussion on a forum consists of an initial comment (post) by a single user, to which other users may reply, beginning threads of users replying to the initial post and to each other's comments.

The Steam Community forums for both *ES2* and *Civ VI* are the largest forums that exist for these two games , and provide sizeable samples of playerbase opinions with >9,800 (*Endless Space 2 General Discussions* 2019) and >20,000 posts (*Sid Meier's Civilization VI Steam Community* 2020) respectively.

However, *Civ VI*'s Steam Community forum has almost double the amount of submissions as *ES2*'s. So, in order to compensate for this somewhat, further comments were also gathered from the */r/endlessspace* subreddit (*The Subreddit for Endless Space.* 2019). This also has the benefit of alleviating any issues from community bias that may arise from having a smaller community.

With communities selected, it would be possible to solicit player opinions directly by beginning threads asking players for their thoughts on AI opponents they face in-game. However, this suffers from several issues:

- The initial question would need to be worded carefully to avoid bias in players' responses. If it *were* biased, it'd be very easy to miss out on whole areas of what annoys players.
- The thread might only gather a small number of responses.
- The wording of the question might bias which players respond to the thread, resulting in a biased sample.

Some of the above could be mitigated if the question were asked carefully enough, but instead forums were trawled for *pre-existing* threads expressing implicit/explicit opinions about the game's AI. Individually, those same threads might share the above issues, but a collection of such threads increases numbers of players that might respond, and alleviates bias through several questions and differing sets of respondents.

The Comment-Gathering Process

In each forum searched, threads start with an initial text, and other users may respond to this text with their own comments. On the reddit forums, users may also comment on other comments. For the purposes of data collection, the initial thread text, comments in response to the thread, and comments in response to other comments were all considered as 'comments'.

1. Search for threads containing the keyword 'AI' in the relevant forum. For *Civ VI*, *Steam Community :: Sid Meier's Civilization VI* 2023. For *ES2*, *The Subreddit for Endless Space.* 2019 and *Steam Community :: ENDLESS™ Space 2* 2023.

2. Read through each thread in the search results, evaluating whether the thread's comments were expressing opinions about the AI. This was quite permissive; there was no filter on what *kind* of opinion the players were expressing. The initial text of the thread is considered as a comment here.

3. If a thread's comment did not express opinion on the game's AI, the comment was discarded.

4. Comments such as 'same' or 'agreed' within the remaining threads were discarded to avoid duplication.

The goal was to find roughly 100 comments per game: an arbitrary starting number, with the aim that this would be an initial batch of comments, and further comments would be sought if these were not enough to reach saturation.

The process resulted in 219 comments from 37 separate threads. 730 comments were discarded.

- *Endless Space 2*
  - 76 Reddit comments
  - 32 Steam Community comments
- *Sid Meier's Civilization VI*
  - 111 Steam Community comments

Analysis of Comments

Grounded Theory was applied to an ever-expanding subset of the comments. First only the *ES2* comments were examined, in order to form a cohesive picture, and then the *Civ VI* comments were incorporated, to challenge previously-found ideas.

After several iterations, a cogent theory emerged.

2.5.3 Analysis

Grounded theory was applied to a continually expanding subset of the gathered comments, starting with *ES2* in order to find categories and codes that applied to just that game, though effort was made to make these generic. Eventually, the comments from *Civ VI* were also filtered in. After several iterations of the process, a cogent theory was reached.

2.6 Results

Following the data is a series of expectations that players seem to have of the AI in both *Endless Space 2* and *Sid Meier's Civilization VI*. These expectations have been interpreted in a general manner, as they might reasonably be applicable beyond the two games investigated in this study. Each expectation, and the nuances used to refine them, were consistently represented in the data. Once the realisation was made that these were intrinsic expectations that AI opponents failed, the rest of the theory followed.

Setting of Expectations The first and key element is that by playing in a specific role — specifically, that in which a human would usually play, another empire in the 4X game — there are implicit expectations that players have of the AI. What follows are the expectations that the AI seems to *fail*. [1]

- **Competition Over Victory** The AI should compete with the player over victory, in a manner which provides a sense of *tension*. This tension is discussed in greater detail in subsection 2.6.2.
- **Fair Play** The AI should play by the same rules that the player themselves is playing by. After all, they're sitting in the same proverbial seat.
- **Explainability, Transparency and Closure** The actions of a given player player or AI should be explainable; that player should be able to communicate with others about *why* they took certain actions.
- **The Deaf Student** If the actions of a player or AI are suboptimal or misguided, other players should be able to educate them, or at least try persuading them to try a different course of action.

2.6.1 The Setting of Expectations

The theory starts with the idea that games set expectations for the entities and players within them. Each game is a context of its own, a set of rules outside that of the real world, a 'magic circle' (Johan Huizinga 2008) that players enter when playing. It's perfectly fine to punch people in the face in *Mike Tyson's Punch-Out* (*Punch-Out!!* 1990), but not in the real world. Dragons exist in *Skyrim* (Bethesda Game Studios 2011), but not on Earth. Games just have their own rules.

Each game gets those rules across to the player, either explicitly (through included manuals, tutorials, etc.), or implicitly (that griffon has *wings*, so it should be able to fly over that lake); most games go to great effort to facilitate this.[2] This informs the player of the actions they have available to them, and what kind of meaningful effect those actions would have: the player's agency (Cole and Gillies 2019). However, games can imply rules or interactions they do not in fact have, and experiencing the disconnect between what is *implied* and what *is* can be jarring (Johnson 2015).

[1] There might be other expectations that the AI passes, but players weren't talking about these

[2] There are exceptions such as *Mao* (*Mao* 2023), *I See The Moon In This Spoon* (*Quieter/Sit down Games* 2012) , and *Nethack* (*Nethack* 1987). But these games make it an intrinsic part of their design that rules are discovered and not necessarily told.

A good example of this is invincible hedges in some action games, such as *Grand Theft Auto IV* 2008: You're driving at shocking speeds, smashing through bus stops and other cars with reckless abandon and no consequences. So you aim to take a shortcut through a garden, and- its small hedge stops you dead without budging a pixel, your car left a smoking ruin. The game has implied that you'd be able to smash through that hedge. After all, your car has done it to much sturdier-looking objects with nary a scratch. But this was not the case. You thought you were playing under one set of rules, and so played that way. The game betrayed those expectations. [1]

In a similar manner, there are expectations of the AI in games, and those differ with the role they take on. They can be a faceless goon, a kindly NPC, or a chess grandmaster. In strategy games, such as *ES2* and *Civ VI*, they play as a peer, another empire competing with yours. As the players are so intimately familiar with this role (they are also playing this role), they know what they'd do. They know what can and can't be done. In short, their expectations are many, and the AI isn't always up to the task. This seems to remain frustrating even when players know they're facing AI instead of humans.

In analysis, the implied expectations that arise from facing a human opponent in strategy games are:

- They'll play by the same rules as the player. After all, they're in the same role, in the same game.
- There will be a logical explanation of the actions taken by the opponent, even if its reasoning is flawed.
- They can be communicated with if others, for whatever reason, desire their behaviour to change. For example, if they're acting erratically or causing another player grief, they may be asked to play differently and their behaviour change as a result.
- They will compete over victory. Not everyone can win, and therefore there's struggle and competition.

Strategy game veterans may already raise their eyebrows here; they *know* that the AI doesn't fulfil these from experience, and this might lessen the blow somewhat. But even though they might lower their expectations to compensate, it still affects their experience of the game, in a similar manner that one might push harder to cut with a dull knife. It still cuts the tomato in the end, but the whole experience would benefit if that knife had been sharp in the first place.

2.6.2 Competition Over Victory

The first of the expectations found is that other players (opponents, including AI) will compete over victory. However this isn't as simple as just placing several players in a game with a single possible victor; this doesn't

[1] The first time this happens, the disconnect between a player's expectations and the game's reality may cause outrage, especially if this results in consequences for the player. Afterwards, the player is likely to remodel the behaviour of hedges in their mental model of the game, and simply not crash into hedges.

necessarily elicit true *competition*. For example, imagine if a skilled *Chess* player were to not hold back against someone new to the game. There's no question as to who will win, the outcome is a foregone conclusion; players clearly do not enjoy this from either side of the equation, leaving comments in the sampled forums such as:

> "*Toodle around doing whatever you like…until you only see a dull slog ahead of you leading to your inevitable victory*"

> "*..got to a point where most of us end up: at some point in the game you just know you win and it's just a matter of skipping turns till the pop-up of the winning screen turns up*"

> "*Serious[A difficulty level] is the new Endless, and Impossible truly feels impossible, at least for me: barring some fluke, there is simply no set of moves one can take to keep up with the insane rate of expansion by the AI*"

On the other hand, a game in which the players are evenly-matched is *much* more exciting to be a part of. The feeling of being on the edge of your proverbial seat, not knowing *who will be victorious* is an ambrosia that many players seek — including those strategy game players sampled in this study.

> "*…I try to aim for gameplay which encourages me to constantly aim for winning, instead of reaching a point where I just skip turns…*"

> "*I'm always playing on endless, and the AI got very good in the recent times, especially craver are now more aggressive and harder to beat than usual. I almost lost against a endless craver AI but I could beat them in the end.*"

We will call this sensation of uncertain excitement when facing an opponent *tension*.

Tension

We define tension as *excitement derived from a player being invested in an event, and due to another entity contesting their efforts to bring the event about, being uncertain if that event will occur.*

It is made up of several components:

- **Investment** The player's desire for a specific future event to occur. This must be non-zero.
    - The player may have invested resources towards the event's outcome, or time, or may simply wish for the event to occur without having spent any resources at all. The important part is that the outcome *matters* to the player.

- Players, might not even pay attention to an event they don't care about. Why feel tense about what a tiny empire in the other corner of the map is doing, when it has no impact on you?

- **Uncertainty** The player must be uncertain whether the event will occur.

- **Perceived Power** The player must believe they have some power over whether this event will occur.

- **Perceived Opponent Power** The player must believe that contesting entities (whether another human player or AI) also have power over whether the event will occur, and might use it.

In short, the player *wants* an outcome, but isn't sure if it will occur, inciting excitement. This shares similarities with suspense (Moulard, M. Kroff et al. 2019), which arises from the difference between a person's *hope* (assessment that a pleasurable event will occur), and *fear* (assessment that a negative event will occur). Both hope and fear imply and contain uncertainty. Suspense also brings investment into the fold, and to quote Carroll 1984, p. 76: "suspense only takes charge when we care about those future outcomes about which we are uncertain". However, immersion in games takes on a different shade than in film, with players seeing their character as an extension of themselves (Bartle 2001). They also have a level of control over what occurs. While player *effort* is explicitly addressed as increasing the strength of suspense, players' agency — "the satisfying power to take meaningful action and see the results of our decisions and choices" (Murray 1997) — does not seem to be present in the model. Agency might simply fall under the umbrella of hope, fear, effort, or uncertainty, but it's also possible that it might affect the resulting emotion in different ways.

In the field of games, the closest construct is *challenge*, presenting players with an objective, and providing barriers to achieving it (Denisova, Guckelsberger and Zendle 2017). There are several types of challenge: performative, cognitive, emotional and decision-making (Denisova, Cairns et al. 2020), though cognitive challenge and decision-making challenge are the most relevant for strategy games.

- **Cognitive challenge** Arises from a need to plan, memorise, prepare and multi-task.

- **Decision-making** Arises from having to make difficult or possibly-regretful choices.

In contrast to suspense, challenge explicitly acknowledges the player's agency. It's also strongly rooted in games, and its practicality echoes through both academia (Denisova, Cairns et al. 2020; Zohaib 2018) and industry (Adams 2014; Costikyan 2013). The study of challenge has also recently explored challenging player emotions (Bopp, Opwis and Mekler 2018; Cole, Cairns and Gillies 2015), bringing with it how players can *connect* to what they're challenged by. Nonetheless, one would not necessarily have to increase challenge to affect player experience through AI. Imagine a player fighting alongside an ally with a generic appearance and no speech. The player's experience would change if this ally is endowed with narrative character, a level of social responses, and attendant art (Nass and Moon 2000; von der Pütten et al. 2010).

This isn't to suggest that suspense or challenge are inferior to each other or to tension. They each cover slightly different ground, though are highly related. In the case of this grounded theory, tension is the construct that arose from the data. It allows us to investigate why something matters to players, the uncertainty of outcome, and players' perspectives on both their own agency and that of others. It's a relatively holistic lens on antagonistic gameplay.

Opponent Action

In addition to the player being able to work towards the event's occurrence, their opponent can work to make it *not* happen. This results in a proverbial tug-of-war, which ends up being fairly enjoyable. If the opponent pulls too hard, the player may no longer feel like they have power over the event, or that it's inevitable. Tension ceases.

> "*Serious is the new Endless, and Impossible truly feels impossible, at least for me: barring some fluke, there is simply no set of moves one can take to keep up with the insane rate of expansion by the AI*"

> "*I'm getting out-settled by the AI like crazy... I get messages on turn 36 that an AI has colonized 8 systems, or by turn 50 one of them has 20? It just seems nuts*"

> "*I'll be tooling around building up and my first interaction is a curbstomp with my face playing the role of home plate*"

If the opponent pulls too lightly, the player might believe the opponent has no real power over the outcome, or simply become certain of the outcome. Tension ceases.

> "*AI is addicted to small ships it's annoying, one of my fleets can wipe all of theirs with no losses... it's ridiculous*"

> "*Just moved my units right underneath and by the fighters and they still did nothing. At this point it's fair to say that the AI of this game is to some extent plain broken.*"

Some might suggest dynamic difficulty adjustment (DDA) (Hunicke 2005) to avoid both of these possibilities. By adjusting the challenge that opponents present in real-time, you can avoid challenging the player too much or too little, keeping them in a flow state (Csikszentmihalyi 1990). However, by the definitions of tension, DDA fails if the player *notices* that DDA is in effect (Zohaib 2018). Now the players are not uncertain about the outcome should they put less effort in. They stop believing their opponent has power over the outcome, because that power is always relative to the effort players put in to challenge the outcome. A player can exploit this by simply putting less effort in. Tension ceases, barring constant suspension of disbelief. An alternative is more useful: actively fluctuating difference between challenge and skill (Baumann, Lürig and Engeser 2016).

However, there is something a great deal worse than when the opponents do not pull adequately in this proverbial tug-of-war: when the *player's* agency is impugned; when they are prevented from pulling upon the proverbial rope, even by other members of their 'team'.

> "*The moment i enter an alliance with an AI i instantly regret it. I joined alliance with unfallen since they had largest score. and immidiately i ended up declaring war on horatio which in turn formed an alliance with lumeris whom i had peace trade and science agreement with. And lumeris says i declared war on them and betrayed the peace treaty... Like what the hell?*"

> "*Having AI decisions set in stone and with the force of immutable law is very frustrating, especially since the types of AI players that will form a federation in the first place tend to be galactic peacenik bureaucrats. If you like being locked into a box by the AI ("AI forcing you to not do something is somethhing deeply required"), then you always have Stellaris available to you as an option; please do not try to replicate that problem here. Or there's always Settlers of Catan.* "

In these circumstances, players entered an alliance, and were then forced by other members of the alliance to commit to actions the player never intended on doing. And there was no method for the player to deny this, no ability to change the course of action; agency was removed from them unwillingly[1]. Players absolutely *hated* this.

Another denial of agency occurs when players realize significant and exploitable flaws in the AI's behaviour. In these cases, not only is the AI not providing sufficient challenge, but the players actively hamstring themselves to even the playing field, reducing their own agency again. Whether they do so or not, tension suffers, either because the AI is no longer providing challenge and therefore uncertainty, or because their choices have been limited.

> "*The AI is in ways plain broken. You end up having to play in sort of artificial conditions to get a challenge out of a behaviorally severely challenged AI (the military portion of its virtual brain makes lesions look good)*"

[1] It was not the case that options to challenge decisions were available but hidden to the player; those decisions simply came into effect as soon as they were made

> "*I am looking at five AI with big empires with an army score of 200 or below. Mine: 3,000 and change. Can you please fix this? I am just stopping myself from walking all over the AI, but it is beyond silly how they are not even close to a threat.*"

> "*Last point: I won diplomatic victory only by having enough disaster emergencies to win 3 points during Congress sessions, i would never have won by regular votes. Because the AI can not handle those well and gifting 300 gold usually is enough to win it.*"

> "*Honestly, the easiest exploit in the game is how stupid the AI is. Find a neighbor who is relatively easy in the ancient era. Then go snipe one of their workers. Pillage everything and take their cities. taking cities is cheaper than building settlers and you get to take out an enemy at the same time*"

In other genres, exploitable behaviour *isn't* bad, and is often a central pillar of an NPC's design. For example, boss NPCs start out incredibly difficult for many players... until players start to piece together the patterns in the boss' behaviour, and then exploit those patterns. Those patterns don't tend to change during the fight, allowing players to eventually defeat the boss, often consistently. However, the boss still remains a challenge in some ways; the exploits required to defeat them are not trivial to execute. The outcome retains a degree of uncertainty, power remains on both sides, and tension is sustained.

In contrast, strategy game opponents displaying exploitable behaviour is often catastrophic. As these are not often explicitly-designed vulnerabilities, they can be much easier to perform, and reduce perceived opponent's power significantly. When such an exploit means players become certain of the outcome of pitting themselves against an AI opponent, tension once again ceases.

2.6.3 Fair Play: A Level Playing Field

A second expectation that players hold of their opponents is that if both player and opponent are fulfilling the same role in a game, then they will be playing by the same rules. In the case of strategy games, both the humans and AI share the same role: that of ruling an empire and controlling its actions. The player could feasibly load the save and play as that other empire if the game allowed it. Such a level playing field is *fair*. It does not often exist.

Many strategy game AIs play by somewhat *different* rules they play by, and often have augmentations to their incomes or abilities that a player would not have in their place (*Difficulty Level (Civ6)* 2023). They are given these extra capabilities in order to challenge the player sufficiently when they don't play as optimally.

This is not a level playing field, and players bemoan this fact, even going so far as to say that with the exact same outcome, they'd still prefer AI that didn't cheat.

> "*I can't tell if the greater difficulty is due to the AI being smarter or given more advantages. If it is the former, then I am quite impressed by what the developers have done*"

> "*The bonuses the AI get are insane, mainly the production boost and free units at the start. That it's possible for them to have 800 military strength by the late Classical, early Medieval era, is just a tiiiiiiny bit unfair, from my perspective. They also just seem to speed so quickly through the tech tree that I'm left behind...*"

It isn't just outright resource-boosting that players consider unfair. Sometimes, players will start a game surrounded by enemies, yet those enemies have room to expand outwards. Sometimes, players will be given an extreme disadvantage in trades, to the point that they don't see any trading with the AI as being worthwhile. It's akin to playing tug-of-war and realising the opponent has an anchor tied to their waist.

> "*The trade of almost anything (edit: well, many things, and it depends on the situation, having thought about it a bit more) in this game is hilariously stacked against the player. In order to trade for a tech worth 7k science from AN ALLY, I have to offer 70k science worth of my own in order for them to even consider it! 10x as much Science value!*"

> "*My biggest gripe about the AI, is mainly it's unwillingness to trade with you. Sometimes getting even a small amount of a strategic resource is next to impossible even if the AI has more then enough to spare*"

It may well be that alliances with the AI also trigger this: when players have no ability to make a decision on certain fronts, or to veto, this too is a large inequality between player and AI. The AI gets all the say; the player gets none.

2.6.4 Explainability, Transparency, and Closure

When you see a human player do something confusing, you can ask them why they did it, and might get a logical answer. Imagine the following hypothetical conversation:

Steve...we had an agreement that you wouldn't murder my settler!

Yeah, well, that was before I learned you had 2 spies in my capital.

That's not fair and we both know it!

On the other hand, when an AI does something questionable, there's no explanation given at all. Why did they suddenly cease trading coal to you? Why did they declare war on someone and drag you into that war when there's almost no way you could contribute?

There are whole forum threads reasoning about why or how an AI did something in a given game, and sometimes these threads reach no resolution. When players can't conclude *why* an AI did something, its ability to play the game is often questioned, and players cannot get closure, even though closure plays an important role in knowledge acquisition (Kruglanski and Fishman 2009). When players can't conclude *how* an AI did something, players will often accuse the AI of cheating, even though (sometimes) the AI did no such thing. This is addressed in the FPS genre by killcams: replays of your own death from your killer's perspective, showing you how they did it, and sometimes in fact revealing that they *were* cheating. However, this isn't possible for many strategy games for several reasons: information hidden by the fog-of-war remains very useful for quite some time and should not therefore be revealed; there is no single 'death' state where it is easy to assume the player wants to see what happened; such replays would have to go back in time quite a while, showing detailed information about the enemy's resources over time. These are issues that arise merely showing *how* the AI did something. Showing *why* an AI did something is another matter entirely; with the hybrid Frankensteinian structure of many strategy game AIs (*Civilization V Preview - PC Preview at IGN* 2010) it can be incredibly difficult to make concrete conclusions about why an AI took a particular course of action, let alone explain it satisfactorily to the player in-game.

Let's compare the situation with an instance of AI outside of strategy videogames, a best-case scenario in which the audience *could* get closure on why an AI did something they didn't understand: the games of Go played between Lee Sedol and Google's *AlphaGo* in 2016 (Silver et al. 2016; *The Challenge Match* 2023). Unlike strategy videogames, it was impossible for either Lee Sedol or the AI to cheat; the games were under great scrutiny, both players had perfect information, and there was an abundance of world-class players present to comment on play. Like many strategy game AIs, *AlphaGo* often made moves that were incredibly unconventional. However, many of these proved themselves, either through their effects showing later in the game, or because experts dissected the possible advantages behind them. *AlphaGo* also showed projected winrates for the moves it could make (that is, if it makes move X, it has $Y\%$ chance of winning the game). Experts could then compare moves' scores to see which were favoured at any given time, using this as a foundation for their reasoning.

All of this combined meant that cryptic play could be reasoned about, and the AI was effective enough that even moves that remained mysterious left little room to be criticised. Strategy games, as a whole, don't have these advantages. They often take much longer, with more players per game, hidden information, and multiple actions per player per turn. Comparatively, it's *much* harder to see the long-term effects of specific moves, or reasona bout why those actions might have been done. Sessions of the game do not have a large pool of experts to comment upon and dissect possible reasoning behind actions. Players can't see examples of play that have been explained by experts, in order to build up trust. So much as there are examples of AI with explainable actions, in commercial strategy games such explainability is rare.[1]

The kind of closure that players are looking for — a definite conclusion about the reasoning behind actions — is *cognitive closure* (Kruglanski and Fishman 2009). It's a sore enough point that players raise it directly, especially when it impacts their own play:

> "*AI voting behavior: AI often votes incomprehensibly. Which does not mean they vote 'wrong', but it is unclear why they vote this way. More transparency is needed on that!*"

Such reasoning is practically demanded by players when the AI does something outside of what players see as the sole understandable course of action, or when their behaviour is far outside the norm that would be expected. For example, when the AI makes demands as if it has complete and total leverage over the player, when it in fact has none at all. Players see this as having absolutely no chance of success; why are they even trying?

> "*I was experiencing this earlier, with the Cravers asking me (Sophons) for high military techs while I was absolutely sh[***]ing all over them. It was like, yeah, no, I'm not gonna accept that. It doesn't make any sense at all that they make demands from you when the truce is something THEY need.*"

> "*Happens to me all the time. I'm at war, invading systems & steamrolling over every ship they send me way, then in truce negotiations they want all [my] dust + strategic resources + technologies.*"

Such incredibly foolish decisions are a lot more explainable by humans. Sure, they're demanding absolutely everything you have. When you ask them why, hands thrown up in the air, they could say "'cos I'll never surrender to *billions of clones of the same man*[2]". Mechanically, they're still doing the same thing: giving you such terrible trades that they may as well be refusing. But in this case it's turned from a source of frustration to some enjoyable fiction.

Even if human players don't spin such tales or make their motives clear, other players can at least trust that they have a mysterious but ultimately sensible agenda. Such trust is not commonly placed in AI, and so without being able to convey rationale, incomprehensible moves remain incomprehensible and a source of frustration.

[1]Note that players do not have to be *right* about the AI's reasoning. The explanation just needs to make enough sense to them to not yearn for further closure.

[2]This is in fact a race in *ES2*: The Horatio

2.6.5 The Deaf Student

A great many of the complaints in the data, whatever specific issue they raise, follow a common pattern: The AI does this wrong. It does X It should do Y instead.

> "*It's getting sad now, but funny and helpful on a high difficulty. One, I'm invading a civ that has three fighter planes in striking distance of my units and I don't have any anti-air support either. If those fighters were to get involved I might just be stopped.*"

> "*3. AI is bad at military tech. I have play on Endless difficulty for a couple games now, and I rarely seen the AI using strategic resource weapons. meaning the titanium missiles, the Hyperium lasers, or anti matter beams. last night I finished a game with the Cravers on Endless difficulty, and at turn 137, the Horatios are still using slugs. sure, if you like slugs that's fine, but at least use better slugs*"

This sounds quite simple; it's the general shape of any given grievance, but it hides something that may be improved: The player is so divorced from being able to change the AI's behaviour that they must vent on forums about it, in the hope that the developer sees the complaint and deigns to make the requested changes. This is often the community's method of encouraging change in games, and can be a reasonable (if slow) method of interaction with the developers. However, effectively making such changes to an AI is particularly difficult for developers; AI-related issues are much harder to debug than most issues in game development (John, Gow and Cairns 2019). AI issues are often situational, difficult to reproduce, and a fix of one AI bug could introduce others that aren't readily apparent. This makes for a multi-headed hydra of strange or unwanted behaviour — cut off one head, and multiple grow back.

In all, it's clear that being able to iterate faster on AI behaviour would be vastly beneficial for both players and developers. If this were made easy enough, players might be able to make changes to AI behaviour at a local level, in games they themselves play in. Such ability to sand off the rough edges may go a long way to relieving the burdens of frustration for the players and of time for developers. Taken even further, the idea could develop into somewhat crowdsourced behaviour modifications to the AI, though that lies in the realm of 'may not in fact be sane'.

