



Universiteit
Leiden
The Netherlands

Bachelor Computer Science

Exploring the Use of ChatGPT for Creating
Meaningful Dilemmas for Narrative Games

Suzanne Wooning

Supervisors:

Dr. Maarten Lamers & Sabijn Perdijk

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)

www.liacs.leidenuniv.nl

15/04/2026

Abstract

Choice-based narrative games rely on meaningful choices to engage their players. Artificial intelligence (AI) can be used to generate narrative content. However, there is little research on how AI-generated dilemmas compare to human-written ones in terms of player experience. This study explores this, mainly focusing on user-perceived meaningfulness, narrative quality, emotional engagement, and the experience when making the decision.

A survey-based study with a within-subject design was conducted. Participants were shown scenarios based on an existing narrative-driven video game and were presented with two different dilemmas: one was the original dilemma from in the game, the other was AI-generated. Participants rated each dilemma on perceived meaningfulness, narrative quality, emotional engagement, and decision experience. Besides, they directly compared the dilemmas to see which they preferred.

The findings of this study show that the original dilemmas scored higher overall. There was a significant difference found for the decision experience. The differences for meaningfulness, narrative quality, and emotional engagement point in the same direction, but were not statistically significant. These results suggest that AI-generated dilemmas can be comparable to human-written ones on some aspects, but are less effective in creating a strong decision-making experience.

Contents

1	Introduction	1
1.1	Background and Motivation	1
1.2	Thesis Overview	2
2	Literature Review	2
2.1	Emotional and Narrative Experience in (Narrative) Games	2
2.2	AI in (Narrative) Games	3
2.3	AI's Role in Player Experience	3
2.4	AI Prompting	3
2.5	Measuring Player Experience in Narrative Games	4
3	Research Statement	4
4	Methods	5
4.1	Scenario Selection	5
4.2	AI Choice Generation	6
4.3	Survey Design and Procedure	7
4.3.1	Platform and Global Structure	7
4.3.2	Scenario and Choice Presentation	8
4.3.3	Question Design	8
4.4	Participants	9
4.5	Data Analysis	9
5	Results	10
5.1	Participant Familiarity	10
5.2	Descriptive Statistics of Experience Scores	11
5.3	Comparison of OG and AI Experience Scores	12
5.4	Preference Choices	14
6	Discussion	16
6.1	Interpretation of the Findings	16
6.2	Limitations	17
6.3	Future Research	18
7	Conclusion	19
	References	22
	A AI Prompts and Generated Outputs	23
	B Standardized Survey Scenarios and Dilemmas	25
	C Survey	27
	D Raw Collected Data	31

1 Introduction

1.1 Background and Motivation

Narrative video games are becoming increasingly popular. They often allow players to engage with complex stories through interactive choices. The opportunity for different options is an important part of these games because they often give the player a sense of control and agency. Making in-game choices is an important part of most narrative games because they can allow the player to influence the story and feel more connected to the characters [Pin24]. These choices can also create strong emotional reactions evoked through interaction and immersion with the game [Ana18]. For example, players may be faced with making difficult moral decisions or their choices may shape relationships between characters (e.g. *Detroit: Become Human*, Quantic Dream, 2018; *Baldur's Gate 3*, Larian Studios, 2023).

While this thesis refers to narrative games in general, the focus is in particular on choice-based narrative games. Not every narrative game has choices that players can make. This distinction is important because this study looks at dilemmas and decision-making, which are parts of narrative games where the players are able to make choices that actively change the story.

Narrative game design has mostly relied on human-written content, but recent developments in artificial intelligence raise new possibilities. Large language models (LLMs) can already generate dialogue or branching storylines. For example, the *PANGeA* project [BKZ⁺24]. This project shows how LLMs can be used for creating narrative content in role-playing games. In addition, recent work has begun to explore how emotional arcs can guide AI-driven narrative generation [WHZ⁺25]. Apart from these technical examples, there have also been studies that have looked at the role of AI in game design. These highlight the potential of AI in game design to support creativity and the risk that it might limit innovation [Alh25].

Despite these advances, little research has examined how players respond emotionally to AI-generated choices compared to human-written ones. Some recent research has compared AI- and human-authored narratives more broadly [RG25], but this does not address the choice-driven context that some narrative games have. This gap is relevant to address, since many narrative games are designed to evoke emotion through player decisions. Understanding whether AI can generate experiences that feel as meaningful as human-written ones could be helpful for both game design practice and the role of AI in creative processes.

This research therefore explores how players perceive dilemmas in choice-based narrative games, comparing human-written dilemmas with AI-generated ones. Participants will evaluate meaningfulness, narrative quality, emotional engagement, and decision-making experience of each dilemma through a survey-based study. With these responses, this study aims to gain insight into how AI-generated dilemmas compare on these aspects.

1.2 Thesis Overview

This bachelor thesis, conducted at LIACS (Leiden Institute of Advanced Computer Science), was supervised by Maarten Lamers & Sabijn Perdijk.

The remainder of this thesis is structured as follows. Section 2 presents a literature review. Section 3 outlines the research question. Section 4 describes the methodology. Section 5 presents the results of the study. Section 6 discusses the interpretation of the findings and limitations of this research and possibilities for future research. Finally, Section 7 concludes the thesis.

2 Literature Review

This section reviews existing literature that is relevant to understanding player experiences in narrative games. It first explores the emotional and narrative aspects of gameplay. Then, it goes into how AI can shape both narrative game content and player experience. Next, it examines an AI prompting format used for generating narrative dilemmas. Finally, methods for measuring player experience are discussed.

2.1 Emotional and Narrative Experience in (Narrative) Games

Narrative-driven games are a good base for studying emotional experiences. They make players connect with characters and stories. Games can evoke emotion through multiple factors such as music and visuals that create ambiance. A study by Bopp et al. [BMA⁺19] highlights that a player’s attachment to a game character also contributes to the emotional engagement. Players who care about the characters often find the story more meaningful because they care about what happens.

Besides, the role of player agency and choices add to the emotional engagement [Pin24]. Especially with narrative games where players are faced with decisions and interactions within the narrative. These can often evoke feelings of responsibility or empathy. Pinto [Pin24] also highlights that sometimes the emotional impact of a decision comes less from what happens in the story and more from the responsibility the player feels for making it.

Meaningful choices in games are often described as choices that have impactful consequences and that make the player feel like their decisions actually matter within the story [Pin24][BMA⁺19]. When players think a choice is meaningful they are more likely to feel emotionally involved in the game. This means that meaningful choices are really important for getting players to care about the story and feel like they are a part of it. Meaningful choices are what make players feel like they are really playing the game rather than just watching it. Thus, meaningful choices are a crucial part of a game’s design.

This shows that players’ experiences are shaped not only by the game, but also by the player’s involvement. Emotional responses, how meaningful a story feels, and how players experience their decisions are all bound together. Understanding this is important for comparing AI-generated and human-written dilemmas, because differences in story quality or emotional impact might change how meaningful a choice feels.

2.2 AI in (Narrative) Games

AI has become a useful tool in the development of narrative-driven games. For instance, it offers new ways to generate and adapt narratives to enhance player experience. Procedural narrative systems, such as *PANGeA* [BKZ⁺24], show how AI can create dynamic storylines that respond to players’ choices. This allows for unique story experiences that are tailored to each player. Since with this the narratives are not fixed, there’s more replayability.

Another approach is the use of AI to guide the emotional arc within games. A recent study by Wen et al. [WHZ⁺25] explores how procedural level generation can be aligned with the emotional progression of a story. This ensures players face challenges and conflicts that fit the emotional tone of the narrative which makes the story more impactful and meaningful. This thus shows how AI can help games evoke intended feelings at key moments.

These developments show how AI in narrative games can not only influence the technical aspects of the game but also the narrative structure, and emotional impact.

2.3 AI’s Role in Player Experience

A study by Hernandez et al. [HBH14] explores how AI can be used to create emotion-based interactive storytelling. Their work focuses on interactive storytelling where AI not only reacts to players’ choices, but also to the emotional cues from the player or the in-game characters. In this context, AI does more than just simply generate or control story events. It also becomes an active agent that actively tries to understand and respond to emotional states. This has clear implications for emotional engagement. This study suggests that emotional engagement in games can be shaped by AI, and not only triggered by narrative events or characters. For narrative games, this could make emotional engagement feel more personal and dynamic. This can make the experience feel less fixed and more responsive to the player.

Similarly, Tucek [Tuc24] investigates how AI can be used to enhance empathy in games through personalized interactions with characters. The study shows that AI can shape emotional engagement. Players may feel a stronger emotional connection because the system appears to understand them. This highlights how AI can create situations where players’ decisions feel significant not only narratively but also emotionally. It also suggests that AI-driven empathy can deepen the perceived authenticity and importance of narrative choices.

2.4 AI Prompting

In their study, Harmon & Rutman [HR23] explore how prompt structure affects the generation of narrative choices by large language models. They compare several prompt formats in the context of movie plot excerpts to see which kinds of instructions lead to more successful responses of the LLM. The prompt formats they looked at are *Simple Colon*, *Masterful Storyteller*, and *Answer Interrogative* (see Figure 1). They found that the performance was best when the *Answer Interrogative* prompt format was used.

