

UNDERSTANDING CONGRUENCE BETWEEN PLAYER
AND CHARACTER BELIEFS IN DIGITAL GAMES

A dissertation submitted to the
College of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Computer Science
University of Saskatchewan
Saskatoon

By
Jason T. Bowey

©Jason T. Bowey, August 2021. All rights reserved.

Unless otherwise noted, copyright of the material in this thesis
belongs to the author.

PERMISSION TO USE

In presenting this dissertation in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this dissertation in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my dissertation work or, in their absence, by the Head of the Department or the Dean of the College in which my dissertation work was done. It is understood that any copying or publication or use of this dissertation or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my dissertation.

DISCLAIMER

Reference in this dissertation to any specific commercial products, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement, recommendation, or favoring by the University of Saskatchewan. The views and opinions of the author expressed herein do not state or reflect those of the University of Saskatchewan, and shall not be used for advertising or product endorsement purposes.

Requests for permission to copy or to make other uses of materials in this dissertation in whole or part should be addressed to:

Head of the Department of Computer Science
176 Thorvaldson Building, 110 Science Place
University of Saskatchewan
Saskatoon, Saskatchewan S7N 5C9 Canada

OR

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan S7N 5C9 Canada

ABSTRACT

Game researchers and designers both know that a key element of making a game enjoyable is for the player to identify with the game’s protagonist. A large majority of research focuses on studying how the physical appearance of the protagonist influences player identification. We propose that elements of the character’s personality are equally, if not more, important when creating bonds between player and protagonist.

When approaching protagonist design, there is a trade-off between character depth and identification. In order to have more engaging and deep protagonists, certain subsets of players will not identify with the protagonist due to conflicting beliefs and values. On the other hand, game designers can make characters shallow and generic to avoid isolating players, but this leads to a less interesting narrative experience.

One solution to this problem is to tailor the beliefs and values exhibited by the protagonist to align with those of each individual player. In order to achieve this, however, the game needs to have a way of knowing the real-world beliefs of the player. This can be accomplished through validated questionnaires, but this is not ideal for players who are playing a game for enjoyment.

In this work we address the problem that we don’t have validated methods of measuring real-world beliefs of the player through game mechanics and make the following main contributions:

1. Understand how a congruence (and incongruence) of beliefs between the player and protagonist relates to identification and enjoyment
2. Create and test the effectiveness of using text-based narrative prototypes to evaluate narrative elements of games
3. Develop and evaluate a method of using a player’s in-game dialogue choices to construct a behaviour model of the player
4. Demonstrate that in-game dialogue choices can be used to accurately measure the player’s real-world beliefs

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Regan, for her patience, wisdom, and guidance throughout my studies. Also, my co-authors and lab-mates for your contributions, feedback, and friendship.

I would like to thank my wife, Emma, for your love and support. Also, my Daughter, Naomi, for your silly distractions.

CONTENTS

Permission to Use	i
Abstract	ii
Acknowledgements	iii
Contents	iv
List of Tables	viii
List of Figures	x
List of Abbreviations	xi
1 INTRODUCTION	1
1.1 Problem and Motivation	1
1.2 Solution	3
1.3 Methodological Approach	5
1.3.1 Experimental Design	6
1.3.2 Validated Questionnaires	6
1.3.3 State and Trait Variables	6
1.3.4 Statistical Tests	7
1.3.5 Technical Terms	7
2 MANUSCRIPT A: DON'T TALK DIRTY TO ME: HOW SEXIST BELIEFS AFFECT EXPERIENCE IN SEXIST GAMES	9
2.1 Introduction to Manuscript A	9
2.1.1 Publication and Individual Contribution	10
2.2 Abstract	10
2.3 Introduction	10
2.4 Related Work	12
2.4.1 Gender Roles and Sexism in Games	13
2.4.2 Player Experience	14
2.4.3 Transportation and Identification	15
2.5 Experiment Design	16
2.5.1 Game	16
2.5.2 Prestudy	19
2.5.3 Cover Story	19
2.5.4 Measures	19
2.5.5 Participants and Deployment Platform	20
2.5.6 Procedure	21
2.5.7 Follow-Up Study	21
2.5.8 Data Analyses	21
2.6 Results	22
2.6.1 Gender and Sexist Beliefs	22
2.6.2 The Role of Identification in Predicting Enjoyment	22
2.6.3 The Role of Gender and Sexist Beliefs on Identification	23
2.6.4 The Role of Sexist Beliefs in Predicting Enjoyment	24
2.7 Discussion	25
2.7.1 Summary of Findings	25

2.7.2	Explanation of Findings	25
2.7.3	Limitations and Future Work	28
2.8	Conclusion	29
3	MANUSCRIPT B: THOSE ARE NOT THE STORIES YOU ARE LOOKING FOR: USING TEXT PROTOTYPES TO EVALUATE GAME NARRATIVES EARLY	31
3.1	Introduction to Manuscript B	31
3.1.1	Publication and Individual Contribution	32
3.2	Abstract	32
3.3	Introduction	32
3.4	Related Work	34
3.4.1	Player Experience	34
3.4.2	Narrative Experience	35
3.4.3	Iterative Development in Interfaces and Games	36
3.5	Industry Approaches to Game Narratives	37
3.5.1	Integrated Development of Story and System	38
3.6	Constructing a Narrative Prototype	39
3.7	Study 1	40
3.7.1	Game Design	40
3.7.2	Measures	42
3.7.3	Results	43
3.7.4	Study 1 Discussion	44
3.8	Study 2	45
3.8.1	Game Design	45
3.8.2	Measures	47
3.8.3	Results	47
3.8.4	Study 2 Discussion	48
3.9	Discussion	49
3.9.1	Implication of Results	49
3.9.2	Industry Use Cases	49
3.9.3	How to Integrate Prototypes into the Design Process	49
3.9.4	Limitations and Future Opportunities	50
3.10	Conclusion	51
4	MANUSCRIPT C: RED OR BLUE PILL: FOSTERING IDENTIFICATION AND TRANSPORTATION IN RPGS	54
4.1	Introduction to Manuscript C	54
4.1.1	Publication and Individual Contribution	54
4.2	Abstract	55
4.3	Introduction	55
4.4	Related Work	57
4.4.1	Perceived Agency and Control	57
4.4.2	(Tailoring) Meaningful Choices in Games	58
4.4.3	Narrative Experience	58
4.5	System Description	59
4.5.1	Morality Model	59
4.5.2	Narrative System	59
4.5.3	Dialogue Choices	60
4.6	Recruitment, Outliers, and Bot Detection	62
4.7	Study 1	62
4.7.1	Measures	63
4.7.2	Data Analyses	64
4.7.3	Results	64

4.7.4	Study 1 Discussion	65
4.8	Study 2	66
4.8.1	Methods	66
4.8.2	Results	67
4.8.3	Study 2 Discussion	68
4.9	Study 3	68
4.9.1	Methods	68
4.9.2	Results	69
4.9.3	Study 3 Discussion	69
4.10	General Discussion	70
4.10.1	Summary of Results	70
4.10.2	Explanation of Results	70
4.10.3	Reliability of the Prediction Model for Tailored Choices	71
4.10.4	Autonomy vs. Agency & Trait vs. State	71
4.10.5	Game Design Considerations	72
4.10.6	Limitations	72
4.10.7	Future Research Directions	73
4.11	Conclusion	74
5	MANUSCRIPT D: PREDICTING BELIEFS FROM NPC DIALOGUES . . .	75
5.1	Introduction to Manuscript D	75
5.1.1	Publication and Individual Contribution	75
5.2	Abstract	76
5.3	Introduction	76
5.4	Related Work	77
5.4.1	In-game Assessments: Dialogue and Choices in Games	77
5.4.2	Player Choice in Dialogues and Games	78
5.4.3	Gender and Sexism in Games	79
5.4.4	Gender and Sexism in Avatar Identification	80
5.4.5	Summary	80
5.5	Methods: Dialogue Creation and User Study	80
5.5.1	Game/Narrative	81
5.5.2	Dialogue Creation	81
5.5.3	Prestudy	81
5.5.4	Main Study Design	82
5.5.5	Participants	83
5.6	Results	83
5.6.1	Did they play as themselves or an external character?	83
5.6.2	Do the dialogue models predict players' real-world beliefs?	84
5.6.3	Do participants prefer the Direct or Indirect methods for dialogue creation?	84
5.7	Discussion	85
5.7.1	Summary of Results	85
5.7.2	Game Design Applications	86
5.7.3	Using Sexism as an Alignment Measure	87
5.7.4	Ethics and Privacy	87
5.7.5	Limitations and Future Opportunities	88
5.8	Conclusion	88
6	GENERAL DISCUSSION	91
6.1	Summary of Results	91
6.1.1	Manuscript A	91
6.1.2	Manuscript B	91
6.1.3	Manuscript C	92
6.1.4	Manuscript D	94

6.2	Contributions	95
6.2.1	Industry Contributions	96
6.2.2	Methodological Contributions	97
6.3	Ethics	97
6.3.1	Ethics of Data Collection	97
6.4	Reflections on Manuscripts	98
6.4.1	Manuscript A	98
6.4.2	Manuscript B	99
6.4.3	Manuscript C	100
6.4.4	Manuscript D	101
7	CONCLUSION	102
	References	104
	Appendix A Manuscript A Consent Form	119
	Appendix B Manuscript B Consent Form	120
	Appendix C Manuscript C Consent Form	121
	Appendix D Manuscript D Consent Form	122
	Appendix E Ambivalent Sexism Inventory Questionnaire Items	124
	Appendix F Homophily Questionnaire Items	125
	Appendix G Intrinsic Motivation Inventory Questionnaire Items	126
	Appendix H Male Role Attitudes Questionnaire Items	127
	Appendix I Moral Foundations Questionnaire	128
	Appendix J Narrative Engagement Questionnaire Items	129
	Appendix K Player Experience of Needs Satisfaction Questionnaire Items	130
	Appendix L Player Identification Scale Questionnaire Items	131
	Appendix M Sense of Agency Questionnaire Items	132
	Appendix N Transportability Questionnaire Items	133
	Appendix O Transportation Questionnaire Items	134

LIST OF TABLES

3.1	Study 1 results - Means and SD for Low-Fidelity and High-Fidelity conditions, as well as F, p, and β values from a Multivariate Analysis of Variance (MANOVA) for Condition, Transportation, and the interaction.	53
3.2	Study 2 results - Means and SD for Low-Fidelity and High-Fidelity conditions, as well as F, p, and β values from a Multivariate Analysis of Variance (MANOVA) for Condition, Transportation, and the interaction.	53
4.1	Estimated Means \pm SE for Studies 1,2, & 3. Values of co-variates for each study are shown in bottom right.	69
5.1	Interactive dialogue for the Direct condition, derived from the Male Role Attitudes Scale.	89
5.2	Interactive dialogue for the Indirect condition, derived from the Male Role Attitudes Scale.	89
5.3	Means and standard deviations for all measures used in the main study.	90
5.4	Regressions results with explained variance at the second level (R^2), standardized regression coefficients (β), and p values for regressions predicting questionnaire scores, using dialogue-model scores (DMS), dummy-coded condition (direct vs indirect, C), and the interaction ($DMS * C$).	90
E.1	Individual questionnaire items from the Ambivalent Sexism Inventory. Questionnaire presented in a Likert scale ranging from <i>Strongly Disagree</i> to <i>Strongly Agree</i> . Reverse-coded questionnaire items indicated by (-).	124
F.1	Individual questionnaire items from the Homophily questionnaire. Questionnaire presented in a Likert scale ranging from <i>Strongly Disagree</i> to <i>Strongly Disagree</i>	125
G.1	Individual questionnaire items from the Intrinsic Motivation Inventory questionnaire. Questionnaire presented in a Likert scale ranging from <i>Strongly Disagree</i> to <i>Strongly Agree</i> . Reverse-coded items indicated by (-).	126
H.1	Individual questionnaire items from the Male Role Attitudes questionnaire. Questionnaire items presented in a Likert scale ranging from <i>Strongly Disagree</i> to <i>Strongly Agree</i>	127
I.1	Individual questionnaire items from the Moral Foundations Questionnaire. Questionnaire items marked with * were presented in a Likert scale ranging from <i>Not at all Relevant</i> to <i>Extremely Relevant</i> . Otherwise, items presented on Likert scale ranging from <i>Strongly Disagree</i> to <i>Strongly Agree</i>	128
J.1	Individual questionnaire items from the Narrative Engagement questionnaire. Questionnaire items presented on Likert scale ranging from <i>Strongly Disagree</i> to <i>Strongly Agree</i> . Reverse-coded items indicated by (-).	129
K.1	Individual questionnaire items from the Player Experience of Needs Satisfaction questionnaire. Questionnaire items presented on Likert scale ranging from <i>Strongly Disagree</i> to <i>Strongly Agree</i> . Reverse-coded items indicated by (-).	130
L.1	Individual questionnaire items from the Player Identification questionnaire. Questionnaire items presented on Likert scale ranging from <i>Strongly Disagree</i> to <i>Strongly Agree</i>	131

M.1	Individual questionnaire items from the Sense of Agency questionnaire. Questionnaire items presented on Likert scale ranging from <i>Strongly Disagree</i> to <i>Strongly Agree</i>	132
N.1	Individual questionnaire items from the Transportability questionnaire. Questionnaire items presented on Likert scale ranging from <i>Strongly Disagree</i> to <i>Strongly Agree</i> . Reverse-coded items indicated by (-).	133
O.1	Individual questionnaire items from the Transportation questionnaire. Questionnaire items presented on Likert scale ranging from <i>Strongly Disagree</i> to <i>Strongly Agree</i> . Reverse-coded items indicated by (-).	134

LIST OF FIGURES

2.1	Screenshots from the three main levels in Fantasy Adventure: Hack n' Slash (left), Infinite Runner (middle), and Shooting Gallery (right).	16
2.2	Neutral (left) and gendered (right) characters. From left to right: Harlot, Knight, Barbarian, Princess.	18
2.3	Distribution of sexist beliefs by player gender.	22
2.4	Linear regression lines showing showing the prediction of sexist beliefs on avatar identification for the three game versions.	24
3.1	Screenshots for low-fidelity (left) and high-fidelity (right) versions of Space Adventure	41
3.2	low-fidelity (top-left), high-fidelity Level 1 (top-right), Level 2 (bottom-left), and Level 3 (bottom-right) Screenshots	46
4.1	Screenshot of the interface design of dialogue system used in Manuscript C studies.	60
6.1	Mediation relationships as computed in Manuscript A. Arrows indicate a significant mediation and dotted lines indicate an insignificant mediation.	92
6.2	Mediation relationship between Choice, Effort, and Identification as computed in Manuscript C. Arrows indicate a significant mediation and dotted lines indicate an insignificant mediation.	94
6.3	Mediation relationship between Choice, Effort, and Transportation as computed in Manuscript C. Arrows indicate a significant mediation and dotted lines indicate an insignificant mediation.	94

LIST OF ABBREVIATIONS

ANOVA	Analysis of Variance
GDC	Game Developer's Conference
GUR	Games User Research
HiFi	High-Fidelity
HIT	Human Intelligence Task
HUD	Heads Up Display
ID	Identification
IMI	Intrinsic Motivation Inventory
LoFi	Low-Fidelity
MANOVA	Multivariate Analysis of Variance
MRAS	Male Role Attitudes Scale
MTurk	Amazon Mechanical Turk
NES	Narrative Engagement Scale
NPC	Non-Player Character
PENS	Player Experience of Need Satisfaction Scale
PIS	Player Identification Scale
RPG	Role Playing Game
QA	Quality Assurance
SDT	Self Determination Theory

1 INTRODUCTION

1.1 Problem and Motivation

In media, identification between the audience and the protagonist is a fundamental factor in producing an engaging experience that transports the audience into the world in which the narrative is taking place. In interactive media, in which the audience member is actively participating in the experience, this concept of identification with the protagonist is especially important because the player is directly involved in the narrative, as opposed to passively experience a written work or a film.