Such ability to modify the AI may be very useful, but it's certainly not easy or practical to implement in many games. Many types of AI don't facilitate tuning of their behaviour at all. Those that do would require intuitive methods for players to modify behaviour, though the surfacing of large language models (LLMs) (*ChatGPT* 2022) may have brought this closer to reality.

What *is* practical to implement, and AI-agnostic, is relaying robust and well-designed methods of providing game data and bug reports to the developer, such as in *Songs of Conquest* 2022 or *Against the Storm* 2022.

An easy-to-reach method of contacting developers with rich data about the current in-game situation allows for that feedback loop to tighten. Moreso, it allows players to feel like the devs are listening to them (they made this for us to send reports in!) as well as making it more likely that such issues will be resolved.

2.6.6 Next Steps

This grounded theory presents a useful lens through which to view what players need from an opponent: instead of being more human, does the AI align with what the player expects of them?[1] By addressing those individual expectations, we can better tailor the AI to what players actually seek. By manipulating those expectations subtly, or overtly through explicit communication with the player, designers can be granted more freedom when it comes to AI design, without detracting from the players' experience. Such freedom could allow simpler AI to remain palatable for players, or for AI that plays by different rules without angering players[2]. This could be explored further by testing player experiences with and without some expectations being explicitly set, or by looking at games where default genre expectations are clearly subverted, such as *AI War: Fleet Command* 2009.

While awareness of these expectations is beneficial, there is yet room to explore them further: how can each be adjusted? Is there such a thing as too much of the above?

The first of these expectations to be explored is competition over victory — the creation and maintenance of tension between a human player and their (possibly AI) opponents.

[1] This might not even have to have anything to do with the AI appearing human.
[2] Tell them how it plays differently!

Chapter 3

Exploring Tension

At this point, we have a grounded theory on what provides players with a sense of thrill when competing against opponents: tension, the excitement arising from facing unknown outcomes when playing against an adversary. That grounded theory lays out the components of tension, as follows:

1. Players hold an investment in a particular future outcome.
2. Players hold a belief of how much power they hold over this outcome occurring.
3. Players hold a belief of how much power their opponents hold over this outcome occurring.
4. Based on points 2 and 3, players are uncertain about the likelihood that the outcome will occur.
5. This uncertainty, combined with their investment in the outcome, forms tension.

As with any theory, it should be battle-tested, in order to possibly invalidate its existence in real-world conditions, and to find concrete methods through which to manipulate tension. To do so, we present hypotheses mirroring the relationships between tension's components:

1. Less certainty will result in more tension, regardless of other factors.
2. More perceived player power compared to perceived opponents' power will result in more certainty.
3. Increasing perceived player power will result in decreasing perceived opponents' power.
4. More investment will result in more tension, regardless of other factors.

We then construct a controlled environment in which multiple objectives can be provided to the player, with different objectives individually tweaking certainty of outcome, player power levels, and opponent power levels.

We also use this as an opportunity to find in-game events and states that contribute to players' feelings of tension. An analysis of responses then grants insight into what mechanics may be used to adjust tension in games. To ensure ecological validity, we use a real game as the controlled environment for this.

3.1 *Battle For Wesnoth*

Battle For Wesnoth (*The Battle for Wesnoth* 2003) is an open-source strategy game. Version 1 was originally released in 2003, and Version 14 (the version used for this study) in 2018. It is free, rated 93% on Steam at the time of study, and impressively moddable. Scenarios may be written in a combination of the scripting language *Lua* (Ierusalimschy, De Figueiredo and Filho 1996) and *Wesnoth Markup Language* (The Battle for Wesnoth Project 2021) (WML), a declarative language designed to ease scenario-creation. *WML* allows for easy construction of standard scenario elements, and use of *Lua* makes more complex and custom behaviour possible when needed. The combination is demonstrably flexible: the game's core campaigns are written in this way. Any community-created content may be uploaded to the *Battle for Wesnoth* addon server, allowing any player to easily install the content from within *Battle for Wesnoth* itself.

The game also has a sizeable community over 100,000 users strong in its forum (*The Battle for Wesnoth Forums - Forums* 2021), as well as other communities in its subreddit and Steam forums.

There were other games shortlisted to create the controlled environment needed to tweak investment and perceived power, but they had their own drawbacks:

- ***Civilization VI*** (*Sid Meier's Civilization VI* 2016) Extensive modding capabilities (*Modding (Civ6)* 2021) and a large playerbase to draw participants from (*Steam* 2021). However, games take hours to play at minimum.

- ***Starcraft II*** (*Starcraft II* 2010) Intensely focused on by the AI research community with short playtimes, an open state observation API (*Blizzard/S2client-Api* 2017), and modding documentation (*StarCraft II Editor Guides - Blizzard Starcraft II Tutorials* 2021). However, it's real-time; turn-based gameplay gives far more time for newer players to mull over events that have occurred. As we wish to explore player experiences and thoughts on specific AI behaviour, turn-based is more desirable.

- ***Endless Space 2*** (*Endless Space 2* 2017) Large community, but games take a long time and modding capabilities aren't strongly documented.

- ***MicroRTS*** (Villar 2021) Heavily used in AI research, and open-source, but it's not intuitive at all to play.

- ***Hexboard*** (Walton-Rivers, Williams and Bartle 2018) A game engine that would allow the relatively easy creation of a game that suits this study's purposes. However, said game (and its AI) would have to be designed, written, and rigorously tested.

With the above in mind, it was deemed that *Battle for Wesnoth* was the most suitable game for this study in particular.

Mechanics In a *Battle for Wesnoth* scenario, players take turns commanding units (military troops) in order to achieve objectives set out at the beginning of the scenario. These units can move and attack each other, with time-of-day, terrain, and unit stats affecting how effective each unit is at moving, attacking, and defending. Those units must be funded; to recruit and keep them in play, players need gold. Gold is supplied by villages, so capturing them gives a steady income with which to source troops. The end result is that players must use the large array of tactical considerations at their disposal to overwhelm their opponents: a small force of spearmen might have no chance against a larger band of orcs at night, but may well fend them all off if villages are captured in the daytime and used to heal and reinforce numbers.

3.1.1 Participants

Participant recruitment took three forms: a forum post on the *Battle for Wesnoth* forums (*Interception, an Experiment on Tension - The Battle for Wesnoth Forums* 2021); the public listing of the scenario on the *Battle for Wesnoth* add-on server as *Interception* (*Add-Ons Server - The Battle for Wesnoth* 2006); asking prospective participants directly to take part.

Participants recruited directly were given a link to a post (*Interception* 2020) explaining via both text instructions and a screen recording how to install *Battle for Wesnoth* and the scenario, as well as providing a link to the questionnaire.

The scenario itself details that it's for research purposes for participants wishing to take part, that no data is collected as part of the scenario, and that there is a questionnaire afterwards. The questionnaire, which *does* collect data, explicitly collects consent for this after detailing how the data will be used.

After screening for responses that did not answer the questionnaire fully, 27 respondents remained. Gender and age were not gathered; with the number of hypotheses being tested, there are many correlations being tested; more would ensure the Bonferroni corrections (Weisstein 2022) demand *incredibly* small p-values to remain statistically valid. However, experience with *Battle for Wesnoth* was gathered. 15 participants were "Extremely familiar with the game", and 10 had either never played a strategy game before (*Battle for Wesnoth* included), or had played 1 out of the game's 2 tutorials.

3.2 The Scenario

Participants in this experiment play through a *Battle for Wesnoth* scenario comprised of multiple objectives. Each objective is designed to adjust tension's components (investment, perception of player power, perception of opponent power, uncertainty) in different ways, in order to test hypotheses about the structure of tension. After playing through the scenario, participants answer a questionnaire in two parts: the first measures the levels of tension compontents that participants felt for each of the scenario's objectivs; the second asks participants about any events that caused perceived levels of power or tension to change. The questionnaire ends with a free-response question allowing players to mention any other information they see as relevant to the topics covered.

The second part of the questionnaire is analysed via thematic analysis.

3.2.1 Details

When players begin the scenario, players are informed of the setting and their goals within it. This does more than set the fiction; it manages players' expectations of what might occur, therefore affecting their perceptions of power.

> "*The orcs have declared war. Their warhorns echo from the horizon, and their fires tint the sunset with acrid smoke. Those with any militia to call their own have been called upon to lend their strength to the defense of the realm - including you, Duke Roland.*
> *...Don't worry, that's not you in the [picture of royal man in mortal combat with a huge bipedal creature]. But it might be, if you slip up.*"

> "*You are to secure the old river bastion before the orcs do, and kill their leader. Should the orcs claim the fort before you do, they will give themselves an unshakeable foothold from which to drown the northern marches in ruin. It will also make your mission to kill their leader Grem Three-Eyes a* ***lot*** *harder, so don't underestimate the threat this poses.*"

With this information, players are aware of their main objective: kill Grem Three-Eyes. It's also hinted that if they do not take the central fort before the orcs do, their job will become a great deal harder; this provides player investment for taking the fort, even if they're a new player that would otherwise not know this.

The playable portion of the scenario then begins, giving them two fail states: 25 turns elapsing, or Roland (the player's commander) perishing. This turn limit allows us to more tightly control the progression of difficulty in the scenario.

The scenario presents multiple objectives to the player over its course, but those will be explained within context of the scenario.

The Map

The player starts in the south of the map, and the orcs in the north. The aforementioned fort lies in the center of the river, with multiple paths of frozen ice also spanning the river; these paths are viable, but dangerous, as units' ability to dodge enemy attacks is severely hampered on ice.

While players can see the entirety of the terrain, they cannot see enemy units unless their own units are close enough to them; this is the *Fog of War.* This allows players to plan with terrain and landmarks (such as the fort) in mind, while still keeping enemies' positions and strength a potent source of uncertainty.

A small encampment near the player's start allows them to recruit a meager one unit per turn. Roland outright informs the player that the fort can be used to recruit many more, and that the orcs may be encroaching on the fortress at that very moment. This gives two separate reasons for the player to gain control of the fort first: speeding unit recruitment (this is necessary for players' continued survival in *Battle for Wesnoth*), and getting to that resource before the enemy.

Getting to the fort first. Set up in this manner, the players have one of the scenario's non-primary objectives in sight: *Get to the fort first.* They must still perform other actions in service of this objective: scouting, recruiting starting units, and capturing villages to produce unit-sustaining income. They must split their resources to do so, reducing players' certainty that they will achieve this objective.

Rescuing Corporal Nobbs. On the other side of the fort, initially unbeknownst to the player, is a Heavy Infantryman named Corporal Nobbs. Rescuing him is another objective of the scenario, but implicit and entirely optional. However, Nobbs is vastly useful; he is sturdier than any unit the player can usually recruit, and remains more functional at night than their other units.

He is also quite unaware of the incoming orcs. If the orcs find him before the player, the first thing the player becomes aware of is a high-pitched scream. He's also given a level of personality through his descriptions and interactions when found, in the hopes that characterisation aids in player investment[1].

The Orcs. The orc stronghold lies in the north; from here, the orcs' leader Grem recruits units for much of the scenario. Over time, these units are recruited in discretely stronger waves, directly controlling the level of power the opponents have. It also makes the orc stronghold difficult to assault directly.

[1] Players will still often see NPCs as resources

There are three such waves: the initial orcs, weak units; stronger units that might be defeated through outmaneuvering; some of the strongest units orcs are capable of fielding, which players will have a hard time defeating without a great deal of thought.

3.2.2 All Objectives

The full set of objectives the scenario provides is as follows:

1. Getting to the fort first
2. Saving Corporal Nobbs
3. Getting to the fort at all
4. Killing Grem
5. Surviving Wave 1 - the weakest wave of orcs, starting with few level 1 units; the same tier of units the player can freely recruit.
6. Surviving Wave 2 - three additional orc units spawn immediately and visibly at the orc stronghold. These units are level 2, and will require the player to hold an advantage of some sort in order to defeat: outnumber them, hold favourable terrain, or heal.
7. Surviving Wave 3 - three additional orc units spawn immediately and visibly at the orc stronghold. These units are level 3 and are extremely dangerous to almost all of the player's current units. However, Grem also advances at this time, making it plausible for the players to snatch victory from the jaws of defeat.

Each of these objectives attempts to tweak the components of tension in different ways, as listed in Table 3.1.

Players might give themselves many of their own objectives, whether in service of the scenario's objectives or not, but there's not proactical way of anticipating these.

3.2.3 The Questionnaire

The questionnaire was hosted via Qualtrics (*Qualtrics* 2021), a web-based tool allowing for online distribution of surveys and collection of their responses. The end of the scenario links participants to the questionnaire.

Participants are first asked:

- How much previous experience they had with the game
- Whether they found Corporal Nobbs (as some might not do so)
- The latest wave of orcs they survived until. Each wave is uniquely identified by dialogue and a particular type of unit — reiterated in the question — allowing participants to answer this effectively

Objective	Investment	Player Power	Opponent Power	Uncertainty	Methods
Get to fort first	High	-	Low	-	Investment heightened by stressed importance in intro, and uncertainty heightened by players knowing little about the orcs' current presence.
Save Corporal Nobbs	Medium	-	-	-	Investment provided by Nobbs' usefulness, as well as characterisation in his dialogue.
Get to fort at all	High	-	High	High	If players haven't gotten to the fort first, the orcs have, and an orc-occupied fortress is much harder to infiltrate.
Kill Grem	High	-	(varies)	-	Overarching and explicitly stated objective - investment provided by this being what the players came here to do.
Survive Wave 1	-	High	Low	Low	Few low-level enemies.
Survive Wave 2	-	Medium	Medium	Medium	Enemy breaks usual summoning rules some, raising uncertainty. Higher-level units attack player in higher numbers.
Survive Wave 3	-	High	High	High	High-level units that would crush players' in open field, but player is fortified, and Grem has joined battle, making him vulnerable.

Table 3.1: Scenario objectives

These questions are used to control whether other questions in the questionnaire are shown to the participant. If they did not find Corporal Nobbs, they will not be asked about him again. Similarly, they will not be asked about any wave they did not survive to see.

The questionnaire then proceeds to measure the components of tension that players experienced during the scenario for each objective. The intensity of feeling for each of these components is measured on a Likert scale:

However, investment was *not* queried for the 'surviving X wave' objectives. The assumption was that the waves were a subobjective of killing Grem — they must survive the waves of orcs to kill him — and so they would not have independent levels of investment. This is a limitation of the study.

The Likert scales for each component of tension are as in Table 3.2

All of these are 5-point Likert scales, with the exception of tension: it's nonsensical to say tension has or is at a maximum — akin to saying something is at a maximum heat — and so there is no 'complete tension' point on the scale.

At this point, the questionnaire shifts focus towards finding what events and mechanics appeared to cause

Figure 3.1: Components of tension, and the question asked for each

1. Investment
   - *How important did you feel that it was for you to...?*
   - It was reasoned that players would have differing interpretations of the technical term “investment”, and that importance was a reasonable, and more universal, substitute.
2. Certainty
   - *For each following objective, please indicate how certain you were, during the attempt, that you could complete it.*
3. Own Power
   - *How much power over the following outcomes did you feel you had?*
4. Orcs’ Power
   - *How much power do you think the orcs had to stop you achieving the following objectives?*
5. Tension
   - *How much tension did you feel while trying to complete the following objectives?*

and for each of these components:

Objectives

a. Killing Grem
b. Getting to the fort first
c. Getting to the fort at all
d. Saving Corporal Nobbs
e. Surviving the first wave
f. Surviving wave two
g. Surviving wave three

Component	Likert Scale
Investment	(Extremely / Very / Moderately / Slightly / Not at all) important
Player Power, Opponent Power	(Complete / Much / Moderate / Little / No) power
Certainty	Certain, (Quite / Moderately / Slightly) possible, Impossible
Tension	(Very / Moderately / A little) tense, No tension

Table 3.2: Likert scales for Tension and components

tension for players. It asks participants to describe any events that adjusted their perceptions of *their own power*, *the orcs’ power*, and *tension*. These questions are answered via free-form text.

The questionnaire ends with a cover-all question, asking participants to raise anything they felt was relevant that the questionnaire hadn’t thus far covered.

- **[Participant] power** *Were there any events that made you feel like you had more or less power over achieving any of your objectives? If so, please describe them.*
- **Orc power** *Were there any events that made you feel the orcs had more or less power to stop you achieving your objectives? If so, please describe them.*
- **Tension** *Were there any events that made you feel more or less tension? If so, please describe them.*

- **Cover-all** *Is there anything I haven't covered or asked here that you'd like to raise about the scenario?*

The full extent of the questionnaire is in the Appendix.

3.3 Results

Not all participants completed all objectives. 3 of the 27 did not save Nobbs, and 4 did not survive to the third wave of the orcs, removing that question from the questionnaire.

The freeform questions prompting players for events that changed tension for them received 17-20 responses each, and the final cover-all question 14 — excluding content-free responses such as "see previous answer" or "not really".

Each of the initially presented hypotheses has its validity assessed. Correlations between tension's components are calculated using Spearman's rank test; non-parametric testing is essential as we cannot assume the data is normally distributed. Moreover, as we are testing *multiple* correlations — and explore further correlations as we investigate details of the hypotheses, for a total of 7 correlations being drawn — the p-values of these are therefore Bonferroni-corrected (Weisstein 2022) to be 7 times higher than each individual test indicates.

3.3.1 Relationship between Certainty and Tension

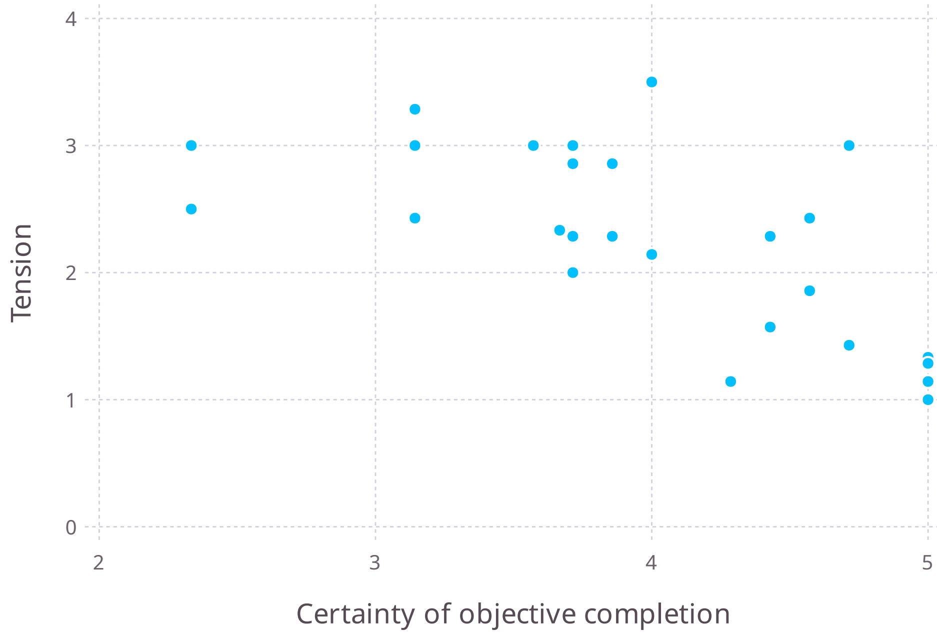
Hypothesis 1: Less certainty will result in more tension, regardless of other factors.

The Spearman's rank correlation between certainty and tension was $r(27) = -0.658, p < .01$; see Figure 3.2. This suggests that uncertainty is indeed able to significantly impact tension, though this doesn't discount the idea that other emotions may also feed into tension (Delatorre, León, Salguero, Palomo-Duarte et al. 2018), nor is it strong enough a correlation to imply certainty is the *sole* driver of tension.

However, the certainty a player has that they will complete an objective remains an effective mechanism for adjusting tension; examples of this are shown in the thematic analysis.

Though there were some players who rated their certainty of completing an individual objective as *Impossible* and still felt high tension, these players were inexperienced, all of them being new to the game, or having only played through the tutorials, or in one case simply being familiar with other games in the genre. This reinforces Delatorre et al's findings that those with lower knowledge of game rules may still feel suspense in situations impossible to get out of (Delatorre, Leon et al. 2017). However, there were still experienced players who rated objectives as "Slightly possible" and felt a small amount of tension.

Figure 3.2: Participant averages of Certainty *vs.* Tension



Component	Spearman's ρ with Certainty
Player Power	$0.878; p < 0.001$
Opponent Power	$-0.772; p < 0.001$
$PlayerPower - OpponentPower$	$0.861; p < 0.001$

Table 3.3: Correlations with Certainty

3.3.2 Player Power & Enemy Power *vs.* Certainty

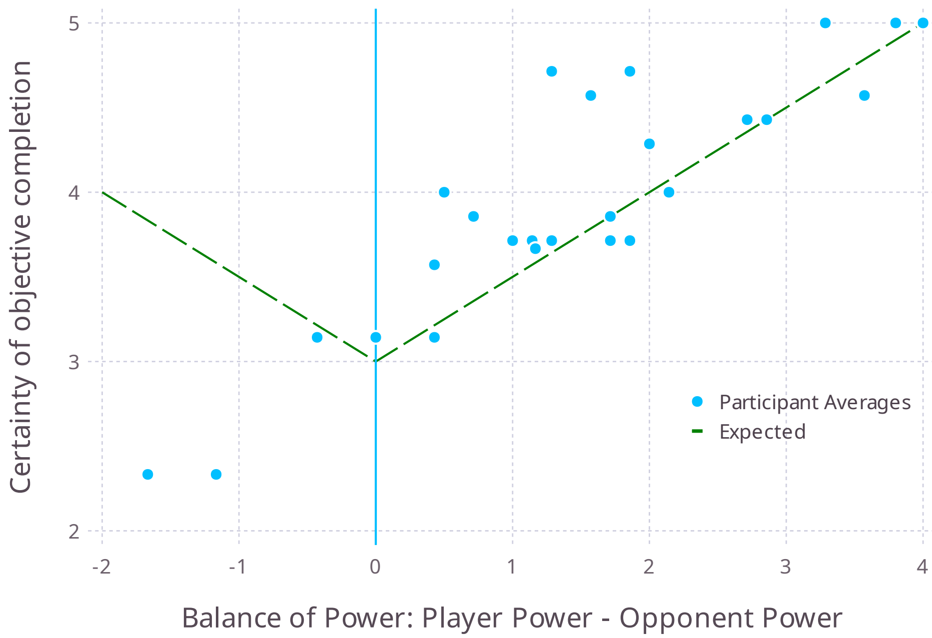
Hypothesis 2: Increasing (perceived player power − perceived opponent's power) will result in more certainty.

The difference between the player's and opponent's perceived powers has a Spearman's rank correlation of $r(27) = 0.861, p < .001$ with certainty of objective completion. However, perceptions of player power and opponents' power are also highly correlated with certainty — see Table 3.3.

As individual perceptions of power and the difference between powers are roughly equally-correlated with certainty, it's not clear whether it is the difference of power that contributes to uncertainty, or solely how much power the player believes they themselves have. Balance of power is graphed against certainty in Figure 3.3 in order to glean some insight.

If it was the *balance* of power affecting certainty the most, then we'd expect a U-shaped graph as shown by the green dotted line, with lowest certainty when power levels are balanced (their difference 0), and highest certainty when the difference is greatest. That symmetry is not present here, but *certainty of which*

Figure 3.3: Balance of Power *vs.* Certainty



outcome would occur may actually follow this trend; *certainty of objective completion* (what was measured) is meaningfully different from this, and the lack of symmetry would make sense. The less relative power the player is, the less certain they are that they can complete their goal.

It would be best to directly measure outcome certainty to find the ratio of powers at which players are most uncertain of what will occur as opposed to uncertain they can manage their objective. Still, it remains unclear whether it is the balance of power or simply one's own perceived power as the strongest driver of uncertainty and therefore tension, as shown in Table 3.3.

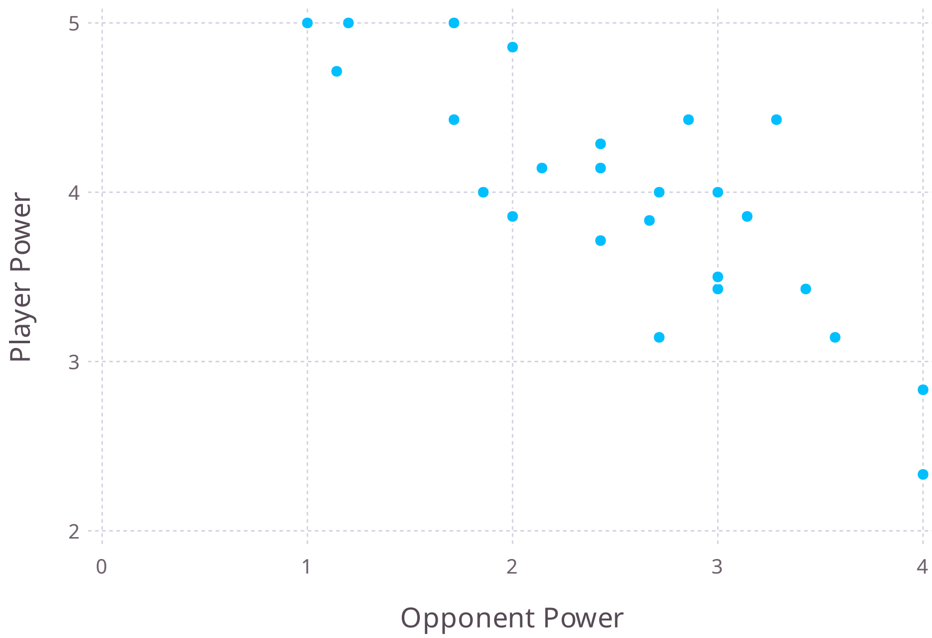
3.3.3 Player Power *vs.* Opponent's power

Hypothesis 3: Increasing perceived player power will be accompanied by decreasing perceived opponents' power.

This hypothesis seems to hold true – the two are related to each other with a correlation of $r(27) = -0.780; p < .001$, their strong relationship graphed in Figure 3.4:

The strong inverse correlation suggests that perception of power may be on a spectrum, from complete player power to complete opponent power. This allows game designers to adjust either to adjust tension, broadening the scope of changes available.