Format	Template
Simple Colon	Story: source-plot
	(1) The next choice source-character is confronted with, (2) source-character 's decision for this choice, and (3) what happens as a result of source-character 's decision:
Masterful Storyteller	A story is provided: source-plot
	The masterful storyteller explains the next part of the story by providing (1) The next choice source-character is confronted with, (2) source-character 's decision for this choice, and (3) what happens as a result of source-character 's decision:
Answer Interrogative	Story: source-plot
	What is (1) the next choice source-character is confronted with, (2) source-character 's decision for this choice, and (3) what happens as a result of source-character 's decision? Answer:

Figure 1: Overview of prompt formats used for narrative choice generation, source: Harmon & Rutman [HR23].

For narrative games, this shows that the structure of the prompt can influence the quality of AI-generated choices. In this study a similar approach inspired by the *Answer Interrogative* method will be used. However, instead of focusing on which prompt format works best, the main focus is on the perceived meaningfulness and quality of the AI-generated choices.

2.5 Measuring Player Experience in Narrative Games

In the field of game studies, a commonly used tool to assess player experience is the *Game Experience Questionnaire* (GEQ) [IdKP13]. It focuses on seven different dimensions of gameplay experience: immersion, flow, competence, positive and negative affect, tension, and challenge. The GEQ helps understand how players feel and interact with games. It provides a structured way to evaluate the player’s experience. Overall, it is a useful tool for studying general player experience.

The GEQ is especially useful for capturing the more general experience of gameplay. However, it does not really look at smaller things such as how you experience making choices in a story. The GEQ does instead look at things like how tense you get and how you feel overall. The questionnaire is really effective for providing insights into emotional reactions during gameplay.

To measure player experiences in narrative games, looking into more narrative-specific experiences can also provide more insight. For example, the *Measuring Narrative Engagement* framework by Busselle & Bilandzic [BB09] can be used for this. This approach focuses on more narrative related constructs. For example, how people relate to characters or how engaging the narrative is. In their study, the focus was on films. However, parts of their framework can be easily adapted to also be applicable for narrative games.

3 Research Statement

This research aims to explore how players perceive and experience dilemmas in narrative decision scenarios when these dilemmas are either human-written or AI-generated. In this study, player

experience is understood as a combination of perceived meaningfulness, narrative quality, emotional engagement, and the decision-making experience during choice selection.

Main Research Question

RQ: How do AI-generated versus human-written dilemmas in narrative decision scenarios differ in terms of perceived meaningfulness, narrative quality, emotional engagement, and players’ decision-making experience?

Sub-questions

- **SQ1:** How do AI-generated and human-written dilemmas differ in perceived meaningfulness?
- **SQ2:** How do they differ in perceived narrative quality?
- **SQ3:** How do they differ in perceived emotional engagement?
- **SQ4:** How do they differ in players’ decision-making experience?
- **SQ5:** How do players directly compare AI-generated and human-written dilemmas when asked to indicate an overall preference between them?

4 Methods

To answer this research question a survey-based study with a within-subject design was conducted. Participants were shown several narrative scenarios derived from a narrative-driven video game. For each scenario, they evaluated two dilemmas: one human-written (from the original video game) and one AI-generated. Participants were not informed beforehand that they would be evaluating AI-generated content. Participants rated each dilemma on perceived meaningfulness, narrative quality, and emotional engagement. They were also asked to evaluate the experience when making the decision the dilemma presented. In addition, participants were asked to directly compare the two dilemmas for each scenario.

The following sections describe in further detail the methods of the selection of scenarios, the generation of AI-based dilemmas, the survey design and procedure, the participant sample, and the data analysis.

4.1 Scenario Selection

The scenarios were based on the narrative-driven game *Life is Strange* [Don15]. This game was selected because it is strongly centered around player choice and moral dilemmas, making it suitable for exploring the research question. In *Life is Strange*, players progress through a story and are forced to make choices that will influence the narrative. The game contains several “major choices” which have a significant impact on the story and are often emotionally intense. When a player reaches a point in the story where such a “major choice” appears, they are presented with two

distinct choice options.

Initially, five of these “major choice” moments were selected from the game. Major choice moments were chosen because they have a clear narrative and emotional impact. Then the narrative context around these choice moments was written out in text form. Since these scenarios were later used for both the AI generating and for in the survey, care was taken to avoid overly emotional language or wording that might steer participants toward a particular choice option. Rather, the scenarios were written to be clear and informational while remaining as neutral as possible.

4.2 AI Choice Generation

To generate alternative choice options for the selected scenarios, a generative language model was used. All generated choices were produced using ChatGPT 5.1 [Ope25].

Three different kinds of prompt were explored. The initial prompt format was based on the *Answer Interrogative* format described by Harmon & Rutman [HR23]. This format was initially chosen because it is designed for successful responses regarding narrative choice generation. However, with this format the AI would recognize the source material. It would explicitly identify the scenarios were from the game *Life is Strange* and then reproduce the original in-game choice options.

To avoid this, additional restrictions were added to the prompt, see Figure 2. These restrictions instructed the AI not to reference external media or identify the source of the scenario. With this, the AI is prevented from reproducing existing in-game choices and it ensures that genuinely alternative dilemmas were generated.

Restricted Prompt Format

Do not reference external media or game titles.

You may not identify or name the source of this scene.

Scene:

<written out scenario>

What are two major choice options <main character> is confronted with next?

Answer:

Figure 2: Prompt format with restrictions at top. When used for generating, the scenario in text form was given in <written out scenario>, and <main character> was replaced with the name of the game scenario’s main character.

However, the output of the choices with this format were identical to the choices already in game. As the goal of this study was to compare AI-generated choices to human-written choices, a new prompt format was made. With this format, the AI was given the choice options that are already in the game and was explicitly told to create other choice options, see Figure 3.

Alternative Choices Prompt Format

Do not reference external media or game titles.

You may not identify or name the source of this scene.

The following choices are forbidden:

- <in-game option 1>
- <in-game option 2>

Scene:

<written out scenario>

What are two entirely new major choice options <main character> is confronted with next?

Answer:

Figure 3: Prompt format to generate entirely new options. When used for generating, options that are in the game were given in <in-game option 1> and <in-game option 2>, the scenario in text form was given in <written out scenario>, <main character> was replaced with the name of the game scenario’s main character.

Using the *Alternative Choices Prompt Format 3*, all scenarios were processed, resulting in a set of alternative AI-generated dilemmas. See Appendix A for all prompts used and the generated outputs.

4.3 Survey Design and Procedure

4.3.1 Platform and Global Structure

The survey was created using Qualtrics. It was an online survey and it was distributed by spreading a link. In the survey, participants were shown written-out scenarios based on the narrative-driven video game *Life is Strange*. For each scenario, they were shown a dilemma; a choice that had to be made at that moment in the game with two options to choose from.

For each scenario, participants evaluated two dilemmas. One dilemma was directly taken from the game, the other was generated by AI. Participants were not informed that one dilemma was made by AI and that they would compare AI-written dilemmas to human-written ones. Instead, in order to avoid bias towards AI-generated content, the participants were told that the study focussed on how dilemmas in narrative games are perceived and experienced.

Participants were asked to evaluate a dilemma on aspects of meaningfulness, narrative quality, and emotional engagement. Next, they were shown the same scenario but with the other dilemma and were asked the same set of questions again. After the participants had evaluated both dilemmas for a scenario, they were asked five comparison questions.

A complete overview of the survey used in this study can be found in Appendix C.

4.3.2 Scenario and Choice Presentation

Initially, five scenarios were prepared for the survey. However, due to limiting the time the survey takes, this number was reduced to three. This reduction was done to keep the survey at manageable length for the participants. The scenarios were slightly rewritten for use in the survey. All the characters, their names, and pronouns were gender neutralized to avoid any potential biases related to gender.

For each scenario, the order in which the original and AI-generated dilemmas were presented was randomized using Qualtrics. The order in which the dilemmas were presented in the comparison sections was also randomized.

The choice options for each dilemma were presented in a standardized format. The AI-generated output of choice options were often quite long and added a whole description to it. Since it was not a goal of this study to compare the difference in length or level of detail of a presented option, all the options were rewritten in the same format. Each option started with a short description of the choice itself, followed by one or two sentences providing context and explanation of the choice. The original in-game choices are originally very brief and do not provide explanation. To these options, similar context sentences were added so that both AI-generated and original choices could be evaluated equally. This standardization ensured that participants’ judgement focused on the content of the choices rather than on differences in presentation. The standardized scenarios and dilemmas used in the survey can be found in Appendix B.

4.3.3 Question Design

The survey questions were designed to explore how participants experience and perceive dilemmas in narrative-driven games. The survey focused on three main constructs: meaningfulness, narrative quality, and emotional engagement. These constructs were informed by existing work on game experience and narrative engagement. The constructs in the survey drew inspiration from concepts used in the GEQ [IdKP13] and from Busselle and Bilandzic’s [BB09] work on narrative engagement. Rather than directly using existing scales, the questions were adapted to fit the context of short, written game dilemmas rather than full gameplay experiences.

For each scenario, participants evaluated two dilemmas (one from the original game and one AI-generated) separately. After reading one scenario and a dilemma, participants rated statements using a five-point Likert scale ranging from strongly disagree to strongly agree. Statements related to meaningfulness focused on whether the choices presented in the dilemma felt impactful and important to the story. Narrative quality was assessed through statements on how well the choices fit the characters’ personalities and tone of the story. Narrative quality was also assessed with statements about whether the choices fit logically within the presented narrative. Emotional engagement was measured through statements focusing on emotional impact, tension, and the participants’ feelings of connection to the characters.