According to literature, there are three main ways in which an audience member can identify with the protagonist in media [64]: embodied presence (experience the media through the protagonist’s eyes), wishful identification (seeing the protagonist as an idealistic version of themselves), and similarity identification (seeing similarities between themselves and the protagonist).

While much of the literature surrounding the concept of identification relates to static media (i.e., text and film) there has been a recent interest in the concept relating to interactive media among games researchers [32, 61, 163, 258].

The majority of this work among games researchers focuses on the physical similarities between player and protagonist [13, 76, 121, 173, 256] (e.g., the character looks like the player, therefore the player experiences higher similarity identification), rather than deeper psychological similarities such as personality. As a result of this, previous work has shown that allowing players to customize their character’s appearance and personality positively contributes to identification with the character, and through that increased identification and increased enjoyment of the game [255].

This method of avatar customization is now commonly used in a wide variety of games; the mechanic is most commonly found in role playing games (RPGs) such as the Mass Effect series¹, Elder Scrolls V: Skyrim², and Pokémon: Sword/Shield³, but avatar customization is also found in other genres of games such as The Sims 4⁴. In these types of games, the protagonist is often a blank slate both visually and in terms of personality, allowing the player to project their own feelings and experiences onto the protagonist to maximize identification.

¹Mass Effect, Electronic Arts, 2007-2017

²Elder Scrolls V: Skyrim, Bethesda Softworks, 2011

³Pokemon: Sword / Pokemon: Shield, Nintendo, 2019

⁴The Sims 4, Electronic Arts, 2014

Even among games in which the player cannot traditionally customize the protagonist’s appearance, such as the Super Mario game series⁵ and The Legend of Zelda series⁶, the protagonist is narratively boring in terms of their personality and motivations, and their appearances are cartoony and generic, allowing the possibility for a wide variety of players to project themselves onto the character. While the popularity of these games show that this is an effective way to make players connect with their protagonists, having flat characters limits the narrative potential of a story, and the importance of character depth and richness has been previously argued [8].

Some games on the other hand, have protagonists with a much more defined personality. Protagonists in games like Gears of War⁷, Halo⁸, and God of War⁹ have well-defined protagonists expressing hypermasculine traits, both physically and psychologically. Conversely, games such as Tomb Raider¹⁰ and Bayonetta¹¹ contain hyperfeminine protagonists. Other games have protagonists that lie more in the middle of the gender spectrum but are still strongly fleshed out characters in terms of their personality, motivations, and experiences, such as Gordon Freeman from the Half-Life series¹², Joel and Ellie from The Last of Us¹³, and Cal Kestis from Star Wars Jedi: Fallen Order¹⁴. The thing all of these characters have in common is the fact that they are not a blank slate upon which the player can project their own beliefs and experiences. These characters are all, for the most part, physically and psychologically designed by the developers, giving players very few options to customize the protagonists. Despite this lack of freedom, these games are all critically acclaimed (Half Life received a review score of 9.5/10 [3], The Last of Us received a score of 10/10 [4], and Jedi: Fallen Order received a score of 9/10 [5]), and many have spawned several sequel titles following the protagonists on further adventures.

These two methods of designing protagonists leads to a tradeoff between allowing players to project themselves onto the protagonist (designing characters as a blank slate) and having a more fleshed out character providing a more complex and interesting narrative. The problem with the latter option goes back to identification: if the character’s appearance and personality are set in stone by the developers, while this may provide a more complex and interesting story, many players will not identify with the protagonist, which may limit elements of player experience such as enjoyment.

Allowing players to customize aspects of the protagonist’s appearance allows them to more strongly identify with the character, but that pushes the game back along the spectrum towards

⁵Super Mario, Nintendo, 1985-2015

⁶The Legend of Zelda, Nintendo, 1986-2021

⁷Gears of War, Xbox Game Studios, 2006

⁸Halo: Combat Evolved, Xbox Game Studios, 2001

⁹God of War, Sony Computer Entertainment, 2005

¹⁰Tomb Raider, Square Enix, 1996

¹¹Bayonetta, Sega, 2009

¹²Half-Life, Valve Corporation, 1998

¹³The Last of Us, Sony Interactive Entertainment, 2013

¹⁴Star Wars Jedi: Fallen Order, Electronic Arts, 2019

the blank slate. One solution would be to find a middle ground between the two extremes: allow the player to customize some parts of their character to increase their identification with the protagonist while allowing the parts relevant to the game narrative to remain fixed. Another solution is for the game to know something about the player in advance and tailor the protagonist in such a way that the player will more strongly identify with them, without needing to explicitly specify how they want the character to look and behave.

Video games do not currently provide players with both a psychologically complex protagonist (i.e., well-defined beliefs and values) while simultaneously providing appeal to a wide range of players through identification. One solution is that game designers could measure player's real-world beliefs and tailor the protagonist's design to better align with this model. However, measuring player beliefs, personality or values is challenging inside of a video game because the current best practices of modelling these types of individual measures are limited to gathering data through lab-based questionnaires. These questionnaires, while effective, are not suited to be integrated into an entertainment product such as a video game since they are boring and repetitive to complete. As such, the specific problem we address in this dissertation is how game designers can effectively measure players' real-world beliefs inside of a game setting, and how tailoring characters to player's beliefs affects game experience. There are many ways in which beliefs and values influence character identification, but in video games there is a prevalent and historical divide on how gendered characters are designed and perceived. As such, we use sex and gender as the domain in which we explore tailoring.

1.2 Solution

In order to effectively measure player's real-world beliefs inside a game setting, there are several sub-problems that must be investigated:

1. **Understand identification in pre-authored protagonists as driven by congruence of individual beliefs and demographics**

As discussed in the previous section, most of the research into player identification focuses primarily on physical appearances: if the protagonist looks similar to the player, or expresses an idealized version of the player's appearance, they will strongly identify with that character [13, 76, 256]. While this is true, this approach only captures part of how a player can identify with the protagonist. As described in the previous section, a character's beliefs, values, and backstory can greatly contribute to the way in which a player identifies with the character (e.g., if the protagonist holds similar beliefs as to those held by the player, the player should identify strongly with that protagonist).

In order to create characters that the player strongly identifies with, we must better understand how players identify with pre-authored characters (i.e., characters with a fixed appearance and pre-determined beliefs and values). What role do demographic factors, such as gender, play in identification? How does the congruence or incongruence of beliefs between the player and the protagonist affect identification?

In particular, this will focus on beliefs relating to gender, such as benevolent sexism [106] and male role attitudes [62].

2. Investigate the importance of graphic fidelity in game narratives

A problem tangentially connected to the overall problem of how to create complex characters that a wide range of players can identify with is to understand the role graphic fidelity plays in the identification process. We know that giving players the freedom to customize their own avatar can increase identification [32, 255] but there is no research, to the best of our knowledge, examining narrative experience isolated from graphic fidelity. As an example, in order to study identification through character personality and beliefs, current methodologies would require researchers to construct a graphically complex game with a custom dialogue system along with engaging mechanics. While researchers could use this to compare different dialogue techniques by keeping everything else in the game fixed, the problem is that this approach requires a large investment in terms of development time, prohibiting most researchers from making progress.

I propose and test a better methodology for evaluating narrative-specific aspects of games by using text-prototypes to measure player experience without the need for graphically complex games.

3. Understand identification when players are allowed to customize the protagonist's personality through in-game choices

As discussed previously, a large amount of the research into player identification focuses on allowing players to physically customize their characters. Other types of customization are commonly found in games, most often found in role playing games. These customization mechanics often involve either allocating skill points after acquiring a required amount of experience points or making narrative decisions on behalf of the protagonist. While allocating skill points could play a factor in how the player identifies with the protagonist (e.g., allocating points to the strength skill because the player sees themselves as strong) these stats almost always play a strong role in the actual game mechanics (e.g., the strength skill often determines how powerful the protagonist's attack is) and therefore splits the player's motivations between making the character more like them and effectively playing the game.

Dialogue choices, on the other hand, are the primary way that role playing games allow players to exert their influence over the game world, and through this agency, the player is able to shape their character's personality. Unlike allocating skill points, which are usually tightly coupled to the game mechanics, these types of dialogue choices drive the direction of the narrative but do not usually affect the actual gameplay.

In this step I use the text-prototypes proposed in Step 2 to provide players with the opportunity to customize the protagonist's morality through dialogue choices, similar to how games allow players to customize their character's appearance through an avatar creation menu.

The main goal of this step is to better understand the mechanisms of how dialogue choices affect overall player experience. In particular, I construct a system that uses the player's dialogue choices to model their in-game behaviour and uses this model to feed back into the game and dynamically change the morality of the provided dialogue options. I construct models to show the relationships between the choices themselves (i.e., how varied the possible options are, the amount of effort players perceive putting into making the decision) and narrative experience (i.e., identification with the protagonist, transportation into the narrative world).

4. Design and test alternative approaches for measuring player's real-world beliefs

In Step 1 I examine how the design of the protagonist's personality affects individual players' identification, through whether or not the player's beliefs are congruent with those displayed by the protagonist. In it, I use external questionnaires to measure the player's real-world beliefs. In an actual game scenario, however, players would not be willing to fill out pages of questionnaires prior to playing a game. Thus, I address the problem of how to accurately measure individual players' real-world beliefs through observed behaviour within the game itself.

Specifically, I design an alternative way of measuring the player's real-world beliefs using dialogue choices embedded in the game, as opposed to using external questionnaires. I show that the dialogue choices, which will be derived from the validated questionnaires, can effectively measure the player's real-world beliefs to a similar degree of accuracy as the validated questionnaires. Additionally, I demonstrate that the majority of players made choices consistent with their own personal beliefs and values, as opposed to role playing as an external character.

1.3 Methodological Approach

This section will describe the general methodological approaches used throughout this dissertation. In particular, I will include descriptions of commonly used statistical tests, experimental structures

and methods, and clarify technical terms used.

1.3.1 Experimental Design

The research presented in this dissertation primarily uses online studies and follows a between-subjects experimental design. In these experiments participants are assigned to one of several experimental conditions which in turn corresponds to a particular version of the primary stimulus (in most cases, a specific version of a game or interactive system). In these experiments, participants fill out basic demographic questionnaires prior to playing through the game or interactive system. After engaging with the system, participants fill out subjective, validated questionnaires to capture their perceived experience on various aspects of play experience such as enjoyment, identification with the protagonist, or transportation into a narrative world, to name a few.

1.3.2 Validated Questionnaires

Validated questionnaires are usually created to ask the same thing multiple different ways (e.g., asking the player if they felt competent while playing the game, and later asking them if they felt capable and effective while playing). This approach provides researchers with a confidence that the data gathered is accurate, since participants whose responses for a particular measured construct are deemed internally inconsistent may be removed from further data analysis. A single variable is computed for each construct by calculating the mean average of the participant's responses for the appropriate questions. Sometimes, validated questionnaires include an item that is opposite of the sentiment behind the construct being measured (e.g., asking the player if they didn't feel competent while playing the game). This practice is called *reverse coding*, and the player's response is reversed prior to calculating the variable.

1.3.3 State and Trait Variables

Variables measured through validated questionnaires, as described in the previous section, can be categorized as one of two types: state or trait variables. A state variable represents a construct that describes a snapshot of experience in response to an internal or external stimulus. Conversely, a trait represents a construct that remains relatively stable in a participant in the long term, regardless of stimuli. In the research presented in this dissertation, both state and trait variables are used. State variables are generally used to compare experiential measures between multiple conditions (e.g., compare enjoyment between condition A and condition B) to determine the effects the experimental conditions had on the variable in question. In experimental terminology, these outcome variables are referred to as dependent variables, since the values they hold may be dependent on the experimental condition.

Trait variables, on the other hand, are used in several ways. These variables are sometimes used as a way of subdividing the data set into multiple groups (e.g., compare between individuals who are highly susceptible to narrative immersion and those who are not). These variables can also be used as covariates in statistical tests to account for possible variations in these measures between conditions. The practice of using covariates in this way can often improve the accuracy and reduce error in the final statistical model [1]. In experimental terminology, these variables are referred to as independent variables, since their value is not expected to vary between experimental conditions.

1.3.4 Statistical Tests

Here I briefly summarize the statistical tests that were used in the research presented in this dissertation.

Two statistical tests used commonly in this dissertation are the analysis of variance (ANOVA) and the multivariate analysis of variance (MANOVA). An ANOVA is a statistical test that compares the means of a single dependent variable from two or more groups to determine whether there is a statistically significant difference between any two groups in the statistical model [86]. Similarly, a MANOVA tests the same thing but considers multiple dependent variables in a single test [86]. As mentioned in the previous section, covariates may be added to these models to improve their accuracy and reduce the chance of errors.

Another commonly used statistical test in this dissertation is linear regression. Linear regression is a method for determining if one variable (X) predicts another variable (Y) [86]. For example, this test can be used to determine whether one experience variable, such as avatar identification, predicts another experience variable, such as enjoyment. Another way of describing this is through correlation: when players experience higher avatar identification they also experience higher enjoyment.