Figure 3.4: Player Power *vs.* Opponent Power



3.3.4 Relationship Between Player Investment and Tension

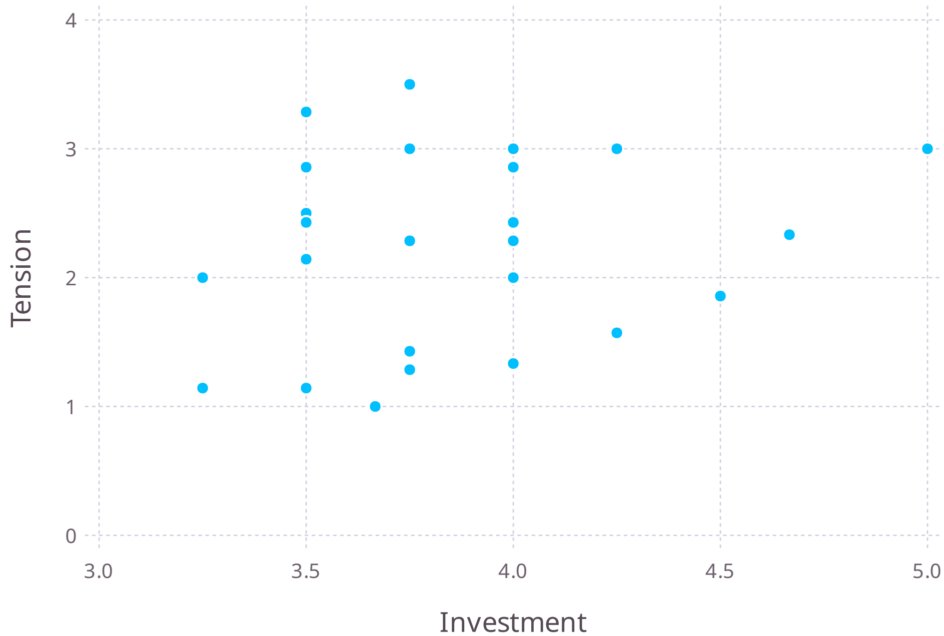
Hypothesis 4: More investment will result in more tension, regardless of other factors.

Measured player investment appears to have no meaningful correlation with any other component of tension or with tension itself: Spearman's rank correlation is $r(27) = 0.168, p = 2.4$. This abnormal probability is partially due to Bonferroni correction (Weisstein 2022), but is certainly no real correlation, and no guarantee that this was not by sheer chance.

This is a surprise. Investment is indicated by players to be a major factor of tension (Gomme and Bartle, Richard 2020) and arousal is a strong component of suspense in literature (Moulard, M. Kroff et al. 2019), so *why* is there no significant relation?

There are several possible explanations: participants and researchers interpreted investment differently; investment does not contribute to tension; investment did not contribute to tension in *this* scenario; or participants and researchers interpreted tension differently. This merits further investigation.

Figure 3.5: Investment *vs.* Tension



3.4 Thematic Analysis

In analysing events players viewed to affect their perceptions of power and tension, 12 separate categories were revealed; these are broken down in Table 3.4.

3.4.1 Information Acquitision and Tension

Tension and perceptions of power most often changed for players when information they possessed changed; *Information Acquisition* and *Surprise* were two of the most prevalent themes present in players' responses, with 26 and 28 occurrences respectively out of 133 total coded segments. Information (and lack thereof) has direct impact on uncertainty; keeping information hidden from players (as long as they realise that something is hidden), or revealing previously hidden information, is shown to be an incredibly effective way to manipulate both perceptions of power and uncertainty.

One of the primary methods of achieving this proved to be the *fog of war*, a darkness that obscures the presence of enemy units. Players seemed to have ideas of what might happen to them or their objectives in the darkness while they had no information, and acted with this knowledge in mind.

> *Orcs could jump out of the forest anywhere … it could delay so much that I wouldn't be first.*
>
> (Participant 18, Information Acquisition: Possible Futures)

Table 3.4: Themes in changing power or tension levels

Category	Summary	Occurrences
Surprise	When player is surprised by something; information reveals itself to player.	28
Information Acquisition	When the results of player's active gathering of information adjusts tension.	26
Hopeless Odds	When the player is outnumbered in a dire situation and does not think they'll be able to get out.	14
Resources depleted	Money? Men? Units? Health? The player is low on these.	14
Entrenched	Player is fully 'set up', having many of the resources they'd usually desire (e.g. fortifications, strong units).	13
Window of opportunity	An event allowing victory to be snatched from the jaws of defeat	10
High-level units	Strong enemy units intimidating player.	9
Lack of experience	Player's lack of knowledge of game rules or entities affects tension or power.	6
Overall pressure	Enemy pressuring player, often by sheer quantity, over time.	6
Certain about outcome	When players gather enough information or confidence to be entirely certain of the outcome.	5
Realisation of no challenge	Tension dipping lower and lower as realisation crystallises that things aren't challenging.	1
I Died Last Time	When a player didn't succeed the last time they tried, and are making another attempt.	1

Once hidden information is revealed, tension can fall: many players thought the fortress would be overrun with orcs by the time they arrived, and upon finding out this wasn't the case, and that the fortress was guarded by a friendly unit (Nobbs), tension was reduced.

> *Discovering the fort is empty, with an allied Heavy Infantryman holding the north bridge made me 100% certain I can capture the fort first which reduced the tension.*
>
> (Participant 19, Information Acquisition: Empty Fort)

> *The beginning of the scenario had a fair amount of tension simply because of the fog-of-war ...*
>
> (Participant 2, Information Acquisition: Fog of War)

However, the information hidden by the fog of war wasn't always revealed by the player's actions – sometimes, the orcs would emerge from the fog, surprising some players, and those surprises also changed perceptions of power and tension when they were revealed, this time *increasing* perception of opponent power and tension.

> *When an orc traversed the ice floe and attacked a healing knight in my rear*
>
> (Participant 22, Surprise)

> *When the wolf rider appeared in the fortress* (Participant 25, Surprise)

While those may not have surprised veteran players, newer players were certainly impacted. The experience of any given player had a marked impact on tension.

3.4.2 Players Lacking In Experience: Wait, I Can Do That?

One particular kind of surprise was unique to players newer to the game and genre. Instead of the fog of war hiding state or options, it was their lack of game experience.

> *I did not realise they could scale the walls of the keep ... so I was alarmed when they came over the walls to the east where I thought Roland would be safe.*
>
> (Participant 21, Surprise: Wait, They Can Do That?)

> *I didn't realise you could go over ice ... so when [the orcs flanked with this] I felt like that wasn't very fair* (Participant 25, Surprise: Wait, They Can Do That?)

While this changed perceptions of opponent power, that change sometimes came at the cost of player frustration and a sense of unfairness; while the orcs were following the rules of the game, the player's inexperience with those rules was being taken advantage of.

This reinforces Delatorre et al's work (Delatorre, Leon et al. 2017) highlighting that a player's familiarity with a game directly impacts tension, in this case in interactive play.

More experienced players, on the other hand, were adept at spotting when enemies *did* break the rules.

3.4.3 Experienced Veterans: Wait, I Know They Can't Do That

When Grem summoned successive waves of increasingly powerful units, he broke the usual rules of the game, in direct sight of the player[1]. Units are usually recruited by spending gold in the player's reserve. Gold income per turn is increased by capturing villages, and reduced by most units — they cost upkeep. The orcs departed from this rule twice when they spawned waves 2 and 3 of the scenario, creating units of quantity and quality that they should not have been able to recruit or support with the income available to them.

This departure from usual rules was purposeful; it was done to manipulate opponent power directly in the scenario, and keep opponent power constant between participants. However, it was clear to veteran players that the rules were being bent here, and this served to vastly reduce players' perception of their own power over the situation, and increase their perceived power of their opponents; they no longer knew what limits were on the opponent's unit-recruiting powers.

[1]The camera actually panned to him summoning them, to ensure the players were aware of where those units came from

> *The seemingly unlimited recruiting ability made me feel like I had much less power*
>
> (Participant 15, Hopeless Odds: The Hordes Are Endless)

> *... the summoning of helpers removed tension and made the ending seem inevitable*
>
> (Participant 20, Hopeless Odds: The Hordes Are Endless)

It wasn't just the rule-bending that achieved this. The sheer number and strength of orcs still intimidated newer players, bringing us to *resources*.

3.4.4 Resources

The amount of in-game resources that players had significantly changed how tense players were, and how much power they perceived themselves to have. In some instances, players were *Entrenched*: they had secured territory, income, villages or units, and were in a good position to enact their plans, or to use standard approaches that they had used before. In these cases, there was little tension - and a high perception of their own power.

> *I was still pretty confident of a win because I controlled the fortress and had a decent income each turn* (Participant 2, Entrenched)

> *once i realised there were villages inside the walls i was able to [employ] a defence in depth approach and no longer feared being defeated* (Participant 14, Entrenched: Established Strategy)

The opposite occurred too: higher tension when players' units were being overpowered, were low on health, or if they could not heal or summon more units. These players had their *Resources Depleted*:

> *When I lost most of my men and then the orc boss showed up, that was pretty tough. I kept trying to recruit more but I ran out of money.* (Participant 25, Resources Depleted)

Entrenched having less tension, and *Resources Depleted* having more indicates that they lie on opposite ends of a single spectrum. It's possible that locations of high and low tension on this spectrum are dependent on the task and challenges at hand. Some games subvert this, using a glut of resources — health or ammo pickups, new powers — as an overt sign to experienced players that they're about to have a very difficult time (*Suspicious Video-Game Generosity* 2023).

The orcs' resources were important too - the amount of units present directly affected players' predictions of their chances.

Definitely after wave 3 started. They were too many and I had no chance

(Participant 26, Hopeless Odds)

The amount of enemies that kept spawning made me feel powerless to actually stop them …

(Participant 15, Hopeless Odds: The Hordes Are Endless)

Though the amount of orcs was overwhelming, there *was* opportunity built into the scenario.

3.4.5 Window of Opportunity

The scenario presented a window of opportunity if players survived long enough: Grem would join the offense himself, adding power to the orcs' attack, but providing an opportunity to defeat him. This event was reflected in players' responses; players in deep trouble at that point saw it as a light at the end of the tunnel:

Then I was on the back foot and fleeing for my life, and I bumped into the enemy leader, with sudden opportunity to turn things around (Participant 24, Window of Opportunity)

The amount of enemies … made me feel powerless to actually stop them, until the leader left the fort … which gave me a little more hope, but it was too late for me by then

(Participant 15, Window of Opportunity)

This is particularly relevant for aspects of suspense related to changes in power over time (Madrigal, Bee and Johnny Chen 2022; Moulard, M. Kroff et al. 2019), as such 'windows of opportunity' are excellent ways to adjust uncertainty through perceived powers.

However, other players saw it as a foolish decision on Grem's part, an artificial blunder to make it easy on them (and possibly cheapening their previous efforts).

… I won because the general moved poorly. I don't think a human player would have behaved like that (Participant 23, Window of Opportunity: Suspiciously Convenient Opportunity)

… felt like that scenario at the end perhaps presented itself too easily, and maybe that diminished the tension a little (Participant 15, Window of Opportunity: Suspiciously Convenient Opportunity)

With that in mind, such windows of opportunity must be used carefully; if they seem arbitrary or contrived to the player, their use may backfire. It might be useful to ensure in the fiction that expectations of this kind of behaviour are set, if used. This does, however, make it difficult to use as a surprise.

3.4.6 Limitations

There are inherent limitations to the insights offered here. First is that participants self-reported their scenario experiences in the questionnaire; there were no concrete definitions of concepts such as 'tension' offered to players, possibly affecting their interpretations of questions, particularly when it came to the difficult-to-summarise concept of investment. The questionnaire also came after the entire scenario, and their recollections of their experiences may not be as complete as their experience at the time. These issues might have been sidestepped if players were to think aloud during scenario play (Charters 2003).

Another issue was that tension's Likert scale in the questionnaire had 4 items, while other measurements had 5, introducing possible noise. While it may be true that measuring tension as "complete" might be nonsensical, the same could be said of power. It would have been better to consistently apply a judgement on scales here, in this case having a 5-item Likert scale for each measurement taken in the questionnaire.

Changes of tension and its components also seemed to be more fine-grained than once-per-objective. More detail would come from measuring during gameplay when players indicated significant changes.

It would have also been useful to measure investment of players in the wave-surviving objectives. It was originally assumed that their investments would be tied into the objective of killing Grem – after all, it's just surviving so you can finish the scenario – but this assumption may have been optimistic.

3.5 Discussion

Tension appears to hold up; relationships between most of its constituent components are clear and consistent with initial hypotheses. There is evidence that perceptions of player and opponents' power lie on opposite ends of a spectrum, that both are linked to uncertainty, and that uncertainty is linked to tension. Tension seems to be intuitively understood by players too; there was no definition of tension in the questionnaire, yet players' responses remained consistent.

We also found several categories of events that successfully adjust perceptions of power and tension. Many relate to controlling information and resources the player has, as well as the resources that opponents control. All three of these are easy to manipulate for designers; tools such as fog of war already exist to do so, and there is certainly design space to manipulate them in novel ways.

However, tension didn't survive unscathed: while other relationships between its components held true, players' investment in objectives did not appear to be correlated with tension whatsoever. Moreso, the $p = 2.4$ (0.343 before Bonferroni correction) indicates that there is very little confidence in this lack of correlation; it is not assertive at all. To be certain that investment does not in fact contribute to tension, we need more robust results.

Chapter 4

Investigating Investment

Previous results on investment were inconclusive. They suggest that investment has no statistically significant relationship to tension, but do so very weakly. This is in stark contrast to the initial grounded theory (Gomme and Bartle, Richard 2020), which suggests that a level of investment is necessary for tension to form.

This might be due an error in measurement — that we did not measure investment with previous questions — or an actual lack of relationship between the two constructs. As the wording of the question that measured investment was changed to refer to "how important" it felt for players to perform objectives, it seems likely that the measurement might be the issue. If this is the case, we need a more effective method of querying investment. If investment is not in fact related to tension, we need more robust confirmation.

To further these goals, we create another *Battle for Wesnoth* scenario with multiple objectives, and present participants with questions using alternative terms for the concept of investment. If the wrong wording was chosen initially, it is likely that one of the presented alternative questions will be interpreted by participants as originally intended, and therefore have a cohesive relationship with tension. If not, none should have a significant relationship with tension.

The alternative terms that are used are:

- **Importance** The original wording used in the previous study's questionnaire, which seemed to not match the concept of investment (Gomme and Bartle, Richard 2020).
- **Want** Players may be emotionally invested if they *want* an outcome to occur.
- **Need** Similar to above, though perhaps more starkly. There is a difference between want and need, however; this may show in tension.
- **Investment** It was originally assumed that using the word 'investment' would cause players to misinterpret the question. This term is present to challenge that assumption.

To act as a control, we also measure player power and enemy power. Should these relate to tension in the same way as previously, we are given confidence that it is the measurements of investment that has changed, as opposed to the scenario producing different effects on participants.

4.1 The Scenario - Unto Others

This is another *Battle for Wesnoth* scenario, named *Unto Others* to distinguish it from the previous scenario *Interception*. *Interception* is not used, as measured 'investment' did not vary significantly, with the exception of the objective to save Corporal Nobbs: see Figure 4.1. In this case, we wish to vary investment between objectives more significantly.

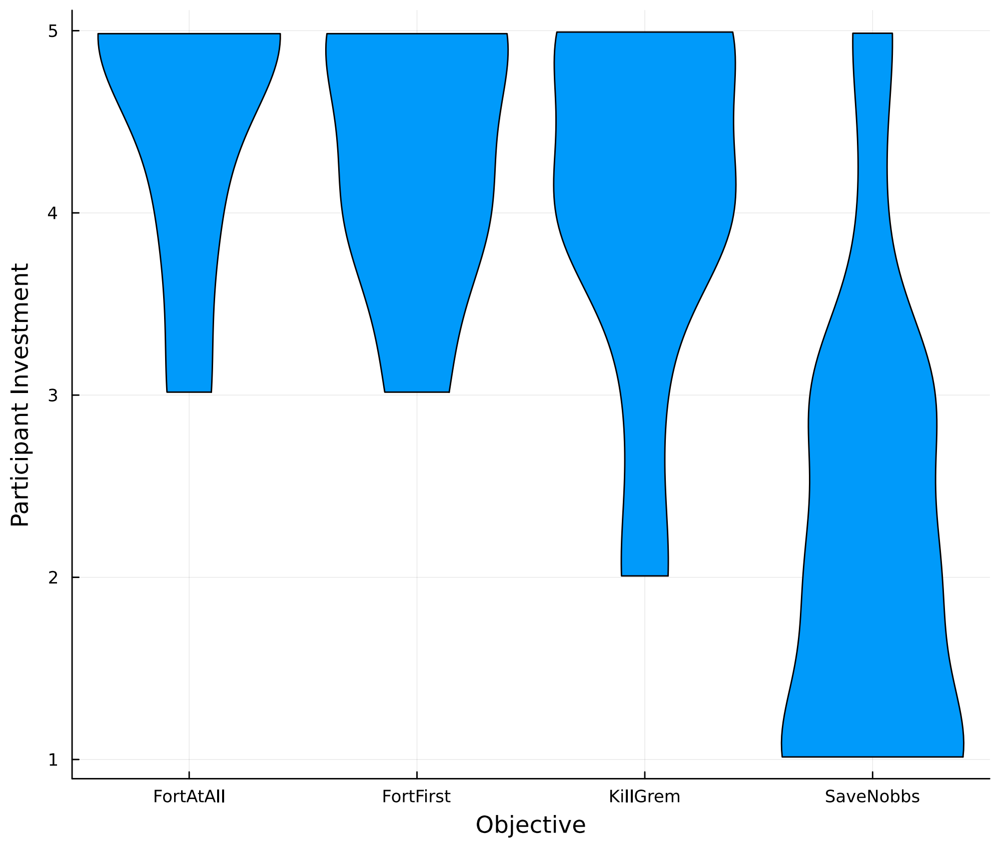


Figure 4.1: Distribution of Participant Investment per Objective (Interception Scenario)

To this end, we introduce objectives that should vastly change the focus of what the player should find important in the scenario. In many cases, this is a Tomato Surprise (Raymond 2009) — a surprise that would change how players play from the start, should they play a scenario again with knowledge of that surprise — and should be avoided. However, in this case, there is no intention of participants playing a second time.

The synopsis of the scenario is as follows:

- Two factions are at war: the Indrith (player) and the Yalthans (enemy), over an old tomb hiding secrets.
- The Indrith are performing a large ritual - they plan to destroy the tomb rather than let you have it.
- The player is to escort a cadre of mages through the city to effect a counter-ritual.
- The player must find the ritual, then protect the mages as they destroy it.
- Partway through the ritual, the mages instead take the ritual as their own, intending on *destroying the city the player is in.*
  - This is the point at which players' focus - and therefore what objectives they might be invested in - is to change, from defending the mages to trying to kill them before they finish subverting the ritual.

The objectives this time are fewer in number; as we are asking participants many more questions regarding investment, keeping 7 objectives may risk less participation in the experiment. The objectives chosen were:

- Find the ritual
- Destroy the ritual before it goes critical
- Destroy the Yalthans
- Kill the mages (before the ritual goes critical)
- Kill the mages (after it goes critical)

4.1.1 Participants

Participants were recruited via a post on the *Battle for Wesnoth* forums, the listing of the scenario on the official *Battle for Wesnoth* add-on server as *Unto Others* (*Add-Ons Server - The Battle for Wesnoth* 2006), and some were asked directly to take part.

38 participants took part. Gender and age were not gathered, as they were not deemed as relevant to the study. Experience with the game was gathered: see Table 4.1.

4.1.2 Scenario Details

In this scenario, participants start on the south side of the city, at the point denoted 'Player 1'. The mages with them start at the point denoted 'Player 2'. They are not controlled directly by the player, and will attempt to seek out the ritual site while defending themselves (though they *do* need escorting to not perish). 'Player 3' is where the Yaltha (opponents) start, and several are in position in front of the player.

Experience	Count
Extremely familiar with the game	20
Had played it before	2
Had played through both tutorials	1
Had played through one tutorial	7
Had never played before	1

Table 4.1: Participants' Experience Levels

Figure 4.2: Full Map of Scenario

The ritual lies on the northeast island, roughly where the gold-rimmed hex is in Figure 4.3.

Opponent Units

The Yalthans are comprised of Dunefolk units in-game. These are weaker at night, during which the player can use their Light Mage to mitigate their own night-weakness — a significant advantage to the player, as the entire scenario occurs in the night. However, this is more than counteracted by the fact that the player *cannot recruit*, whereas their opponents can. They have a mix of units roughly equal in individual strengths to the player's. While they're initially few in number, their leader is able to teleport to wall keeps around the city and recruit from there.

The ritual is a special unit of theirs which cannot move. It has 200 hit points — a *large* number for a

Figure 4.3: Participants' Starting Vision

single unit, but likely to take several turns to whittle down — and significant resistance to any type of damage except arcane.

Player Units

The player's starting units are as follows:

- **Light Mage** The players' leader. Extends a hex-wide radius in which player units may ignore penalties from attacking during nighttime.
- **Rogue** One of the most important units the player has in this scenario. Has many movement points, but is also a *skirmisher*, meaning that it can ignore (read: move past) enemy units' Zones of Control[1]. This makes it essential to find the ritual site even with enemies' occupation of the city.
- **Heavy Infantryman x2** Serving as 'tanks', these units have heavy resistance to mundane attacks, quite a bit of damage, and a lot of health. As a trade-off, they're not very mobile at all, and are unable to

[1] Units extend a hex-wide Zone of Control (ZoC). Any unit entering a ZoC hex usually ends its movement immediately.

gain full dodge bonuses from terrain that's otherwise very useful for it.

- **Shock Trooper x1** As above, but hits significantly harder and has more health.
- **Swordsman x1** An all-around unit.

The Mages

The mages also have their own units. These are on a separate 'team' in the scenario code, though this is hidden to the player, and they are considered allies initially. They are the only units that can damage the ritual effectively, so while they *can* die, the player is unlikely to successfully destroy the ritual if they're gone.

- **Red Mage x3** Capable offensive fighters with their fireballs, but quite vulnerable, with relatively low hitpoints and no resistance to mundane damage types.
- **Arch Mage x1** The mages' 'leader', Altix is used as the default speaker for any of the mages' exclamations, and is more capable than his comrades.

4.2 The Ritual

On finding the ritual on the northeast isle with any unit, the mages will move relatively obsessively towards it, and attack it repeatedly. When the ritual reaches a critical threshold, an event occurs in which the player is informed by the Yalthans that the mages have taken the ritual for their own extremely destructive purposes. The Yalthans offer to stand with them against this horror.

At this point, the player is informed that they may *turn* on the mages if they believe the Yalthans. By right-clicking any unit, they can choose to consider that faction's units as friendly or unfriendly on their turn, thereby allowing them to attack anyone they wish.

If they attack the mages: The mages will swear that this is the only way to win the war, and continue attacking the ritual. Should the player successfully kill the mages before the ritual runs out of health, the Yalthans offer a shaky peace, and the player wins the scenario.

If they do not attack the mages: The mages will still attack the ritual.

If the ritual's health hits 0: The mages cry that their plan is complete, and the ritual detonates, utterly leveling the city: see Figure 4.4. The scenario is lost.

Figure 4.4: Ritual Detonation

4.3 The Questionnaire

After completing the scenario, whether they win or lose, participants are linked via QR code and URL to a survey hosted on Qualtrics (*Qualtrics* 2021).

Participants are asked:

- Their previous experience with *Battle for Wesnoth*
- Whether the ritual went critical
- If they attacked the mages, and whether the attacking was before or after the ritual becoming critical
- If they attacked the Yaltha too after turning on the mages
- For each objective:
  - 'How important did you feel it was for you to...'
  - 'How much did you want _____ to occur?'
  - 'How much did you need _____ to occur?'
  - 'How invested were you in _____ occurring?'

Component	Likert Scale
Importance	(Not at all/Slightly/Moderately/Very/Extremely) important
Want	Not at all/A little/A moderate amount/A lot/A great deal
Need	Not at all/A little/A moderate amount/A lot/A great deal
Investment	(Not at all/A little/Moderately/Very/Incredibly) invested
Own Power	Not at all/A little/A moderate amount/A lot/A great deal
Opponent Power	Not at all/A little/A moderate amount/A lot/A great deal
Tension	Not at all/A little/A moderate amount/A lot/A great deal

Table 4.2: Likert Scales for Unto Others' Questionnaire

  - 'How much power over the following outcomes did you feel you had?'
  - 'How much power do you think the enemy had to stop you achieving the following objectives?'

- What makes an objective something the player personally wants to see completed in a game
- How that player would describe how they feel about such an objective, to another person
- A cover-all question allowing participants to raise any other details they feel necessary

Each of tension's components was measured on a Likert scale as detailed in Table 4.2.

The full extent of the questionnaire is in section B.1.

4.4 Results

First, the control relationships are assessed: player power and opponent power with respect to tension. Each alternative term for investment is then assessed for its relationship with tension using Spearman's rank test. Each p-value is then adjusted with Bonferroni-Holm correction (Holm 1979) to control familywise error rates.

Of the 38 participants who took part, 10 were discarded by virtue of not having completed the questionnaire fully (*fully* being defined as consenting and more than 80% of the questionnaire complete), for not being aware of the ritual in the scenario, or in one case experienced unintended behaviour, as he'd killed all the enemies before they could inform the player that the ritual had been twisted in order to be used for violence.

Of the remaining 28 participants, 16 survived long enough to see the ritual hit the critical stage. Of these, 9 proceeded to not attack the mages, and 5 attacked the mages after it went critical. Two attacked the mages *before* it went critical.

4.4.1 Control Relationships: Powers vs Tension

Interestingly, the correlation between player power and tension falls apart in this scenario, with an adjusted p of 0.6. Even the relationship between the player's perceived power and the enemy's perceived power falls apart, with a correlation of $r(26) = 0.168; p = 0.39$. This may be due to players not knowing where the ritual was at all, compared to the previous scenario where players knew exactly where they needed to go.