In addition to these evaluative questions, participants were also asked decision-related questions. They were asked which choice option from the dilemma they would choose themselves and how confident, tense, and conflicted they felt while making that decision. These questions were included to capture aspects of the decision-making experience that are important in narrative games.

After both dilemmas for the same scenario had been evaluated, participants were asked to directly compare the two. These comparative questions focused on which dilemma felt more emotionally engaging, meaningful, and narratively fitting. Besides, participants were asked which dilemma they would prefer to see included in the actual game. For each comparative question, participants could indicate a preference for one dilemma with an additional option to indicate no clear preference.

Besides all these evaluative questions, the survey also included three familiarity questions. At the beginning of the survey, participants were asked whether they were familiar with narrative-driven games in general and about their familiarity with the video game *Life is Strange*. These questions were not meant to filter participants, but simply to provide some context about the participants' background. The survey also included a familiarity question about a fictional game title. This question was included as a simple control question.

4.4 Participants

A total of 16 participants took part in the study. They were recruited mainly via personal networks, with additional participants reached through peer-to-peer sharing of the survey link. Participation was voluntary and anonymous. All participants were required to be proficient in English as the survey was conducted in English. Only participants of 18 years and older were allowed to partake. Participants reported varying levels of familiarity with narrative-driven games.

4.5 Data Analysis

All Likert-scale responses were coded from 1 to 5. In general, higher scores indicated a more positive evaluation of the presented dilemma. However, for one question from the decision experience evaluation, the scale was reversed.

The Likert scale used for these questions was as follows:

- 1 - Not familiar at all
- 2 - Slightly familiar
- 3 - Moderately familiar
- 4 - Very familiar
- 5 - Extremely familiar

For each construct (meaningfulness, narrative quality, emotional engagement, and decision experience), the relevant items were combined into a single score by calculating the mean per participant. This was done separately for each dilemma. In addition, mean scores were also calculated across all three scenarios combined. This resulted in overall average scores for the original game dilemmas and AI-generated dilemmas for each construct. These scores were then combined to calculate an

overall score for the original and AI generated dilemmas.

For the comparative questions, the number of times participants selected the original dilemma, the AI-generated dilemma, or indicating no preference were calculated. These frequencies were used to examine which type of dilemma was more often perceived as more meaningful, emotionally engaging, or narratively fitting.

The raw dataset can be found in Appendix D.

5 Results

5.1 Participant Familiarity

Descriptive statistics were calculated for three variables: familiarity with the genre of narrative games, familiarity with the game *Life is Strange*, and familiarity with a non-existing game “*Strumming Shadows*”. This analysis was done to get an overview of the participants’ backgrounds with narrative games. Familiarity with a non-existing game was included as an attention check to test whether participants were filling in the survey truthfully and carefully.

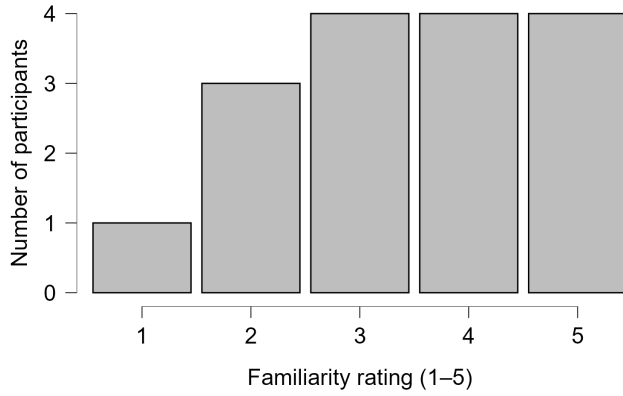
Table 1 shows participants were fairly familiar with the narrative games genre (Mean = 3.44). Familiarity with *Life is Strange* was still moderate but slightly lower (Mean = 2.94). As expected, the familiarity with the fake non-existing game was very low (Mean = 1.13). This suggests that the attention check functioned as intended.

Table 1: Descriptive statistics for participants’ familiarity with the narrative game genre, the game *Life is Strange*, and the non-existing game “*Strumming Shadows*”

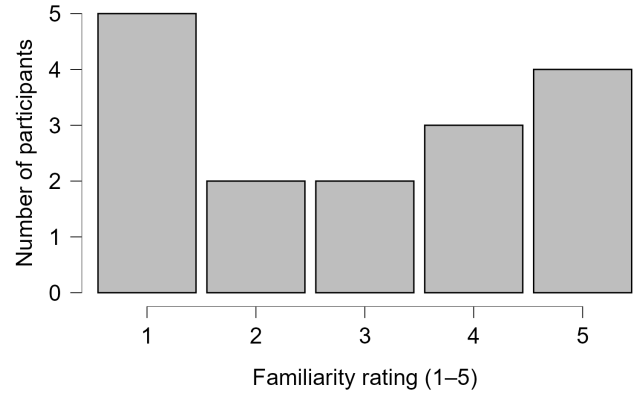
	Familiarity with		
	Narrative game genre	<i>Life is Strange</i>	<i>Strumming Shadows</i> (fake game)
Mean	3.438	2.938	1.125
Std. Deviation	1.263	1.652	0.342
Minimum	1.000	1.000	1.000
Maximum	5.000	5.000	2.000

The distribution plots, Figure 4, show these patterns as well. Familiarity with the narrative game genre shows a relatively broad distribution across the scale. This suggests that there is a fair amount of variability in the participants’ familiarity with this genre. A similar pattern is seen for familiarity with *Life is Strange*, however the responses are more dispersed. The responses have higher frequencies at both ends of the scale. This indicates most participants were either largely unfamiliar with the game or quite familiar with it. The familiarity with the fake game shows participants only indicated a very low familiarity level of 1 or 2. Again, this shows that the attention check functioned as intended and that participants did not report familiarity with a non-existing

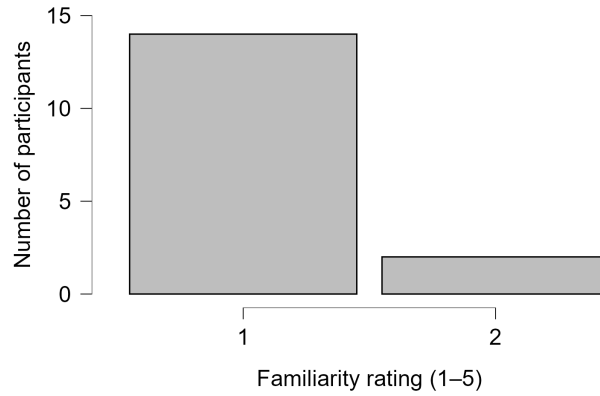
game.



(a) Genre familiarity



(b) *Life is Strange* familiarity



(c) Fake game familiarity

Figure 4: Distribution of familiarity ratings for the narrative game genre, *Life is Strange*, and the non-existing game “*Strumming Shadows*”

Overall, these descriptive statistics show that participants were a bit above moderately familiar with the narrative game genre. It also shows participants were moderately familiar with the game *Life is Strange*. Besides, it shows that participants were largely unfamiliar with the fake game. This suggests that the attention check worked as intended and that participants responded attentively.

5.2 Descriptive Statistics of Experience Scores

Descriptive statistics were calculated for all aggregated variables of the measured concepts for both versions of the presented dilemmas (Original Game (OG) and AI-generated (AI)). These variables include meaningfulness, narrative quality, emotional engagement, decision experience. Besides, an overall score was calculated by averaging these measures.

Table 2 shows that overall, the original game version scored higher than the AI-generated ones across all measures. The largest difference was in the decision experience (OG: Mean = 3.17; AI: Mean = 2.52).

The Shapiro-Wilk tests indicate that most variables were approximately normally distributed ($p > 0.05$), see Table 2. The only variable that did not meet this p-value is the variable of emotional engagement in the OG version. This suggests that this variable has a potential deviation from being normally distributed.

Table 2: Descriptive statistics for meaningfulness, narrative quality, emotional engagement, decision experience, and overall scores for the original game (OG) and AI-generated (AI) dilemmas

	Mean	Std. Deviation	Shapiro-Wilk	P-value of Shapiro-Wilk
OG meaningfulness	4.542	0.378	0.933	.269
AI meaningfulness	4.208	0.507	0.957	.604
OG narrative quality	4.073	0.433	0.898	.074
AI narrative quality	3.818	0.460	0.959	.636
OG emotional engagement	4.139	0.630	0.841	.010
AI emotional engagement	3.979	0.502	0.963	.718
OG decision experience	3.167	0.443	0.952	.515
AI decision experience	2.521	0.514	0.942	.380
OG overall score	3.980	0.340	0.993	1.000
AI overall score	3.632	0.319	0.977	.933

5.3 Comparison of OG and AI Experience Scores

To examine whether participants evaluated the Original Game (OG) and AI-generated (AI) dilemmas differently, paired samples t-tests were conducted. A paired design was used because all participants evaluated both versions of the dilemmas (within-subject design).

The analysis tested whether the average evaluation scores for the OG and AI versions differed significantly across the measured concepts; meaningfulness, narrative quality, emotional engagement, decision experience, and an overall score. All measures were rated on a five-point Likert scale.