Related to linear regression, moderated mediation analyses were also conducted. This type of analysis investigates whether a latent mediating variable (M) explains the variance predicted between a predictor (X) and a dependent measure (Y) [118]. In other words, if linear regressions indicate that both M and X predict Y, and X predicts M, mediation analysis provides a way of representing these relationships in a single model and better understand the causation between them. A significant mediation in this case would indicate that even though X predicts Y, the reason for this relationship is due to X predicting M, which in turn predicts Y. In this scenario, the model would be described as M mediating the relationship between X and Y.

1.3.5 Technical Terms

There are several technical terms used commonly throughout this dissertation, and some of these terms are used in different ways depending on the context. For these reasons, these terms are

described here to clarify their use throughout the dissertation.

The term *model* is used in several different ways throughout this dissertation. In the context of statistical tests, this term is used to describe the relationship between variables as a result of a statistical test. For example, a significant relationship as a result of a linear regression, as described in the previous section, would be described as a model, in which one variable is shown to predict another.

Alternatively, the term model is sometimes used in a computational sense, as a description of a data structure used in a piece of software or interactive system. This use of the term is primarily used in Manuscript C and D, when describing the data structure that is created dynamically in response to a player's dialogue choices.

A pair of terms used commonly throughout this dissertation are *belief* and *value*. These two terms are generally used together, and describe something that an individual holds to be true, sometimes through a lack of evidence and sometimes in opposition to evidence. These beliefs may be privately held (i.e., the individual holds them to be true but they do not physically act on them) or they may be openly held (i.e., the individual considers them when taking actions). The main beliefs and values used in this dissertation are sexist beliefs, primarily benevolent sexism [105] and male role attitudes [206], which represents beliefs relating to traditional gender roles for women and men, respectively.

2 MANUSCRIPT A: DON'T TALK DIRTY TO ME: HOW SEXIST BELIEFS AFFECT EXPERIENCE IN SEXIST GAMES

2.1 Introduction to Manuscript A

As described in the Introduction, games often contain characters that are almost entirely designed by the developers, with very little customization options provided to the players engaging with the game. While this has the potential to create more complex and interesting narratives, the fact that these characters' appearances and personalities are static creates the possibility that players very different from the presented protagonist may not identify as strongly with the playable character, and through this experience reduced enjoyment.

Understanding the mechanisms relating to this, namely whether an incongruence between the player's appearance and beliefs with those exhibited by the protagonist results in reduced enjoyment of the game, is important for game designers to understand when creating protagonists. For example, many games create hyper-masculine or hyper-feminine protagonists whose appearances reflect very few members of their target audience. Alternatively, a game whose protagonist holds and exhibits sexist beliefs throughout the game could potentially isolate a large demographic of the game's potential audience that do not agree with the protagonist's beliefs. In both of these cases, it is important to understand how this congruence, or lack thereof, between player and protagonist affects identification with the protagonist, and through this identification, how enjoyment of the game is affected.

In this chapter we show that player identification predicts overall enjoyment. We also show that gender congruence between the player and the playable character does not significantly affect identification or enjoyment. Conversely, we show that belief congruence between the player and playable character significantly influences both player identification and enjoyment, and that player identification significantly mediates the relationship between belief (sexism) and enjoyment. Finally, our results suggest that narrative elements are more salient for studying player identification than graphic fidelity.

2.1.1 Publication and Individual Contribution

This work was published as the following: [40]. Although a multi-authored contribution, I was responsible for leading the research. My roles included: designing the experiments and constructing the interactive systems used in the experiments (i.e., building the game, modelling the character’s clothing for the different gendered conditions, and the character rating system used in the prestudy to determine the perceived genders of characters). Additionally, I was involved in the data analysis and contributed to the literature review and writing of the publication. I also presented the publication at the conference at which it was published. I contributed approximately 65% of the total work for this project and publication.

2.2 Abstract

Research on sexism in digital games has suggested that women self-select out of playing sexist games; however, assuming a homogeneous gender-based response does not account for the diversity of identities within a gender group. Gender-incongruent responses to recent events like #gamergate implies that the gender of the participants is not paramount to experience, but that their beliefs about gender roles are. To explore the role of sexist beliefs on experience in sexist games, we created three versions of a game that were identical except for the presence of sexist imagery and/or dialogue. We show that enjoyment of sexist games is not predicted by player gender, but by the player’s pre-existing beliefs about gender. Furthermore, avatar identification is the pathway through which enjoyment is facilitated. Finally, sexist dialogue does not improve the play experience for anyone - rather it harms experience for players of all genders who do not hold sexist beliefs.

2.3 Introduction

Sex sells. In the context of digital games – which is a \$100 billion industry worldwide [79, 265] – this translates into many popular games that portray female characters as hypersexual (e.g., Red Dead Redemption [74]) or portray male characters as hypermasculine (e.g., Gears of War [98]). Female characters are often treated as rewards to be obtained (e.g., The Witcher [210]), as damsels in distress who need to be rescued (e.g., The Legend of Zelda [192]), as a prop in service of the male protagonist’s characters development - often referred to as “woman in the refrigerator” (e.g., God of War [239], [240]), or simply as background decoration (e.g., Hitman:Absolution [124]) [223, 224]. And even when the main character is a strong female with a developed personality and attributes, her body and clothing (or lack thereof) is often hypersexual (e.g., Lara Croft [72], Bayonetta [99]). Previous literature has argued that sexist characters roles can be attractive and useful in fostering avatar identification [119, 130]; however, other literature suggests that female players are offended

by highly sexualized female characters and high levels of violence [116, 275], may be irritated by sexist design choices intended to target male gamers [61], and in many cases may make an active choice to not play games with sexist content [88, 193].

Feminist game critics have long criticized the game industry for catering to the presumed straight male player [56, 133, 223, 224]. The vitriolic response from a group of mostly male gamers escalated into threats of rape, violence, and even death against women who spoke out against sexism in games and the game industry [19, 182, 273]. The debate around sexism in the games industry got massive public exposure through the recent #gamergate and #notyourshield controversies [241], which helped to shed light on many aspects of sexism within games and the game industry, including how players – particularly women – experience sexist game content [59]. Discussions around sexist game content sometimes suggest an ‘us vs them’ dynamic between male and female gamers [55]; however, the issue is not as simple as one of pitting men against women. For example, many men oppose the sexist rhetoric of #gamergate supporters [241] – an open letter in support of diversity in the game community was started by a male feminist and was quickly signed by 2500 people in response to #gamergate [277]. Conversely, there are women who challenge the idea of women being offended by sexist games and support #gamergate as a fight for masculinity and for male spaces in games [36, 234, 235]. That men and women do not react to rhetoric around sexism in games as two distinctly homogenous groups is not surprising, as feminist scholars have long suggested that it is a mistake to divide men and women into two *dichotomous* groups and assume corresponding attitudes, rather than allowing them to choose and perform their own individual identity and beliefs that might include a *dualistic* balance of the masculine and feminine [50, 71, 193, 236].

As applied to an understanding of play experience, this all suggests that it may not be as simple as suggesting that women self-select out of playing games with sexist content, imagery, or themes [90, 193]. We propose in this paper that the gender of the participant is not paramount to game experience, but that their *beliefs about gender roles and sexism* are. To explore the relationships between gender, sexist beliefs, and sexism in games, we conducted an experiment to determine how people who hold a range of sexist beliefs respond to games that contain sexist content.

We created a short role-playing game based on a prologue, 3 playable levels with a single common mechanic in each (i.e., fighting, infinite runner, shooting), and an epilogue. We created a game (rather than adapting an off-the-shelf game) so that we could completely control and isolate the source of the sexist content in the game. To create sexist game versions, we added sexist imagery – i.e., characters that visually reinforced sexist stereotypes (helpless damsel, sexualized harlot, hyper-masculine barbarian, manly hero) – in one version and both sexist imagery and dialogue – i.e., through using gendered pronouns, sexual innuendo, diminutive nicknames, and objectification – in another.

Participants played one game version and answered several questionnaires about their beliefs around gender roles, sexist beliefs, and play experience. As it has been suggested that sexist characters can aid in avatar identification [119,130], and that avatar identification facilitates game enjoyment [119], we measured avatar identification and resulting game enjoyment. We have four main findings.

- The gender of the player does not predict avatar identification; however, their beliefs about gender roles and benevolent sexism do predict avatar identification.
- Lower avatar identification translates into less game enjoyment.
- Sexist dialogue harms play experience: adding sexist dialogue harms experience for people who do not hold sexist beliefs, and also does not improve experience for those who do.
- Including visually sexist character stereotypes does not harm identification and enjoyment; however, sexist character stereotypes reinforced through both imagery and dialogue harms both identification and enjoyment.

Our work explored player response to sexist characters and behaviour in a role-playing game. We refute the notion that men and women form distinct groups that respond in a predictable manner to sexist content in games. Player response is determined not by their gender, but by the beliefs about gender that they hold.

2.4 Related Work

We use *sex* to refer to the biological sex of a character or player, and *gender* to refer to their identification as male, female, or non-binary [203]. Because we asked participants their gender, rather than their sex, we refer to the players as having genders. *Gender roles* describe beliefs about appropriate behaviour based on gender [105,206] and *sexist beliefs* refers to beliefs that are based on sex and/or gender.

Sexist content in games can be derived from multiple sources, including from graphics, interactions, narratives, or themes. The presence of sexual content does not by definition make a game sexist; we use sexism to refer to content (sexual or otherwise) that promotes gender inequality [129,212] between male and female characters or players. Gender inequality is generally derived from sex-based differences in power (i.e., the ability to influence the behaviour or another person), violence (i.e., behaviour intended to harm another person who tries to avoid being harmed), or objectification, which either emphasized someone’s instrumentality (i.e., treating someone as a tool for achieving one’s own goals) or dehumanizes (i.e., denying that people have thoughts and feelings) them [145]. As such, we use *sexism* similarly to [90] to refer to *attitudes and actions that are*

based on gender and/or sex and promote inequality between men and women. Although sexism is often of a hostile nature (e.g., violence toward women), sexism is also often of a benevolent nature; for example, the view that women are more helpful, empathetic, or intimacy-seeking than men or that women require paternal protection [105]. Although framed in subjectively positive traits, this benevolent form of sexism is an attitude of inequality based on gender or sex [105] and is sexist as it places women on a pedestal while simultaneously reinforcing their subordination [244].

2.4.1 Gender Roles and Sexism in Games

Male characters are represented in games much more often than female [78,127,271]. When female characters are portrayed in video games, they are often depicted in sexualized ways such as partially nude, with an unrealistic body image, or wearing sexually revealing clothing or inappropriate attire [48,75,127]. This tendency towards sexualizing women in video games has repeatedly led to criticism of the game industry for only catering towards presumed straight male players [88,109,224]. Research on the consequences of sexualized imagery indicates that men show a greater likelihood to harass women after play [75,274]. The harassment towards female users are usually sexist remarks ranging from traditional sexism (e.g., “get back in the kitchen”) to sexual harassment (e.g., “show me your tits”) [88]. Fox and Tang [90] conclude that sexualized imagery therefore might promote the growing problem of harassment of women in games. In addition to this, research suggests that female users experience diminished self-efficacy [22] and increased self-objectification [88,89] after play. The effect that sexualized imagery of women has on both men and women has led to the conclusion that, when exposed to these portrayals, women may self-select out of such environments [88], which in turn reinforces the idea of games being a “man’s world” dominated by masculine disclosure [220].

Issues of masculinity in games and the games industry have been explored by feminist scholars for several decades [132]. A common perspective focuses on how to make games more gender-inclusive to be attractive to girls as a leisure activity [56,132]. Based on the notion that games provide early exposure and comfort with digital technologies, the arguments often centre around how the male-centric games industry disadvantages girls from gaining that early exposure to STEM topics [55,65] and from the stealth-learning that digital games provide [133]. Books, such as Graner Ray’s *Gender Inclusive Game Design: Expanding the Market (2003)* [209] present arguments on how to design games that appeal to the playstyle and preferences of female players. Cassell and Jenkins’ *From Barbie to Mortal Kombat* (1998) [56] similarly focused on the emergence of the ‘girls game’ movement, albeit with a more inclusive perspective of gender preferences.

Beyond this foundational work on female gamers, literature on sexism in games can be distilled into three main trends. First, the majority of the research focuses on how women experience sexism in games differently than men. The distinguishing variable that influences the effect of sexist content

in games is almost always gender conceptualized as the binary “men vs. women” [78,88,271]. This dichotomous approach asserts that ‘women’ are a group with a set of similar characteristics and beliefs that is fundamentally different from the group ‘men’. Contemporary theories about gender might argue that this dichotomous distinction is reinforcing a deterministic view of gender relations in which people are divided into clear-cut groups (i.e., women and men [50], rather than a dualistic distinction in which people embody varying amounts of the masculine and feminine, regardless of their biological sex [193]. Furthermore, it may be more relevant to distinguish players who adhere to the concepts of gender that sexist games elicit (e.g., strong, dominant men and sexualized, helpless women) from those who don’t. Adherence to these more traditional concepts of gender might be a better predictor for who is bothered by sexist game content and who is not.

Second, most studies investigated sexualized imagery [78]. While the visual depiction of female characters is certainly a powerful example of how games depict genders differently, there are other aspects of games that are, while being more subtle, just as sexist. Though imagery is an important part of games, it should not be the only factor considered when examining sexist content. In particular, the structure of a game’s narrative, or the story trope, is crucial because of its power to objectify male and female characters [153]. In addition to the structure of tropes, language itself can have a powerful effect on sexism. Swim et al. [244] considered the role of sexist language and how it relates to subtle sexism. McConnel et al. [175] showed that adding male suffixes to occupation titles led perceivers to interpret a social target’s personality as more masculine. Finally, language is a dominant means by which stereotypes are defined, communicated, and assessed [165], so language and dialogue in media should not be overlooked when it comes to sexist content and the effects it can have.

Third, to our knowledge, all studies investigating the effect of sexist content in games considered consequences in areas like tolerance for sexual harassment [75], beliefs about benevolent sexism [274], and agreement with gender attitudes and stereotypes [22], rather than directly examining the effects of sexism on game experience. In our work, we propose to examine the role of sexist game content on player experience, as it is the driving factor behind game enjoyment and greatly influences purchasing decisions.

2.4.2 Player Experience

Research in understanding the experience of players draws from traditions in psychology, games user research, and human-computer interaction. The more prevalent model that explains player experience is the satisfaction of needs during play [219], which is part of the broader context of self-determination theory [217]. Self-determination theory suggests that we are intrinsically motivated to engage in an activity because of our pure desire to do it, outside of the context of any external rewards [216] and has been shown to describe the experience of people playing games [219].