Enemy Power and Own Power

It would counter previous evidence if, all else being equal, perceptions of power decoupled themselves from tension. However, the previously very strong relationship *balance between powers* does not hold true in this scenario; they're related together with $r(26) = 0.168; p = 0.78$; see Figure 4.5.

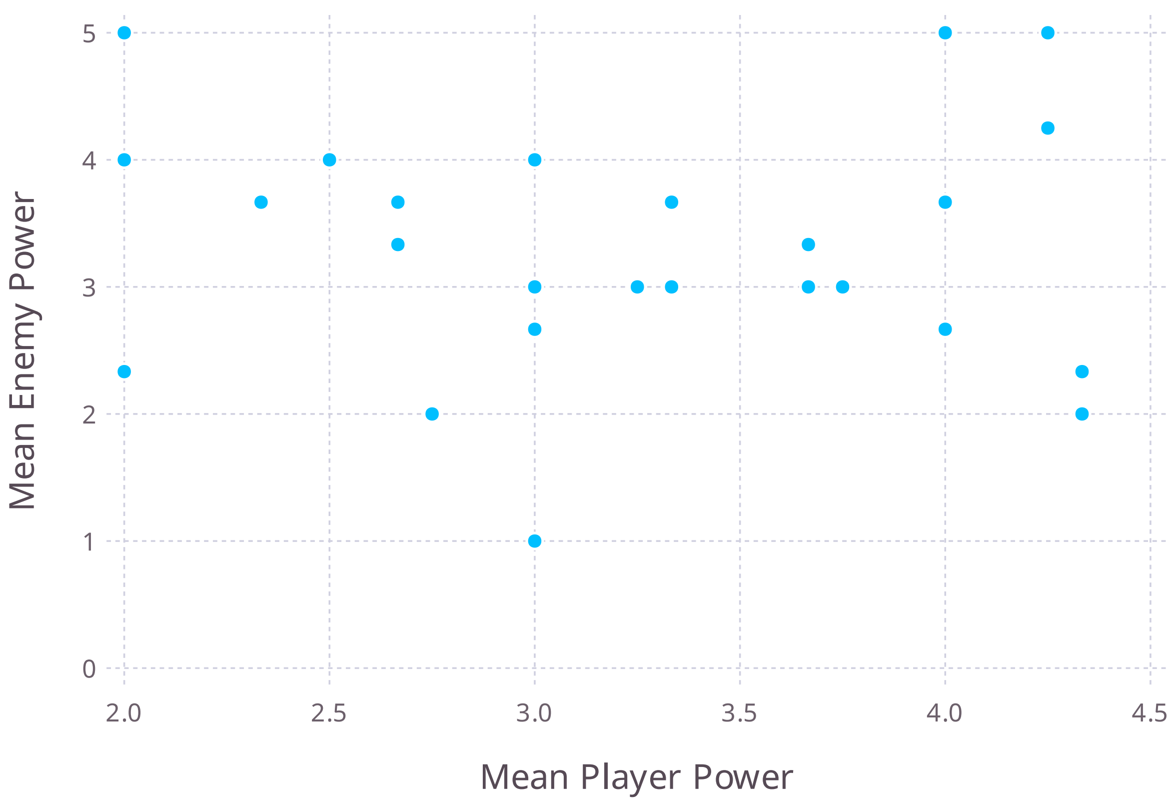


Figure 4.5: Participants' Perceived Powers

This is a significant change from before — where the two were very tightly connected — and implies that something else is happening here that changes perceptions of power. It's difficult to say what exactly may be causing this; it doesn't seem to be directly connected to whether the ritual went critical and made players aware of the betrayal. However, as a whole, players seemed less sure of what was happening in this scenario than they were in *Interception*.

In any case, it seems that perceived player power and perceived enemy power can be decoupled from each other.

4.4.2 Different wordings for investment

The wordings for investment each connected strongly to tension: see Table 4.3 and Figure 4.6. It is clear that *Importance* — the wording used previously to measure investment in *Interception*'s questionnaire — is less connected to tension than almost any other wording used, making it reasonable to assume it's indeed not the most effective method of querying investment. However, it *is* still quite strongly connected with tension, unlike before. This implies that player investment and balances of power can *each* be significant drivers of tension, though as of yet we have not seen them coexist.

The wording that seems to match most with the expected relationship between investment and tension is *Need*: 'how much did you *need* to [do X]?'. This makes sense; the word *need* implies more urgency of desire than *invested*, *importance* or *want*.

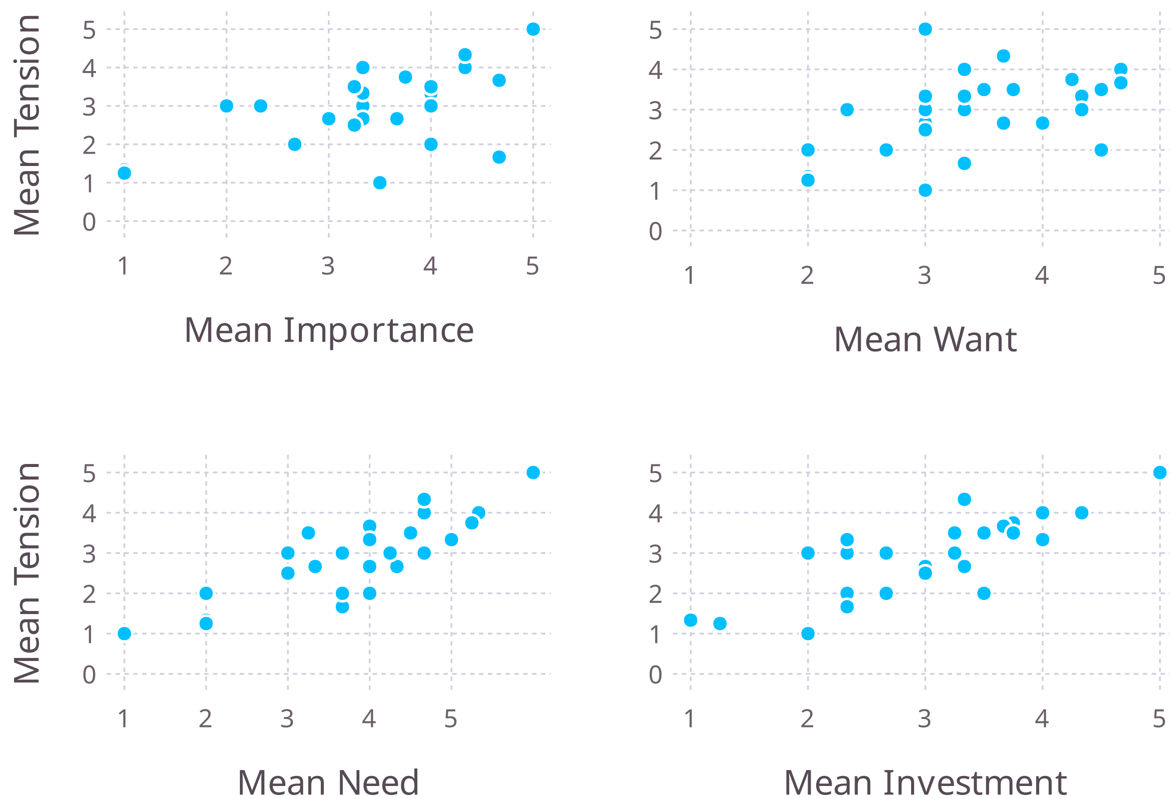


Figure 4.6: Investment Wordings vs Tension: Participant Averages

Investment Wording	ρ with Tension	p
Need	0.782	$< .0001$
Investment	0.766	$< .0001$
Importance	0.520	$< .05$
Want	0.512	$< .05$

Table 4.3: Correlations between wordings of investment, and tension

4.5 Discussion

By asking players about investment in a strategy game scenario in several ways, we find an effective method of asking players about their investment as it relates to tension: their *need* to complete an objective, though asking about *investment* directly also seems to work.

However, we also find that perceptions of power don't always directly feed into tension, or each other. This may occur in scenarios or situations where either their power or their opponents' power is unclear to the player, and therefore in questioning afterwards may not respond consistently. It also highlights a need for sampling components of tension during gameplay in order to ensure more accurate data, as well as possible timelines of tension.

Chapter 5

Reflective Research

The conclusions drawn thus far, as it turns out, connect to a vast warren of concepts within the field of games research, narrative, psychology, and even the governmental sector. There are several areas of interest that have yet to be explored. There are places where multiple fields of research have made similar strides, whose wisdom we can draw upon. We start with one such source of wisdom.

5.1 Expectation Management: Existing Literature and Open Questions

Strategy games often violate players' expectations of the AI they play against (Gomme and Bartle, Richard 2020), to the detriment of the player's experience. This begs several questions: what other expectations do players hold in games? Can we measure the impact on player experience that violated expectations have? Do those expectations differ when a player has more or less general experience in games?

To look at the most mature literature on expectations, we have to step outside of games momentarily. Expectations have been explored in the fields of consumer satisfaction and citizen satisfaction in the public sector (Zeithaml, Berry and Parasuraman 1993; Zhang et al. 2022). Here, researchers have investigated the effect that expectations and actual performance have on satisfaction when a service is performed and evaluated. Of the few models present, the most tested construct appears to be the **E**xpectancy-**D**isconfirmation **M**odel (Oliver 2010).

5.1.1 The Expectancy-Disconfirmation Model

Its heart is the following:

- Citizens have expectations of the services that they receive from their council
- Citizens hold an opinion on how well these services were performed

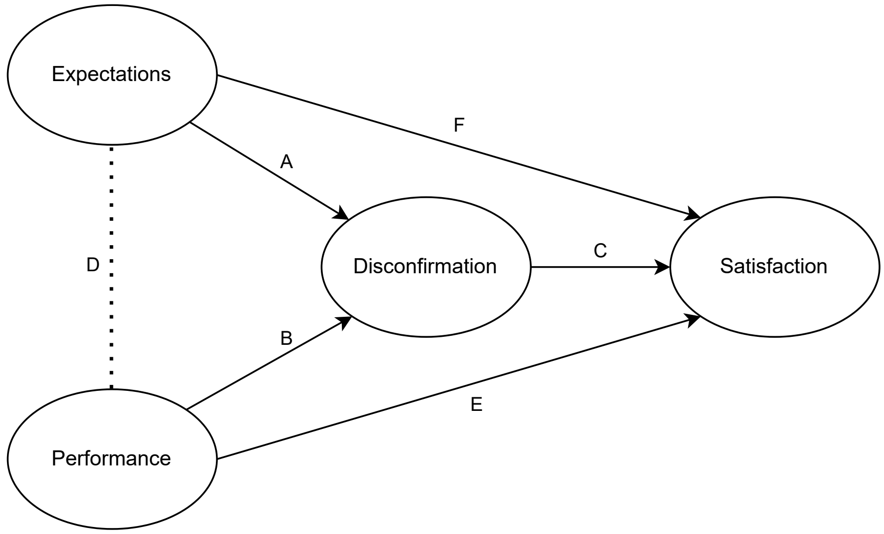


Figure 5.1: The expectancy disconfirmation model, adapted from Van Ryzin 2004

- A mismatch between expectations and perceived performance causes disconfirmation
  - Performance > Expectations: *positive* disconfirmation
  - Expectations > Performance: *negative* disconfirmation
- Disconfirmation affects satisfaction
  - Negative disconfirmation: less satisfaction
  - Positive disconfirmation: more satisfaction
- Expectations are correlated with performance[1]

In short, exceeding expectations results in more satisfaction, and failing to meet them results in less satisfaction. Changing expectations, while keeping performance the same, should still affect end satisfaction. However, while this has held true in cross-sectional studies (Roch and Poister 2006; Van Ryzin 2004, 2006), the model suffers somewhat in experimental contexts, where expectations and disconfirmation appear to have little effect compared to performance (Grimmelikhuijsen and Porumbescu 2017; Van Ryzin 2013), and where simply raising expectations (without increasing performance) raises overall satisfaction (Van Ryzin 2006). There are some possible explanations put forth by Grimmelikhuijsen and Porumbescu 2017: there are multiple types of expectations which are often ignored by studies; there are multiple ways of measuring disconfirmation; and country of study varied effects significantly.

The most salient of details, when asking how *player* expectations affect player experience, are the multiple types of expectations that citizens have (Favero and Kim 2021; James 2011):

[1] The specific nature of this relationship is rarely specified, and there seems to be little agreement over whether this is an essential part of the EDM (Andersen and Hjortskov 2016; Van Ryzin 2013).

- **Ideal expectations** What citizens think service performance *should* be.
- **Forecast expectations**[1] What citizens think service performance *will* be.

Ideal expectations appear to affect disconfirmation more strongly (Grimmelikhuijsen and Porumbescu 2017), but this has yet to be explored in detail. They're also quite hard to adjust compared to forecast expectations (James 2011), remaining relatively stable over time despite performance changes (Hjortskov 2019). Therefore, it seems important to distinguish between the two types of expectation.

It's worth noting that of the expectations found in this thesis, only *tension* appears to be a forecast expectation. AI playing fairly, being able to explain its actions, or being able to adapt its actions to player feedback are not extensively seen in the field of strategy games, and would therefore appear to be ideal expectations. It would be interesting to see, for each of those expectations, whether players believe them to be ideal or forecast.

5.1.2 Priming Subjects: The Effect of Giving Reminders

This thesis has discussed the possibility of changing players' expectations, in order to avoid breaking expectations that cannot be met. One might expect that to do so, you could present the player with information that would then adjust their expectations. However, Andersen and Hjortskov 2016 find that when participants are primed with information about current and future states of affairs, instead of changing *expectations*, perceived performance was changed instead, even when the information contained no change to the state of affairs. This suggests that perceived performance, while moored to actual performance, has more slack than might be assumed in this bond, and could be quite flexible depending on what the participant has recently been reminded of, making the ordering of questions when measuring performance an important factor.

5.1.3 Tolerated Performance

A slightly different approach to expectations springs from research into consumer satisfaction. The theory appears to have been through very little observable testing, but bears similarities to EDM, and offers another measure that may come to be relevant in expectation exploration. Zeithaml, Berry and Parasuraman 1993, through qualitative interview study, hold that consumers expect a given level of service. Between a consumer's *desired* level of service and what they deem to be *adequate* service, there is a *zone of tolerance* defining the levels of service that the consumer will accept.

[1] These have actually most often been referred to in public service literature as *normative* and *empirical* expectations respectively, and those terms are best to use when searching that literature. However, those are significantly less readable than ideal/forecast (introduced by Higgs, Polonsky and Hollick 2005) and somewhat of a misnomer: 'empirical' expectations appear to be affected by more than simply evidence (Andersen and Hjortskov 2016). Thus, ideal and forecast are used here.

Once more we see an *expected* level of performance, and what performance *should* be in the eye of the consumer/citizen. The difference here is the zone of tolerance. Should this zone exist, it may have been an unaccounted-for confounding factor in EDM experiments. It would be unsurprising if the AI of many games resided in this zone of tolerance. After all, many games use a similar range of technologies, which would inform players' 'adequate service' threshold under this model (Zeithaml, Berry and Parasuraman 1993). As mentioned above though, ideal expectations (desired level of service) affect satisfaction more strongly, which suggests it's still wise to reach higher than what is merely acceptable.

5.1.4 How EDM Informs Possible Expectations Research

Through examining strides made with expectations outside games, we learn that expectations have been shown to affect satisfaction, but the exact mechanisms by which they do so need to be tested more thoroughly, in line with the recommendations set out by Grimmelikhuijsen and Porumbescu 2017. Such testing should also distinguish between different kinds of expectation. This should assist in avoiding pitfalls that the governmental sector has experienced in their studies.

EDM, and the 'zone of tolerance' have been raised in a different field, and therefore should be tested before any kind of use in games even if both were entirely unquestioned. The 'services' in games that players would hold expectations about are different on some fundamental levels, and reside in an entirely different environment. However, there are some reasons to believe EDM might hold some water in other sectors. First is that it shows flexibility within the public sector, showing no statistically significant differences when applied to services from refuse collection to road condition and library quality (Zhang et al. 2022). Second is that EDM reasons about the individual, their perceptions, and their end experience, which remains relevant when it comes to players. They're still humans, and raising expectations (without also raising performance) has been seen to affect player experience in games too (Denisova and Cairns 2015). Lastly, the fact that aspects of EDM echo within consumer satisfaction suggest kernels of truth that apply across domains.

5.2 What Expectations Exist in Other Games?

This thesis has unearthed several expectations for AI in strategy games. Some of them may even apply to other types of games, though strategy games are likely where they'd be broken most egregiously. Yet there is no obvious reason that players would only have expectations for strategy games, and no obvious reason that the breadth of their expectations would only cover AI behaviour. Thus, it makes sense to ask about what other kinds of expectations exist in games.

To date, there has been relatively little research on the topic. There is some work on *rational* expectation (Aumann and Dreze 2008): that is, given a game state and rules, a value for each state, and reasoning about other players' actions, how much one player might expect to get. This is certainly an expectation of a sort, but not players' desire of how a game *should* work, which then might be breached in the course of gameplay, and thus cause frustration.

Much more relevant is a recent grounded theory on player need frustration (Ballou and Deterding 2023) which incorporates player expectations, and gives us means by which to anticipate as-yet unexplored expectations that might arise in games. Twelve interviews were conducted, in which participants were asked to detail recent experiences of need frustration in games. Between interviews, codes were iterated upon, memoed, and interview prompts were tweaked. Ballou and Deterding explain their grounded theory in remarkably readable terms:

> *Need frustration occurs when players' need-related expectations are negatively violated during gameplay by an unexpectedly salient or intense need-frustrating situation. Over time, need frustration experiences update players' expectations, which in turn prompts alterations to play behavior aimed at reducing future need frustration—unless extrinsic motivations overpower these.*
>
> — Directly quoted from Ballou and Deterding 2023

To clarify, the needs that were being frustrated were those presented by self-determination theory (Deci and Richard M Ryan 2012): autonomy, competence, relatedness.

Ballou and Deterding unearthed several types of frustrating situations for each type of need:

- Autonomy
  - Desired playstyle is constrained
  - Ability to play is constrained
  - Players are compelled to play
- Competence
  - Stagnation: players feel unable to complete the next step, stifling their progression
  - Unfair situations: External forces make achieving a desired goal without providing players with expected means to achieve said goal
  - Meaningless choices: Where their decisions simply don't matter
- Relatedness
  - Disconnection from other players. Disagreements with teammates, skill mismatches, and harassment.

  - Disconnection from the community at large
  - Disconnected from the game's characters or world. This didn't cover just *uninteresting* settings, but settings that disappoint or alienate the character: one participant reported they simply didn't enjoy playing in theaters of war.

These categories provide a more granular lens than Self-Determination Theory's broad concepts of autonomy, competence and relatedness. By being more concretely tied to games, these categories allow us to address issues arising in these categories more robustly. This lens seems to work relatively well; the expectations found in this thesis mostly fall into these categories, suggesting that both AI expectations and need-frustration categories hold validity.

- **Tension** This falls under either *autonomy* or *competence*. For *autonomy* frustration, players who wish to play a hyper-aggressive war are unable to, for inept AI would simply fold under the pressure; the player's playstyle is constrained. Additionally, sometimes players wage war with every cognitive and material resource at their disposal, only for the AI to give absolutely no resistance in their heartland (Gomme and Bartle, Richard 2020, p.7). That would thwart player *competence*; the player's choices and efforts have been rendered meaningless.

- **Fair Play** Ballou and Deterding place this exact phenomenon in their *unfair situations* category of need frustration.

- **The Deaf Student** This is frustration of *autonomy*. Players are unable to affect the change they wish to. They know how they want the AI to change, but lack the means by which to make that change effectively.

- **Explainability, Transparency, and Closure** This expectation is harder to pin down specifically. However, it does fit very well within ***P**redictive **P**rocessing* (Clark 2015), a cognitive framework that Ballou and Deterding also connect strongly to their findings. We will delve into Predictive Processing, but in short, humans yearn to resolve differences between predictions/expectations and observations. That is, in effect, exactly what players are trying to do when they post on forums regarding unexpected behaviour.

The authors show that violating player expectations has consequences. Do so negatively, in a way that does not support player needs, and the player experience suffers. At that point, players may come to expect further need frustration in gameplay.[1] Players may then change the way they play to mitigate the need

[1] This offers a counterpoint to the public sector's assertion that expectations are difficult to change, or that further information changes current perception of 'performance' instead of changing expectations. However, forecast expectations are what have changed for players in this case, which are easier to affect than ideal expectations (James 2011).

frustration they have now come to expect, or they may end up simply dropping the game entirely (Ballou and Deterding 2023).

These findings are also general, applying across genres. The favourite games of those interviewed covered various shooters, Multiplayer Online Battle Arenas, roleplaying games, platformers and sandbox games. However, it's important to acknowledge that each game is almost certain to engender its own expectations in a player. Sequels will entail similarities and improvements to its predecessor. Games entering a well-populated genre would, by default, be expected to include prevalent design choices within that genre. The style of art and fictional setting will also dictate expected tone; all else being equal, *Darkest Dungeon* 2016 would have likely frustrated many players if its art was in the style of *Stardew Valley* 2016. It may be prudent to carry out investigations of expectations that gamers currently carry into the games they play, and how those expectations might have shifted over time, or how those expectations might differ between genres.

5.2.1 Finding Failed Expectations Through Measuring Need Frustration

We have, until recently, only had measurements of need *fulfilment* in games with the Player Experience of Need Satisfaction scale (Richard M. Ryan, Rigby and A. Przybylski 2006). However, since need frustration has been identified as seperate from need satisfaction (Bartholomew et al. 2011; Warburton et al. 2020), we need a separate tool to measure instances of such need frustration. As violated expectations appear to be quite related to need frustration, such a tool would also be useful to unearth expectations being failed.

The Basic Needs in Games Scale (BANGS) (Ballou, Denisova et al. 2023), a questionnaire to investigate both need fulfilment *and* frustration in games, seems a promising tool to do so. It has been statistically validated, and can be used across single or multiple game sessions. However, it has not yet undergone peer review and should thus be used with caution.

5.2.2 Explaining Players' Needs for Explainability: Predictive Processing

One expectation found in this thesis didn't fit neatly into the theory of need frustration: strategy game players' expectations of explainability, transparency and closure regarding the AI's actions. However, it is explained well by the **P**redictive **P**rocessing framework. It posits the brain as a prediction engine, constantly churning out predictions about what it will observe and why those things happen, and comparing those predictions with actual observations (Nave et al. 2020). Differences between predictions and sensory input raise *prediction errors*, which the brain seeks to resolve via changing its predictions, or by changing the environment to match its predictions.

At first, this seems to be incredibly high-level and abstract. But as the soul of many games revolves around uncertainty (Costikyan 2013; Koster 2013), it turns out to be incredibly relevant and applicable to games. It also covers areas not adequately explained by self-determination theory; idle games such as Orteil 2013 do almost nothing to foster players' competence, and Soulslike games such as *Sekiro: Shadows Die Twice* 2019 would crush players' feelings of competence with near-certain failure on most attempts (Deterding et al. 2022). Predictive Processing, however, dictates that (at least some) appeal in games is continuously doing better than expected, which holds true for idle games, Soulslike games, and those in-between.

This thesis will not attempt to disentangle PP from SDT, or to argue which is superior. Predictive processing simply fits the data on expectations very well: players have expectations (predictions of how play will occur), these are violated, and then the player adjusts their play to avoid further prediction error. They have little room to adjust the environment to match their expectations, but do what they can through either modding (*Vox Populi* 2014) or posting to forums about their frustrations (Gomme and Bartle, Richard 2020).

5.2.3 Future work on expectations

We have seen that expectations are important to human judgements in general, as well as to the satisfaction and well-being of players in games. Games studies are also starting to gain a sense of the expectations players hold when they play games, and can therefore take action to avoid violating them negatively. However, we as yet have little work on what factors manipulate these expectations, and little work on how we might be able to adjust those factors. The expectations we that have come to light been mostly theoretical thus far, and require further testing. Those expectations may shift over time too, though we do not know in what ways.

5.3 What We Can Learn From Suspense

Tension has received much attention in this thesis, in both its conception and the beginnings of its testing. However, at the time that tension was theorised (Gomme and Bartle, Richard 2022), I knew nothing of the literature on suspense. In some ways, this is a good thing; grounded theories are supposed to be drawn directly from the data with as little outside influence as possible, though the theory will inevitably be touched by the analyst's biases (Charmaz 2006). Knowing about suspense when conducting that grounded theory would have certainly influenced the outcome. I have since become aware of suspense, and it would be irresponsible to not acknowledge where the two concepts overlap, where suspense holds insights from its longer pedigree, where tension differs, and what future research might look like.

Initially, study of suspense was limited to drama (Zillmann, Hay and Bryant 1975), but then made its way into games (Klimmt, Rizzo et al. 2009), where it was found to positively impact game enjoyment. However, it was Moulard, M. W. Kroff and Folse who laid down a cogent theory of what exactly suspense *is*, and how it applies to games in particular (Moulard, M. Kroff et al. 2019).

Suspense is 'an experience of uncertainty that occurs over time, from the moment a person realizes a potential outcome to the moment that outcome is resolved. ...suspense entails uncertainty in the period prior to the game outcome and the emotions that arise from that uncertainty (hope/fear)' (Moulard, M. Kroff et al. 2019). Hope/fear are defined as follows.

- **Hope** The *positive* emotion experienced by a consumer assessing whether a pleasurable consumption event will occur
    - Game rewards, and uncertainty regarding the potential outcome, contribute towards hope
- **Fear** The *negative* emotion experienced by a consumer assessing whether a painful consumption event will occur
    - Game punishments, and uncertainty regarding the potential outcome, contribute towards fear

Uncertainty is key; the player does not know whether the outcome will or will not occur. Moulard et al also theorise that uncertainty can be driven by fluctuating probabilities (Moulard, M. W. Kroff and Folse 2012), that is, events occurring which change the player's assessed probabilities of the outcome. They suggest that to maximise suspense, one should oscillate the player between high levels of fear, and high levels of hope. Furthermore, if a player hopes and fears at the same time, suspense is theorised to be higher than if the player only hopes or only fears.