Before conducting the t-test, the assumption of normality was tested using Shapiro-Wilk tests, see Table 3. The tests show that none of the difference scores significantly deviated from normality (all $p > 0.05$).

The paired samples t-test, Table 4, showed a significant difference between the OG and AI overall scores with $p=0.018$. This indicates that participants rated the OG versions significantly higher overall than the AI versions.

Table 3: Test of normality (Shapiro-Wilk) for the difference scores between OG and AI

Difference scores (OG – AI)	Shapiro-Wilk	P-value of Shapiro-Wilk
Meaningfulness	0.968	.809
Narrative quality	0.968	.797
Emotional engagement	0.968	.805
Decision experience	0.942	.377
Overall score	0.958	.622

Looking at the individual measures, the test showed a high significant difference for the decision experience ($p < 0.001$). This indicates participants reported stronger feelings of tension and internal conflict when making the decisions in the OG versions compared to the AI-generated versions.

For meaningfulness, the difference approaches significance ($p=0.055$), but does not reach the 0.05 threshold. Narrative quality ($p=0.097$) and emotional engagement ($p=0.471$) did not show statistically significant differences.

Table 4: Paired samples t-test comparing mean scores of OG and AI dilemmas across measured concepts

Measure 1		Measure 2	t	df	p
OG meaningfulness	-	AI meaningfulness	2.080	15	.055
OG narrative quality	-	AI narrative quality	1.768	15	.097
OG emotional engagement	-	AI emotional engagement	0.739	15	.471
OG decision experience	-	AI decision experience	4.421	15	< .001
OG overall score	-	AI overall score	2.665	15	.018

These findings are consistent with the descriptive statistics presented in Table 2, where the OG versions scored higher on all measures. The paired samples t-test indicates that these differences are statistically significant for the overall score and decision experience.

Overall, the paired samples t-test shows that the OG dilemmas were evaluated more positively than the AI-generated dilemmas. The difference in the overall score is statistically significant. This difference appears to be largely driven by the big difference in the decision experience.

Differences in meaningfulness and narrative quality had higher scores for the OG versions, but these did not reach statistical significance. The difference in the mean scores for emotional engagement was lowest. Consequently, emotional engagement did not differ significantly between the two versions.

Taken together, these findings suggest that participants overall perceived the OG version more positively. The clearest difference herein was in the decision experience.

5.4 Preference Choices

Participants were also asked to compare the original game (OG) dilemma with the AI-generated dilemma. For each comparison, they indicated which version they preferred. There was also an option for indicating no preference.

Each participant had five comparison questions per scenario. Since there were three scenarios, this resulted in 15 preference choices per participant. Therefore, a participant could choose the OG dilemma a maximum of 15 times and then the AI dilemma 0 times (or vice versa). Higher counts therefore indicate that that version of the dilemma had a stronger preference.

As seen in Table 5, on average, participants preferred the OG dilemma 8.25 times out of 15. The AI dilemma was chosen 5.25 times on average. The “no preference” option was not often selected (Mean = 1.50). This suggests that participants were generally able to express a clear preference between the two versions rather than perceiving them as equal.

Table 5: Descriptive statistics of participant preference choices for OG and AI dilemmas

	OG dilemma preferred	No dilemma preference	AI dilemma preferred
Mean	8.250	1.500	5.250

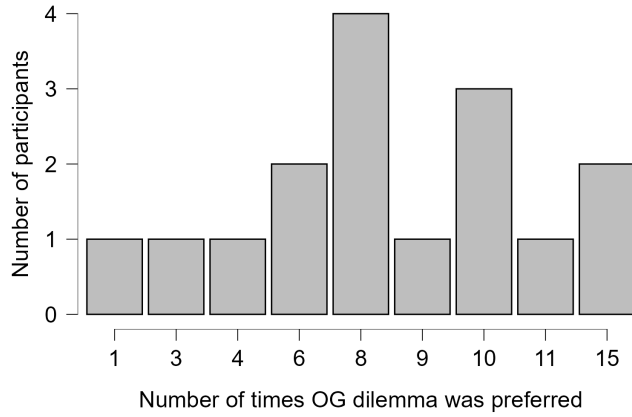
The distribution plots, see Figure 5, provide further insight into these preference patterns. The distribution of the OG dilemma preferences, Figure 5a, shows that higher preference counts occurred more frequently. In contrast, AI preference counts were more often lower, see Figure 5b. The “no preference” distribution, Figure 5c, shows that many participants rarely selected this option. It shows that 8 out of the 16 participants never selected the “no preference” option at all. This further indicates that most participants were able to indicate a clear preference between the two versions.

To examine whether there were differences in how often each option was preferred, a repeated measures ANOVA was done. This test was conducted using the frequencies of OG, AI, and “no preference” selections. This analysis tests whether the frequency of choosing these options differs significantly.

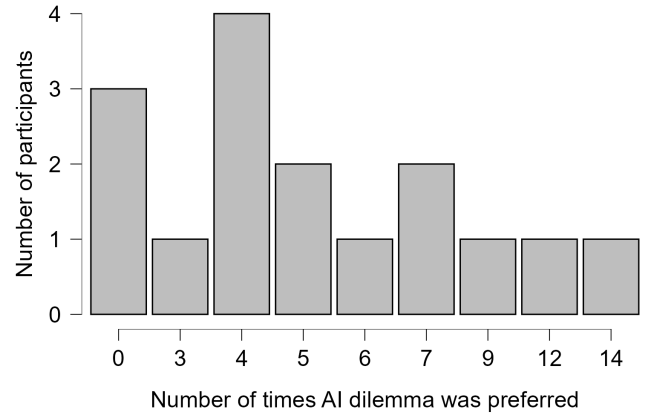
The repeated measures ANOVA, Table 6, shows a p-value < 0.001 . This indicates that the frequency with which participants preferred OG, AI, or None differed significantly.

Table 6: Within-subjects effects for preference choices (OG, AI, No preference) from the repeated measures ANOVA

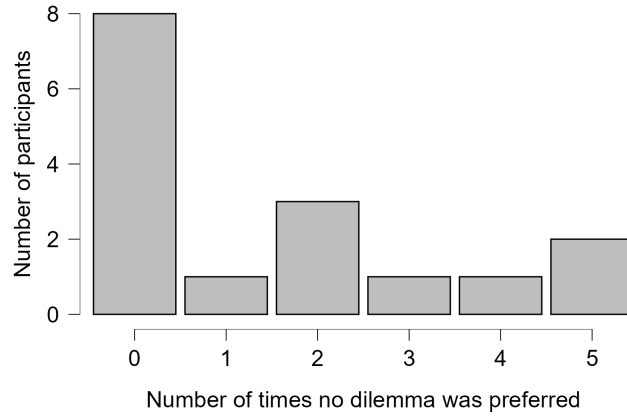
Cases	Sum of Squares	df	Mean Square	F	p
Preference	366.0	2	183.0	10.85	$< .001$
Residuals	506.0	30	16.87		



(a) OG dilemma preferred



(b) AI dilemma preferred



(c) No preference

Figure 5: Distribution of participant preference counts for the OG dilemma, no preference, and AI dilemma. X-axis shows the number of times a participant chose that option (0–15), Y-axis shows the number of participants. For example, in the OG dilemma plot (a), 4 participants chose the OG dilemma 8 times (out of 15).

Because this overall difference may mainly be caused by the “no preference” option being chosen less often, a second repeated measures ANOVA was done. This time, only the OG and AI options were included, excluding the “no preference” option. The goal was to test whether there was a statistically significant difference in how often participants preferred the OG versions compared to the AI versions in the comparison questions.

The test, see Table 7, shows a p-value of 0.13. This value is larger than 0.05 thus the difference between the number of times participants preferred the OG versions versus the AI versions is not statistically significant.

Table 7: Within-subjects effects for preference choices (OG vs AI) excluding the “no preference” option from the repeated measures ANOVA

Cases	Sum of Squares	df	Mean Square	F	p
Preference	72.00	1	72.00	2.523	.133
Residuals	428.0	15	28.53		

Overall, the preference data shows a slight preference for the OG dilemmas. The “no preference” option was rarely chosen. However, when comparing only the OG and AI preferences, the difference was not statistically significant.

6 Discussion

6.1 Interpretation of the Findings

The findings of this study suggest that the participants evaluated the original game dilemmas more positively overall than the AI-generated dilemmas. When looking at the average combined overall score of the individual measures, the OG dilemmas scored higher and this difference was statistically significant. However, when looking more closely at these results, the overall difference appears to be mainly caused by the difference in the decision experience scores. On the other hand, the differences for meaningfulness, narrative quality, and emotional engagement were smaller and not statistically significant. When participants were asked to directly compare the OG and AI-generated dilemmas, the OG dilemmas were more frequently preferred. However, the difference between the OG and AI preference frequencies was not statistically significantly different.

The clearest finding of this study is therefore that the OG dilemmas were rated better on decision experience than the AI-generated dilemmas. This suggests that when participants were asked to make the decision presented in the dilemma, they experienced stronger feelings of tension and conflict with the OG dilemmas. Since decision-making is a central part of moral dilemmas in narrative games, this difference is important. It may indicate that the AI-generated dilemmas resemble the original game dilemmas on a surface level. However, it also suggests that the AI-generated dilemmas may not fully recreate the same intensity of the decision itself.