Intrinsic motivation is comprised of our interest-enjoyment in an activity, the effort-importance we invest on the activity, our perceived competence with the activity, and how much tension-pressure we feel during the activity [174]. Measurement of motivation is accomplished using the Intrinsic Motivation Inventory (IMI) [174]. The satisfaction of our need for competence, autonomy, and relatedness (i.e., mastering challenges under our own volition, while feeling connected to others) work together to foster intrinsic motivation [216]. This relationship has been shown in a range of domains, including video game play [246]. As such, these constructs are often used in evaluating game experience, including the experience of different types of people during game play [34,136,207], of different game mechanics [101], or of different game input devices [31,232,266]. Measurement of need satisfaction is often accomplished using the Player Experience of Need Satisfaction Scale (PENS) [219].

In translating research on need satisfaction to the play context, Ryan et al., [219] added two additional constructs of particular relevance to game play. Intuitive controls describe control of the game that feels natural, whereas immersion/presence describes the experience of being transported into the digital environment [131]. Other researchers have described immersion as being comprised of three components [26]: transportation [110] (i.e., narrative engagement, loss of awareness, and flow); identification [64] (i.e., identifying with a character’s virtual personality and representation); and telepresence [164] (i.e., feeling present inside of a virtual world). Because the immersive aspects of transportation and identification are particularly relevant to the experience of sexism in games [119,130], we describe them in more detail.

2.4.3 Transportation and Identification

Transportation is often described using a traveler metaphor, in which an audience member becomes absorbed and lost in the world created by a piece of media [102]. The audience member loses their sense of self as mental systems become focused on the narrative, excluding attention to the outside world. Similar to transportation, transportability is a personality trait defined as an individual’s susceptibility to be transported into a narrative [70]. Previous work has confirmed that transportability predicts transportation [70], as well as enjoyment [27]. In addition, transportation has been shown to be a causal factor in increased identification with story characters [70]. Identification is a cognitive process through which audience members experiencing a piece of media adopt the point of view of a particular character. The process of identification has been described as a process in which the audience member loses awareness of self to be temporarily replaced with emotional and cognitive connections to a particular character [64].

The process begins with the audience member either: adopting a character’s perspective, feeling fondness towards a character, or realizing some similarity between themselves and the character. As the narrative progresses, the audience member adopts the goals of the identified character and

empathizes with the events of the narrative. Previous work has shown that higher identification with a character in a game leads to higher levels of enjoyment [119]. Also, greater time spent engaging with a story has been shown to result in stronger identification [259].

It has been previously hypothesized that narratives could be used in persuasion [26] and identification has been proposed as an important candidate to be effective at creating persuasive narratives [144]. In particular, identification could be used in persuasion by manipulating the source of a message, rather than the message itself, to make the message more attractive and palatable [64]. In addition, it has been shown that higher identification leads to increased activation of trait characteristics and increases the chance of mimicking behavioural tendencies displayed by the identified character, suggesting that experiencing strong identification for long periods of time could have lasting effects on an individual’s beliefs and values. As such, it is plausible that identifying with a character who displays sexist behaviour could affect the experience and beliefs of players of sexist games.

2.5 Experiment Design

We conducted an online study in which we presented three different versions of a simple role-playing game (RPG) that varied in their degree of sexism. We measured individual traits and player experience using several validated scales.

2.5.1 Game

We designed a 10-minute 2.5D game that was comprised of a prologue, three levels, and an epilogue. The prologue was a simple cut-scene to set up the narrative’s conflict, the levels were each designed around a single straightforward game mechanic, and the epilogue was an interactive conversation between the player and one of the characters in the game (see figure 2.1).



Figure 2.1: Screenshots from the three main levels in Fantasy Adventure: Hack n’ Slash (left), Infinite Runner (middle), and Shooting Gallery (right).

Game Narrative and Mechanics

The main conflict in the game's narrative was based around the 'damsel-in-distress' trope. Players were informed that the Kingdom's Royal Heir has been captured by an evil wizard and must be rescued. Along the way, the player would encounter obstacles, opponents, and dilemmas in order to reach their goal [97].

The prologue was a short cut-scene showing the Royal Heir running in a field with a looming dragon shadow overhead.

In the first level, we employed a side-scrolling fighting mechanic. The player had to progress through a forest - using the 'A' and 'D' keys to move left and right respectively - fighting barbarians they encountered, using the left mouse button to attack. Eventually, the player reached a clearing where a barbarian was watching over a cage with a prisoner. The player is told the prisoner is to be sold to slavers, and challenges the barbarian to a fight. After defeating the barbarian, the prisoner promises to help the player rescue the Royal Heir once all of the barbarians have been defeated. Several more barbarians attack and the player must defeat them all. Having defeated all of the barbarians, the player finds a key to the cage and releases the prisoner. The prisoner tries to go back on their promise to help the player and runs off, forcing the player to follow.

In the second level, we employed an infinite runner mechanic. The player chases the prisoner down a forest path, avoiding obstacles by jumping or swerving left or right by pressing 'W', 'A', and 'D' respectively, trying to reach the prisoner and get the promised information regarding the Royal Heir. The player slowly catches up to the prisoner, until the prisoner finally admits defeat and leads the player to their village.

In the third level, we used the shooting mechanic. The player and the prisoner arrived at the village, but was under attack by barbarians. The player found a crossbow and had to shoot twenty barbarians, while avoiding shooting the villagers, to scare the invaders away and earn the promised information. Pressing the left mouse button would slow down time so the player could be more accurate with their shot, and releasing the mouse button fired the crossbow.

Finally, in the epilogue, the prisoner led the player to the Evil Wizard's castle, only to discover that 'The Princess had been taken to another castle'. The player then had to choose what to do with the prisoner to end the game.

Game Versions

Depending on their assigned condition, participants played one of three versions of the game: Gender Neutral, Sexist Imagery, and Sexist Imagery plus Dialogue. Although the game mechanics did not vary across versions, we manipulated the graphics of the characters (see Figure 2.2) and the sexism present in the dialogue.

Gender Neutral. All four characters were androgynous in appearance, each identical in body



Figure 2.2: Neutral (left) and gendered (right) characters. From left to right: Harlot, Knight, Barbarian, Princess.

type, face, and hair, varying the hair colour, skin colour, and shirt colour to differentiate them. All dialogue in the game was generic and non-gendered, following the 'damsel-in-distress' narrative trope as loosely as possible while maintaining the overall story structure. All gendered pronouns were removed and characters were referred to generically by their titles (Royal Heir, Prisoner, Bandit, Player).

Sexist Imagery. All four characters were highly gendered in their appearance to conform to the Damsel-in-Distress Trope. The Player looked like a knight (masculine), the Bandit looked like a barbarian (masculine), the Heir looked like a princess (feminine) and the Prisoner looked like a harlot (feminine). Character models were custom designed to reinforce sexist stereotypes (helpless damsel, sexualized harlot, hyper-masculine barbarian, and manly hero). The gender-neutral dialogue was also used in this version. To confirm that our characters were perceived by participants in the intended way, we conducted two prestudies (see next section). Our male characters were perceived as highly masculine, our female characters as highly feminine, and our neutral characters as neutral (see Figure 2.2 for final character designs).

Sexist Imagery plus Dialogue. We used the sexist imagery, but also reinforced gender roles through the use of sexist dialogue. First, the dialogue used gendered pronouns (he, she). Second, the dialogue was systematically edited in four ways to reinforce sexist stereotypes, inspired by the tropes and dialogue typically used in modern games and by theories of objectification and instrumentality [195]. We reinforced the idea that women are weak and men are strong (e.g., “That is a frightening weapon!” vs. “One of them dropped a crossbow, use that.”). We included sexual innuendo in reference to the Harlot (e.g., “Wow! You really know how to handle a sword.” vs. “Amazing, your skill with a blade is impressive.”). We added diminutive nicknames in reference to the Harlot (e.g., “Darling, Sweetie”). We also reinforced the idea that women are interchangeable [195] (e.g., “She is ours. Go get your own.” vs. “This is our prisoner.”). We did not change the

plot or add new narrative concepts in any way; the variations were intended to reinforce sexism in the game through dialogue. We did not do additional studies to confirm that the dialogue was perceived as sexist; however, we have some confidence that it had the intended effect based on the “bug report” responses of participants who played this version (participants were under the pretense that they were playtesting a game under development), (e.g., “Make it not a chauvinistic sexist game”, “Remove misogyny”, “I would change the rampant sexism of the game”, and “I feel like I need a shower”).

2.5.2 Prestudy

We conducted two pre-studies to determine that our character designs indeed matched the gender tropes we were trying to match. In the first study (N=80), we tested the gendered characters and several neutral designs, asking participants to rate the character on a continuous sliding scale from masculine (-1.0) to feminine (1.0). One-sample t-tests (comparing to neutral) showed that the knight ($M = -.82, SD = .4, t_{78} = -18.0, p < .001$) and barbarian ($M = -.89, SD = .3, t_{78} = -27.1, p < .001$) were perceived as significantly masculine; the harlot ($M = .89, SD = .3, t_{78} = 23.5, p < .001$) and princess ($M = .89, SD = .3, t_{78} = 25.4, p < .001$) were perceived as significantly feminine. The initial neutral characters were perceived as more masculine, thus we iterated on the model and hairstyles, conducting a follow-up study (N=40). We selected our neutral characters based on the two designs that were perceived as most neutral ($M = -.11, SD = .6, t_{39} = -1.1, p = .260$; $M = -.14, SD = .6, t_{39} = -1.4, p = .169$); see Figure 2.2.

2.5.3 Cover Story

Participants were told they would be playing a short game and evaluating it for enjoyment. After completion of the experiment, participants were informed of the true purpose of the study and presented with the manipulation check that confirmed they read and understood the debriefing.

2.5.4 Measures

We collected both subjective and behavioural measures.

Individual Traits

Transportability. We assessed transportability, which is a personality trait that determines an individual’s susceptibility to being transported or immersed into a narrative, using the Transportability Scale of 20 items [70] (Cronbach’s alpha = .91).

Male Role Attitudes. We assessed pre-existing sexist beliefs by gathering the Male Role Attitudes Scale (MRAS), which measures masculine ideology through 8 items, such as “It bothers

me when a guy acts like a girl” and “A young man should be physically tough, even if he’s not big” [206] (Cronbach’s alpha = .79).

Benevolent Sexism. We also assessed pre-existing sexist beliefs by gathering the benevolent sexism subscales of the Ambivalent Sexism Inventory (ASI), which measures protective paternalism (e.g., “Women should be cherished and protected by men”), complementary gender differentiation (e.g., “Women have a superior moral sensibility”), heterosexual intimacy (e.g., “Despite accomplishment, men are incomplete without women”), and a combined level of general benevolent sexism through 11 items [105] (Cronbach’s alpha = .91).

Player Experience

Enjoyment was measured using the interest-enjoyment subscale of the Intrinsic Motivation Inventory (IMI) [174], which measures how much the player enjoyed the game through five items (Cronbach’s alpha = .94). We chose to use IMI instead of PENS [217] to measure Player Experience because we were more interested in exploring enjoyment, rather than focus on the building blocks that lead to enjoyment, which is what PENS measures.

Identification

Identification was measured using the Player Identification Scale (PIS) [258]: similarity identification, embodied identification, and wishful identification combine to give a general measure of avatar identification through 17 items (Cronbach’s alpha = .97).

2.5.5 Participants and Deployment Platform

We recruited 439 participants to participate in the study through Amazon Mechanical Turk (MTurk), which has been shown to be a reliable research tool [169] in the context of games research [32, 39]. To exclude participants who did not take care in answering the survey questions, we identified noncompliance by response time, response patterns and consistency metrics, as suggested by Mead and Craig [178]. We removed participants who spent less than 1.5 seconds/question on 2 or more questionnaires (N=50). We also detected the variance of responses for each individual subscale and excluded participants with more than 3 standard deviations above the mean on more than 3 subscales (N=36) - this process excluded participants who consistently did not pay attention to reverse-coded items.

This left 353 participants (161 self-identified as female) in our analyses. Ethical approval was obtained from the behavioural research ethics board of our university, and participants provided informed consent. To comply with ethical guidelines, the experiment was only available to workers in the USA who were older than 18. Participants were compensated with \$3 USD; the entire task took approximately 20 minutes.

2.5.6 Procedure

At the beginning of the experiment, participants were assigned to one of three conditions which determined which version of the game they would play: gender-neutral dialogue and visuals (*neutral*)(N=122, 52 female), gender-neutral dialogue with gendered graphics (*sexist imagery*)(N=122, 66 female), and gendered dialogue with gendered graphics (*sexist imagery plus dialogue*)(N=109, 43 female). Other than the different versions of the game, participants all experienced the same experimental protocol.

First, participants were presented with the cover story and asked to provide informed consent to their participation in the study. Next, they provided basic demographic information and completed the Transportability scale.

Participants then played their assigned version of the game in a between-subjects design. Immediately after playing the game, participants were asked to fill out the player experience questionnaires (Identification and Enjoyment) and a questionnaire about sexist beliefs (ASI). Participants were provided with several open-answer questions asking for comments about the game and general feedback. Finally, participants were debriefed on the purpose of the study and were asked to confirm which version of the game they experienced.

2.5.7 Follow-Up Study

Approximately one week after their participation in the study, participants were sent an email inviting them to participate in a follow-up study. Participants were given five days to respond and were compensated \$1 USD for their participation. In total we received follow-up data from 362 participants (82% response rate).

The purpose of the follow-up study was to measure participant's **pre-existing sexist beliefs** without the potential influence or priming caused by the game. The follow-up study included only two questionnaires: the ASI and the MRAS as measures of sexist beliefs. The individual traits were gathered after the main experiment to not prime participants as to the nature of the manipulations prior to play; we waited several days before the follow-up study to ensure that the experimental conditions no longer affected participant's baseline answers.

2.5.8 Data Analyses

Collected data was analyzed using SPSS 24, with the Process macro for SPSS [117]. We had 353 participants in our full sample, and we had pre-existing sexist belief data for 287 of these participants from our follow-up study. We first checked for group differences in the individual traits to ensure that random assignment generated groups that did not differ on our traits of interest. One-way ANOVAs revealed that the groups were comparable in terms of transportability

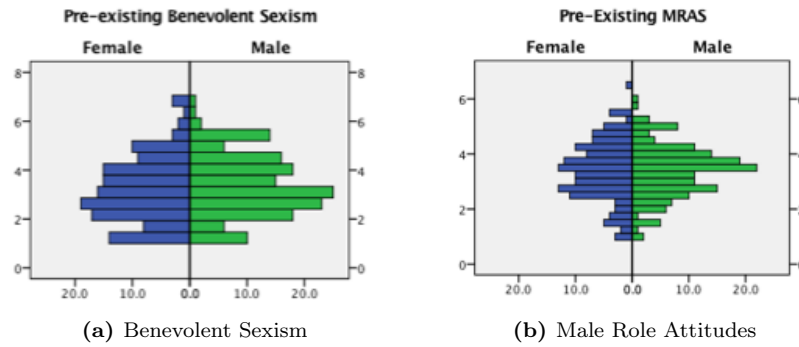


Figure 2.3: Distribution of sexist beliefs by player gender.