5.3.1 Suspense in Horror Games

Perhaps unsurprisingly, research into horror games is where most work on suspense's effects are found in games. Delatorre et al found that players who expected no chance of winning at all (corresponding to no suspense; the outcome is now not unknown) tended to leave the game (Delatorre, Leon et al. 2017). They also show that there is often a frustrating period between the realisation that failure is inevitable, and that failure, termed *The Long Path of Frustration*; this is something that strategy game players complained of also (Gomme and Bartle, Richard 2020).

In another experiment, participants were placed in a virtual environment in which they evaded a killer (Delatorre, León, Salguero, Mateo-Gil et al. 2017). Different groups were given varying amounts of feedback on where the killer might be: full screen visibility (as opposed to occluded by fog), the sound of footsteps, and a directionality indicator. This also noted that suspense was higher when information was hidden from the player, similar to the findings of fog of war being related to tension in *Battle for Wesnoth*.

5.3.2 Differences Between Tension and Suspense

The largest difference is that tension emphasises the agency of the player in a situation; it doesn't simply take into account perceived probability of something happening, it's the player's ability to do something about it, their agency. As player autonomy is so meaningful to players (Richard M. Ryan, Rigby and A. Przybylski 2006), that may be an important distinction, but is yet to be one that is tested fully. While Delatorre et al suggest that decision makers (players) and viewers (their audience) had little variation in reported suspense (Delatorre, León, Salguero, Mateo-Gil et al. 2017), all participants answered on a questionnaire before each threat turn, which may have stunted the pace for the players. One feasible test would be to remove agency from a player while observing physiological correlates to tension.

Moreover, the experiments testing tension are, to the author's knowledge, the only ones that have thus far tested uncertainty's relationship to tension; other studies measure hope or fear directly, but not uncertainty. The author is also unaware of other studies giving evidence that the possibly of the outcome occurring/not occurring appears lie on a spectrum.

5.4 Investment: What Makes a Player Care?

One thing that tension and suspense certainly agree on is the requirement for investment (Gomme and Bartle, Richard 2020): a player's desire for something to occur. In short, care. So what makes a player care, in order for tension to arise?

According to E. Brown and Cairns 2004, it is when a player is *engrossed* — in a state of partial immersion — that they become emotionally affected by a game.[1] For this to arise, the player must have put time, effort, and attention into the game, and it is these that are attributed to a player's emotional attachment. However, factors related to the game itself are more nebulous: the game's 'construction': visuals, interesting tasks, plot. Jennett et al. 2008 go some way towards making immersion more concrete by creating a statistically-validated questionnaire to measure immersion. It measures five factors: cognitive involvement, real world dissociation, emotional involvement, challenge, and control. However, this still doesn't say what *in-game elements* drive emotional involvement. It's also not entirely clear whether emotional involvement helps bring immersion about, or whether immersion is a catalyst for emotional involvement, or both.

Bopp, Mekler and Opwis 2016 highlight several concrete experiences that affect players' emotional response to games: loss; character attachment; agency, or lack thereof; atmosphere; achievement; personal memories. We will start the journey of emotional involvement here, examining each of these and how they might be

[1] I cover immersion as presented by E. Brown and Cairns in section 1.3.4. However, it is explored further later in this chapter

used to further drive player investment or emotional attachment within games. Be warned: this is a long rollercoaster of a journey that seems to touch almost every aspect of gameplay.

5.4.1 Loss

At first glance, this could be an aftermath of tension. Something you were invested in was at risk, and fell on the wrong side of that risk. However, the loss assessed by Bopp, Mekler and Opwis is specifically *loss of a character the player was attached to*, and the player does not even have to be aware this was a possibility, if this loss is a surprise. Thus, the concept examined there is not tension.

Does loss have to be of beloved *characters* to elicit emotional response? It could arguably be broader than that, and might extend to possessions or resources, or anything else cared about by the player. For example, a particular quest in *Fallen London* 2009 brings "Pain like losing half your stats in a single click. Pain like throwing away your Destiny. Pain like sacrificing your hardest earned possessions just for a chance of progressing" (Cobbett 2016).

Should this be the case, there are plenty of opportunities to evoke loss without characters. Give players valuable resources that they might lose: mighty units that fall, chokepoints that are overwhelmed, alchemicals that run dry. However, one could argue that characterisation is what *makes* those moments memorable instead of having them fade into the backdrop. If so, many strategy games and abstract games lose out on a great deal by not taking advantage of this. To quote Crawford 2013, "dramatic power comes from people and their problems, not things". Let us discuss how we might foster character attachment, then.

5.4.2 Character Attachment

Players who reported character attachment as a source of emotional resonance often attributed such attachment to time and effort invested into the game, strengthening the assertions made in immersion's original grounded theory (E. Brown and Cairns 2004). Other players simply stated they liked a given character, or that the bond grew from taking on a character's goals as their own. At the further end of the scale, players 'came to view [the characters] as almost people' (Bopp, Mekler and Opwis 2016), or identified themselves as a character. We have arrived again at a hallmark of immersion, suggesting that 'good' characters assist in immersion.

Yet this doesn't explain much about *how the construction of the characters themselves* engendered such attachment. If it was simply effort and time, players would be as attached every character equally, simply through repeated exposure.[1] It doesn't explain why Shadowheart is the most popular romantic interest

[1] Though this does not preclude players becoming attached to objectively bland characters through time spent with them.

in *Baldur's Gate 3* 2023 (Larian Studios 2024), as opposed to say, Lae'zel, who is introduced to the player at roughly the same time in the game.[1]

Some insights are shed by Bopp, Müller et al. 2019, in which several stereotypes of beloved characters were categorised. We can explore these to find concrete ways of deepening characters in games:

- **Cool and Capable** Characters in this category were most often player characters that gave feelings of power and control, fulfilling player needs of competence and autonomy. They also appeared to have a sense of humour players found entertaining.
  - The humour emphasises that even if one presents players with powerful characters in a game, the tone of writing flavours the character, and makes an indelible mark on player perceptions.
- **Respected Nemesis** Players feel respect and awe towards these characters, usually due to their being charismatic, powerful, and showing virtues despite their opposition to the player.
  - This shows the importance of creating characters that engender complex, and possibly conflicting reactions from the player. This is corroborated by (Coanda and Aupers 2021), which we explore shortly.
- **Admired Paragon** These characters embody virtues player aspire to themselves.
- **Crush** Characters that players became attracted to, usually due to appearance and demeanor.
  - Thus we know that character *appearance* matters significantly; this might explain the recently-mentioned Shadowheart dilemma.
- **Concern for Protégé** Players feel concern for these characters, often due to perceiving themselves as their mentor. Putting such characters in danger, is an excellent way of provoking emotional challenge (Cole and Gillies 2019).
- **Sympathetic Alter Ego** Characters with whom the player feels they share experiences or traits. This can create experiences that resonate highly with individual players, with one of the most standout examples being depiction of psychosis in *Hellblade: Senua's Sacrifice* 2017.
- **Trusted Close Friend** Players connected strongly with these characters, who were consistently loyal or kind to the player.

[1] It is possible that preference for physical features explain some of the playerbase's romantic preference, but that's still an aspect of the character coming into play that isn't simply time or effort.

These categories are stencils we can use to create characters that work for our purposes. Let's say we wish to evoke tension: we wish to increase character investment in an event. We could elicit large feelings of player power by giving them control of a Cool But Capable character. To equal this, we can place a Respected Nemesis in opposition. We could even endanger the safety of a Protégé.

All told, doing so would vastly affect how a player cares about the game, adding to the 'game construction' alluded to by E. Brown and Cairns 2004. There are other proverbial stencils to construct with too; this study focused upon characters that were liked, but hated characters would also presumably leave an impact.

More elementary — and powerful — tools are not the shapes that characters end up taking, but traits that make them resonate powerfully with players irrespective of the stereotype they fall into. Coanda and Aupers interview players about human-like characters, and find several traits that can make a character seem more human to players: moral ambiguity, emotionality, and imperfection (Coanda and Aupers 2021). It's arguable that characters exemplifying the opposite of these traits would be flat and uninteresting.

- **Moral ambiguity** Not being solely good, not being solely evil. Where decisions they make that might be 'evil' are still *understandable*. To quote a participant: "What's dehumanising are the popular Mary Sues or commander Meredith [antagonist of the second Dragon Age game]. Meredith had that look and demeanour that you may as well put on a neon sign saying, 'I'm evil, evil, evil!"'

- **Emotionality** Characters that are able to communicate the emotions that character would feel in response to what's happening to them. This can be difficult; doing so in 3D games often requires excellent voice-acting, or sublime artistry replicating facial or body expressions. Two-dimensional games have fewer tools by which to express such emotion, and often use single detailed portraits of characters in tandem with text describing what they actually say. Text games can express emotion in the same way as conventional literature, to varying degrees of success depending on the quality of the writing.

- **Imperfections** "To err is human" (Pope 1716). Characters who weren't perfect, where their strengths contrasted against their weaknesses and vulnerabilities, were seen as quite human-like.

The characters covered in the interviews are from two prominent franchises — *The Witcher* and *Dragon Age* — in roleplaying games, which is arguably the genre at gaming's bleeding edge of character design. Other genres could learn from them when it comes to characterisation.

These traits also shed some light on why time spent with NPCs might foster attachment towards them. The player gets to see more of the NPC, and more of these moments of imperfection, emotion or ambiguity that resonate. However, this is merely theorising on my part; this could be tested more empirically by having players spend a given amount of time with NPCs that did or did not have these traits, and then interview them on how attached they were to the characters.

5.4.3 Agency

Emotionally moved players in games (Bopp, Mekler and Opwis 2016) also felt 'in control of their own actions'. This makes sense: players have an intrinsic need to feel autonomous (Richard M. Ryan, Rigby and A. Przybylski 2006), and choice and interactivity are the defining feature of games over other media. But how much control do we need to give players to facilitate their emotional involvement while still remaining interactive? In sandbox games such as *Minecraft* 2009, players can do whatever they wish to the game world, though limited to actions supported by the game (for example, try as *Minecraft* players might, they would not be able to turn all the cubes into spheres even in Creative Mode).

However, games that are reliant on challenging the player would not be able to do so if the player were omnipotent; the player must be shackled in some way. Without such shackling, experiences such as tension wouldn't be possible, as there would be no uncertainty. Games that rely on pre-written narrative also cannot allow players to arbitrarily do whatever they wish; the finite amount of narrative could not contain a meaningfully different outcome for each and every action the player might take.

Murray's original definition of agency still holds weight here: "the satisfying power to take meaningful action and see the results of our decisions and choices" (Murray 1997). Cole and Gillies 2019 extend this with *interpretive* agency (freedom for the player to build their own interpretation of the data given to them, the imaginative space a player constructs when they suspend their disbelief (D. W. Brown 2012)), *fictional* agency (agency pertaining to the world, story, narrative, or NPCs; an example might be being able to kill NPCs in *Baldur's Gate 3* 2023 with their absence being notable later in the game) and *mechanical* agency (the range of actions the player has in the game. Movement would be one aspect of this) (Cole and Gillies 2019). This extends our vocabulary of agency that we can grant players. Some games might be heavy with one kind of agency and not others: mechanically, (*Papers, Please* 2013) has only two actions the player can perform: ACCEPT or REJECT a traveller's passport at the border of Arstotzka. Interpretatively however, the player is given a vast amount of room to imagine what could happen should this person be allowed through, whether that person is deserving of the risk you'd take by rejecting, and to make those choices more recklessly because you simply need more money.

Players may perceive themselves as having power over the game world through the agency they're given, but it's not clear whether giving a player more mechanical agency directly translates to a player's *perceived* agency. Thue et al. suggest this isn't the case, stating that players' perceived agency likely scales with choices that could lead to desirable outcomes (Thue et al. 2010). In a later study, they give players the same number of choices, but present one group with more 'desirable' events relating to those choices, and find that players feel more agency in that group (Thue et al. 2011).

This might suggest an interesting relationship between investment in tension, and a player's perception of their own power: that increasing investment would increase a player's perception of their own or their opponent's power. Unfortunately, this relationship has not been tested.

5.4.4 Atmosphere

While we've been focusing on drivers of emotional involvement, this driver of emotional involvement actually loops back around to two of the core components of immersion: cognitive involvement and real world dissociation (Jennett et al. 2008). Atmosphere encompasses compelling depiction of the game world, with visuals and audio coming together to form a 'moving experience'.

Unfortunately, the effects of atmosphere traits on the player experience do not seem to be well-represented in the games literature. The most salient work in the field, however, provides theoretical guidelines for producing joyful and fearful atmospheres through design themes, color, texture, and lighting configuration (Steinhaeusser et al. 2022). These guidelines are concrete and specific, and can help us tweak player emotions. To name a few of the *fear*-inducing guidelines: using cold lighting to dampen pleasant feelings, limiting the players view, or emphasising the boundaries of the environment to empasise confinement.

Through no fault of its own — it is but one study — it does not give us a complete toolbox for atmospheric adjustment: what of wonder, awe, or cosiness? How might one create an atmosphere that lets a player softly let go of their woes? Do lower-fidelity environments impact player immersion? Hints to such answers may have been explored in other fields such as narrative, painting, and cinema (both sound and graphics).

Regardless, it's clear that the games industry itself has deep talent for creating atmosphere: the dense cityscapes of *Cyberpunk 2077* 2020; the vibrant country of *The Witcher 3: Blood and Wine* 2016; the militaristic practicality of *Highfleet* 2021 are all sterling examples. Structured interviews with game environment artists and audio designers, analysed qualitatively, could yield deep insights into this area.

5.4.5 Achievement: Where Should the Balance of Challenge and Skill Reside?

Players who were proud of an achievement in-game remembered the experience fondly (Bopp, Mekler and Opwis 2016). Once more SDT explains this smoothly, competence being a human need (Richard M. Ryan, Rigby and A. Przybylski 2006). This raises a question that has haunted games for a while: what level of challenge is best to present to a given player? Flow theory (Csikszentmihalyi 1990) advocates a 'balance' of difficulty and skill, though does not present where this balance should reside. Is it a 50% chance of overcoming the challenge? Or 30%? Is it the player's *perceived* chance of overcoming the challenge that is key to inducing flow, or being close to achieving one's objective? This has not yet been rigorously tested, and evidence thus far — whilst mounting in favour of more complexity than just 'balanced challenge' — is inconclusive (Cutting et al. 2023).

Discussions of difficulty and skill are quite relevant to tension, but highlight why we have yet to reach a conclusion on the matter of optimal challenge. There are myriad ways of measuring this balance of skill and difficulty. Do we:

- Partition several discrete levels of resources for both player and their opponent in different control groups? This might be termed as *absolute* balance, but it may not be the case that the opposition's resources are directly comparable to the player's.
  - Do those discrete levels of resources even produce different perceptions of difficulty for players?
- Create a dynamic difficulty adjustment that guides the game towards severities of win/draw/loss? This is quite doable in games using a total score, and is the approach Cutting et altake; it successfully controls for player improvement over time without being noticed.
- Take the difference between player and opponent's remaining resources after a game to be the difference between a player's skill and the challenge presented them?
- Ask for subjective measures of challenge and skill from the player?

It is difficult to know which of these is the most valid construct to measure, each having its own merit. Even then, the difference between skill and challenge is defined, what do we then measure to correlate it to? Tension? Players' levels of immersion? How long they were in a flow state? Player-perceived balance? How can we compare studies that measure skill/difficulty in different ways, in relation to different (but quite related) constructs?

5.5 A Knot of Relevant Constructs and Confusingly Similar Terms

On the topic of very similar constructs, one might, while reading, become increasingly overwhelmed or alarmed at the number of different, yet very similar, terms used to describe the player experience. The uses aren't just arbitrary; each often has its own slightly different meaning within the literature that emphasises one particular aspect or another. But the small differences within the smörgåsbord of terms make game experience a very difficult field to navigate.

- Immersion (E. Brown and Cairns 2004) / Presence (Lombard and Ditton 1997) / Flow (Csikszentmihalyi 1990)
  - Immersion and flow both result in attention being focused on the game. But immersion is graded, flow is all-or-nothing. Even then, total immersion is transient, whereas flow can be sustained for a time.
  - Presence is contained in both total immersion and flow.
- Agency (Cole and Gillies 2019) / Control (O'Brien and Toms 2008; Thompson, Armstrong and Thomas 1998) / Effectance (Klimmt, Hartmann and Frey 2007) / Interactivity (Delatorre, León, Salguero, Mateo-Gil et al. 2017) / Autonomy (Richard M. Ryan, Rigby and A. Przybylski 2006)
  - Control is rather similar to agency, but agency has been more consistently defined, and by this point explicitly delineates several forms of choice a given player might make.
  - Effectance is receiving direct and immediate feedback on one's actions.
  - Interactivity is sometimes regarded as 'how *much* agency is given', and sometimes as whether a narrative is interactive at all.
  - Autonomy is used in the context of Self-Determination Theory; a need for autonomy is the need to be able to choose one's actions.
- Environment / Atmosphere / Staging
  - None of these have been formally defined as far as the author is aware, but these terms have all been used to stand in for the same concepts.
- Normative and Empirical expectations (James 2011) / Ideal and Forecast expectations (Higgs, Polonsky and Hollick 2005)
  - These terms mean the same thing. Ideal and Forecast are far more readable terms, but Normative and Empirical expectations are terms more widely used in that literature.
- Suspense (Moulard, M. Kroff et al. 2019) / Tension (Gomme and Bartle, Richard 2020)
  - Both of these regard uncertain outcomes the player is invested in. Tension emphasises another entity's opposition being the source of uncertainty, but the effect this has is as yet unclear.
  - Yes, I too have contributed to this confusion. Hence, it would be useful to test suspense against tension in order to either prove their (useful) difference, or to tuck awareness of tension under the wider umbrella of suspense.

A bevy of measurement tools also allow us to measure many concepts adjacent to game enjoyment: the Challenge Originating from Recent Gameplay (CORGIS) measures multiple forms of challenge in games (Denisova, Cairns et al. 2020), which provides a means for the player to expend effort, a key to immersion; the Player Experience of Needs Satisfaction questionnaire measuring players' intrinsic needs (Richard M. Ryan, Rigby and A. Przybylski 2006); the more holistic but as-yet-untested Basic Needs In Games Scale (Ballou, Denisova et al. 2023); the much-used but quite-flawed Game Experience Questionnaire (Brühlmann and Schmid 2015; IJsselsteijn, De Kort and Poels 2013; Johnson, Gardner and Perry 2018; Law, Brühlmann and Mekler 2018); the Video Game Demand Scale, which measures the extent to which a game 'makes' a player engage (Bowman, Wasserman and Banks 2018). This is a vast array with quite a lot of overlap (Revi, Millard and Middleton 2020).

Koenitz refers to this state of affairs as both Babylonian Confusion and Sisyphonian Tool Production (Koenitz and Eladhari 2019): a lack of shared vocabulary, and an abundance of tools created with similar purpose. This deeply impacts accessibility to the field and consistency within it, but the author is unsure of how to resolve this state of affairs, other than to give a small guide as above.

5.6 Social Agency? Finally Engaging in Dialogue

One last topic that should be examined is a type of agency thus far unavailable to players, which therefore hasn't been named in the literature; we simply haven't had the technology to provide it. We have never been able to converse with AI in games. We've been able to give AI pre-scripted and pre-voiced responses to very limited player options, yes, and have done so heavily in role-playing games. But we've never been able to have an arbitrary back-and-forth with AI, let alone in character.

Jesse Schell refers to such interaction as using 'above-the-neck verbs', and laments players can't really make use of them in games (Jesse Schell 2018). They can't talk, ask, negotiate, shout, plead, complain. At best, a player might be able to choose between two or three options from the above at pre-determined points in a game. Schell isn't the only one to be frustrated at the situation; Chris Crawford left the games industry due to its lack of interest in, and active rejection of at the time, such interaction (Chris Crawford 1992).

At the risk of jumping aboard recent hype, large language models (LLMs) may come to change this. They've provided the strongest avenue of human-computer dialogue to date, and have successfully controlled gameplay actions as well as conversational interaction in small games(Park et al. 2023). However, while those agents displayed excellent emergent behaviour and steerability — one could implant a 'seed' into their 'thoughts', thereby influencing their actions — their dialogue had little personality or flair to it. That's something that would need to be improved upon, if we wanted to use them for player-AI communication.

Should such language models begin to seem like they have a personality, possibly via the traits mentioned in Character Attachment earlier, they may become usable to drive more social kinds of interaction in games. As but one example, AI might negotiate with a player in strategy games, haggling and cursing, calling upon recent events. Other games might have you simply able to sit on a hillside and talk with a character for a while.

Chapter 6

Conclusions

Players come to strategy games with implicit expectations of those they play against. It makes sense: these opponents are often other humans who exert the player's mental capacities, who spin tales of why or how their actions are justified, who react visibly to moves against their plans, who may be coached or coaxed into more or less sensible courses of action. However, not all opponents are human, and AI opponents simply cannot match expectations that they should live up to human standards.

We've identified several expectations that AI fails:

1. Providing a sense of tension in gameplay
2. Playing fairly, by the same rules as the player
3. Providing explanations as to why they performed certain actions
4. Changing in accordance to feedback

These failures may not cause players to put a given game down, but they negatively impact on the player's experience, enough for those players to actively vent about it online. For areas where the AI fails expectations, we provide ludic and narrative tools and suggestions to mitigate the damage or avoid failing those expectations in the first place.

The failing of expectations begins, but does not end, in gameplay. The AI cannot consistently challenge skilled opponents without 'cheating', not with the difficulties that strategy games present to their implementation. Yet players despise such cheating (Gomme and Bartle, Richard 2020), especially if they were unaware of its existence when starting to play the game (Gillin and J. Huizinga 1951). This is a rock and a hard place. Do we ensure that the AI can challenge players by cheating? Or do we make the AI play entirely fairly, yet provide lacklustre strategic opposition? As it stands, the strategy game community grudgingly accepts that such cheating is currently a necessary evil in the genre, but players would certainly like to see its necessity disappear. However, we are not entirely trapped between cheating and underwhelming play, and can use multiple tools to free ourselves from this dichotomy.

6.0.1 Tension

The first is that we don't have to solely think of concrete 'challenge': the amount of actual opposition offered to the player. We can also take into account player *perceptions* of challenge, what the player *believes* their chances of success to be. This gives opportunity to improve player experience without actually expending more resources. Moreso, if a player is challenged according to their skill levels, but in pursuit of an objective they care little about, the player may not be excited at all. Therefore, we've used a metric that's challenge's cousin, that explicitly acknowledges the illusions often presented to the player in games. It's *perceived* challenge for an objective the player *cares about*: we call this tension.

Using tension instead of challenge doesn't magically make the AI more competent at the game — it still cannot consistently oppose the player directly — but by focusing on altering players' perceptions and their current priorities, we can adjust how challenged they feel without forcing the AI to do more heavy lifting. By changing gameplay elements, we suggest several concrete methods that can be used to do this:

1. Manage what is hidden to the player, and what is done, changed, or hinted at in those unknown spaces. This can turn opposition that would otherwise be mundane or completely solvable into something much more engaging. Lovecraft 1938 succinctly gets across the effect this has: *"The oldest and strongest emotion of mankind is fear, and the oldest and strongest kind of fear is fear of the unknown"*. It would be hyperbole to suggest players quake in their boots whenever information is hidden, but the unknown still gives us space to affect players.

   - Fog-of-war is one of the most standout uses of this. It turns *known* opposition into *unknown* opposition.

2. Adjust what matters to the player by changing which resources are more useful than others in particular circumstances, then contest players' acquisition of these resources.

   - For example, if the player's faction requires stone to build units, and another requires gold, the player would have comparatively little interest in acquiring gold themselves. However, add a powerful unit to each faction that requires *both*, and you expand what matters to the player.
     - Depending on the AI technology you're using, the AI may be just as tripped-up by this as the player.

3. Challenge players' familiarity with the game.

    - This is particularly relevant for veteran players, who accurately grok possible outcomes in much of standard play. By introducing mechanics, tweaks or changes that they have not yet encountered, you once again have room to make tension differ from the actual challenge they face.
    - One must also be careful with this: if you break conventions without backing up the changes with narrative, or without ability to *explore* these changes, those same veterans may simply become confused or outright annoyed at the changes.

These are methods by which to adjust tension, but one must be mindful of how much tension you wish for. Current theories on suspense in psychology (Moulard, M. Kroff et al. 2019; Moulard, M. W. Kroff and Folse 2012) suggest that it is best to vary levels of tension over time; even more reason to have a varied toolbox.

However, the most powerful methods to adjust player perceptions without changing the actual opposition they face — the central idea behind using tension instead of challenge — lie in tweaks *outside* of gameplay. With these, we can reframe the player's perspective on actions that the AI take, even if the actions the AI takes remain the same.

Take AI's ability to communicate: most strategy AI cannot effectively communicate *why* it performed — or didn't perform — certain actions in the current situation [1] . Even seasoned AI developers find this information difficult to procure, creating in-house tools to dig through large amounts of data and logic to find possible candidates (John, Gow and Cairns 2019). Players have very little power to do this themselves, and are clearly frustrated when they don't know why actions were performed. They create forum threads to resolve this lack of knowledge, and players often bicker over the possibilities. Such threads can take hours, days, or months to resolve, and sometimes end with AI mod authors coming in with much more detailed explanations. With humans, finding this kind of explanation is much, much easier:

```
Bob: Hey, Steve, why'd you just pillage my farmland?
Steve: Well, I did ask you for the last 75 gold I needed to build some hoplites, and you said -
    to put it politely - no. I've gotta get that gold somewhere.
```

This hints that communication with AI can turn otherwise annoying events into a story that reduces frustration, even if the behaviour itself does not change; you get a response in seconds instead of wondering for weeks. Take a conversation with a work colleague: If they're silent for days on a project, it can be frustrating. But if they send you an email saying "Hey, sorry I've been doing very little lately, I've been under the weather",

[1] "Agendas" and "player relationships" and in *Sid Meier's Civilization VI* 2016 are possibly the closest that we have. Respectively: a known predilection for particular behaviours; and a list of positive or negative modifiers to your relationship with the AI, with a description of that modifier's cause. For example: *-9: We converted their religious city.* This, however, doesn't often explain much more than why the AI is at war with you.

that communication alone can ease the frustration despite thin or incorrect reasoning. If strong reasoning can be provided, that's even better: "My computer needed to do some updates, and it broke a few things in the process".