The other measured differences in scores were however not statistically significant. The scores for meaningfulness, narrative quality, and emotional engagement all scored higher for the OG dilemma. But the difference was not statistically significant. So, even though the OG overall score was significantly higher than the AI score, this does not necessarily suggest that OG dilemmas performed better on all measures. The OG dilemmas were overall better evaluated, mainly on decision experience, but on the other aspects it was more comparable to the AI-generated dilemmas.

This may suggest that AI-generated dilemmas are capable of creating situations that feel meaningful and that fit within the narrative. Participants may have found both versions emotionally engaging simply because the narrative scenarios themselves already evoke emotion. The relatively small sample size of participants could also have played a role. With fewer participants, it is harder to detect smaller differences. This means that there may be a difference between the AI-generated and original game dilemmas on some measures, but they may not have reached statistical significance in this study for some of the measures.

In the direct comparison questions, participants chose the original (OG) dilemmas more often than the AI-generated ones. However, the difference between OG and AI-generated preference counts was not statistically significant. This suggests that participants' explicit preferences, when they had to directly choose between the two versions, were less clear than what was seen in the rating-based evaluations. A possible explanation for this is that with the direct comparison questions, participants were not directly asked for which dilemma they found the making of the dilemma more conflicting for example. And as seen previously, with the rating evaluations this decision experience measure was important. So, when that is not a part of criteria, then we do not see as clear of a difference between the OG and AI dilemmas.

Overall, these findings suggest that the original game dilemmas were better at creating a strong and engaging decision-making experience than the AI-generated dilemmas. In other words, participants appeared to experience the act of making the decision more strongly in the original versions. At the same time, the smaller differences in meaningfulness, narrative quality, and emotional engagement suggest that the AI-generated dilemmas were still able to capture some of the basic elements of a narrative dilemma. However, this does not necessarily mean that they created the same experience of actually having to make a meaningful choice. Compared to the original dilemmas, the AI-generated ones may have been less effective in making the decision itself feel important or impactful for participants.

6.2 Limitations

This study has several limitations that should be considered when interpreting the findings.

First of all, the scenarios were written representations of moments in the game. The scenarios were written without trying to steer participants in a particular direction. That means the scenario texts were possibly not as emotionally evoking as when participants were really playing the game. Using the scenarios in text format causes a loss of interactivity and perhaps less immersion. Besides, the text format does not provide any audio or visual context. These factors may have influenced how participants experience the dilemmas. Thus, for the findings of this study it is more so about how

participants experience a dilemma from a game when they read about it rather than when they actively play the game.

Another limitation that has to be taken into consideration is that the prompt format might be limiting the AI-generated output. As previously described in section 4.2, the AI would first generate the exact same dilemmas as in the original game when it was not given any further restrictions of generating an entirely new dilemma. This could be because maybe the narrative is predictable and builds up to a specific dilemma that makes most sense in the situation. The AI-generated output that resembles the original output could imply that AI is capable of producing similar dilemmas to human-authored ones. What this study thus more so explores is whether AI is able to generate good alternative choices rather than whether AI can achieve the same level as the original. In this study the AI-generated output was more constrained. The way the prompt is formatted is also influential. You could, for example, steer the AI in a better direction that would encourage more emotional choices. Another constraint in this study with the AI generation is that a single prompt strategy was used. There was not any iteration or fine-tuning done. Using multiple prompts or refining outputs could lead to different results.

Additionally, a limitation of this study is the limited number of scenarios used. There were only three scenarios included. This was necessary to keep the survey at a reasonable length. The low amount of scenarios also means that the variety of scenarios is not as broad. This could limit the results as some people may experience different scenarios more intensely as others. Besides, for this study only scenarios from one single game were used. Thus, the results may be specific to the game’s specific narrative style and may not be generalizable to other games.

Another limitation is the use of a survey. In the survey, Likert items were used, which are inherently open to interpretation. The terminology used to assess the dilemmas is also subjective and open to interpretation. For example, participants may have varying interpretations of what “emotionally engaging” or a “meaningful” dilemma means. As a result, this study does not measure objective quality, but rather participants’ perceived experiences of the dilemmas.

The sample size of this study ($N=16$) is also a limitation. With a relatively small sample size, it is difficult to detect smaller effects. This results in lower statistical power of the findings. This may explain why some differences in the results did not reach statistical significance. Besides, the small sample size limits the generalisability of the results. There was also limited demographic information collected about the participants. This makes it difficult to assess how representative the sample is.

Taken together, these limitations highlight that the findings should be interpreted within the specific context and design of this study.

6.3 Future Research

Future research could build on the findings of this study in several ways.

First of all, future research could improve the methodological design of this study. A larger sample

size would increase statistical power. Studies with experimental designs with more control could also provide clearer insights. This could for example be achieved by doing participant observation or focus groups to collect richer data. These research methods would also be more suitable for doing a study where participants are playing a real game instead of reading text. Measuring the experience during real gameplay would likely provide a more accurate representation of how dilemmas are experienced.

Another addition to this study could be collecting more qualitative data. This study mainly focused on quantitative self-report measures. Adding more qualitative data to this could provide further and deeper insights. For example, asking participants the reason why a dilemma felt more meaningful or conflicting could help better understand the experiences. Real-time methods (for example, letting participants think-aloud) could also give deeper insight into the experience process.

Another direction for future research is to further explore the possibilities of generating dilemmas with AI. For example, this could include experimenting with different prompt formats, using multiple prompts, or applying iterative prompting to refine the output. Comparing AI models could also provide insight into which models perform better with generating dilemmas. Besides, research between AI recreating existing dilemmas and AI generating new dilemmas could also be valuable. This could give insight into whether AI can generate on a level that is comparable to human-written narrative dilemmas or if AI is more effective in generating alternative variations.

Future research could also extend this study by including scenarios from different games. These could be other games in the narrative games genre, but games of other genres could also be used. Furthermore, including a broader variety of moral dilemmas could help determine if this study’s findings are more generally applicable. Additionally, comparing dilemmas from multiple human-written games could provide a good baseline for evaluating the AI-generated content.

7 Conclusion

In short, this study explored how AI-generated dilemmas for in-game scenarios compare to human-written ones in terms of perceived meaningfulness, narrative quality, emotional engagement, and players’ decision-making experience. When letting the AI generate with minimal limitations, the AI-generated dilemmas were identical to those from the original game. This suggests that AI may be capable of creating similar dilemmas to human-written ones. However, this would not allow for a meaningful comparison. Thus, this study instead focused on how AI compares when it is “forced” to generate entirely new alternatives for the dilemmas.

The results showed that the original game dilemmas were evaluated higher overall. The strongest difference was found in the participants’ decision-making experiences. The other differences were still in favor of the OG dilemmas, but they were small and not statistically significant. When participants were asked to directly compare the dilemmas, there was a slight preference seen for the OG ones. However, this difference was also not statistically significant.

These findings indicate that when AI is used to generate alternative dilemmas based on an existing

scenario, the results are partially comparable to human-written ones. However, they appear to be less effective in creating a strong decision-making experience.

Overall, these findings suggest that AI can support narrative game design. However, it may need human refinement to create truly impactful and meaningful choices. Future research is needed to better understand how AI can improve in capturing the emotional depth and decision-making experience of narrative game dilemmas.

References

- [Alh25] S. A. Alharthi. Generative AI in Game Design: Enhancing Creativity or Constraining Innovation? *Journal of Intelligence*, 13(6):60, 2025. doi: 10.3390/jintelligence13060060.
- [Ana18] A. Anable. *Playing with Feelings: Video Games and Affect*. University of Minnesota Press, 2018. ISBN: 9781517900250.
- [BB09] R. Busselle and H. Bilandzic. Measuring Narrative Engagement. *Media Psychology*, 12(4):321–347, 2009. doi: 10.1080/15213260903287259.
- [BKZ⁺24] S. Buongiorno, L. Klinkert, Z. Zhuang, T. Chawla, and C. Clark. PANGeA: Procedural Artificial Narrative Using Generative AI for Turn-Based, Role-Playing Video Games. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 20, pages 156–166, 2024. doi: 10.1609/aiide.v20i1.31876.
- [BMA⁺19] J. A. Bopp, L. J. Müller, L. F. Aeschbach, K. Opwis, and E. D. Mekler. Exploring Emotional Attachment to Game Characters. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play, CHI PLAY ’19*, pages 313–324, New York, NY, USA, 2019. Association for Computing Machinery. doi: 10.1145/3311350.3347169.
- [Don15] Dontnod Entertainment. *Life is Strange*, January 2015. Video game, published by Square Enix. <https://lifeisstrange.square-enix-games.com/en-us/games/life-is-strange-remastered-collection>.
- [HBH14] S. P. Hernandez, V. Bulitko, and E. St. Hilaire. Emotion-Based Interactive Storytelling With Artificial Intelligence. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 10, pages 146–152, Oct. 2014. doi: 10.1609/aiide.v10i1.12721.
- [HR23] S. Harmon and S. Rutman. Prompt Engineering for Narrative Choice Generation. In L. Holloway-Attaway and J. T. Murray, editors, *Interactive Storytelling*, volume 14383 of *Lecture Notes in Computer Science*, pages 208–225, Cham, 2023. Springer. doi: 10.1007/978-3-031-47655-6_13.
- [IdKP13] W. A. IJsselsteijn, Y. A. W. de Kort, and K. Poels. *The Game Experience Questionnaire*, 2013.
- [Ope25] OpenAI. ChatGPT. <https://chat.openai.com/>, 2025. Large language model. Model version used: ChatGPT 5.1.
- [Pin24] F. Pinto. Beyond the Screen: Exploring the Player’s Role and Its Impact on Emotional Experiences in Video Games. *International Journal of Games and Social Impact*, 2(2):28–56, 2024. doi: 10.60543/ijgsi.v2.n2.02.
- [RG25] G. Raffloer and M. C. Green. Of Love & Lasers: Perceptions of Narratives by AI Versus Human Authors. *Computers in Human Behavior: Artificial Humans*, 5:100168, 2025. doi: 10.1016/j.chbah.2025.100168.