($F_{2,352} = .95, p = .388$, and pre-existing benevolent sexism ($F_{2,286} = 2.16, p = .118$).

2.6 Results

Our research questions centre around the relative roles of gender and sexist beliefs in the play experience of games that contain sexist imagery and dialogue. We first present results regarding the translation of identification into enjoyment, followed by the relative contributions of gender and sexist beliefs to play experience, and the role of sexist imagery or dialogue in experience.

2.6.1 Gender and Sexist Beliefs

Before presenting the effects of sexist content, we first show the relationship between gender and sexist beliefs. We used a linear regression to determine whether gender predicted sexist beliefs; however gender does not significantly predict benevolent sexism ($\beta = .182, p = .238$) or MRAS ($\beta = .002, p = .988$). Figure 2.3 shows the distribution of benevolent sexism and MRAS by gender.

2.6.2 The Role of Identification in Predicting Enjoyment

We first investigate effects of identification on enjoyment because literature suggests that identification with characters in narratives and games leads to enjoyment [119].

Identification predicts enjoyment

Linear regressions revealed that avatar identification predicts enjoyment in all three versions of the game (Neutral: $\beta = .568, p < .001$; Imagery: $\beta = .667, p < .001$; Dialogue: $\beta = .672, p < .001$). So regardless of whether the game contained sexist imagery or dialogue, identifying with the characters significantly increases game enjoyment.

2.6.3 The Role of Gender and Sexist Beliefs on Identification

Previous literature on sexist games has suggested that women self-select out of playing games with sexist imagery or themes because they don't find it enjoyable to play them [61]; however, previous work has argued against the effectiveness of looking for differences between men and women [84], which suggests that it may not be as simple as assuming that men enjoy sexist themes while women do not. Perhaps men who don't identify with the sexist themes also self-select out of playing while women who do identify with these themes enjoy the games. To determine the relative roles of gender and sexist beliefs on play experience of sexist games, we computed several models.

Predicting identification from gender

We first tested whether a player's gender predicted how much they identified with the protagonist in the different versions. Linear regressions revealed that gender did not significantly predict avatar identification in either the neutral ($\beta = -.254, p = .279$), sexist imagery ($\beta = -.244, p = .330$), or sexist dialogue ($\beta = .079, p = .792$) conditions.

Sexist beliefs predict identification

We next tested whether a player's index of benevolent sexism or affiliation with male role models (MRAS) predicted how much they identified with the game's protagonist. Linear regressions revealed that neither benevolent sexism nor MRAS predicted avatar identification in the neutral condition (benevolent sexism: $\beta = .109, p = .252$; MRAS: $\beta = .158, p = .208$). However, both constructs predicted identification in the sexist imagery (benevolent sexism: $\beta = .293, p = .009$; MRAS: $\beta = .296, p = .044$) and sexist dialogue (benevolent sexism: $\beta = .542, p < .001$; MRAS: $\beta = .536, p < .001$) conditions.

The beta values presented above suggest that sexist beliefs is a more salient predictor of identification in games with sexist dialogue and imagery than with sexist imagery alone. In the game with sexist imagery, a one-point increase on the benevolent sexism scale translates into an increase in identification of one-third of a point. However, in the game with sexist imagery and dialogue, an increase of one point in benevolent sexism translates into an increase of over half a point in identification (see figure 2.4).

Sexist imagery is helpful, but sexist dialogue and imagery together is harmful

Figure 2.4 shows the relationship between sexist beliefs and identification in the different conditions. Identification with the character is similarly high in both sexist games and for those with a high level of benevolent sexism; however, the imagery plus dialogue condition translated into less identification for those low in benevolent sexism. As such, overall, adding sexist dialogue to sexist imagery can

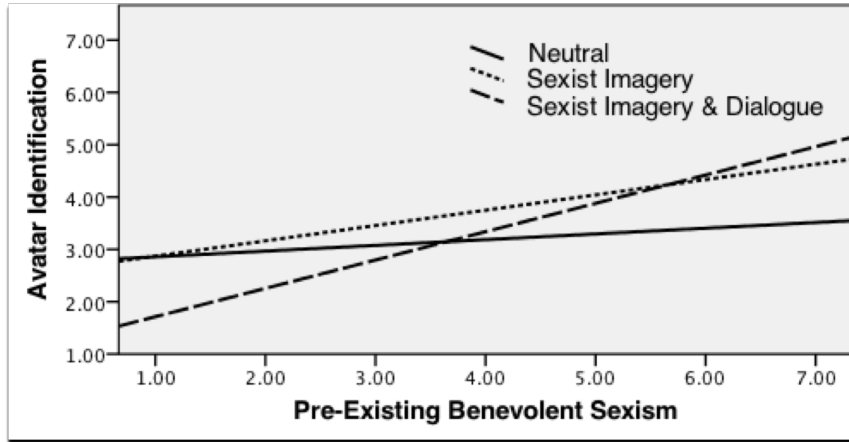


Figure 2.4: Linear regression lines showing the prediction of sexist beliefs on avatar identification for the three game versions.

be considered to be harmful for identification.

On the other hand, when we consider the effect of adding sexist imagery alone, we see a different pattern emerge. Figure 2.4 shows the relationship between sexist beliefs and identification for the neutral condition. For those low in benevolent sexism, there is no difference in identification between the neutral or sexist imagery versions of the game; however, identification increases with increased benevolent sexism in the sexist imagery game. Interestingly, sexist graphics alone actually aid with identification for people who hold sexist beliefs over a neutral version of the game.

The pattern that sexist imagery aids identification, but sexist imagery and dialogue harm identification is also demonstrated by looking at the overall levels of identification in the different games. A one-way ANOVA shows a significant effect of game version on avatar identification ($F_{2,350} = 4.88, p < .008$). Post-hoc comparisons (Tukey’s HSD) show that players in the sexist imagery condition identify more than players in the neutral condition ($p = .045$) and that players in the sexist imagery plus dialogue condition identify less than those who played with sexist imagery alone ($p = .010$).

2.6.4 The Role of Sexist Beliefs in Predicting Enjoyment

Our results show that identification predicts enjoyment in each game, but that people who do not hold sexist beliefs identify less in the sexist versions of the game. So, do the sexist beliefs of players that affect identification in sexist games translate into enjoyment of those games?

The prediction of identification by sexist beliefs in the sexist game translate into enjoyment

We conducted a moderated mediation [117] for the two sexist versions of the game only. The full model is significant ($R = .638, p < .0001$). The prediction of benevolent sexism on enjoyment is

fully mediated by identification ($p < .001$), and the mediation is significant for both the game with sexist imagery and the game with sexist imagery and dialogue (the bootstrapped confidence intervals do not contain zero in either case [117]: imagery: .016 – .372; dialogue: .213 – .541). As such, our results suggest that sexist beliefs affect the enjoyment of sexist games because of the resulting identification with the character.

Women enjoy the games more, regardless of game version

Finally, we conducted a 2 (gender) by 3 (game version) ANOVA on game enjoyment, and found a significant effect of gender ($F_{2,347} = 31.5, p < .0001$), in which women rated their enjoyment as higher than men. However, there was no significant interaction of gender with game version ($F_{2,347} = 1.54, p = .798$) showing that this result did not vary with the presence of sexist imagery or dialogue.

2.7 Discussion

We summarize our findings, situate results in the context of underlying theory, and address limitations and future work.

2.7.1 Summary of Findings

We show results of the relative roles of gender and beliefs about gender on player experience of sexist games.

First, the gender of the player does not predict avatar identification; however, their beliefs about gender roles and sexism do. Second, the differences in avatar identification as a result of the sexist content translated into game enjoyment. Third, the sexist imagery and dialogue harmed avatar identification (and resulting enjoyment) for those players who do not hold sexist beliefs; it did not aid avatar identification over sexist imagery alone for those who do. And fourth, sexist character stereotypes reinforced through graphic representations increased identification and enjoyment; however, sexist character stereotypes reinforced through both graphics and dialogue harms identification and enjoyment.

2.7.2 Explanation of Findings

In this section we will explain our findings and contextualize them in terms of the underlying theory.

Why Identification?

Avatar identification is a key component of becoming immersed in a narrative for various types of media [183]. The importance of identification likely is explained by the act of playing a game. In

text-based media, such as a novel, identification is created through the point of view of a narrating character. The reader finds a connection with the character to facilitate this identification [64] but it heavily depends on the quality and delivery of the work itself. Similarly, in film, identification is primarily created by leveraging the perspective through which the viewer experiences the media, namely the camera. This has been theorized to be a weaker and more difficult connection to forge for cinema-style viewing [64] and extremely difficult for a viewing experience situated in front of a television [64]. In games, however, the player has control over many aspects of the character - from controlling the character, moving and controlling the camera perspective, and in many cases, customizing the character's visual appearance. For these reasons, it would be fair to argue that avatar identification would indeed play a large and important role in the way players experience games. Furthermore, while games likely facilitate stronger identification than other forms of media, it would also make sense for identification to be greatly reduced if there is a divergence between the player's beliefs and the beliefs and actions displayed by the character when outside of the player's control.

Why do sexist visuals help identification?

There are several possibilities why the sexist visuals increased avatar identification. One possibility is that the Knight character's appearance was more attractive and interesting than the androgynous character, which may have led to players liking the Knight more, leading in turn to stronger identification [64]. Second, most players are used to playing as a male knight character, regardless of their own gender, because most games have male protagonists [78]. Positive previous experience with a character can lead to an increased fondness for that particular character, which enables easier identification in similar situations in the future [64], so many players would more easily identify with a character if they have fond memories of other games where they played as a similar character. Third, knight characters are traditionally seen as noble or chivalrous [6]: they are an ideal to strive for. As such, players who strive to be a better person may identify with the Knight character in a wishful way, forging a connection with him that they wish they could be. Previous literature has suggested that tropes can facilitate immersion and enjoyment through familiarity [26], so it is not surprising that adding gender-stereotypical visuals aid identification and enjoyment.

Why is sexist dialogue harmful to identification?

Language is an important factor in perpetuating sexism [165, 175], which means players can easily interpret sexist ideas when they are conveyed through language. Also, for those who do not hold sexist beliefs, a sexist character would be less likable, which is detrimental to the process of identification [64]. Similarly, a sexist knight may be seen as less noble, which may reduce the previously mentioned desire players may have to be similar to him.

Why is identification harmed for those who do not hold sexist beliefs?

We likely did not observe greater identification when sexist dialogue was added because of the asymmetry between creating identification and breaking it. The act of breaking identification can occur easily, from external stimuli such as the phone ringing, to stimuli brought by the story such as changing narrative perspective [64]. Conversely, forming identification is more involved and must be maintained for the audience member to feel identified. This asymmetry would explain why people who do not hold sexist beliefs were more easily turned off by the sexism, and why the opposite does not seem to be true for people who hold sexist beliefs. The sexist dialogue is just one more factor on top of several others that would forge identification for people with sexist beliefs (e.g., agency through controlling the character, relating to the knight visually).

Which beliefs matter for identification and enjoyment?

We measured two constructs to gather information on the participants' preexisting beliefs about gender: benevolent sexism and male role acceptance. For both constructs, we assumed they might differentiate between players whose experience was harmed by sexist content and players who didn't mind the sexist content. While male role acceptance showed similar trends to benevolent sexism, our empirical results are stronger for benevolent sexism. We believe this trend is due to the relationship between the occurring themes in the game and the nature of the preexisting beliefs that were examined. The male role acceptance scale measures beliefs about masculinity and does not include gender excitations for women. Benevolent sexism includes some ideas about masculinity (e.g., "Despite accomplishment, men are incomplete without women") however, focuses mostly on beliefs about women (e.g., "Women should be cherished and protected by men"). Our game, while also prompting traditional concepts of masculinity (i.e., the strong, dominant and aggressive hero), mostly caters to sexist ideas about women (i.e., the scantily-dressed seductive, yet helpless harlot, the pure and innocent princess in distress). Benevolent sexism was likely the stronger predictor in our data because it more accurately speaks to the kind of sexism that is portrayed in our game.

This opens up the interesting follow-up questions: Would we find the same results with a different game as an example? We interpreted our results as exemplary to how preexisting beliefs influence the experience of a game. Benevolent sexism is most likely not always the distinguishing trait to determine who is affected by sexist content and who is not. Sexist content can take on many forms, from overtly misogynistic violence towards women (e.g., Grand Theft Auto [194]) to subversively embedded ideas of traditional gender roles (e.g., Princess Peach in Super Mario Bros. [191], God of War [239]). It is reasonable to assume that there is a set of different constructs that can become relevant depending on the "brand" of sexism portrayed in the game.

Shooting games, for example, often provide a militarized narrative structure that emphasizes hypermasculine traits such as dominance, violence, and lack of emotional expression (e.g., Gears of

War [98]). These types of games are likely more affected by preexisting ideas of masculinity, than benevolent sexism. Similarly, games that portray highly sexualized and objectified female avatars are likely highly affected by a preexisting propensity to sexualize and objectify women (e.g., Duke Nukem [233], Dead or Alive Xtreme Beach Volleyball [190]) and may be more affected by existing ideas about women as sexual objects. Finally, games that contain high levels of aggression toward women are probably affected by a tendency to normalize and justify harassment and physical violence towards women (e.g., Grand Theft Auto [194], God of War [239]), as is measured through scales of hostile sexism [105].

How do our results fit with existing literature?

As presented in the related literature section, there is a significant body of work on gender and games that makes research contributions using a range of methodological approaches - in particular, through qualitative contributions (e.g., [71, 132, 133] or through quasi-experimental approaches (e.g., [90, 193]). We approached the issue of sexism in games from an empirical perspective, looking at the effects of a particular experimental manipulation on player experience. There are many potential problems that arise from playing sexist games (e.g., hostility towards women [75, 88], long-term effects in society [88, 90], and issues of gender equality and representation [78, 127, 271]) that our results do not address; however, our work is in line with the body of work on gender in games that suggests that sexism in games is harmful to play experience for a large proportion of gamers. Our work takes the literature a step further by identifying that the harm of sexist games is not specific to female players, but that sexism in games negatively affects the enjoyment of both female and male players.

2.7.3 Limitations and Future Work

Although we carefully controlled our design and analyses, there are limitations that can be addressed in future work.