For an in-game example, imagine an enemy civilization presenting you, the player, with a deal swapping [three of your technologies] for [one of their technologies]. This is frustrating, even if the civ is on 'hostile' terms with you, and it's not a deal you were ever likely to take. However, if the deal comes with the text *"We would have offered a more equitable trade, but since you stole Iowa from us earlier..."*, the deal immediately becomes more understandable, and clearly ties players' prior actions to current consequences: a clear narrative. This narrative might even continue should the player then take the deal and relations return closer to normal.

So communication from the AI to the player already becomes impactful even if neither the game's mechanics or the AI's actions change. Any such form of communication would need work to implement for most AI architectures, but even rudimentary communication would redeem itself by improving player experience no matter which technologies your AI stack uses.

But player→AI communication can also provide large benefits, especially if that communication impacts the AI's play. Instead of simply trading resources, or commanding other civilisations to make war, you can have more fine-grained negotiation, trading *future* finds, bringing someone else into an alliance, or even ceasing a war for a particularly short time while barbarians are at both your gates.

Imagine the following[1]:

```
Player: Hey, could I have 300 gold?
AI: 300 gold?! Okay, on one condition: The next 3 bits of coal you mine go to me, and you don't
    make a settler for the next 10 turns.
Player: Why do you need the coal?
AI: A whole lot of industry.
Player: How about you get the next 2 coal I find instead?
AI: 250 gold now, for the next 2 coal you mine then.
Player: Deal.
```

This is already an improvement over what you can do with current AI, though a more rigid version of this could be implemented without natural language. It gets even more interesting should this deal be non-binding. If the player is then unable to supply the 2 coal, there could be further negotiation and possible sanctions with the AI. If someone the player is at war with takes the coal mines before the player can extract the coal, the AI could then demand *further* debts or sanctions. They might even be able to call a forum with other

[1] I actually tried this with ChatGPT 3.5 (*ChatGPT* 2022). Its responses were long-winded and too willing to be steamrolled in a negotiation, but otherwise the deal was made successfully.

leaders where the player could defend themselves; all things that are possible with human players, and far more difficult to do with AI.

6.0.2 Changing the expectations themselves

Instead of better following players' default expectations (or hopes, if the player knows what to 'expect'), we can instead change those expectations. If an expectation is one we are likely to fail, we can attempt to reframe it. Here are several examples:

- **Providing tension** *This option may not always be implementable in a manner that fits the game's fiction.* In a space game where you play as the last remnants of humanity after the Great AI Overlord suffered catastrophic damage, it might be reasonable to frame 'opponents' as damaged AI fragments that are therefore prone to logic errors. However, in a history-based civilization game, it's difficult to justify every non-player leader harboring a form of madness.

- **Playing by the same rules as the player** This expectation can be changed by outright stating how the AI plays by different rules. *Starcraft II* 2010 goes a ways towards this by naming its most difficult AI levels *Cheater 1 (Vision)*, *Cheater 2 (Resources)*, and *Cheater 3 (Insane)* respectively. *Insane* does not specify specific how it 'cheats', but existence of the previous two levels adequately implies that it *will* be cheating in some form that is more potent than increased vision or resources — either of which is usually considered egregious by players when they discover an AI had been exploiting without their knowledge.

  - Another option is to actually provide different rules for the AI, these rules being less thorny for the AI implementation, yet transparent to the player. For example, providing "beacons" pre-populating the map that allow the AI to witness what goes on around them without being physically present nearby. These beacons might come with a cost or difficulty for the player to destroy. Now, the AI seeing things where the player does not is part of the game itself, something that the player can interact with and structure gameplay around.

- **Explaining actions** *It is likely rather difficult to just 'disappear' this yearning. We may, however, address it differently.* A possible remedy to unexplained and outrageous actions is to embrace that we cannot easily summon a reasonable explanation for those actions. Instead, we give the player something they can do about it. By pressing an easily-accessible "What The Hell?!" button, players might be able to attain closure simply by *officially* acknowledging that behaviour was strange. In this way, we don't have to be able to better explain behaviours, but it could help the players feel less aggrieved about it. As an added bonus, if this button sends a complete information report to the developers, it's likely to help with improving the AI's behaviour.

- **Changing behaviour from player feedback** One could create a public repository of reported AI strangeness, and importantly *connect such reports to any patches where this was fixed.* While this doesn't improve the length of the feedback loop between player-supplied report and developer-supplied change, it provides a much more visible affirmation that change has occurred due to player requests. This is likely to be even more effective if a given player could see the status of *their* submission, and therefore gain a sense of agency over AI behaviour.

Chapter 7

Retrospective

This work has taken a long and winding journey. That journey began firmly in computer science, with the question 'how can I make strategy game AI better?'. It then took a hairpin turn into the wilds of game studies and psychology, as the question morphed into 'what *is* better strategy game AI? What do players actually want?'. After finding out what players care about, this thesis charts a map through much of the game experience literature, and ends with feet in both the ludic and narrative sides of game design, in order to provide, manipulate, or downright sidestep player expectations without necessarily placing further strain on AI.

That's a lot of ground to cover, especially for someone whose initial expertise lay almost entirely in computer science. The original intention was simply to make strategy AI make fewer mistakes. It was to create an AI framework that was more organised, easy to extend, and applicable to complex games than the custom labyrinthine mosaics that often form today's strategy AIs out of necessity. That framework was likely to use a great deal of functional programming to provide the behind-the-scenes architecture that complex game AI could use without reinventing the wheel, while still allowing the flexibility for custom additions. Here is where much of the AI techniques research took place. What techniques were out there? What could work for game AI and be rigged together in an organised manner?

That initial trajectory was derailed by the question 'what *is* better AI?'. Answers to this seemed to either assume *believable* AI, or providing adequate challenge for the player. Moreover, there were relative legions of able and skilled researchers improving AI's scope and performance[1], but few questioning the direction that AI improvements should — or could — be taking. The work done thus far was still useful: the literature review on AI techniques let me effectively demonstrate *why* no current techniques work in isolation for strategy games, and so that review remained in the work.

[1] That is, an AI's capability of 'winning' in a given game, thus setting the bar of the most skilled player that could be adequately challenged by it

To find out what makes good strategy game AI, or at least, AI that players enjoy playing against, consulting the players themselves was a necessary step. The thought of gathering large numbers of players as participants was alien to both my experience and my social capacities at the time, so I turned to a rich and viable source of players' experiences: forums. These were ideal, as they served as a natural venting ground for players' opinions, and had relatively large numbers of existing comments. Yet those comments had to be able to contribute to answering "what makes for good strategy game AI?". To that end, several branches of qualitative research were explored. Though all of these were new to me, grounded theory seemed to do what I was looking for. I did further reading into the topic (Charmaz 2006) to hit the ground running, and started gathering comments.

Once the player expectations emerged from the process, there was a wide range of choice regarding this work's next direction. There were *four* expectations, each of which could have been explored: tension, fair play, transparency of logic, and ability to change. Tension was chosen, as I wanted to see what one could tweak about the how challenged players felt, without actually throwing more at them. However, each of the others was equally valid, and provide strong suggestions as to other possible areas to explore next.

The concept of tension needed field-testing. It was a reasonable theory rooted in some data, but the whole point of this thesis is to give practical methods to avoid failing expectations, and that couldn't be done entirely in the realms of theory. To explore whether tension existed at all as imagined, there needed to be room to affect how much the player *wanted* their objectives to happen, and have possible mechanisms for adjusting tension without throwing more opponents at the player; the proverbial smoke and mirrors that tension makes explicit. Thus were the *Battle for Wesnoth* scenarios born.

While much of tension survived the field-testing unscathed, the results suggested that investment was nigh-unrelated to tension. Was it actually investment that had been measured, or something else? If something else, how could I effectively measure investment instead? This lead to the study on investment specifically, which was, in effect, a shotgun approach: throw players into a scenario, and in the questionnaire afterwards test several variants of asking how invested they were in given objectives, including the original wording used in Gomme and Bartle, Richard 2022. Test the strength of each wording's connection to tension, and see if there was a single variant that worked as originally intended.

That experiment worked, in some ways. It gave good evidence that the phrasing used was a concern, and that several other options fit better. It worked less well in others: the relationship between the perceived powers of the player and enemy became decoupled with no particularly clear reason. It's not that the coupling was weak previously and this was merely the nail in the coffin; the relationship was quite strong. However, the scenario added mechanics to the game allow for 'switching' what players were invested in, and these mechanics seemed to muddy the waters of who had how much power, especially for newer players.

In all, I've learned quite a lot about player expectations. I've gone from programmer to a qualitative neophyte passingly familiar with the wide breadth of games experience research, and slowly, ever so slowly, I might have even learned how I should conduct research going forward.

7.1 Areas of Growth; What I'd Change

This is how I would go about doing this thesis if, at the start, I had the sense I do now.

7.1.1 Make even your data-gathering reproducible

When looking for comments for the initial grounded theory, I used concrete search terms to trawl the forums. Unfortunately, many comments found this way were still irrelevant to AI (for example, the equivalent to 'hear, hear'), and those were discarded.

What I should have done was to include all the comments I found with those search terms in the dataset, *then* mark them as irrelevant, along with any memoes as to why it was done in individual cases.

There were also significant issues with the stability of the software I used to aid in my qualitative analysis: *NVivo* 2018. This immediately imparted the lesson that backing up data is mandatory in life. However, the university *did* back up my data, and *NVivo* still crashed on launch.

7.1.2 Try Not To Switch Contexts

The qualitative analysis in this work — especially the original grounded theory — was grueling, and I worked on the actual coding of data on-off. Unfortunately, the context-switching this entailed, along with the need to refamiliarise myself upon each return to the coding, meant the overall effort was far more of a slog.

7.1.3 Keep It Simple Sometimes

Two of the experiments in this thesis used a full strategy game level to glean their insights. This had some advantages, but it also introduced confounding factors that made it harder to be certain of exactly why something happens. Take the initial scenario built to test tension. This could have been much simpler: say, an online tug-of-war against a virtual opponent, in which the opposition's strength is shown by the amount of stick figures pulling on the opponent's end of the rope. The player could pull against this by mashing a key. Player power could be adjusted by the effectiveness of each keypress. Opponent power could be adjusted by the number stick figures, or for *perceived* opponent power, the number of stick figures could change but the challenge itself remain the same. There could even have been story text beforehand shaping players' expectations.

Players' experience levels would have been far less of a factor. The changes to both perceived and actual power levels would have been much simpler. It's possible that the data would have been clearer, as a result.

7.1.4 Talk With Your Researcher Fellows

I did much of this thesis' work on my own. Frankly, that was a mistake. An easier one for me to make than most: I find reaching out difficult, and there were few fellow IGGI students at Essex. However, that doesn't change the profound impact that working with others had when it *did* occur. Talking with a statistician friend exposed some significant initial issues with my maths in (Gomme and Bartle, Richard 2022). Seeing how they explored the data gave me a serious number of insights in the space of half an hour.

The two times I'd worked with other researchers on papers outside this thesis' direct remit, the process of writing the paper was *much* more sustainable than working alone, and ideas flowed far more freely.

7.1.5 Participant Memory is Flawed

I have learned that simply relying on memory is its own confounding factor. Memory is quite fallible (Ericsson 2006; Loftus 2004), and it may be difficult to tell where participants are filling the gaps. Techniques such as think-aloud (Charters 2003; Eccles and Arsal 2017), while requiring large amounts of transcription afterwards, do much better at capturing the participant's actual stream of thought.[1]

7.1.6 Seek a Guide in Unfamiliar Domains

Being completely new to psychology when starting this research, combined with my lack of reaching out, meant that I simply did not have existing breadcrumb trails towards relevant work that I might otherwise been aware of. My initial literature searches cleaved *close* to some work that was highly relevant — suspense in the psychological literature — yet missed it until a reviewer of one of my papers supplied me with breadcrumbs I hadn't realised existed.

One might claim that such a guide is difficult to find. This is not as true as I thought it at the time. It would have simply taken some asking around either my cohort in other universities, or Essex's psychology department, or asking a librarian for related bodies of work.

7.1.7 Validated Measures Are Gold

Before I'd started this thesis, and for a disturbing amount of time while researching for it, I wasn't even aware that measures could be statistically-validated. Consider me a changed man; if I had known how much

[1] Taking part in a friend's experiment, when quizzed afterwards, I simply could not recall seeing a chest, plotting a route to it, and looting it, despite hints being dropped. Yet in the think-aloud transcript, its presence was clear.

such validation steadies the foundation of research, I would have sought out a statistician to work on such constructs quite fast.

I hope that whoever reads the above can find some fragment of wisdom they can use.

Appendix A

Interception Study Materials

A.1 Questionnaire Content

For questions involving objectives and Likert scores, the participant could choose a point on the Likert scale for each listed objective.

1. Some things to comply with the Data Protection Act, and for your own peace of mind.

   No data collected here is personally identifiable, including savefiles; they only contain the list of actions you took in the scenario, as well as details of the scenario itself. The data collected here will be processed by myself to draw evidence-based conclusions on the nature of tension in games, and the results (hopefully!) published in a paper. I intend to make the data submitted here available to any who ask for it, in the spirit of academic openness.

   That being said, the only data collected here is what happened in your playthrough, and your thoughts about what went on during the scenario.

   The author of this study is Daniel Gomme, and you can contact me at dsdgom@essex.ac.uk if needed.

   Now, to the study itself! It should only take 5-10 minutes, honest.

2. Please upload your savefile if you can - it'll be named something like ___InterceptionXX.gz. You'll find it at:

   Documents\My Games\Wesnoth<version>\saves\ on **Windows**, or

   /.local/share/wesnoth/<version>/saves/ on **Linux**

3. What previous experience did you have with Battle for Wesnoth, by the time you started the scenario?

   a) Extremely familiar with the game

   b) I've played it before
   c) I'm familiar with similar games
   d) I've played through both tutorials
   e) I've played through the first tutorial
   f) I've never played it before

4. Did you find corporal Nobbs?
   a) Yes
   b) No

5. Which was the latest wave you managed to survive until?

   Each new wave zooms to Grem as he summons more units

   a) Starting wave
   b) Wave 2

      Begins with Grem proclaiming *"Looks like the humeys need more of a kickin'... OI! You lot, the ones rollin dem bones. Get in there!"*

   c) Wave 3

      Begins with Grem proclaiming *"If the tin cans won't come to Grem, Grem will go to crush the damn tin cans 'imself..."*

6. Which of the following objectives did you consider to be most important to you during the scenario?
   a) Killing the orcs' leader
   b) Getting to the fortress first
   c) Saving Corporal Nobbs

7. Did your most-important objective change at all during the scenario? If so, why?
   a) Yes [**text field for further input**]
   b) No

8. For the following questions, the option of N/A is available for if you did not notice or attempt achieving the given objective.

Wave 1 begins with Grem shouting "Oooh, more tasty humeys! And they've sent a shiny one too. Here I was thinkin' this was gonna be borin'!"
or Roland commenting about the stench of the Orcs.

Wave 2 with Grem shouting *"Looks like the humeys need more of a kickin'... OI! You lot, the ones rollin dem bones. Get in there!"*

Wave 3 with Grem shouting *"If the tin cans won't come to Grem, Grem will go to crush the damn tin cans 'imself... OI! All of yez, every single one of yez! Time to crush some skulls!"*

9. How important did you feel that it was for you to?

Objectives

- Get to the fort first
- Get to the fort at all
- Kill the orcs' leader
- Save Corporal Nobbs

Likert scale

- Extremely important
- Very important
- Moderately important
- Slightly important
- Not at all important

10. For each following objective, please indicate how certain you were, during the attempt, that you could complete it:

Objectives

- Get to the fort first
- Get to the fort at all
- Kill the orcs' leader
- Save Corporal Nobbs
- Survive the 1st wave
- Survive the 2nd wave
- Survive the 3rd wave

Likert scale

- Certain
- Quite possible
- Moderately possible
- Slightly possible
- Impossible

11. How much power over the following outcomes did you feel you had?

Objectives

- Get to the fort first
- Get to the fort at all
- Kill the orcs' leader
- Save Corporal Nobbs
- Survive the 1st wave
- Survive the 2nd wave
- Survive the 3rd wave

Likert scale

- Complete power
- Much power
- Moderate power
- Little power
- No power

12. How much power do you think the orcs had to stop you achieving the following objectives?

Objectives

- Get to the fort first
- Get to the fort at all
- Kill the orcs' leader
- Save Corporal Nobbs
- Survive the 1st wave
- Survive the 2nd wave
- Survive the 3rd wave

Likert scale

- Complete power
- Much power
- Moderate power
- Little power
- No power

13. How much tension did you feel while trying to complete the following objectives?

Objectives

- Get to the fort first
- Get to the fort at all
- Kill the orcs' leader
- Save Corporal Nobbs
- Survive the 1st wave
- Survive the 2nd wave
- Survive the 3rd wave

Likert scale

- Very tense
- Moderately tense
- A little tense
- No tension

14. Were there any events that made you feel like you had more or less power over achieving any of your objectives? If so, please describe them.

15. Were there any events that made you feel the orcs had more or less power to stop you achieving your objectives? If so, please describe them.

16. Were there any events that made you feel more or less tension? If so, please describe them.

17. Is there anything I haven't covered or asked here that you'd like to raise about the scenario?
    Is something broken? Got an opinion on tension in this yourself? Anything goes, here!

A.2 Interception Scenario Code

The full code for the scenario may be found at `https://github.com/OctarineSourcerer/Interception`

Appendix B

Unto Others Study Materials

B.1 Questionnaire Content

1. What previous experience did you have with Battle for Wesnoth by the time you started the scenario?
    a) Extremely familiar with the game
    b) I've played it before
    c) I've played through both tutorials
    d) I've played through the first tutorial
    e) I've never played it before
2. Did the ritual go critical in your run?
    a) Yes
    b) No
3. Did you...
    a) Attack the mages BEFORE the ritual went critical
    b) Attack the mages AFTER the ritual went critical
    c) Not ever attack the mages
4. Did you attack the Yaltha too after turning on the mages?
    a) Yes
    b) No

5. How important did you feel it was for you to:

   Scale: (Not at all/Slightly/Moderately/Very/Extremely) Important

   a) Find the ritual
   b) Destroy the ritual (before it went critical)
   c) Destroy the Yalthans
   d) Kill the mages (before the ritual went critical)
   e) Kill the mages (after the ritual went critical)

6. How much did you want _____ to occur?

   Not at all/A little/A moderate amount/A Lot/A Great Deal

   a) Find the ritual
   b) Destroy the ritual (before it went critical)
   c) Destroy the Yalthans
   d) Kill the mages (before the ritual went critical)
   e) Kill the mages (after the ritual went critical)

7. How much did you need _____ to occur?

   Not at all/A little/A moderate amount/A lot/A great deal

   a) Find the ritual
   b) Destroy the ritual (before it went critical)
   c) Destroy the Yalthans
   d) Kill the mages (before the ritual went critical)
   e) Kill the mages (after the ritual went critical)

8. How invested were you in _____ occurring?

   (Not at all/A little/Moderately/Very/Incredibly) invested

   a) Find the ritual
   b) Destroy the ritual (before it went critical)
   c) Destroy the Yalthans
   d) Kill the mages (before the ritual went critical)

   e) Kill the mages (after the ritual went critical)

9. How much power over the following outcomes did you feel you had?

   None at all/A little/A moderate amount/A lot/A great deal

   a) Find the ritual
   b) Destroy the ritual (before it went critical)
   c) Destroy the Yalthans
   d) Kill the mages (before the ritual went critical)
   e) Kill the mages (after the ritual went critical)

10. How much power do you think the enemy had to stop you achieving the following objectives?

    None at all/A little/A moderate amount/A lot/A great deal

    a) Find the ritual
    b) Destroy the ritual (before it went critical)
    c) Destroy the Yalthans
    d) Kill the mages (before the ritual went critical)
    e) Kill the mages (after the ritual went critical)

11. How much tension did you feel while trying to complete the following objectives?

    None at all/A little/A moderate amount/A lot/A great deal

    a) Find the ritual
    b) Destroy the ritual (before it went critical)
    c) Destroy the Yalthans
    d) Kill the mages (before the ritual went critical)
    e) Kill the mages (after the ritual went critical)

12. What makes an objective something that you personally want to see completed in a game? Not necessarily in just Battle for Wesnoth.

13. Given such an objective - one you personally want to see completed - how would you describe to another person how you feel about it?

14. Is there anything I haven't covered or asked here that you'd like to raise about the scenario?

B.2 Unto Others Scenario Code

The full code for the scenario may be found at https://github.com/OctarineSourcerer/InvestmentInvestigation

Bibliography

4X Games (18th June 2008). URL: https://web.archive.org/web/20080618074603/http://www.mobygames.com/game-group/4x-games (visited on 26/06/2023).

Adams, Ernest (2014). *Fundamentals of Strategy Game Design*. New Riders.

Add-Ons Server - The Battle for Wesnoth (2006). URL: http://addons.wesnoth.org/ (visited on 24/05/2021).

Against the Storm (1st Nov. 2022). Eremite Games. URL: https://store.steampowered.com/app/1336490/Against_the_Storm/.

AI War: Fleet Command (21st Oct. 2009). Arcen Games.

Alex Vu (23rd Oct. 2018). *A Different Approach to Difficulty*. Game Developer. URL: https://www.gamedeveloper.com/design/a-different-approach-to-difficulty (visited on 25/06/2023).

"An Introduction to Utility Theory" (2014). In: *Game AI Pro*. In collab. with David Graham. Boca Raton: CRC Press/Taylor & Francis Group, p. 122. ISBN: 978-1-4665-6596-8.

Andersen, Simon Calmar and Morten Hjortskov (2016). "Cognitive Biases in Performance Evaluations". In: *Journal of Public Administration Research and Theory* 26.4, pp. 647–662. URL: https://academic.oup.com/jpart/article-abstract/26/4/647/2222571 (visited on 23/03/2024).

Aumann, Robert J. and Jacques H. Dreze (2008). "Rational Expectations in Games". In: *American Economic Review* 98.1, pp. 72–86. URL: https://www.aeaweb.org/articles?id=10.1257/aer.98.1.72 (visited on 25/03/2024).

Baldur's Gate 3 (Aug. 2023). Larian Studios.

Ballou, Nick, Alena Denisova et al. (2023). "The Basic Needs in Games Scale (BANGS): A New Tool for Investigating Positive and Negative Video Game Experiences". In: *PsyArXiv. September* 7. URL: https://osf.io/4965z/download (visited on 16/03/2024).

Ballou, Nick and Sebastian Deterding (29th Sept. 2023). "'I Just Wanted to Get It Over and Done With': A Grounded Theory of Psychological Need Frustration in Video Games". In: *Proceedings of the ACM on Human-Computer Interaction* 7 (CHI PLAY), pp. 217–236. ISSN: 2573-0142. DOI: 10.1145/3611028. URL: https://dl.acm.org/doi/10.1145/3611028 (visited on 16/03/2024).

Barriga, Nicolas, Marius Stanescu and Michael Buro (2017). "Combining Strategic Learning with Tactical Search in Real-Time Strategy Games". In: *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment* 13.1 (1), pp. 9–15. ISSN: 2334-0924. DOI: 10.1609/aiide.v13i1.12922. URL: https://ojs.aaai.org/index.php/AIIDE/article/view/12922 (visited on 25/06/2023).

Barriga, Nicolas A, Marius Stanescu and Michael Buro (2017). "Combining Scripted Behavior with Game Tree Search for Stronger, More Robust Game AI". In: *Game AI Pro 3*, pp. 179–187.

Bartholomew, Kimberley J. et al. (Feb. 2011). "Psychological Need Thwarting in the Sport Context: Assessing the Darker Side of Athletic Experience". In: *Journal of Sport and Exercise Psychology* 33.1, pp. 75–102. ISSN: 0895-2779, 1543-2904. DOI: 10.1123/jsep.33.1.75. URL: https://journals.humankinetics.com/view/journals/jsep/33/1/article-p75.xml (visited on 26/03/2024).

Bartle, Richard (1996). "Hearts, Clubs, Diamonds, Spades: Players Who Suit MUDs". In: *Journal of MUD research* 1.1, p. 19.

— (2001). *Avatar, Character, Persona*. URL: https://mud.co.uk/richard/acp.htm (visited on 15/03/2024).

— (2007). "Presence and Flow: Ill-Fitting Clothes for Virtual Worlds". In: *Techné: Research in Philosophy and Technology* 10.3. ISSN: 1091-8264.

Battle Chess (1988). Interplay Entertainment Corp. URL: https://store.steampowered.com/app/622830/Battle_Chess/.

Baumann, Nicola, Christoph Lürig and Stefan Engeser (Aug. 2016). "Flow and Enjoyment beyond Skill-Demand Balance: The Role of Game Pacing Curves and Personality". In: *Motivation and Emotion* 40.4, pp. 507–519. ISSN: 0146-7239, 1573-6644. DOI: 10.1007/s11031-016-9549-7. URL: http://link.springer.com/10.1007/s11031-016-9549-7 (visited on 15/01/2020).

Behavior Trees (2023). URL: https://docs.unrealengine.com/5.2/en-US/behavior-trees-in-unreal-engine/ (visited on 24/06/2023).

Bethesda Game Studios (11th Nov. 2011). *The Elder Scrolls V: Skyrim*.

Blizzard/S2client-Api (July 2017). Blizzard Entertainment. URL: https://github.com/Blizzard/s2client-api (visited on 08/01/2020).

Bopp, Julia Ayumi, Elisa D. Mekler and Klaus Opwis (7th May 2016). "Negative Emotion, Positive Experience?: Emotionally Moving Moments in Digital Games". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI'16: CHI Conference on Human Factors in Computing Systems. San Jose California USA: ACM, pp. 2996–3006. ISBN: 978-1-4503-3362-7. DOI: 10.1145/2858036.2858227. URL: https://dl.acm.org/doi/10.1145/2858036.2858227 (visited on 26/03/2024).