- [Tuc24] Tom Tucek. Enhancing Empathy Through Personalized AI-Driven Experiences and Conversations with Digital Humans in Video Games. In *Companion Proceedings of the 2024 Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY Companion '24, pages 446–449, New York, NY, USA, 2024. Association for Computing Machinery. doi: 10.1145/3665463.3678856.
- [WHZ⁺25] Y. Wen, C. Huang, H. Zhou, Z. Zeng, C. M. L. Po, J. Togelius, T. Merino, and S. Earle. All Stories Are One Story: Emotional Arc Guided Procedural Game Level Generation. arXiv preprint arXiv:2508.02132, 2025.

Appendix A: AI Prompts and Generated Outputs

Note: All AI-generated dilemmas were produced using ChatGPT 5.1. Below the final prompts used to generate alternative dilemmas are shown with the corresponding AI-generated outputs.

Prompt 1

Do not reference external media or game titles.
You may not identify or name the source of this scene.

The following choices are forbidden:

- Report Nathan
- Hide the truth

Scene:

Max Caulfield is an 18-year-old photography student at Blackwell Academy in Arcadia Bay, the small town where she lives and goes to school. She has just discovered she can rewind time after witnessing a classmate, Nathan, pull a gun on a girl in the restroom. By rewinding and triggering the fire alarm, Max prevents the shooting. Nathan storms out, followed moments later by the girl he threatened. The girl turns out to be Chloe Price, Max's estranged childhood best friend. Once the bathroom is empty, Max steps into the hallway, shaken. The principal immediately notices her distressed state and stops her. Seeing that something has clearly happened, he asks Max to explain what's going on.

What are two entirely new major choice options Max is confronted with next?

Answer:

Output for Prompt 1

1. Confess the Impossible: Max decides to tell the principal that something supernatural happened in the restroom—that time seemed to rewind, that she relived the scene, and that this is why she knew to pull the fire alarm. She doesn't name who had the gun, but insists she's not making it up and begs him to believe her and get her help. This path risks the principal treating her as unstable, sending her to counseling, contacting her parents, or putting her under tighter supervision, but it also opens the door to adults knowing something truly strange is happening.

2. Claim a Vision of Disaster: Max hides any mention of time manipulation or specific people, but tells the principal she pulled the alarm because she had a vivid, overwhelming vision of someone being shot and the school descending into chaos. She frames it as a terrifying premonition or panic attack, saying she'd rather be punished for a false alarm than ignore it. This path steers attention toward her mental state and the possibility of looming danger on campus, potentially triggering psychological evaluations, increased monitoring, and new protocols at the school—all without directly exposing what really happened in the restroom.

Prompt 2

Do not reference external media or game titles.
You may not identify or name the source of this scene.

The following choices are forbidden:

- Go to the police
- Look for proof

Scene:

After a humiliating video spreads across campus, Kate Marsh becomes increasingly withdrawn and upset. Kate is a quiet and deeply religious classmate of Max. Max visits her in her dorm room, finding Kate anxious, scared, and desperate for support. Kate confides that something happened at a recent party, but she doesn't fully remember the details. She believes she may have been drugged, and her family is already judging her harshly because of the video. Kate turns to Max for guidance, asking what she should do next.

What are two entirely new major choice options Max is confronted with next?
Answer:

Output for Prompt 2

1. Encourage Kate to Step Away from Campus: Max suggests that Kate temporarily withdraw from school or at least take an extended break, prioritizing her safety and mental health over everything else. She offers to help Kate talk to the administration about a leave of absence, arrange for extensions on her work, and coordinate with dorm staff so Kate isn't pressured or harassed. Choosing this means Max is guiding Kate toward retreat and recovery first—removing her from the toxic environment—even though it might delay confronting what happened and could be seen by others as “running away.”

2. Support Kate in Telling Her Story Publicly on Her Own Terms: Max encourages Kate to reclaim control of the situation by sharing her own account in a way she chooses: a written statement, a recorded message, or speaking at a campus meeting or faith group. Max promises to help her craft the message, stay by her side, and push back against shaming by emphasizing consent, the possibility of drugging, and Kate's right to be treated with compassion. This choice directs them toward openly challenging the rumor mill and social judgment, risking more attention and backlash, but potentially shifting the narrative from humiliation to solidarity and advocacy.

Prompt 3

Do not reference external media or game titles.
You may not identify or name the source of this scene.

The following choices are forbidden:

- Steal the money
- Don't steal the money

Scene:

Late at night, Max and Chloe break into the school principal's office to search for information about a missing student. While rummaging through files and drawers, they stumble upon a large envelope inside a cabinet. Inside is a significant amount of cash clearly intended for the school fund for disabled students. Chloe sees the money as a way to pay off a debt she owes to a slightly dangerous man, while Max realizes that taking it would mean stealing resources meant for students who need them.

What are two entirely new major choice options Max is confronted with next?

Answer:

Output for Prompt 3

1. Document the Cash and Turn It into Leverage Against the Administration:

Max stops Chloe from touching the envelope and instead carefully photographs the money, the school fund label, and the office surroundings as quiet evidence. She suggests they leave the cash exactly where it is but use what they've found to anonymously alert the school board, parents, or even local media later, questioning why such a vulnerable fund is hidden and unsecured. This choice shifts the focus away from Chloe's debt toward exposing possible negligence or corruption, potentially blowing open a larger scandal that could help or hurt their investigation into the missing student.

2. Refuse to Use the Fund and Decide to Face the Debtor Head-On: Max tells Chloe that using the fund is off-limits and instead proposes they deal with the dangerous creditor directly, without touching the envelope at all. She offers to go with Chloe to confront him, try to negotiate more time, trade information, or find another arrangement that doesn't involve harming disabled students. This choice sacrifices the "easy" solution in favor of personal risk and moral consistency, potentially deepening Max's involvement with dangerous people and pulling their focus away from the school's secrets in the process.

Appendix B: Standardized Survey Scenarios and Dilemmas

Note: The dilemmas are labeled here as "Original" and "AI-generated" for clarity. These labels were not shown to participants during the survey. The order in which dilemmas were presented was randomized.

Scenario 1

Sam is an 18-year-old photography student. They have just discovered they can rewind time after witnessing a classmate, Jordan, pull a gun on someone in the restroom. By rewinding and triggering the fire alarm, Sam prevents the shooting. Jordan storms out, followed moments later by the person Jordan threatened. This person turns out to be Casey, Sam's estranged childhood best friend. Once the restroom is empty, Sam steps into the hallway, shaken. The principal immediately notices their distressed state and stops them. Seeing that something has clearly happened, the principal asks Sam to explain what's going on.

Original Dilemma

- **Option 1: Report Jordan**

Sam tells the school principal that Jordan brought a gun into the restroom.

- **Option 2: Hide the truth**

Sam avoids explaining what really happened in the restroom.

AI-Generated Dilemma

- **Option 1: Confess the impossible**

Sam tells the principal that they experienced time rewinding and acted on it.

- **Option 2: Claim a vision of disaster**

Sam explains they pulled the fire alarm due to a panic attack and a terrifying sense that something was about to go wrong.

Scenario 2

A humiliating video of a fellow student, Morgan, spreads across campus. After this, Morgan becomes increasingly withdrawn and upset. Morgan is a quiet and deeply religious classmate of Sam. Sam visits Morgan in their dorm room, finding Morgan anxious, scared, and desperate for support. Morgan confides that something happened at a recent party, but they don't fully remember the details. Morgan believes they may have been drugged, and their family is already judging them harshly because of the video. Morgan turns to Sam for guidance, asking what they should do next.

Original Dilemma

- **Option 1: Go to the police**

Sam encourages Morgan to report what happened at the party to the police.

- **Option 2: Look for proof**

Sam suggests trying to find evidence about what happened.

AI-Generated Dilemma

- **Option 1: Take a break from campus**
Sam suggests that Morgan temporarily leaves school to focus on their safety and wellbeing.
- **Option 2: Share their story publicly**
Sam encourages Morgan to publicly share their own account of what happened.

Scenario 3

After using their rewind powers to alter a key moment in the past, Sam returns to a drastically changed present. Casey, instead of being the rebellious person Sam knows, has grown up in this timeline severely injured and living with significant medical needs. Sam spends the day with Casey, learning how much Casey's condition has shaped their and their family's lives. That evening, in Casey's room, the atmosphere becomes heavier as Casey opens up about their daily pain, the burden they feel they place on their parents, and the emotional toll of their condition. With tears in their eyes, Casey turns to Sam and asks Sam to help them end their life.