First, our game was relatively short (10 minutes) compared to the way that games - especially role-playing games - are normally played. These short play times likely help to explain our small effect sizes. If players spent longer interacting through and with the characters, we would likely observe greater identification through increased interaction [259]. We would like to explore the role of sexist beliefs on identification in games with longer play times and depth.

Second, we introduced sexist dialogue almost immediately. Identification is a cognitive process in which players are less likely to question their own beliefs by being immersed in the story [64]; however, in our game, the players were immediately faced with sexist dialogue, giving no chance to first identify with the knight. Our results may have been different if the sexism was added after the player had already undergone the process of identification. For example, a game in which

a player builds a strong connection and empathy with a character early on may be less likely to alienate players if the character suddenly began displaying behaviours that went against the player’s own personal beliefs because the process of identification was already underway. In this way, game designers could leverage identification in a persuasive manner to force a player to see something from a different perspective, with their defensive shields down and more willing to listen to points of view contrary to their own. We would like to explore the persuasive aspects of character identification.

Third, we focused on the effects of sexist games on player experience and avatar identification. We would like to extend the empirical control we used to the exploration of how sexist elements in games normalize sexist thoughts and behaviours [22,274] by unpacking how sexist content from different sources (i.e., graphics, dialogue) influence participants’ sexist beliefs in the short term.

Fourth, we focused on a particular type of sexism - the objectification of female characters alongside their helplessness and interchangeability. We would like to further explore the role of hypermasculine characters on game experience for players who hold different beliefs.

Fifth, the characters in the sexist imagery condition had light skin; two of the characters in the neutral version had blue skin. We did not examine how sexist depictions interact with race of the character or player and would be interested in exploring the interaction of sexist imagery with unfavorable depictions of other marginalized groups.

Finally, the characters in our game were assigned, rather than created by the players. Increasing player agency in the game may allow the players to construct the character’s persona more similarly to their own, which is likely to increase identification. We would like to explore how increasing player agency affects experience.

2.8 Conclusion

Research on sexism in digital games has suggested that women self-select out of playing sexist games; however, we suggest based on literature that the gender of the participant is not paramount to experience, but that their beliefs about gender roles are. To explore the role of sexist beliefs of experience in sexist games, we created three versions of a game that were identical except for the presence of sexist imagery or sexist imagery and dialogue. We conducted an experiment with 353 players and showed that enjoyment of sexist games is not predicted by the gender of the player, but by pre-existing beliefs about gender that the player holds. In addition, we demonstrate that avatar identification is the pathway through which enjoyment is facilitated for players who hold different sexist beliefs. Furthermore, we show that adding sexist dialogue to sexist games harms experience for people who do not hold sexist beliefs, but does not improve experience for those who do. Finally, we show that including visually sexist character stereotypes can aid in identification

and enjoyment; however, sexist character stereotypes that are reinforced through both imagery and dialogue harms identification and enjoyment.

Our work explores player response to sexist characters and behaviour in a role-playing game. We refute the notion that men and women form distinct groups that respond in a predictable manner to sexist content. Player response is not about gender, but is determined by the beliefs about gender roles that players hold.

3 MANUSCRIPT B: THOSE ARE NOT THE STORIES YOU ARE LOOKING FOR: USING TEXT PROTOTYPES TO EVALUATE GAME NARRATIVES EARLY

3.1 Introduction to Manuscript B

In the previous chapter, I used a custom-built game to study the relationship between the player's real-world physical attributes (i.e., self-identified gender) and beliefs (i.e., benevolent sexism, male role attitudes) and those exhibited by the game's protagonist.

In that study, the largest bottleneck in completing the research was the development time required to custom build the game. While many free and reasonably priced assets were available to speed up the development, the game still took approximately eight months of part-time development to complete. While this was partly due of a lack of game development experience at the time, the game required the characters' physical appearances to reflect the sexist tropes of each character, multiple versions of the game dialogue, as well as systems to present different versions of the characters and dialogue based on experimental condition, in addition to game mechanics that were engaging enough to maintain participants' interest long enough to gather the data.

Because of the nature of the study (i.e., online deployment, dynamic character models and dialogue depending on condition) using an off-the-shelf game was not an option. As such, the development time was necessary, but limited the amount of research that can be conducted by a single researcher in this area. This is a problem for many researchers conducting similar research: either one or two researchers with a wide range of skills in graphic design, 3D modelling, programming, and game design spends months creating the game, or a large team of designers and developers need to be hired to create the experimental stimuli in a reasonable amount of time.

While some studies require a game that is at this level of graphic fidelity, a key take-away from our results indicated that the most salient element of our experiment was the sexist dialogue, and not the gendered character designs. As such, this chapter deviates away from the main goal of understanding the relationships between players and player characters to instead investigate this

phenomena, for future use in addressing our main problem in this thesis.

The goal of this chapter is to propose and evaluate a new method of prototyping narrative games using text interfaces. In this chapter I show that game developers can use text prototypes to adequately measure the general experience a player would feel in a graphically fleshed-out version of the game, using the same narrative. This tool should help solve the problem described above by allowing game researchers to reduce total development time by reducing the graphic fidelity of a game and still be able to accurately research narrative-related research questions. This tool will be used in subsequent chapters to allow quicker development and evaluation of game narratives.

3.1.1 Publication and Individual Contribution

This work was published as the following: [43]. Although a multi-authored contribution, I was responsible for leading the research. My roles included: designing the experiments and constructing the interactive systems used in the experiments (i.e., the high-fidelity versions of the two games used in the studies, the text-based prototypes), data analysis, literature review, and writing the majority of the publication. I also presented the publication at the conference at which it was published. I contributed approximately 80% of the total work for this project and publication.

3.2 Abstract

In game design, evaluation is important to do early and often; however, evaluating game narratives early in development is an open problem. We don't know how the evaluation of a game narrative will be affected when it is experienced outside of the context of the game's mechanics and graphics. In this paper, we test the plausibility of using a text-based narrative prototype, by evaluating player experience and narrative experience in two studies using different game genres. In both studies, we compare the narrative evaluation of low- and high-fidelity graphics, but in study 1 (N=78), we kept interaction and mechanics intact, and in study 2 (N=124), we removed game interaction in the text prototype. We observed no significant differences in player experience or narrative engagement in either study, indicating that text-based narrative prototypes could be an effective playtesting tool for game studios to integrate into their development cycle early.

3.3 Introduction

Game evaluation is an important component for a game's eventual critical and commercial success, and as such, should be done early in the development lifecycle and often throughout [162]. Some aspects of games are evaluated easily early in development (e.g., basic mechanics and procedures); however, other aspects are more difficult to evaluate outside of the context of a complete prototype

(e.g., accessibility, balance) [97]. Because of the benefits of early evaluation, methods have been proposed for play-testing prototypes of games under development, such as physical prototypes [97] or low-fidelity digital prototypes [97]. Furthermore, researchers have been exploring how the fidelity of the prototype affects the player’s experience of the game mechanics and have shown that the evaluation of core mechanics in low-fidelity prototypes of casual games is reliable [100]. Although games user research (GUR) tends to focus on methods that evaluate a game’s mechanics and dynamics [123, 125], there are other factors that contribute to play experience - such as narrative experience - that are less well understood or represented in GUR.

The evaluation of game narratives is an important aspect of playtesting because incoherent, inconsistent, or boring stories can result in poor critical and commercial reception of a game. In a series of blog posts, interviews, and talks [2, 177, 237], Bioware, a game studio praised for their successful merging of game and narrative, discuss how they employ narrative prototypes to evaluate the story early in a development process in which the writing team is well integrated with the development team. Bioware partially attribute their critical and commercial success of narrative-based games to this early evaluation and the integration of the different teams. In a counter-example, a blog post-mortem of the game *Destiny* [47], developed by Bungie [226], revealed that the game narrative was rebooted just months before the original release date, leading to delays, and a story that was criticized by reviewers for its vague plot, thin characters, and opaque dialogue. These examples lend support to the idea that narratives should be evaluated early in the development lifecycle of a game.

Questions around the process of early evaluation of game narratives are being asked by games user researchers in both academia and industry. At the 2016 GamesUR Summit [113], there was a panel on the evaluation of narratives that questioned whether or not they could be evaluated early and what effects there might be from evaluating narratives in isolation from the completed game. However, there is little guidance for game designers on the effects of evaluating narratives on prototypes; games user researchers do not currently understand the effects of evaluating an interactive narrative early in the design process - that is, outside of the context of a completed game. For example, will a game narrative be experienced differently if it is disconnected from the immersive and transportative aspects of completed games, such as the high-fidelity graphics? Will a game narrative be experienced differently if disconnected from the control scheme used to interact with the game’s mechanics?

In this paper, we ask how game narrative evaluation is affected when embedded in early game prototypes. Specifically, we systematically investigate how the game narrative evaluation is affected if it is disconnected from the high-fidelity graphics of a completed game (Study 1) and the mechanics used to control interaction (Study 2).

In a first study, we investigated the player experience differences in a narrative-based game with

low-fidelity and high-fidelity graphics. All other aspects of the games (i.e., interaction and game control) were kept consistent between the two versions to avoid confounding factors that might cause differences in experience. The game used was similar to an interactive narrative (e.g., *The Stanley Parable* [52]). We found no statistically significant differences on any experiential measures and no interactions between a median split of how transportable a player tends to be and fidelity on any experiential measure.

In a second study, we used an adventure game that included more non-narrative interaction (i.e., fighting, running, and shooting). Furthermore, we did not mirror all of the controls and interactions as we did in study 1, but rather prototyped only a very minimal amount of game interaction to replicate a situation closer to what would be seen in industry when games are under development. We also added measures of narrative engagement and transportation. Again, we found no statistically significant differences on experiential measures or narrative engagement. There were also no significant interactions between player transportability and fidelity on any measure. We did, however, see a minor difference in the transportation of players, which showed that players were slightly more transported when playing the low-fidelity version rather than the high fidelity version, consistent with ideas suggested in the literature for other types of narrative immersion, such as avatar identification [64].

Our results showed that there were no significant differences in player experience or narrative engagement as a result of playing a narrative prototype as opposed to the full game. These results hold regardless of whether or not the game interaction itself was prototyped and in two different game genres. There was a small positive effect on experienced transportation in the narrative prototype of study 2. Our results suggest that text prototypes of game narratives can be used early in the development of a game, increasing the likelihood of integrating narrative development with system development, and maximizing a game narrative’s chance for a successful reception.

3.4 Related Work

We first present player and narrative experience, then discuss iterative development and prototyping in games.

3.4.1 Player Experience

Research in studying how players experience games is primarily rooted in psychology. In particular, a common model used as a basis for explaining player experience is needs satisfaction, which is part of the larger theory called Self Determination Theory (SDT) [217]. SDT explains intrinsic motivation, i.e., doing something for the pure desire to do it, as opposed to being driven by the promise or threat of external rewards [217]. When engaging in an activity, such as a game, an

individual’s intrinsic motivation to engage in the activity can be broken down into four dimensions: interest-enjoyment, effort-importance, perceived competence, and the amount of tension-pressure felt during the activity [174]. The building blocks that lead to the experience of intrinsic motivation are based on our basic psychological needs: competence, autonomy, and relatedness. In using needs satisfaction in the specific context of play, Ryan et al. added two additional constructs: intuitive controls and immersion-presence [217].

3.4.2 Narrative Experience

Researchers have recently begun extending constructs and ideas from media psychology into the realm of digital games [64, 110]. For the purposes of this paper, we will distinguish between narrative engagement, experiences which pertain primarily to the game’s narrative [49] and narrative immersion, which depend more strongly on non-narrative game elements to create an immersive experience.

Narrative Immersion

Two types of narrative immersion have been used in the context of games: identification [119] and transportation [61]. Both types of immersion involve a similar cognitive process in players: a deep immersion into the narrative aspects of the game in which the player loses awareness of the outside world.

A state of immersion through identification involves the creation of a connection between the player and a character in the game so that the player loses awareness of their outside self, experiencing the game from a different point of view [64]. There are three main ways a player can identify with a character [258]: *similarity identification* — identifying through perceived physical similarities with a character; *embodied presence* — seeing the narrative world through the perspective of a character; and *wishful identification* — when a player sees a character as a desirable version of themselves. Previous work has shown that aspects of player experience (e.g., enjoyment) are predicted by the amount of identification [32, 40] experienced.

Transportation is often described using a traveller metaphor, in which the player is so immersed in the narrative world of a game that they lose sense of their physical surroundings [110]. Transportation is facilitated through the interaction and presentation of the game (i.e., the depth, complexity, and attractiveness of the narrative world) but also through the personality trait *transportability*, which has been shown to reliably predict transportation [70].

Narrative Engagement

The Narrative Engagement Scale was created by drawing items from a wide selection of narrative measures to create four new subscales measuring narrative-specific experiences: narrative under-

standing, attentional focus, narrative presence, and emotional engagement [49].

3.4.3 Iterative Development in Interfaces and Games

We describe the iterative cycle (design, development, and evaluation) of interactive systems and games.

Prototyping

In interactive systems and interface design, it is common practice to start with low-fidelity prototypes and iteratively work up to high-fidelity versions [11, 107]. This process allows designers to quickly explore a wide array of different design directions and possibilities, throwing out those ideas that didn't work without losing a significant amount of time and effort. Various types of prototyping methods have been used, each with differing levels of fidelity: paper prototypes [77], digital mockups and wireframes [257], and interactive prototypes [82]. There is also research that directly compare prototypes of varying fidelity that have found no difference in the number of usability problems between prototyping fidelity [57, 262, 270], and that users tended to prefer working with a computer prototype compared to a paper prototype [229].

Game designers have used similar approaches to prototype their ideas because games can be seen as a specific type of interactive system. However, previous work has argued that there is more to test in a game than usability, such as playability, which requires game designers to use slightly different prototyping methods [222]. Fullerton describes two levels of fidelity in prototyping: physical paper prototypes and digital prototypes, as well as various tools to create them [97]. Game designers often prototype by starting with simple low-fidelity assets, such as capsules for characters and boxes for objects in the environment. Previous work has shown that playtesters find the same amount of usability issues in low-fidelity and high-fidelity graphic levels [149].

Evaluation

As mentioned, both interface and game designers follow an iterative design strategy. Part of this iterative process is the notion that the product starts at a low-fidelity stage and increases in fidelity through iteration, but the other essential part of this iterative cycle is the evaluation component that feeds into the next iteration of design.