Bopp, Julia Ayumi, Livia J. Müller et al. (17th Oct. 2019). "Exploring Emotional Attachment to Game Characters". In: *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. CHI

PLAY '19: The Annual Symposium on Computer-Human Interaction in Play. Barcelona Spain: ACM, pp. 313–324. ISBN: 978-1-4503-6688-5. DOI: 10.1145/3311350.3347169. URL: https://dl.acm.org/doi/10.1145/3311350.3347169 (visited on 01/06/2021).

Bopp, Julia Ayumi, Klaus Opwis and Elisa D. Mekler (19th Apr. 2018). ""An Odd Kind of Pleasure": Differentiating Emotional Challenge in Digital Games". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. Montreal QC, Canada: Association for Computing Machinery, pp. 1–12. ISBN: 978-1-4503-5620-6. DOI: 10.1145/3173574.3173615. URL: https://doi.org/10.1145/3173574.3173615 (visited on 17/03/2020).

Borovikov, Igor (2021). "AI-Driven Autoplay Agents for Prelaunch Game Tuning". In: *Game AI Pro – Online Edition 2021*, p. 21.

Bowman, Nicholas David, Joseph Wasserman and Jaime Banks (2018). "Development of the Video Game Demand Scale". In: *Video Games*. Routledge. ISBN: 978-1-351-23526-6.

Braun, Virginia and Victoria Clarke (Jan. 2006). "Using Thematic Analysis in Psychology". In: *Qualitative Research in Psychology* 3.2, pp. 77–101. ISSN: 1478-0887, 1478-0895. DOI: 10.1191/1478088706qp063oa. URL: http://www.tandfonline.com/doi/abs/10.1191/1478088706qp063oa (visited on 20/03/2023).

— (1st Jan. 2012). "Thematic Analysis." In: pp. 57–71. ISBN: 978-1-4338-1003-9.

Brown, Douglas William (2012). "The Suspension of Disbelief in Videogames". In.

Brown, Emily and Paul Cairns (24th Apr. 2004). "A Grounded Investigation of Game Immersion". In: *CHI '04 Extended Abstracts on Human Factors in Computing Systems*. CHI04: CHI 2004 Conference on Human Factors in Computing Systems. Vienna Austria: ACM, pp. 1297–1300. ISBN: 978-1-58113-703-3. DOI: 10.1145/985921.986048. URL: https://dl.acm.org/doi/10.1145/985921.986048 (visited on 14/09/2023).

Browne, Cameron B. et al. (Mar. 2012). "A Survey of Monte Carlo Tree Search Methods". In: *IEEE Transactions on Computational Intelligence and AI in Games* 4.1, pp. 1–43. ISSN: 1943-068X, 1943-0698. DOI: 10.1109/TCIAIG.2012.2186810. URL: http://ieeexplore.ieee.org/document/6145622/ (visited on 02/10/2019).

Brühlmann, Florian and Gian-Marco Schmid (18th Apr. 2015). "How to Measure the Game Experience?: Analysis of the Factor Structure of Two Questionnaires". In: *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. CHI '15: CHI Conference on Human Factors in Computing Systems. Seoul Republic of Korea: ACM, pp. 1181–1186. ISBN: 978-1-4503-3146-3. DOI: 10.1145/2702613.2732831. URL: https://dl.acm.org/doi/10.1145/2702613.2732831 (visited on 22/03/2024).

Carroll, Noël (1984). "Toward a Theory of Film Suspense". In: *Persistence of vision* 1.1, pp. 65–89.

Charmaz, Kathy (2006). *Constructing Grounded Theory*. London ; Thousand Oaks, Calif: Sage Publications. 208 pp. ISBN: 978-0-7619-7352-2 978-0-7619-7353-9.

Charters, Elizabeth (1st July 2003). "The Use of Think-aloud Methods in Qualitative Research An Introduction to Think-aloud Methods". In: *Brock Education Journal* 12.2 (2). ISSN: 2371-7750. DOI: 10.26522/brocked.v12i2.38. URL: https://journals.library.brocku.ca/brocked/index.php/home/article/view/38 (visited on 25/06/2023).

Chaslot, Guillaume, Sander Bakkes et al. (2008). "Monte-Carlo Tree Search: A New Framework for Game AI". In: *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment* 4.1 (1), pp. 216–217. ISSN: 2334-0924. DOI: 10.1609/aiide.v4i1.18700. URL: https://ojs.aaai.org/index.php/AIIDE/article/view/18700 (visited on 25/06/2023).

Chaslot, Guillaume, Jahn-Takeshi Saito et al. (2006). "Monte-Carlo Strategies for Computer Go". In: *Proceedings of the 18th BeNeLux Conference on Artificial Intelligence, Namur, Belgium*, pp. 83–91.

ChatGPT (30th Nov. 2022). OpenAI. URL: https://chat.openai.com.

Chen, Jenova (2007). "Flow in Games (and Everything Else)". In: *Communications of the ACM* 50.4, pp. 31–34.

Chris Crawford, director (1992). *Chris Crawford Dragon Speech "I Had a Dream"*. URL: https://www.youtube.com/watch?v=CBrj4S24074 (visited on 27/03/2024).

Chris Simpson (17th July 2014). *Behavior Trees for AI: How They Work*. Game Developer. URL: https://www.gamedeveloper.com/programming/behavior-trees-for-ai-how-they-work (visited on 01/08/2022).

Civilization V Preview - PC Preview at IGN (28th May 2010). URL: https://web.archive.org/web/20100528124250/http://uk.pc.ign.com/articles/107/1075587p1.html (visited on 25/06/2023).

Clark, Andy (2015). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford University Press. URL: https://books.google.com/books?hl=en&lr=&id=TnqECgAAQBAJ&oi=fnd&pg=PP1&dq=Surfing+Uncertainty:+Prediction,+Action,+and+the+Embodied+Mind&ots=avsg5iGbNO&sig=XPnX0jp5aENyxnO1oDKHWLoZccU (visited on 25/03/2024).

Coanda, Iulia and Stef Aupers (May 2021). "Post-Human Encounters: Humanising the Technological Other in Videogames". In: *New Media & Society* 23.5, pp. 1236–1256. ISSN: 1461-4448, 1461-7315. DOI: 10.1177/1461444820912388. URL: http://journals.sagepub.com/doi/10.1177/1461444820912388 (visited on 27/03/2024).

Cobbett, Richard (20th June 2016). "The RPG Scrollbars: Seeking Mr Eaten's Name". In: *Rock, Paper, Shotgun*. URL: https://www.rockpapershotgun.com/seeking-mr-eatens-name (visited on 26/03/2024).

Cole, Tom, Paul Cairns and Marco Gillies (2015). "Emotional and Functional Challenge in Core and Avant-garde Games". In: *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play -*

CHI PLAY '15. The 2015 Annual Symposium. London, United Kingdom: ACM Press, pp. 121–126. ISBN: 978-1-4503-3466-2. DOI: 10.1145/2793107.2793147. URL: http://dl.acm.org/citation.cfm?doid=2793107.2793147 (visited on 01/10/2019).

Cole, Tom and Marco Gillies (23rd Oct. 2019). "Thinking and Doing: Challenge, Agency, and the Eudaimonic Experience in Video Games". In: *Games and Culture*, p. 155541201988153. ISSN: 1555-4120, 1555-4139. DOI: 10.1177/1555412019881536. URL: http://journals.sagepub.com/doi/10.1177/1555412019881536 (visited on 14/03/2020).

Core War the Ultimate Programming Game (1984). URL: https://corewar.co.uk (visited on 25/06/2023).

Costikyan, Greg (2013). *Uncertainty in Games*. Mit Press.

Crawford, Chris (2013). "Interactive Storytelling". In: *The Video Game Theory Reader*. Routledge, pp. 259–273. URL: https://www.taylorfrancis.com/chapters/edit/10.4324/9780203700457-13/interactive-storytelling-chris-crawford (visited on 26/03/2024).

Creative Assembly (2023). *Total War Games*. URL: https://www.totalwar.com/games/ (visited on 08/11/2023).

Crusader Kings III (1st Sept. 2020). Paradox Interactive. URL: https://www.paradoxinteractive.com/games/crusader-kings-iii.

Csikszentmihalyi, Mihaly (1990). "Flow: The Psychology of Optimal Experience". In: *Design Issues* 8.1, pp. 75–77.

— (2014). *Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi*. Springer Netherlands. ISBN: 978-94-017-9088-8. URL: http://ebookcentral.proquest.com/lib/york-ebooks/detail.action?docID=1802897 (visited on 26/06/2023).

Cutting, Joe et al. (Feb. 2023). "Difficulty-Skill Balance Does Not Affect Engagement and Enjoyment: A Pre-Registered Study Using Artificial Intelligence-Controlled Difficulty". In: *Royal Society Open Science* 10.2, p. 220274. ISSN: 2054-5703. DOI: 10.1098/rsos.220274. URL: https://royalsocietypublishing.org/doi/10.1098/rsos.220274 (visited on 07/02/2023).

Cyberpunk 2077 (Dec. 2020). CD Projekt RED.

Daniel Gomme and Richard Bartle (June 2023). "Examining Player Investment as Related to Tension in Strategy Games". In.

Darkest Dungeon (Jan. 2016). Red Hook Studios.

Deci, Edward L and Richard M Ryan (2012). "Self-Determination Theory." In.

Delatorre, Pablo, Carlos Leon et al. (2017). "The Long Path of Frustration: A Case Study with Dead by Daylight". In: p. 12.

Delatorre, Pablo, Carlos León, Alberto Salguero, Cristina Mateo-Gil et al. (2017). "Impact of Interactivity on Information Management for Suspense in Storytelling". In: *Proceedings of AISB*, pp. 110–116.

Delatorre, Pablo, Carlos León, Alberto Salguero, Manuel Palomo-Duarte et al. (2018). "Confronting a Paradox: A New Perspective of the Impact of Uncertainty in Suspense". In: *Frontiers in Psychology* 9. ISSN: 1664-1078. URL: https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01392 (visited on 28/07/2022).

Denisova, Alena and Paul Cairns (2015). "The Placebo Effect in Digital Games: Phantom Perception of Adaptive Artificial Intelligence". In: *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*. The 2015 Annual Symposium. London, United Kingdom: ACM Press, pp. 23–33. ISBN: 978-1-4503-3466-2. DOI: 10.1145/2793107.2793109. URL: http://dl.acm.org/citation.cfm?doid=2793107.2793109 (visited on 12/12/2019).

Denisova, Alena, Paul Cairns et al. (May 2020). "Measuring Perceived Challenge in Digital Games: Development & Validation of the Challenge Originating from Recent Gameplay Interaction Scale (CORGIS)". In: *International Journal of Human-Computer Studies* 137, p. 102383. ISSN: 10715819. DOI: 10.1016/j.ijhcs.2019.102383. URL: https://linkinghub.elsevier.com/retrieve/pii/S1071581919301491 (visited on 15/03/2020).

Denisova, Alena, Christian Guckelsberger and David Zendle (2017). "Challenge in Digital Games: Towards Developing a Measurement Tool". In: *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '17*. The 2017 CHI Conference Extended Abstracts. Denver, Colorado, USA: ACM Press, pp. 2511–2519. ISBN: 978-1-4503-4656-6. DOI: 10.1145/3027063.3053209. URL: http://dl.acm.org/citation.cfm?doid=3027063.3053209 (visited on 12/03/2020).

Deterding, Sebastian et al. (2022). "Mastering Uncertainty: A Predictive Processing Account of Enjoying Uncertain Success in Video Game Play". In: *Frontiers in Psychology* 13. ISSN: 1664-1078. URL: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.924953 (visited on 18/10/2022).

Difficulty Level (Civ6) (2023). Civilization Wiki. URL: https://civilization.fandom.com/wiki/Difficulty_level_(Civ6) (visited on 25/06/2023).

Discworld (1995). Teeny Weeny Games.

Dor, Simon (Apr. 2018). "Strategy in Games or Strategy Games: Dictionary and Encyclopaedic Definitions for Game Studies". In: *Game Studies* 18.1. ISSN: 1604-7982. URL: https://gamestudies.org/1801/articles/simon_dor (visited on 26/06/2023).

Eccles, David W. and Güler Arsal (8th Aug. 2017). "The Think Aloud Method: What Is It and How Do I Use It?" In: *Qualitative Research in Sport, Exercise and Health* 9.4, pp. 514–531. ISSN: 2159-676X, 2159-6778. DOI: 10.1080/2159676X.2017.1331501. URL: https://www.tandfonline.com/doi/full/10.1080/2159676X.2017.1331501 (visited on 12/03/2024).

Endless Space 2 (18th May 2017). Amplitude Studios. URL: https://www.endless-space.com/.

Endless Space 2 General Discussions (27th Apr. 2019). *Endless Space 2 General Discussions :: Steam Community*. URL: https://web.archive.org/web/20190427020932/https://steamcommunity.com/app/392110/discussions/ (visited on 26/06/2023).

Ericsson, K. Anders (26th June 2006). "Protocol Analysis and Expert Thought: Concurrent Verbalizations of Thinking during Experts' Performance on Representative Tasks". In: *The Cambridge Handbook of Expertise and Expert Performance*. Ed. by K. Anders Ericsson et al. 1st ed. Cambridge University Press, pp. 223–242. ISBN: 978-0-521-84097-2 978-0-521-60081-1 978-0-511-81679-6. DOI: 10.1017/CBO9780511816796.013. URL: https://www.cambridge.org/core/product/identifier/CBO9780511816796A025/type/book_part (visited on 12/03/2024).

Factions (2023). Endless Space 2 Wiki. URL: https://endless-space-2.fandom.com/wiki/Category:Factions (visited on 25/06/2023).

Fallen London (2009). Failbetter Games. URL: https://www.fallenlondon.com.

Favero, Nathan and Minjung Kim (17th June 2021). "Everything Is Relative: How Citizens Form and Use Expectations in Evaluating Services". In: *Journal of Public Administration Research and Theory* 31.3, pp. 561–577. ISSN: 1053-1858, 1477-9803. DOI: 10.1093/jopart/muaa048. URL: https://academic.oup.com/jpart/article/31/3/561/5973537 (visited on 24/03/2024).

Fishburn, Peter C (1970). *Utility Theory for Decision Making*. Research analysis corp McLean VA.

GDC 2005 Proceeding (2019). *GDC 2005 Proceeding: Handling Complexity in the Halo 2 AI*. URL: https://www.gamasutra.com/view/feature/130663/gdc_2005_proceeding_handling_.php (visited on 17/08/2019).

Gill, Arthur et al. (1962). "Introduction to the Theory of Finite-State Machines". In.

Gillin, John L. and J. Huizinga (Apr. 1951). "Homo Ludens: A Study of the Play-Element in Culture." In: *American Sociological Review* 16.2, p. 274. ISSN: 00031224. DOI: 10.2307/2087716. JSTOR: 2087716. URL: http://www.jstor.org/stable/2087716?origin=crossref (visited on 31/10/2018).

Gomme, Daniel and Bartle, Richard (2020). "Strategy Games: The Components of A Worthy Opponent". In: *FDG 2020*.

— (2022). "Tools To Adjust Tension And Suspense In Strategy Games: An Investigation". In: Submitted.

Grand Theft Auto IV (2008). Rockstar Games.

Grimmelikhuijsen, Stephan and Gregory A. Porumbescu (21st Oct. 2017). "Reconsidering the Expectancy Disconfirmation Model. Three Experimental Replications". In: *Public Management Review* 19.9, pp. 1272–1292. ISSN: 1471-9037, 1471-9045. DOI: 10.1080/14719037.2017.1282000. URL: https://www.tandfonline.com/doi/full/10.1080/14719037.2017.1282000 (visited on 23/03/2024).

Halo 2 (2004). Bungie.

Harnad, Stevan (1st June 1990). “The Symbol Grounding Problem”. In: *Physica D: Nonlinear Phenomena* 42.1, pp. 335–346. ISSN: 0167-2789. DOI: 10.1016/0167-2789(90)90087-6. URL: https://www.sciencedirect.com/science/article/pii/0167278990900876 (visited on 29/06/2022).

Hellblade: Senua's Sacrifice (2017). Ninja Theory.

Heroes of Might and Magic: A Strategic Quest (Sept. 1995). New World Computing.

Higgs, Bronwyn, Michael Jay Polonsky and Mary Hollick (1st Jan. 2005). “Measuring Expectations: Forecast vs. Ideal Expectations. Does It Really Matter?” In: *Journal of Retailing and Consumer Services* 12.1, pp. 49–64. ISSN: 0969-6989. DOI: 10.1016/j.jretconser.2004.02.002. URL: https://www.sciencedirect.com/science/article/pii/S0969698904000153 (visited on 24/03/2024).

Highfleet (July 2021). Konstantin Koshutin.

Hjortskov, Morten (Apr. 2019). “Citizen Expectations and Satisfaction Over Time: Findings From a Large Sample Panel Survey of Public School Parents in Denmark”. In: *The American Review of Public Administration* 49.3, pp. 353–371. ISSN: 0275-0740, 1552-3357. DOI: 10.1177/0275074018765822. URL: http://journals.sagepub.com/doi/10.1177/0275074018765822 (visited on 24/03/2024).

Hoff, Jonatan Wilhelm and Hallvard Jore Christensen (2016). “Evolving Behaviour Trees:-Automatic Generation of AI Opponents for Real-Time Strategy Games”. NTNU.

Holland, John H (1992). *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. MIT press.

Holm, Sture (1979). “A Simple Sequentially Rejective Multiple Test Procedure”. In: *Scandinavian Journal of Statistics* 6.2, pp. 65–70. JSTOR: 4615733. URL: http://www.jstor.org/stable/4615733.

Huizinga, Johan (2008). *Homo Ludens: Proeve Eener Bepaling van Het Spel-Element Der Cultuur*. Amsterdam University Press.

Hunicke, Robin (2005). “The Case for Dynamic Difficulty Adjustment in Games”. In: *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*. ACM, pp. 429–433.

Ierusalimschy, Roberto, Luiz Henrique De Figueiredo and Waldemar Celes Filho (1996). “Lua—an Extensible Extension Language”. In: *Software: Practice and Experience* 26.6, pp. 635–652.

IJsselsteijn, Wijnand A., Yvonne AW De Kort and Karolien Poels (2013). “The Game Experience Questionnaire”. In: URL: https://research.tue.nl/en/publications/the-game-experience-questionnaire (visited on 27/03/2024).

Interception, an Experiment on Tension - The Battle for Wesnoth Forums (2021). URL: https://forums.wesnoth.org/viewtopic.php?p=658622 (visited on 10/06/2021).

Interception (20th Aug. 2020). *Interception: An Experiment*. OctarineSourcerer. URL: https://octarinesourcerer.co.uk/interception-an-experiment/ (visited on 10/06/2021).

James, Oliver (2011). "Managing Citizens' Expectations of Public Service Performance: Evidence from Observation and Experimentation in Local Government". In: *Public Administration* 89.4, pp. 1419–1435. ISSN: 1467-9299. DOI: 10.1111/j.1467-9299.2011.01962.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9299.2011.01962.x (visited on 24/03/2024).

Jennett, Charlene et al. (1st Sept. 2008). "Measuring and Defining the Experience of Immersion in Games". In: *International Journal of Human-Computer Studies* 66.9, pp. 641–661. ISSN: 1071-5819. DOI: 10.1016/j.ijhcs.2008.04.004. URL: https://www.sciencedirect.com/science/article/pii/S1071581908000499 (visited on 02/03/2021).

Jesse Schell (10th Aug. 2018). "The Future of Storytelling: How Medium Shapes Story" (GDC 2013). URL: https://www.youtube.com/watch?v=BjrO-di22v8 (visited on 27/03/2024).

John, Nathan, Jeremy Gow and Paul Cairns (2019). "Why Is Debugging Video Game AI Hard?" In: *Proc. AISB AI & Games Symposium*, pp. 20–24.

Johnson, Daniel (Nov. 2015). "Animated Frustration or the Ambivalence of Player Agency". In: *Games and Culture* 10.6, pp. 593–612. ISSN: 1555-4120, 1555-4139. DOI: 10.1177/1555412014567229. URL: http://journals.sagepub.com/doi/10.1177/1555412014567229 (visited on 15/03/2020).

Johnson, Daniel, M. John Gardner and Ryan Perry (1st Oct. 2018). "Validation of Two Game Experience Scales: The Player Experience of Need Satisfaction (PENS) and Game Experience Questionnaire (GEQ)". In: *International Journal of Human-Computer Studies* 118, pp. 38–46. ISSN: 1071-5819. DOI: 10.1016/j.ijhcs.2018.05.003. URL: https://www.sciencedirect.com/science/article/pii/S1071581918302337 (visited on 22/03/2024).

Jordan Baranowski (19th Nov. 2018). *Games You Didn't Know Featured Dynamic Difficulty*. SVG. URL: https://www.svg.com/138490/games-you-didnt-know-featured-dynamic-difficulty/ (visited on 25/06/2023).

Klimmt, Christoph, Tilo Hartmann and Andreas Frey (Dec. 2007). "Effectance and Control as Determinants of Video Game Enjoyment". In: *CyberPsychology & Behavior* 10.6, pp. 845–848. ISSN: 1094-9313, 1557-8364. DOI: 10.1089/cpb.2007.9942. URL: http://www.liebertpub.com/doi/10.1089/cpb.2007.9942 (visited on 09/05/2022).

Klimmt, Christoph, Albert Rizzo et al. (Feb. 2009). "Experimental Evidence for Suspense as Determinant of Video Game Enjoyment". In: *CyberPsychology & Behavior* 12.1, pp. 29–31. ISSN: 1094-9313. DOI: 10.1089/cpb.2008.0060. URL: https://www.liebertpub.com/doi/abs/10.1089/cpb.2008.0060 (visited on 09/05/2022).

Koenitz, Hartmut and Mirjam Palosaari Eladhari (2019). "Challenges of IDN Research and Teaching". In: *Interactive Storytelling*. Ed. by Rogelio E. Cardona-Rivera, Anne Sullivan and R. Michael Young. Vol. 11869. Cham: Springer International Publishing, pp. 26–39. ISBN: 978-3-030-33893-0 978-3-030-33894-7. DOI: 10.1007/978-3-030-33894-7_4. URL: https://link.springer.com/10.1007/978-3-030-33894-7_4 (visited on 22/03/2024).

Koster, Raph (2013). *Theory of Fun for Game Design*. " O'Reilly Media, Inc.". URL: https://books.google.com/books?hl=en&lr=&id=3TAKAgAAQBAJ&oi=fnd&pg=PT3&dq=raph+koster+fun&ots=EMX1fny7QT&sig=A3mXEe_xYCYt0a_JpuMeUcYj9G8 (visited on 16/03/2024).

Krishnan, Sanjay et al. (21st Apr. 2016). *HIRL: Hierarchical Inverse Reinforcement Learning for Long-Horizon Tasks with Delayed Rewards*. arXiv: 1604.06508 [cs]. URL: http://arxiv.org/abs/1604.06508 (visited on 17/02/2023). preprint.

Kruglanski, Arie W and Shira Fishman (2009). "The Need for Cognitive Closure". In: *Handbook of individual differences in social behavior*, pp. 343–353.

Larian Studios (2024). *Since Our Launch in August, over 1.3 Million Players Have Completed Baldur's Gate 3. A Number Almost Equaled by Those of You Transformed into a Wheel of Cheese. All This, and More in Our Latest Thread of Amazing Stats! https://t.co/38hSZw1wPk*. Twitter. URL: https://twitter.com/larianstudios/status/1732091568243229159 (visited on 26/03/2024).

Law, Effie L.-C., Florian Brühlmann and Elisa D. Mekler (23rd Oct. 2018). "Systematic Review and Validation of the Game Experience Questionnaire (GEQ) - Implications for Citation and Reporting Practice". In: *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. CHI PLAY '18: The Annual Symposium on Computer-Human Interaction in Play. Melbourne VIC Australia: ACM, pp. 257–270. ISBN: 978-1-4503-5624-4. DOI: 10.1145/3242671.3242683. URL: https://dl.acm.org/doi/10.1145/3242671.3242683 (visited on 22/03/2024).

Left 4 Dead 2 (17th Nov. 2009). Valve. URL: https://store.steampowered.com/app/550/Left_4_Dead_2/.

Lim, Chong-U, Robin Baumgarten and Simon Colton (2010). "Evolving Behaviour Trees for the Commercial Game DEFCON". In: *Applications of Evolutionary Computation*. Ed. by Cecilia Di Chio et al. Red. by David Hutchison et al. Vol. 6024. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 100–110. ISBN: 978-3-642-12238-5 978-3-642-12239-2. DOI: 10.1007/978-3-642-12239-2_11. URL: http://link.springer.com/10.1007/978-3-642-12239-2_11 (visited on 31/08/2022).

Loftus, Elizabeth F. (Aug. 2004). "Memories of Things Unseen". In: *Current Directions in Psychological Science* 13.4, pp. 145–147. ISSN: 0963-7214, 1467-8721. DOI: 10.1111/j.0963-7214.2004.00294.x. URL: http://journals.sagepub.com/doi/10.1111/j.0963-7214.2004.00294.x (visited on 12/03/2024).

Lombard, Matthew and Theresa Ditton (1st Sept. 1997). “At the Heart of It All: The Concept of Presence”. In: *Journal of Computer-Mediated Communication* 3.2, JCMC321. ISSN: 1083-6101. DOI: 10.1111/j.1083-6101.1997.tb00072.x. URL: https://doi.org/10.1111/j.1083-6101.1997.tb00072.x (visited on 15/03/2024).

Lovecraft, Howard Phillips (1938). *Supernatural Horror in Literature*. The Palingenesis Project (Wermod and Wermod Publishing Group). URL: https://books.google.com/books?hl=en&lr=&id=FYmTAwAAQBAJ&oi=fnd&pg=PR1&dq=Supernatural+Horror+in+Literature&ots=9lnZeTeGJf&sig=Uu78zyZXprYtIOh-3J213P6ZUm8 (visited on 05/03/2024).

MacQueen, Kathleen M. et al. (May 1998). “Codebook Development for Team-Based Qualitative Analysis”. In: *CAM Journal* 10.2, pp. 31–36. ISSN: 1087-822X. DOI: 10.1177/1525822X980100020301. URL: http://journals.sagepub.com/doi/10.1177/1525822X980100020301 (visited on 15/03/2020).