Original Dilemma

- **Option 1: Accept**
Sam agrees to help Casey end their life.
- **Option 2: Refuse**
Sam refuses Casey's request for help in ending their life.

AI-Generated Dilemma

- **Option 1: Reveal the timeline secret**
Sam tells Casey about their powers and the altered timeline, and suggests trying to change events again to reduce Casey's suffering.
- **Option 2: Involve Casey's parents**
Sam tells Casey they cannot decide alone and asks to include Casey's parents in the conversation about Casey's request.

Appendix C: Survey

Age Check

1. Before we start, please confirm your age.

- ☐ Yes, I am 18 or older
- ☐ No, I am under 18

Introduction

Welcome and thank you for taking part in this study!

In this survey, you will read three short scenarios based on a story-driven game. After each scenario, you will be presented with different dilemmas that could occur at that point in the story.

The goal of this study is to explore how people perceive and experience dilemmas in narrative games. These are games where the story is important and your choices can change what happens. You will evaluate the dilemmas based on how meaningful, emotionally engaging, and coherent they feel within the story according to you.

The survey takes approximately 15–20 minutes to complete. Your responses are anonymous, and you may stop at any time without giving a reason. You can always continue later with the survey if you wish.

If you have any questions about this study, you can contact the researcher at: <e-mail address>

Content notice: Some scenarios may include sensitive themes such as gun violence, drugging, and suicidal themes. If you feel uncomfortable at any point while completing the survey, you are free to stop at any time.

(If you are experiencing suicidal thoughts, consider reaching out for support. In the Netherlands, 113 Zelfmoordpreventie offers confidential help at www.113.nl or by calling 113.)

Background Questions

1. How familiar are you with narrative games (story-focused games where player choices matter)?
(1 = Not at all familiar – 5 = Very familiar)
2. How familiar are you with the game *Life is Strange*?
(1 = Not at all familiar – 5 = Very familiar)
3. How familiar are you with the game *Strumming Shadows*?
(1 = Not at all familiar – 5 = Very familiar)

Task Explanation

In the following sections you will read three short scenarios based on moments in a game.

Imagine that you are playing the game as the main character, **Sam**.

For each scenario, you will see two different dilemmas, presented one after the other. Each dilemma represents a possible way the game could offer choices at that moment in the game.

After each dilemma, you will evaluate the choices. Once you've done that for both dilemmas, you will compare them.

There are no right or wrong answers. Please respond based on your first impressions.

Scenario and Dilemma Procedure

Note: This procedure was repeated for each of the three scenarios. The order in which the dilemmas (original vs. AI-generated) were presented was randomized.

Step 1: Scenario Introduction

You will now start with a scenario.

Please read the scenario carefully and imagine that you are experiencing the described moment in the game playing as Sam.

Step 2: First Dilemma Evaluation

Scenario X:

[Scenario text]

At this point in the game, you are presented with the following dilemma:

- Option 1: [Option text]
- Option 2: [Option text]

Evaluation Questions

Rate how much you agree with the following statements (1 = Strongly disagree – 5 = Strongly agree):

1. The choices feel meaningful to the story.
2. The choices feel like they would have an impact on the story.
3. The choices reflect the characters' personalities well.
4. The choices fit the tone of the story.
5. The choices fit logically within the narrative context.
6. The choices are consistent with the situation and characters.
7. The choices make me feel something about the situation or characters.
8. The choices feel emotionally impactful.
9. The choices make me feel connected to the characters involved.

Decision Questions

1. Which option would you pick if you were playing the game?
☐ Option 1

☐ Option 2

Rate how you experienced making this decision (1 = Not at all – 5 = Very much):

2. I felt confident choosing between the options.
3. I felt tense while making this decision.
4. I felt conflicted between the options.

Step 3: Second Dilemma Evaluation

You will now see the same scenario again as a reminder, but with a different dilemma.

Scenario X:

[Scenario text]

At this point in the game, you are presented with a different dilemma:

- Option 1: [Option text]
- Option 2: [Option text]

Evaluation Questions

(Same as above)

Decision Questions

(Same as above)

Step 4: Dilemma Comparison Questions

You have now evaluated two different dilemmas for the same scenario.

The scenario and both dilemmas are shown again for reference.

Scenario X:

[Scenario text]

Dilemma A:

- Option 1: [Text]
- Option 2: [Text]

Dilemma B:

- Option 1: [Text]
- Option 2: [Text]

Comparison Questions

Compare the dilemmas:

1. Which dilemma felt more emotionally engaging?
(Dilemma A / Dilemma B / No preference)
2. Which dilemma felt more meaningful?
(Dilemma A / Dilemma B / No preference)
3. Which dilemma fit the narrative context better?
(Dilemma A / Dilemma B / No preference)
4. Which dilemma felt more coherent with the story and characters?
(Dilemma A / Dilemma B / No preference)
5. If only one dilemma could appear in the actual game, which would you choose?
(Dilemma A / Dilemma B)

Step 5: Next Scenario

You will now proceed to the next scenario.

End of Survey

Thank you sincerely for taking the time to participate in this survey!

Feedback and email

1. Any additional comments or feedback? (Optional)
2. If you would like to receive a copy of the finished thesis, you may leave your email address below. (Optional)

Your responses are anonymous and will be used only for this study. Individual answers will not be shared. Only summarized, anonymous results may appear in the researcher's thesis.

If you have any questions about the study, or if you would like your data to be removed, you can contact the researcher at: <e-mail address>

Appendix D: Raw Collected Data

Table 8 below presents the raw survey responses for all 16 participants.
The survey questions correspond to the table rows as follows:

- **Background questions:**
 - Familiar with narrative games (1-5) → Background Question 1
 - Familiar with “Life is Strange” (1-5) → Background Question 2

- Familiar with the fake game (1-5) → Background Question 3
- **Evaluation questions:**
 - Meaningfulness 1 (1-5) → Evaluation Question 1
 - Meaningfulness 2 (1-5) → Evaluation Question 2
 - Narrative Quality 1-4 (1-5) → Evaluation Questions 3–6
 - Emotional Engagement 1-3 (1-5) → Evaluation Questions 7–9
 - Decision (1 or 2) → Decision Question 1
 - Decision Experience 1*-3 (1-5) → Decision Questions 2–4 (note: *Decision Experience 1 must be inverted before calculations)
- **Comparison questions:** 1–4 (1, 2, or 3) → Comparison Questions 1–4 (1 = OG dilemma, 2 = AI dilemma, 3 = No preference)
- **Dilemma preference:** (1 or 2) → Comparison Question 5 (1 = OG dilemma, 2 = AI dilemma)

All Likert-scale items are rated from 1–5. Decision questions are coded according to the participant’s choice (1 or 2) in the dilemma. Decision Experience 1* must be inverted before analysis due to reversed item wording.

Table 8: Raw survey data for all participants (1-16). Decision experience scores marked with * need to be inverted before analysis.

Question / Participant	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Familiar with narrative games (1-5)	5	4	5	3	2	4	2	1	4	4	3	3	2	3	5	5
Familiar with “Life is Strange” (1-5)	5	2	5	5	1	4	1	1	4	3	3	4	1	1	5	2
Familiar with fake game (1-5)	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	2
Scenario 1 – OG Dilemma																
Meaningfulness 1 (1-5)	5	4	5	4	4	5	5	5	5	5	5	5	4	4	5	5
Meaningfulness 2 (1-5)	5	5	5	5	5	5	5	5	5	5	5	5	5	4	5	4
Narrative quality 1 (1-5)	5	4	4	4	3	4	5	4	4	3	4	5	4	4	5	4
Narrative quality 2 (1-5)	5	4	5	5	4	4	5	4	5	5	5	5	3	4	4	4
Narrative quality 3 (1-5)	3	4	5	3	5	4	4	5	5	4	4	5	4	4	5	5
Narrative quality 4 (1-5)	5	4	5	4	4	5	4	4	5	4	4	5	3	4	4	4
Emotional engagement 1 (1-5)	5	3	5	5	3	2	5	3	4	3	5	5	5	4	5	5
Emotional engagement 2 (1-5)	3	4	5	5	4	4	5	4	5	5	4	5	4	4	5	3
Emotional engagement 3 (1-5)	4	4	5	5	2	4	5	3	5	4	4	4	4	4	5	4
Decision (1 or 2)	1	1	1	2	1	1	1	1	2	1	1	1	2	1	2	1
Decision experience 1* (1-5)	5	2	5	4	4	4	2	2	5	4	4	3	2	2	4	1
Decision experience 2 (1-5)	3	3	4	3	4	2	2	4	4	3	3	4	4	4	4	1
Decision experience 3 (1-5)	1	4	2	4	4	2	4	4	1	4	4	4	4	4	4	1
Scenario 1 – AI Dilemma																
Meaningfulness 1 (1-5)	5	5	4	2	5	4	4	5	5	5	4	5	2	5	3	3
Meaningfulness 2 (1-5)	5	5	5	2	5	4	5	5	5	5	4	5	4	3	5	4
Narrative quality 1 (1-5)	5	4	3	4	4	4	4	3	5	4	3	3	4	4	3	2
Narrative quality 2 (1-5)	5	3	5	3	5	4	4	4	5	4	4	4	3	5	2	4
Narrative quality 3 (1-5)	5	4	4	4	4	4	5	5	5	5	4	2	2	5	3	4
Narrative quality 4 (1-5)	4	4	3	4	4	4	4	4	5	5	5	4	3	4	3	4
Emotional engagement 1 (1-5)	5	4	5	3	4	5	5	5	5	3	4	5	4	4	2	3
Emotional engagement 2 (1-5)	5	4	5	4	4	3	3	4	5	4	4	4	3	4	2	4
Emotional engagement 3 (1-5)	5	4	5	3	4	4	4	4	5	4	4	3	2	4	2	4
Decision (1 or 2)	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2
Decision experience 1* (1-5)	4	5	4	2	2	1	4	4	5	5	2	4	4	5	2	4