For interfaces, evaluation may include formal usability inspections [140], heuristic evaluations [140], and cognitive walkthroughs [140] to name a few examples. For games, a different type of user testing can be required because of the increased interaction, which can lead to different kinds of bugs and design problems, often unforeseen by developers. Evaluation strategies have been created for games that have adapted strategies used in interface design — for example heuristic evaluation [204] and usability testing [125], whereas others are more unique to games, such as analytics [80] and

playtesting [97]. More advanced playtesting approaches have used modelling techniques to evaluate experience using physiological data [167, 168], biometric storyboards [181], and analysis of facial expressions [247]. Additionally, specific tools have been created or adapted from other areas to empirically measure various aspects of players and player experience in games [70, 110, 174, 219, 258].

The main idea behind iterative evaluation is for designers to test their current prototype with users to make sure they like it and are able to use it in the intended way. These goals differ for games and interfaces, but the common thread is that users test the prototype and designers use the feedback to make it better. This process is repeated throughout the entire design and development process with quick bursts of development and testing and updating the design to make the product better and easier to use [156].

3.5 Industry Approaches to Game Narratives

Because industry practices are rarely published in academic venues, and development processes can be a closely guarded secret for game studios, there is not a large amount of publicly-available information regarding the actual processes that studios use to prototype, test, and integrate narrative structures into their games. To gain knowledge on how some prominent companies do this, we searched for blog posts written by developers and game designers, post-mortem analyses of games, and conference talks (e.g., Game Developer’s Conference — GDC) to try to piece together the kinds of processes and systems that are used to evaluate narratives.

Bioware, a studio well-known for their use of narrative structures in their games, seems to be the most forthcoming about their narrative design process, though there are only a few public resources that give an indication of what they do. Combing through the statements and interviews indicate three important development practices that demonstrate their commitment to producing excellent game narratives: 1) the close integration of their quality assurance (QA) team with the development team, 2) their use of narrative prototyping to iteratively test and develop game narrative, and 3) the merging of story design and system design.

Across several blog posts, team leads at Bioware have discussed several important QA practices that the studio has embraced over the years to improve their final products, in both narrative and design. In one such post [177], a QA lead talked about how the QA department plays an increasingly important role in the development cycle, moving away from reactionary to proactive involvement in the development process from the beginning. QA leads began attending designer and programmer meetings to be more informed about the game, and the QA team was eventually split into two separate disciplines — Tech, which supported the programmers, and Design, which supported designers and writers. In another post [2], the lead describes the role QA has at Bioware compared to other large studios and reiterates the integration of QA with development.

In 2007 at The Australian Game Developers Conference, a lead writer for Bioware, Mike Laidlaw, gave a presentation describing the early part of the production process for a Bioware game [148]. Laidlaw noted that the writers are part of the design process from day one, as a part of the design team. He also describes three phases of development the team undergoes: prototyping — exploring ideas, setting, and story work; pre-production — systems, pipelines, and story arcs are defined; and production — “a flood of content”.

Laidlaw further describes the writing process for their game *Mass Effect* [28] from early-stage prototyping to late-stage integration. In the early-stage prototyping, the writing team designed the dialogue system and tested how players would interact with it, the tone and shape of the game was formed, and the ESRB rating was decided so the writers knew what type of content to create. In the next phase, the game story was written by the writing team. The written content was then iteratively tested using a low-fidelity narrative prototype, to test for technical issues with the story content. The end result of the narrative prototyping stage was a playable level with full dialogue in text format. This allowed for major rewrites, because the content was text, which was cheap to rewrite as needed. After this initial prototyping phase, the text was passed to QA, who looked for plot holes, bugs, and inconsistencies. From there, the text moved to voiceovers, cinematics, and integration into the game.

3.5.1 Integrated Development of Story and System

A talk at GDC 2015 [237] discussed Bioware’s process of melding story with game systems in the production of *Dragon Age: Inquisition* [30]. The design team split in half to explore different design techniques. The first team created narrative content and the second team designed standalone systems. What they found was that the narrative team’s content scaled poorly, and the system team’s content had a disconnected impact. The main conclusion was that story and technical systems developed separately cause problems and dissonances in the final product. The solution was to develop story and systems together, in tandem.

Bungie’s *Destiny* [47] is an example of a game that was received poorly by players due to the separation of narrative and system teams. The details of what led to the game’s lukewarm reception after the large pre-release hype are fuzzy, but the general idea is documented in a post-mortem of the game [226]. After three years of development, and only months before the game was initially supposed to be released, the lead writer put together the ‘supercut’ for the game — a video that included the game’s cinematics and major story points, and showed it to the studio’s leadership. The senior staff at Bungie were not satisfied with the story and scrapped it, deciding to start from scratch. The result was a hacked together story, criticized by reviewers for its vague plot, thin characters, and opaque dialogue. Furthermore, there were system elements that could not be scrapped as too much development had gone into them and it was too close to release, so they did

not hang together well with the final version and seemed to be disconnected.

Fans blame *Destiny's* [47] poor reception on its lackluster story, which can likely be attributed to the separation of the development of story and game systems, the same problem Bioware found with *Dragon Age: Inquisition* [30]. That the leadership had not seen the story and cinematics prior to their completion, combined with a mindset that they could develop and implement a full story within the restrictions of existing game systems as a final touch, rather than a fundamental part of the game, resulted in a poorly constructed and inconsistent narrative experience.

The important part of this is that game studios with successful narrative games, such as Bioware, are successful because of how the story is integrated into their process. They develop the story early on, in parallel with the other game systems. The story and game need to fit well together without creating inconsistencies between them and in order to achieve that, iterative testing of these two systems together early and throughout development is important. Many studios, like Bungie, developed these two branches separately, which led to miscommunication between departments and an inconsistent, dissonant narrative experience, even though the game mechanics were innovative and well designed.

3.6 Constructing a Narrative Prototype

In this paper, we conducted two studies in which we created a low-fidelity narrative prototype based off of a finished high-fidelity game. In this section we will document the process we used to construct our narrative prototypes.

The first step was to write the overall story, conveying through text the information that would normally be understood from the visuals. The goal was to try to get the same information across as would be in the final game.

The second step was to simulate player actions and game mechanics using a simple text interface. Player actions refer to simple things like “Walk over to the computer”. Game mechanics refer to interactions such as combat, in which the player has a wider range of options and the interaction is more complex. For these types of game mechanics, we presented options to the player, but in a very simplified version, (e.g., “Attack”).

The final step is to implement the narrative. We used a simple tool [205] that we customized to convert a game script (scene descriptions and character dialogue) into an interactive system. Our tool allowed us to display the text in a game-like environment — we displayed the character portraits alongside their dialogue in study 2, but not in study 1 — and use mouse and keyboard controls to move through the prototype. However, a different medium-fidelity tool (e.g., PowerPoint) could just as easily have been used to achieve the same functionality.

3.7 Study 1

Using Bioware’s narrative prototypes as inspiration, we wanted to systematically measure how effective such a technique is for narrative games, and the feasibility of using it as an iterative development tool. That is, we wanted to determine how the player experience of the narrative could be evaluated early in the development lifecycle. In particular, we wanted to know whether a text-based version of a game would result in reduced player experience, when compared to the same game with high-fidelity graphics.

To investigate this, we used the above-mentioned process to create a text-based version of an existing research game and compared the low-fidelity text version with the high-fidelity graphical version. We performed a between-subjects experiment and collected measures of experience, in-game behavioural motivation, and demographic surveys to assess player traits.

3.7.1 Game Design

We created *Space Adventure*, a narrative game involving an accident on a space ship; participants play the game from the perspective of the ship’s Captain and explore different rooms through a surveillance system to discover the events that occurred on board. The game was implemented in C# using the Unity Game Engine.

Space Adventure is played in a full 3D game environment similar to immersive narrative-based commercial games (Figure 3.1). Players started by designing their customized avatar: they were given the choice to select sex, hairstyle, hair colour, shirt, shirt colour, pants, and pants colour. The player moved around the inside of a spaceship, experiencing the game’s narrative and optionally discovering the backstory that led to the events in the game.

Narrative/Procedure

In the game, the player awakens after the ship’s engines had crashed. One of the ship’s soldiers instructs the player about the controls for moving around the environment, before guiding the player to investigate the cause of the failure. First, players are introduced to the *Surveillance Room* where they can access various cameras around the ship to view footage from different rooms of the ship, shown to the player through a text-based description. They then teleport to the *Hacking Room*, which is necessary in the game to unlock the cameras one at a time for viewing in the *Surveillance Room*. The game primarily consists of players teleporting back and forth between the *Hacking Room* and the *Surveillance Room* throughout the narrative, unlocking a camera in the *Hacking Room*, then viewing a piece of the story by looking through the camera in the *Surveillance Room*, then teleporting back to hack the next camera. This is done eight times throughout the narrative, each time revealing a different piece of the story. In addition, while in the *Hacking Room*,

the player can unlock more of the game’s backstory by optionally hacking a console each time they return to the room. Pressing the right arrow key reveals more information about the backstory one word at a time, which is revealed while the player-character is waiting to gain access to the next camera in the surveillance room. Once a console is hacked, the player teleports back to the *Surveillance Room* and the cycle continues.

Using cameras as a narrative technique allowed us to guide participants through a narrative from a local point of action, reducing the need for navigation through the virtual world. After hacking the last console, participants could either escape the ship right away by pressing a button, or alternately run down the hallway to escape the ship.

Controls, Interface, and Versions

Participants controlled the avatar using ‘W’ to move forward, ‘A’ to rotate the avatar to the left, and ‘D’ to rotate the avatar to the right. A 3rd-person camera was positioned in a fixed location directly behind the avatar.

We included a heads-up-display (HUD) on which players could hack and teleport by clicking on the *hacking* and *teleport* buttons, using the left mouse button when the options were available (see Figure 3.1). A mini-map was provided in the bottom right corner with highlighting of the current room so the player could see where the fictional camera was looking (see Figure 3.1).

We used two versions of the game: the high-fidelity version as described, and a text-based low-fidelity version (see Figure3.1).

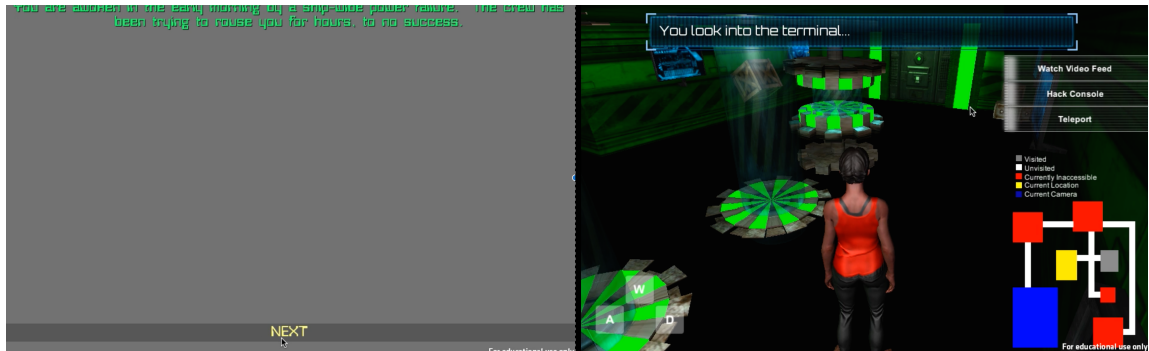


Figure 3.1: Screenshots for low-fidelity (left) and high-fidelity (right) versions of Space Adventure

Participants and Deployment Platform

We recruited participants by telling them that they would be evaluating a game under development. We had 81 participants (48% female) with an average age of 22 (SD = 15.9) participate in our study through Amazon Mechanical Turk (MTURK), a platform that acts as a broker between parties offering a range of *Human Intelligence Tasks* (HITs) and paid workers. Participants received

\$6 USD as compensation and the study took on average 40 minutes (15 minutes of gameplay). Although it has been shown that MTurk is a reliable research tool [169], we measured the time spent per questionnaire to evaluate task performance and ensured that participants were attentive despite the online setting. As a compliance check, we removed participants who spent less than 1.5 seconds per item on two or more questionnaires, as these participants likely did not carefully read and consider the questions. Three participants were detected and removed using this filter, leaving a total of 78 participants to be used in the final analysis ($N_{low-fidelity} = 40, N_{high-fidelity} = 38$).

Ethical Compliance

Ethical approval was obtained from the *University of Saskatchewan* behavioural ethics board; participants were asked to give informed consent at the beginning and debriefed on the experiment at the end. To comply with ethical guidelines, the HIT was only made available to workers in the USA who were older than 18. Also, only workers with an approval rate about 90% were offered the HIT as a means of quality control. After the experiment, we asked participants to respond to simple questions about the story to ensure that the experiment was understood.

3.7.2 Measures

In this section we present the measures used for study 1. Unless otherwise specified, all items were presented to participants using a 7-point Likert scale.

Needs Satisfaction (i.e., competence, autonomy, relatedness, presence, intuitive controls) was measured using the Player Experience of Needs Satisfaction Scale (PENS, [219]). Participants rated their agreement to a series of statement, e.g., “I feel competent at the game”. (Cronbach’s $\alpha = .75, .85, .65, .94, .712$).

Enjoyment, Effort, and Tension were measured using the Intrinsic Motivation Inventory (IMI, [174]), e.g., “I enjoyed the game very much”. (Cronbach’s $\alpha = .94, .85, .85$).

Identification was measured using the Player Identification Scale (PIS, [258]). Participants reflected on their experience and answered several questions on how they identified with the character they played. The PIS consists of three subscales: *similarity identification*, *embodied presence*, and *wishful identification*. (Cronbach’s $\alpha = .95, .96, .91$).

Transportability is a personality trait that describes an individual’s susceptibility to become transported into a narrative world. The transportability scale [70] was presented prior to playing the game to measure this trait. Participants rated their tendency to become transported on a series of items, (e.g., “I have vivid images of the events in the story”) (Cronbach’s $\alpha = .93$). High- and low-transportability were calculated using a median split.

Motivation was measured using the in-game behavioural measurements from the hacking terminals. At each terminal, the player could reveal the backstory, but could leave the task at any

time with no negative consequences to the game. We calculated the motivation score by combining the number of words revealed in each terminal and normalizing the score to an overall average percentage. The purpose of this measure was to get an idea of how much effort participants put into the task, since as crowd workers they could primarily be motivated by the money rather than the task.

3.7.3 Results

The purpose of this study was to investigate differences in player experience of narrative-based games with high or low graphic fidelity. Based on previous research [100] that investigated the experiential effects of high-fidelity and low-fidelity casual games, we expected there to be no main effects by condition on any measure. Also, because the game was primarily an interactive narrative, we expected there to be significant differences on all dependent measures between participants who rated high and low in transportability. Based on this, we came up with the following hypotheses:

H1: There will be no differences in any game experience measures due to graphical fidelity.