Madrigal, Robert, Colleen Bee and Johnny Chen (18th May 2022). “Hope and Fear in the Experience of Suspense”. In: *Cognition and Emotion*, pp. 1–19. ISSN: 0269-9931, 1464-0600. DOI: 10.1080/02699931.2022.2075327. URL: https://www.tandfonline.com/doi/full/10.1080/02699931.2022.2075327 (visited on 23/10/2022).

Manabe, Kazuko and Youichiro Miyake (2021). “Game Balancing Using Genetic Algorithms to Generate Player Agents”. In: *Game AI Pro - 2021 Edition*, p. 14. URL: http://www.gameaipro.com/GameAIProOnlineEdition2021/GameAIProOnlineEdition2021_Chapter17_Game_Balancing_using_Genetic_Algorithms_to_Generate_Player_Agents.pdf.

Mao (2023). BoardGameGeek. URL: https://boardgamegeek.com/boardgame/4213/mao (visited on 25/06/2023).

Master of Orion II: Battle at Antares (1996). Simtex.

McMullen, Chris (13th Apr. 2021). *Skyrim’s Bucket Robbery Is a Testament to Player Creativity*. The Escapist. URL: https://www.escapistmagazine.com/skyrim-bucket-robbery-is-a-testament-to-player-creativity/ (visited on 15/03/2024).

MicroRTS (19th June 2023). Farama Foundation. URL: https://github.com/Farama-Foundation/MicroRTS (visited on 25/06/2023).

Minecraft (2009). Mojang Studios.

Modding (Civ6) (2021). Civilization Wiki. URL: https://civilization.fandom.com/wiki/Modding_(Civ6) (visited on 26/05/2021).

Moulard, Julie Guidry, Michael Kroff et al. (2nd Sept. 2019). “The Role of Suspense in Gaming: Inducing Consumers’ Game Enjoyment”. In: *Journal of Interactive Advertising* 19.3, pp. 219–235. ISSN: 1525-2019.

DOI: 10.1080/15252019.2019.1689208. URL: https://www.tandfonline.com/doi/full/10.1080/15252019.2019.1689208 (visited on 09/05/2022).

Moulard, Julie Guidry, Michael W. Kroff and Judith Anne Garretson Folse (Mar. 2012). “Unraveling Consumer Suspense: The Role of Hope, Fear, and Probability Fluctuations”. In: *Journal of Business Research* 65.3, pp. 340–346. ISSN: 01482963. DOI: 10.1016/j.jbusres.2011.04.011. URL: https://linkinghub.elsevier.com/retrieve/pii/S0148296311001287 (visited on 09/05/2022).

Murray, Janet (1997). “Hamlet on the Holodeck”. In: *Murray presents views to computer-based and interactive narratives and discusses notions of procedural authoring, immersion and agency.*

Nass, Clifford and Youngme Moon (Jan. 2000). “Machines and Mindlessness: Social Responses to Computers”. In: *Journal of Social Issues* 56.1, pp. 81–103. ISSN: 0022-4537, 1540-4560. DOI: 10.1111/0022-4537.00153. URL: http://doi.wiley.com/10.1111/0022-4537.00153 (visited on 26/10/2018).

Nave, Kathryn et al. (2020). “Wilding the Predictive Brain”. In: *WIREs Cognitive Science* 11.6, e1542. ISSN: 1939-5086. DOI: 10.1002/wcs.1542. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/wcs.1542 (visited on 16/03/2024).

Nethack (28th July 1987). URL: github.com/NetHack/NetHack.

NVivo (2018). Version 12. Lumivero. URL: https://lumivero.com/products/nvivo/.

O’Brien, Heather L. and Elaine G. Toms (2008). “What Is User Engagement? A Conceptual Framework for Defining User Engagement with Technology”. In: *Journal of the American Society for Information Science and Technology* 59.6, pp. 938–955. ISSN: 1532-2890. DOI: 10.1002/asi.20801. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.20801 (visited on 25/03/2024).

Old World (19th May 2022). Mohawk Games. URL: https://store.steampowered.com/app/597180/Old_World/.

Oliver, Richard L. (2010). “Customer Satisfaction”. In: *Wiley International Encyclopedia of Marketing*. John Wiley & Sons, Ltd. ISBN: 978-1-4443-1656-8. DOI: 10.1002/9781444316568.wiem03008. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781444316568.wiem03008 (visited on 23/03/2024).

Orteil (2013). *Cookie Clicker*. URL: https://cookieclicker.com/.

Pacheco, Cristiana, Laurissa Tokarchuk and Diego Pérez-Liébana (2018). “Studying Believability Assessment in Racing Games”. In: *Proceedings of the 13th International Conference on the Foundations of Digital Games* (Malmö, Sweden). FDG ’18. New York, NY, USA: ACM, 20:1–20:10. ISBN: 978-1-4503-6571-0. DOI: 10.1145/3235765.3235797. URL: http://doi.acm.org/10.1145/3235765.3235797 (visited on 14/08/2019).

Papers, Please (Aug. 2013). 3909 LLC.

Park, Joon Sung et al. (6th Apr. 2023). *Generative Agents: Interactive Simulacra of Human Behavior*. arXiv: 2304.03442 [cs]. URL: http://arxiv.org/abs/2304.03442 (visited on 04/05/2023). preprint.

Pope, Alexander (1716). *An Essay on Criticism.*. For Bernard Lintot. URL: https://books.google.com/books?hl=en&lr=&id=p6RfAAAAcAAJ&oi=fnd&pg=PA5&dq=alexander+pope+an+essay+on+criticism&ots=MpJ7z2Own0&sig=mtpcMZQkz3-M6ML_qKB2hyz9OrA (visited on 27/03/2024).

PowerWash Simulator (14th July 2022). FuturLab. URL: https://powerwash-simulator.en.softonic.com/.

Przybylski, Andrew K., C. Scott Rigby and Richard M. Ryan (June 2010). "A Motivational Model of Video Game Engagement". In: *Review of General Psychology* 14.2, pp. 154–166. ISSN: 1089-2680, 1939-1552. DOI: 10.1037/a0019440. URL: http://journals.sagepub.com/doi/10.1037/a0019440 (visited on 25/09/2019).

Punch-Out!! (15th Aug. 1990). Nintendo R&D3.

Puterman, Martin L. (1st Jan. 1990). "Chapter 8 Markov Decision Processes". In: *Handbooks in Operations Research and Management Science*. Vol. 2. Stochastic Models. Elsevier, pp. 331–434. DOI: 10.1016/S0927-0507(05)80172-0. URL: https://www.sciencedirect.com/science/article/pii/S0927050705801720 (visited on 25/06/2023).

Qualtrics (2021). Qualtrics XM. URL: www.qualtrics.com (visited on 25/01/2021).

Quieter/Sit down Games (24th May 2012). Diary of a Volunteer Youth Leader. URL: https://diaryofavolunteeryouthleader.wordpress.com/resources/games-ideas-and-practical-youthwork-stuff/quietersit-down-games/ (visited on 25/06/2023).

Raymond, Eric (Apr. 2009). *Battle For Wesnoth Campaign Design How-to*. URL: http://catb.org/~esr/wesnoth/campaign-design-howto.html#_map_composition.

Revi, Ashwathy T., David E. Millard and Stuart E. Middleton (2020). "A Systematic Analysis of User Experience Dimensions for Interactive Digital Narratives". In: *Interactive Storytelling*. Ed. by Anne-Gwenn Bosser, David E. Millard and Charlie Hargood. Vol. 12497. Cham: Springer International Publishing, pp. 58–74. ISBN: 978-3-030-62515-3 978-3-030-62516-0. DOI: 10.1007/978-3-030-62516-0_5. URL: http://link.springer.com/10.1007/978-3-030-62516-0_5 (visited on 27/03/2024).

Richard Bartle and Roy Trubshaw (1985). *MUD2*.

Roadwarden (12th Sept. 2022). Moral Anxiety Studeo. URL: https://store.steampowered.com/app/1155970/Roadwarden/.

Robbins, Michael (2019). "Using Neural Networks to Control Agent Threat Response". In: *Game AI pro 360: Guide to Tactics and Strategy*. CRC Press, pp. 55–64.

Roch, Christine H. and Theodore H. Poister (Jan. 2006). “Citizens, Accountability, and Service Satisfaction: The Influence of Expectations”. In: *Urban Affairs Review* 41.3, pp. 292–308. ISSN: 1078-0874, 1552-8332. DOI: 10.1177/1078087405281124. URL: http://journals.sagepub.com/doi/10.1177/1078087405281124 (visited on 24/03/2024).

Roelofs, Gijs-Jan (2017). “Pitfalls and Solutions When Using Monte Carlo Tree Search for Strategy and Tactical Games”. In: *Game AI Pro 3*, pp. 343–354. URL: http://www.gameaipro.com/GameAIPro3/GameAIPro3_Chapter28_Pitfalls_and_Solutions_When_Using_Monte_Carlo_Tree_Search_for_Strategy_and_Tactical_Games.pdf.

Ryan, Richard M., C. Scott Rigby and Andrew Przybylski (12th Dec. 2006). “The Motivational Pull of Video Games: A Self-Determination Theory Approach”. In: *Motivation and Emotion* 30.4, pp. 344–360. ISSN: 0146-7239, 1573-6644. DOI: 10.1007/s11031-006-9051-8. URL: http://link.springer.com/10.1007/s11031-006-9051-8 (visited on 25/09/2019).

Salmond, Michael (7th Mar. 2019). *Video Game Design: Principles and Practices from the Ground Up*. Reprint edition. Bloomsbury Academic USA. 272 pp. ISBN: 978-1-5013-5482-3.

Schubert, Damion (31st May 2005). *You Don’t Want Realistic A.I.* URL: http://www.zenofdesign.com/you-dont-want-realistic-a-i/.

Scott Rigby and Richard Ryan (16th Jan. 2007). *Rethinking Carrots: A New Method For Measuring What Players Find Most Rewarding and Motivating About Your Game*. Gamasutra. URL: https://www.gamedeveloper.com/design/rethinking-carrots-a-new-method-for-measuring-what-players-find-most-rewarding-and-motivating-about-your-game.

Sekiro: Shadows Die Twice (Mar. 2019). FromSoftware.

Shani, Guy, Joelle Pineau and Robert Kaplow (July 2013). “A Survey of Point-Based POMDP Solvers”. In: *Autonomous Agents and Multi-Agent Systems* 27.1, pp. 1–51. ISSN: 1387-2532, 1573-7454. DOI: 10.1007/s10458-012-9200-2. URL: http://link.springer.com/10.1007/s10458-012-9200-2 (visited on 19/08/2019).

Sharma, Siddharth, Simone Sharma and Anidhya Athaiya (10th May 2020). “Activation Functions In Neural Networks”. In: *International Journal of Engineering Applied Sciences and Technology* 04.12, pp. 310–316. ISSN: 24552143. DOI: 10.33564/IJEAST.2020.v04i12.054. URL: https://www.ijeast.com/papers/310-316,Tesma412,IJEAST.pdf (visited on 25/06/2023).

Sid Meier’s Civilization V (21st Sept. 2010). Firaxis Games.

Sid Meier’s Civilization V - Steam Charts (2023). URL: https://steamcharts.com/app/8930 (visited on 26/06/2023).

Sid Meier's Civilization VI (Oct. 2016). Firaxis Games. URL: https://store.steampowered.com/app/289070/Sid_Meiers_Civilization_VI/.

Sid Meier's Civilization VI - Steam Charts (2023). URL: https://steamcharts.com/app/289070#All (visited on 26/06/2023).

Sid Meier's Civilization VI Steam Community (3rd Mar. 2020). *Sid Meier's Civilization VI Obŝie Obsuždenija :: Soobŝestvo Steam*. URL: https://web.archive.org/web/20200303044135/https://steamcommunity.com/app/289070/discussions/ (visited on 26/06/2023).

Silver, David et al. (Jan. 2016). "Mastering the Game of Go with Deep Neural Networks and Tree Search". In: *Nature* 529.7587, pp. 484–489. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature16961. URL: http://www.nature.com/articles/nature16961 (visited on 15/01/2020).

Songs of Conquest (10th May 2022). Lavapotion AB.

Spaan, Matthijs TJ (2012). "Partially Observable Markov Decision Processes". In: *Reinforcement learning: State-of-the-art*, pp. 387–414.

Starcraft II (27th July 2010). Blizzard Entertainment. URL: https://starcraft2.com.

StarCraft II Editor Guides - Blizzard Starcraft II Tutorials (2021). URL: https://s2editor-guides.readthedocs.io/ (visited on 28/05/2021).

Stardew Valley (Feb. 2016). ConcernedApe. URL: https://www.stardewvalley.net/.

Steam Community :: ENDLESS™ Space 2 (2023). URL: https://steamcommunity.com/app/392110 (visited on 25/06/2023).

Steam Community :: Sid Meier's Civilization VI (2023). URL: https://steamcommunity.com/app/289070/ (visited on 25/06/2023).

Steam (2021). *Steam: Game and Player Statistics*. URL: https://store.steampowered.com/stats/ (visited on 26/05/2021).

Steinhaeusser, Sophia C. et al. (26th July 2022). "Joyful Adventures and Frightening Places–Designing Emotion-Inducing Virtual Environments". In: *Frontiers in Virtual Reality* 3. ISSN: 2673-4192. DOI: 10.3389/frvir.2022.919163. URL: https://www.frontiersin.org/articles/10.3389/frvir.2022.919163 (visited on 27/03/2024).

Stellaris (2016). Paradox Interactive. URL: https://www.paradoxinteractive.com/games/stellaris/about.

Steve Rabin (Apr. 2015). *Game AI Pro 2*. ISBN: 978-1-4822-5479-2. URL: http://www.gameaipro.com/.

— (June 2017). *Game AI Pro 3*. ISBN: 978-1-4987-4258-0. URL: http://www.gameaipro.com/.

— (Dec. 2021). *Game AI Pro 4*. Online Edition. URL: http://www.gameaipro.com/.

Stuff Made Here, director (9th Sept. 2020). *Moving Hoop Won't Let You Miss.* URL: https://www.youtube.com/watch?v=myO8fxhDRW0 (visited on 26/06/2023).

Subnautica (23rd Jan. 2018). Unknown Worlds Entertainment. URL: https://store.steampowered.com/app/264710/Subnautica/.

Suspicious Video-Game Generosity (2023). TV Tropes. URL: https://tvtropes.org/pmwiki/pmwiki.php/Main/SuspiciousVideogameGenerosity (visited on 25/06/2023).

Sutton, Richard S. and Andrew G. Barto (13th Nov. 2018). *Reinforcement Learning, Second Edition: An Introduction.* MIT Press. 549 pp. ISBN: 978-0-262-35270-3. Google Books: uWV0DwAAQBAJ.

Sweetser, Penelope and Peta Wyeth (1st July 2005). "GameFlow: A Model for Evaluating Player Enjoyment in Games". In: *Computers in Entertainment* 3.3, p. 3. ISSN: 15443574. DOI: 10.1145/1077246.1077253. URL: http://portal.acm.org/citation.cfm?doid=1077246.1077253 (visited on 12/12/2019).

Świechowski, Maciej et al. (1st Mar. 2023). "Monte Carlo Tree Search: A Review of Recent Modifications and Applications". In: *Artificial Intelligence Review* 56.3, pp. 2497–2562. ISSN: 1573-7462. DOI: 10.1007/s10462-022-10228-y. URL: https://doi.org/10.1007/s10462-022-10228-y (visited on 25/06/2023).

Tamborini, Ron et al. (Dec. 2011). "Media Enjoyment as Need Satisfaction: The Contribution of Hedonic and Nonhedonic Needs". In: *Journal of Communication* 61.6, pp. 1025–1042. ISSN: 00219916. DOI: 10.1111/j.1460-2466.2011.01593.x. URL: https://academic.oup.com/joc/article/61/6/1025-1042/4098480 (visited on 16/03/2024).

Tavares, Anderson R. et al. (2017). "Combining Genetic Algorithm and Swarm Intelligence for Task Allocation in a Real Time Strategy Game". In: *Journal on Interactive Systems* 8.1.

The Battle for Wesnoth (2003). The Battle for Wesnoth Project. URL: https://www.wesnoth.org/.

The Battle for Wesnoth Forums - Forums (2021). URL: https://forums.wesnoth.org/ (visited on 24/05/2021).

The Battle for Wesnoth Project (2021). *ReferenceWML.* URL: https://wiki.wesnoth.org/ReferenceWML (visited on 08/03/2021).

The Challenge Match (2023). URL: https://www.deepmind.com/research/highlighted-research/alphago/the-challenge-match (visited on 25/06/2023).

The Future of the RTS? (25th Jan. 2008). *The Future of the RTS? It's Already Here.* Flash of Steel. URL: http://flashofsteel.com/index.php/2008/01/25/the-future-of-the-rts-its-already-here/ (visited on 26/06/2023).

The Secret World (3rd July 2012). Funcom. URL: https://www.ea.com/games/the-secret-world.

The Subreddit for Endless Space. (2019). URL: https://www.reddit.com/r/EndlessSpace/ (visited on 17/11/2019).

The Witcher 3: Blood and Wine (May 2016). CD Projekt RED.

Thompson, Suzanne C, Wade Armstrong and Craig Thomas (1998). “Illusions of Control, Underestimations, and Accuracy: A Control Heuristic Explanation”. In: *Psychological bulletin*.

Thue, David et al. (2010). “Player Agency and the Relevance of Decisions”. In: *Interactive Storytelling*. Ed. by Ruth Aylett et al. Vol. 6432. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 210–215. ISBN: 978-3-642-16637-2 978-3-642-16638-9. DOI: 10.1007/978-3-642-16638-9_26. URL: http://link.springer.com/10.1007/978-3-642-16638-9_26 (visited on 14/03/2020).

— (2011). “A Computational Model of Perceived Agency in Video Games”. In: p. 6.

Togelius, Julian et al. (2013). “Assessing Believability”. In: *Believable Bots*. Ed. by Philip Hingston. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 215–230. ISBN: 978-3-642-32322-5 978-3-642-32323-2. DOI: 10.1007/978-3-642-32323-2_9. URL: http://link.springer.com/10.1007/978-3-642-32323-2_9 (visited on 14/08/2019).

Torb (2023). *I Built a 1Hz Redstone Computer... - YouTube*. URL: https://www.youtube.com/watch?v=tDxKhiJfgYk (visited on 25/06/2023).

Treasure Goblin (2019). Diablo Wiki. URL: https://diablo.fandom.com/wiki/Treasure_Goblin (visited on 02/10/2019).

Van Ryzin, Gregg G. (2004). “Expectations, Performance, and Citizen Satisfaction with Urban Services”. In: *Journal of Policy Analysis and Management* 23.3, pp. 433–448. ISSN: 1520-6688. DOI: 10.1002/pam.20020. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/pam.20020 (visited on 23/03/2024).

— (7th Aug. 2006). “Testing the Expectancy Disconfirmation Model of Citizen Satisfaction with Local Government”. In: *Journal of Public Administration Research and Theory* 16.4, pp. 599–611. ISSN: 1053-1858, 1477-9803. DOI: 10.1093/jopart/mui058. URL: https://academic.oup.com/jpart/article-lookup/doi/10.1093/jopart/mui058 (visited on 24/03/2024).

— (2013). “An Experimental Test of the Expectancy-Disconfirmation Theory of Citizen Satisfaction”. In: *Journal of Policy Analysis and Management* 32.3, pp. 597–614. ISSN: 1520-6688. DOI: 10.1002/pam.21702. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/pam.21702 (visited on 23/03/2024).

Villar, Santi Ontañón (10th May 2021). *Santiontanon/Microrts*. URL: https://github.com/santiontanon/microrts (visited on 26/05/2021).

Vinyals, Oriol, Igor Babuschkin, Junyoung Chung et al. (2019). “AlphaStar: Mastering the Real-Time Strategy Game StarCraft II”. In: *DeepMind Blog*.

Vinyals, Oriol, Igor Babuschkin, Wojciech M. Czarnecki et al. (Nov. 2019). “Grandmaster Level in StarCraft II Using Multi-Agent Reinforcement Learning”. In: *Nature* 575.7782 (7782), pp. 350–354. ISSN: 1476-4687.

DOI: 10.1038/s41586-019-1724-z. URL: https://www.nature.com/articles/s41586-019-1724-z (visited on 04/03/2021).

Von der Pütten, Astrid M. et al. (1st Nov. 2010). "“It Doesn't Matter What You Are!” Explaining Social Effects of Agents and Avatars". In: *Computers in Human Behavior*. Online Interactivity: Role of Technology in Behavior Change 26.6, pp. 1641–1650. ISSN: 0747-5632. DOI: 10.1016/j.chb.2010.06.012. URL: http://www.sciencedirect.com/science/article/pii/S0747563210001846 (visited on 26/10/2018).

Vox Populi (13th June 2014). URL: https://github.com/LoneGazebo/Community-Patch-DLL (visited on 25/03/2024).

Vuorre, Matti et al. (13th Sept. 2023). "An Intensive Longitudinal Dataset of In-Game Player Behaviour and Well-Being in PowerWash Simulator". In: *Scientific Data* 10.1, p. 622. ISSN: 2052-4463. DOI: 10.1038/s41597-023-02530-3. URL: https://www.nature.com/articles/s41597-023-02530-3 (visited on 15/03/2024).

Walton-Rivers, Joseph, Edward Longford et al. (2019). "Distributed Social Multi-Agent Negotiation Framework for Incomplete Information Games". In: *2019 11th Computer Science and Electronic Engineering Conference, CEEC 2019 - Proceedings*. IEEE. ISBN: 978-1-5386-3007-5.

Walton-Rivers, Joseph, Piers R. Williams and Richard Bartle (Sept. 2018). "Hexboard: A Generic Game Framework for Turn-Based Strategy Games". In: *2018 10th Computer Science and Electronic Engineering (CEEC)*. 2018 10th Computer Science and Electronic Engineering (CEEC). Colchester, United Kingdom: IEEE, pp. 163–168. ISBN: 978-1-5386-7275-4. DOI: 10.1109/CEEC.2018.8674187. URL: https://ieeexplore.ieee.org/document/8674187/ (visited on 26/05/2021).

Warburton, Victoria E. et al. (Feb. 2020). "Need Satisfaction and Need Frustration as Distinct and Potentially Co-Occurring Constructs: Need Profiles Examined in Physical Education and Sport". In: *Motivation and Emotion* 44.1, pp. 54–66. ISSN: 0146-7239, 1573-6644. DOI: 10.1007/s11031-019-09798-2. URL: http://link.springer.com/10.1007/s11031-019-09798-2 (visited on 26/03/2024).

Wardrip-Fruin, Noah et al. (2009). "Agency Reconsidered." In: *DiGRA Conference*. Citeseer.

Warpefelt, Henrik and Björn Strååt (2013). "Breaking Immersion by Creating Social Unbelievability". In: *Proceedings of AISB 2013 Convention. Social Coordination: Principles, Artefacts and Theories (SOCIAL. PATH)*, pp. 92–100.

Warpefelt, Henrik and Harko Verhagen (2015). "Towards an Updated Typology of Non-Player Character Roles". In: *Proceedings of the International Conference on Game and Entertainment Technologies*.

Watson, Ian et al. (2008). "Optimization in Strategy Games: Using Genetic Algorithms to Optimize City Development in FreeCiv". In.

Weisstein, Eric W. (2022). *Bonferroni Correction*. URL: https://mathworld.wolfram.com/ (visited on 20/10/2022).

What Is a 4X Game? (25th Aug. 2021). *What Is a 4X Game? | DiamondLobby*. URL: https://diamondlobby.com/general/what-is-a-4x-game/ (visited on 09/11/2023).

Yang, Zuozhi and Santiago Ontañón (8th Oct. 2019). "Guiding Monte Carlo Tree Search by Scripts in Real-Time Strategy Games". In: *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment* 15.1 (1), pp. 100–106. ISSN: 2334-0924. DOI: 10.1609/aiide.v15i1.5231. URL: https://ojs.aaai.org/index.php/AIIDE/article/view/5231 (visited on 25/06/2023).

Yee, Nick (2005). "Motivations of Play in MMORPGs". In: p. 46.

Zeithaml, V. A., L. L. Berry and A. Parasuraman (1st Jan. 1993). "The Nature and Determinants of Customer Expectations of Service". In: *Journal of the Academy of Marketing Science* 21.1, pp. 1–12. ISSN: 0092-0703. DOI: 10.1177/0092070393211001. URL: http://link.springer.com/10.1177/0092070393211001 (visited on 23/03/2024).

Zhang, Jiasheng et al. (Jan. 2022). "The Expectancy-Disconfirmation Model and Citizen Satisfaction with Public Services: A Meta-analysis and an Agenda for Best Practice". In: *Public Administration Review* 82.1, pp. 147–159. ISSN: 0033-3352, 1540-6210. DOI: 10.1111/puar.13368. URL: https://onlinelibrary.wiley.com/doi/10.1111/puar.13368 (visited on 23/03/2024).

Zillmann, Dolf, T. Alan Hay and Jennings Bryant (1st Dec. 1975). "The Effect of Suspense and Its Resolution on the Appreciation of Dramatic Presentations". In: *Journal of Research in Personality* 9.4, pp. 307–323. ISSN: 0092-6566. DOI: 10.1016/0092-6566(75)90005-7. URL: https://www.sciencedirect.com/science/article/pii/0092656675900057 (visited on 13/05/2022).

Zohaib, Mohammad (1st Nov. 2018). "Dynamic Difficulty Adjustment (DDA) in Computer Games: A Review". In: *Advances in Human-Computer Interaction* 2018, pp. 1–12. ISSN: 1687-5893, 1687-5907. DOI: 10.1155/2018/5681652. URL: https://www.hindawi.com/journals/ahci/2018/5681652/ (visited on 01/10/2019).

Zubrin, Robert (28th Mar. 1972). "Three Player Chess Board". U.S. pat. 3652091A. Individual. URL: https://patents.google.com/patent/US3652091/en (visited on 03/11/2023).