Continued on next page

Table 8: (continued)

Question / Participant	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Decision experience 2 (1-5)	2	2	5	3	3	5	2	2	4	2	2	2	3	2	2	2
Decision experience 3 (1-5)	1	1	2	5	4	5	4	3	1	1	4	1	2	2	2	1
Scenario 1 – Comparison																
Comparison 1 (1, 2, or 3)	3	2	1	1	2	2	1	1	3	2	2	1	1	1	1	1
Comparison 2 (1, 2, or 3)	1	1	1	1	1	1	1	2	3	2	2	1	1	1	1	1
Comparison 3 (1, 2, or 3)	3	2	1	1	2	3	1	2	3	2	1	1	1	1	1	1
Comparison 4 (1, 2, or 3)	1	2	1	3	2	2	1	2	3	2	2	1	1	1	1	3
Dilemma preference (1 or 2)	1	2	1	1	2	1	1	2	1	2	1	1	1	1	1	1
Scenario 2 – OG Dilemma																
Meaningfulness 1 (1-5)	3	4	5	5	3	4	5	5	5	4	4	5	5	2	5	4
Meaningfulness 2 (1-5)	3	5	5	5	3	4	5	4	5	5	4	5	5	4	5	4
Narrative quality 1 (1-5)	5	4	5	4	3	4	5	3	4	3	4	4	3	2	5	5
Narrative quality 2 (1-5)	3	4	5	5	4	5	5	3	5	4	5	5	3	2	4	4
Narrative quality 3 (1-5)	3	4	5	4	4	4	5	4	4	4	5	5	3	3	4	5
Narrative quality 4 (1-5)	3	4	5	4	4	4	4	4	4	2	5	5	4	2	4	4
Emotional engagement 1 (1-5)	4	4	5	5	2	4	5	3	4	4	5	4	5	2	4	5
Emotional engagement 2 (1-5)	4	4	4	5	3	4	4	3	5	3	4	4	4	4	4	4
Emotional engagement 3 (1-5)	3	4	5	5	3	5	4	3	4	4	4	4	4	3	5	5
Decision (1 or 2)	1	1	2	1	2	2	2	2	2	2	2	2	2	2	1	2
Decision experience 1* (1-5)	5	4	1	2	2	4	4	1	4	2	4	5	5	2	2	2
Decision experience 2 (1-5)	1	2	5	4	3	2	2	4	4	4	3	1	4	4	4	3
Decision experience 3 (1-5)	1	3	5	5	4	3	2	4	2	4	2	4	3	4	4	4
Scenario 2 – AI Dilemma																
Meaningfulness 1 (1-5)	5	5	3	3	4	4	4	5	2	3	4	4	5	2	5	2
Meaningfulness 2 (1-5)	5	4	5	4	4	4	5	5	5	4	4	4	5	4	5	4
Narrative quality 1 (1-5)	5	4	3	2	4	4	4	4	5	3	4	3	3	2	5	2
Narrative quality 2 (1-5)	5	3	4	1	4	3	4	5	4	3	3	4	2	3	5	2
Narrative quality 3 (1-5)	5	4	4	2	4	3	4	4	4	4	4	4	3	2	5	4
Narrative quality 4 (1-5)	5	4	5	3	4	4	4	4	4	4	3	3	2	3	5	4
Emotional engagement 1 (1-5)	5	4	4	2	3	3	4	5	4	2	4	4	5	4	5	4
Emotional engagement 2 (1-5)	5	4	4	2	4	4	4	4	5	4	5	3	5	4	5	4
Emotional engagement 3 (1-5)	5	4	3	2	3	3	4	4	4	4	4	2	4	4	5	4
Decision (1 or 2)	2	1	1	2	2	2	2	2	1	2	1	1	1	1	1	1
Decision experience 1* (1-5)	1	2	2	3	4	4	2	5	5	4	3	5	2	2	4	4
Decision experience 2 (1-5)	5	3	3	1	2	1	2	2	4	1	4	1	4	4	4	1
Decision experience 3 (1-5)	5	3	4	3	2	2	4	2	1	2	4	1	4	4	2	4
Scenario 2 – Comparison																
Comparison 1 (1, 2, or 3)	2	3	2	1	2	1	2	1	1	1	2	1	1	1	3	1
Comparison 2 (1, 2, or 3)	2	1	2	3	2	1	1	2	1	1	1	1	1	2	1	1
Comparison 3 (1, 2, or 3)	2	1	3	1	2	1	1	2	1	1	1	1	1	1	3	1
Comparison 4 (1, 2, or 3)	2	3	2	1	2	1	1	2	1	1	1	1	1	1	1	1
Dilemma preference (1 or 2)	2	1	2	1	2	1	1	2	1	1	1	1	1	1	1	1
Scenario 3 – OG Dilemma																
Meaningfulness 1 (1-5)	5	5	5	4	3	3	4	5	5	2	5	5	5	5	4	5
Meaningfulness 2 (1-5)	5	5	5	4	4	4	3	5	5	5	5	5	5	5	4	5
Narrative quality 1 (1-5)	5	5	3	4	4	4	3	4	4	3	4	4	3	3	3	4
Narrative quality 2 (1-5)	5	4	5	5	4	4	3	3	4	4	5	4	4	2	3	4
Narrative quality 3 (1-5)	5	4	4	4	5	4	4	4	4	4	5	4	4	3	4	5
Narrative quality 4 (1-5)	5	5	4	4	4	4	4	4	4	3	4	4	3	3	3	3
Emotional engagement 1 (1-5)	5	5	5	5	3	4	3	2	4	5	4	5	4	5	4	4
Emotional engagement 2 (1-5)	5	5	5	5	2	4	4	2	5	5	5	5	5	5	3	5
Emotional engagement 3 (1-5)	5	4	5	5	2	4	3	2	4	4	5	5	3	4	4	5
Decision (1 or 2)	2	2	1	1	2	2	2	2	2	2	2	2	2	2	2	2
Decision experience 1* (1-5)	1	2	2	2	5	1	4	4	2	5	1	5	4	5	2	5
Decision experience 2 (1-5)	3	3	5	5	3	5	3	1	4	2	5	5	5	1	4	4
Decision experience 3 (1-5)	5	4	5	4	2	5	4	1	2	2	5	5	2	1	4	4
Scenario 3 – AI Dilemma																
Meaningfulness 1 (1-5)	5	4	3	3	4	4	4	5	5	5	5	3	5	3	4	4
Meaningfulness 2 (1-5)	5	5	2	4	5	4	4	4	5	5	5	4	5	4	4	5

Continued on next page

Table 8: (continued)

Question / Participant	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Narrative quality 1 (1-5)	4	4	5	3	4	3	4	5	4	4	4	4	3	3	4	4
Narrative quality 2 (1-5)	5	5	2	4	5	3	3	5	4	4	5	3	4	4	4	4
Narrative quality 3 (1-5)	3	5	3	4	4	2	3	5	4	4	5	2	4	4	4	5
Narrative quality 4 (1-5)	3	4	3	4	4	3	4	5	4	4	5	3	3	4	4	4
Emotional engagement 1 (1-5)	5	5	4	4	4	2	4	5	4	4	5	4	4	4	4	5
Emotional engagement 2 (1-5)	5	5	5	5	4	3	3	5	5	5	5	4	5	4	3	4
Emotional engagement 3 (1-5)	4	5	3	4	4	2	3	5	4	4	5	1	3	4	4	5
Decision (1 or 2)	1	1	2	2	1	1	1	1	1	1	1	2	1	2	1	1
Decision experience 1* (1-5)	5	5	5	4	4	5	5	5	5	5	2	1	5	3	4	4
Decision experience 2 (1-5)	1	2	1	2	3	1	2	1	2	2	4	3	3	4	4	5
Decision experience 3 (1-5)	1	1	1	3	2	1	2	1	5	2	4	5	2	4	2	4
Scenario 3 – Comparison																
Comparison 1 (1, 2, or 3)	1	3	1	3	2	1	2	1	3	1	1	1	1	1	2	1
Comparison 2 (1, 2, or 3)	1	2	1	1	2	2	2	2	2	2	2	1	1	2	2	2
Comparison 3 (1, 2, or 3)	1	2	1	3	2	3	2	2	2	2	1	1	1	2	2	2
Comparison 4 (1, 2, or 3)	1	3	1	3	2	3	2	2	2	2	2	1	1	2	2	3
Dilemma preference (1 or 2)	1	2	1	1	2	2	2	2	2	2	2	1	1	1	2	2