H2: There will be significant differences in experiential measures between high and low transportable participants.

We conducted an a priori power analysis using g*Power [83] to ensure that we recruited enough participants to find medium to large effect sizes, if they existed. We performed the test expecting an effect size of 0.33, a significance level of 0.05, a β of 0.8, and two different groups in our between-subjects experiment design. We decided that medium-large effect sizes were sufficient, because we only need to show that any differences between dependent measures are quite small, rather than statistically identical. The power analysis suggested a sample size of 37 participants per group, which we rounded up to 40. For all of our analyses, we use $p < .05$ as a threshold for accepting statistical differences, and $\beta > .80$ as a threshold for accepting statistical similarities.

Differences in Experiential Measures

We conducted a multivariate analysis of variance (MANOVA) in SPSS 24 using all of the dependent measures and using both condition and a median split of transportability as between-subject factors. We used the application g*Power [83] to calculate the post-hoc power for each of the analyses performed and derived the β values. We present the means and standard deviations for both factors, as well as the F, p, and β values for both factors and the interactions in Table 3.1.

Effects of Fidelity. There were no significant differences between conditions for any experiential measure, and the β values were sufficiently high enough to determine that there were no differences between competence, intuitive controls, relatedness, effort, tension, wishful identification, and behavioural motivation. Autonomy, presence, enjoyment, similarity identification, and embodied presence did not have a high enough post-hoc statistical power to say there was no dif-

ference, but their non-significant differences were still very small (about 0.5 points on a seven-point Likert scale). These results support H1, with the exception of the measures that didn't have enough statistical power to determine that there was no difference.

Effects of Transportability. There were significant differences between players with high and low transportability for all dependent measures, except for relatedness ($p=.09$) and motivation ($p=.12$). With the exception of these two non-significant differences, our results support H2.

Additionally, there were no significant interactions between fidelity condition and transportability, showing that the lack of differences between low-fidelity and high-fidelity versions did not depend on the transportability of the players.

3.7.4 Study 1 Discussion

Our results suggest that for a short narrative game, player experience does not change very much between a low-fidelity prototype and a high-fidelity version with the same story and interactions. The narrative immersion measures of identification were also not statistically different.

Because the goals of this study were to compare play experience between different levels of graphic fidelity, our measurements were primarily focused on measures of play experience, such as need satisfaction and game enjoyment. We decided to include avatar identification as a measure of narrative immersion, but we did not specifically evaluate the players' experience of the narrative itself.

We also decided in this study to keep the controls consistent between the two versions — the game interaction was made as similar as possible by minimizing the necessary navigation and exploration in the high-fidelity condition and making the interaction in the hacking mini-game identical between conditions. As a result, the high-fidelity game was essentially an interactive narrative rather than a complete polished game. While this means the results are not generalizable to games in general, the game we created was fairly representative of a particular genre of narrative-based games, which are often focused on following a story and being present in a 3D world (e.g., walking simulators, such the *The Stanley Parable* [52]). Furthermore, we know that the lack of game interaction or control did not confound our results.

While the results from study 1 indicate that graphic fidelity may not be as important to certain kinds of games, the short play time is not enough to make a general claim. What we can say, however, is that in a short amount of time playing this type of game, graphic fidelity had very little to do with the overall play experience and narrative immersion. This indicates that this type of low-fidelity text interface could be useful for game developers when designing and testing game narratives early in development, before graphics and mechanics are fully implemented.

In our second study, we addressed the limitations from study 1 to help generalize to commercial narrative games.

3.8 Study 2

In study 1 we showed that measures of experience remain similar when comparing low-fidelity and high-fidelity graphics. We made several changes in study 2. First, we included measures of narrative engagement and removed PENS [219] to reduce the study complexity and shift the focus to narrative experience. We kept the IMI [174] as a means of replicating the results from study 1 and understanding effects on experience. Second, we added simple, controlled game mechanics to generalize our results beyond an interactive narrative. Third, we made the low-fidelity version with simplified game controls to test if the fidelity of game interaction affects experience.

3.8.1 Game Design

We created a simple, 2.5D fantasy game with 5 main levels: a prologue, three levels, each with a single unique game mechanic, and an epilogue. The prologue showed a Princess running through a forest, chased by the shadow of a dragon, accompanied by text setting up the story. The first level used sword-based combat, the second level was an infinite runner, and level 3 used a shooting mechanic. The epilogue was a conversation to complete the narrative arc. The game was built initially to study sexism in games [40] however, a neutral version was used in the present study.

Narrative/Procedure

The underlying narrative of the game followed a damsel-in-distress trope. The player took on the role of a knight character on a quest to rescue a Princess captured by an evil wizard. In the prologue, the Princess's kidnapping is shown, and the premise of the story is described to the player through text scrolling across the screen. In level 1, the Knight is investigating the forest where the Princess was last seen when he runs into a woman trapped in a cage in the middle of a bandit camp. The knight fights off the bandits and rescues the prisoner from her prison. The woman, named Victoria, thanks him, but runs off into the forest, prompting the knight to follow her and learn what happened to the Princess. In level 2, the knight runs along a forest path, trying to catch up to Victoria for find out where the Princess was taken. Catching up to her, she promises to lead him to the wizard's castle if he rescues her village from a bandit raid. In level 3, the knight sits outside the village walls with a crossbow and must shoot bandits, while avoiding the villagers, as they run out through the main gate one at a time. Once the knight has killed enough bandits, Victoria leads home to the wizard's castle, as promised. In the epilogue, the knight and Victoria arrive at the castle only to find it abandoned, with Victoria claiming that "The Princess is in another castle", before deciding to permanently join the knight in his quest to find the Princess.



Figure 3.2: low-fidelity (top-left), high-fidelity Level 1 (top-right), Level 2 (bottom-left), and Level 3 (bottom-right) Screenshots

Low-Fidelity Version

The low-fidelity version of the game follows the same story with identical dialogue as the high-fidelity version (Figure 3.2). As placeholders for gameplay mechanics, this version used simple text-based interactions to simulate the actions the player undertook in the high-fidelity version. For example, the knight fighting off the bandits was replaced with player text options such as “Perform sword slash” and the infinite runner level had options such as “Dodge to the left” or “Jump over the fallen tree” to provide players with high-level actions to replace the actions taken in the high-fidelity condition. Unlike in study 1, the game controls in the low-fidelity version were not intended to be comparable to the high-fidelity version and were more representative of what might be possible to prototype very early in the development of a game, when the details of the mechanics are still being developed.

Participants, Deployment, and Ethics

We used identical deployment methods as with study 1, except we chose to recruit 125 participants (46.8% female), one of which was removed using the outlier filter from study 1 ($N_{low-fidelity} = 65$, $N_{high-fidelity} = 59$). The study took an average of 19 minutes to complete (10 minutes of gameplay), and participants were compensated \$5 USD.

3.8.2 Measures

Player Experience was measured as in study 1 using the enjoyment, effort, tension, and perceived competence subscales from the Intrinsic Motivation Inventory (IMI, [174]). Because our goal with this study was to further investigate narrative experience, we chose to leave out PENS [219].

Transportability and Identification were measured using the same scales as in study 1 [110, 258] (Cronbach's $\alpha = .94, .94, .96, .91$ for similarity identification, embodied presence, wishful identification, and transportability).

In addition, we added the following measures in study 2.

Transportation was measured using the Transportation Scale [110] which measures an individual's transportation into a narrative world during their experience with a particular narrative (e.g., "I wanted to learn how the narrative ended"). Transportation is indicative of the quality of a narrative that is experienced (Cronbach's $\alpha = .83$).

Narrative Engagement was measured using the Narrative Engagement Scale (NES, [49]), which measures four aspects of narrative engagement: *narrative understanding* — reverse coded items to measure the understanding of a narrative, e.g., "I had a hard time recognizing the thread of the story", *attentional focus* — reverse coded items asking participants about how well they were able to mentally focus on the story, e.g., "I had a hard time keeping my mind on the game", *narrative presence* — how strongly they felt present in the narrative, e.g., "At times during the game, the story world was closer to me than the real world", and *emotional engagement* — how emotionally connected they felt to the events of the story, e.g., "The story affected me emotionally". To maintain consistency across the four subscales, results from narrative understanding and attentional focus are presented reversed, so for all four a higher value refers to stronger narrative engagement (Cronbach's $\alpha = .79, .89, .75, .82$).

In the results and discussion below, we distinguish between *player experience*, *narrative engagement*, and *narrative immersion* as discussed in the related work section.

3.8.3 Results

We conducted an identical analysis as was done with study 1. We present the means and standard deviations for both independent factors, as well as the F, p, and β values from the MANOVA by condition, a median split of transportability, and the interaction between the two in Table 3.2.

Differences in Experiential Measures

Play Experience. In terms of play experience, there was enough statistical power to assert that there were no significant differences between conditions for enjoyment, effort, tension, and competence.

Narrative Engagement and Immersion. There was enough statistical power to assert that there were no significant differences between narrative understanding, attentional focus, and embodied presence. Narrative presence and emotional engagement were also not statistically different and had a relatively high power, but didn't meet the threshold of 0.8 ($\beta = 0.71$ and $\beta = 0.72$, respectively) to accept the null hypothesis.

There was, however, a significant difference between the low-fidelity high-fidelity versions in terms of transportation ($p = .04, \eta_p^2 = .034$), similarity identification ($p = .05, \eta_p^2 = .03$), and wishful identification ($p = .02, \eta_p^2 = .04$), although the actual difference between means were small (less than 0.5 on a seven-point Likert scale). An independent-samples t-test showed no significant differences in the transportability between the groups ($t_{22} = 0.18, p = .859$). This shows that the differences cannot be explained by random assignment to the game versions. The results showed that experienced transportation was slightly higher in the low-fidelity version than the high-fidelity version.

Effects of Transportability. Comparing high and low transportability revealed similar results as in study 1 — there were significant differences between highly-transportable and low-transportable players for all dependent measures, with only experienced tension failing to reach significance. Additionally, there were no significant interactions between condition and transportability, showing that the lack of differences between the LoJFi and high-fidelity versions did not depend on the transportability of the players.

3.8.4 Study 2 Discussion

The main interpretation of results is very similar to study 1. For all of the measures of play experience there were no significant differences between conditions, indicating that low-fidelity narrative prototypes can be used to assess the game narrative in early stage development. Furthermore, there were no significant differences in experienced narrative engagement, showing that the quality of the narrative could be assessed outside of the context of the full game experience (i.e., high-fidelity graphics and game interactions).

Interestingly, the two measures of immersion, transportation and identification, were two of the exceptions to this pattern, though the differences were very small in terms of the big picture of overall experience. This could be an indicator that for story immersion, while the story itself is important, there are other factors beyond the story that go into it, such as the presentation of the narrative world and the interaction with and presentation of different characters.

3.9 Discussion

In this section, we will discuss the implications of our results, industry use cases, and how game studios could integrate narrative prototypes into their existing design processes. We will also discuss the limitations of our studies and future opportunities.

3.9.1 Implication of Results

While we did see similarities between many measures of player experience and narrative immersion, these similarities were often not significantly similar and, in some cases, were significantly different.

From a practical point of view, the most useful results from a narrative prototype is the narrative engagement. In Study 2, we showed that level of fidelity had no effect on narrative understanding and attentional focus, and narrative presence and emotional engagement were very similar between the conditions. This shows that the addition of visual graphics and game mechanics had very little effect on how well participants understood the narrative, the level of focus they placed on the story, their presence in the narrative, and the emotional engagement they felt from the story. These four measures suggest that early testing of a story in a text-based format would be useful for game designers to test for these particular narrative properties early in the story's conception, since they appear to be fairly independent from the game mechanics and visuals.

3.9.2 Industry Use Cases

There are two different situations in which the ideas underlying our proposed narrative prototypes could be used and applied to game studios. The most obvious, and the main motivation for this paper, is a tool to be integrated into the iterative process that most game studios now use.

A second application for narrative prototyping is to cheaply evaluate game ideas before committing to a project. While more study is needed to be able to confidently say that our technique would apply across games of different types and genres, our results show that, at least for a certain type of narrative game, we can get similar levels of narrative experience from a simple, text-based narrative compared to a complete experience. This would allow game studios to create fast and cheap prototypes for several story-based games in a short amount of time and quickly test them with users to get a feel for which game, or games, would be the best to invest further resources into.

3.9.3 How to Integrate Prototypes into the Design Process

Here, we discuss how a narrative prototype could be embedded in the iterative design process of games, based on the current use of low-fidelity prototypes, and using our understanding of industry

practices. In our approach, we adhere to the following three guidelines for effectively making use of a narrative prototype: close integration with the current state of the game, iterative testing and feedback, and increasing fidelity over time.

A lesson Bioware learned from working on *Dragon Age: Inquisition* [30], and the lesson Bungie learned with *Destiny* [47], is the need for story to be developed in tandem with the game itself. If the narrative team doesn't consider how players will interact with the game, and if the game designers don't consider the game narrative while developing their content, they may introduce dissonances between the narrative and the game. The goal, then, is to match the narrative prototype with the game's current level of development. This is similar to the goal of interactive system prototypes in which the fidelity of the sketch or prototype should match the maturity of the idea [51].

Early on in the development cycle, the narrative prototype would be similar to the text-based prototypes used in our studies. The goal would be to test and evaluate the story, looking for narrative dissonances, plot holes, and travel through branching and fold-back story structures. As the story is tested and refined, the game would be developed in parallel, constructing low-fidelity versions of the systems and interactions. As the game gets fleshed out, the narrative prototype should slowly add in the assets and mechanics as they become available (for example, how we added the static character image in study 2). This stage can also be used to test multiple options with users (e.g., different avatars or character personalities). As a game level gets constructed, the descriptions in the narrative prototype could be replaced by the actual level, placing the prototype inside the level, with the tester now focused more on the dialogue, and ensuring the setting makes sense with that version of the story. As mechanics get developed, e.g., a sword attack, these can also be added to the prototype, replacing the early-stage text-based interactions. The goal with iteratively developing the story in this way is to always be aware of how the story fits with the game and vice versa. By always testing the latest version of the story, game developers have a better chance of finding dissonances between the two, as with *Destiny* [47], as early as possible, instead of waiting until the game is nearing completion.

3.9.4 Limitations and Future Opportunities

In this section we will discuss limitations to our study design, and future research opportunities.

Limitations

In both studies, the two conditions differed in graphical fidelity and interaction. While we tried to minimize differences in terms of game interaction for study 1, there were still slight differences due to the closely-coupled nature of graphics and interaction. We argue that this prioritizes external validity, as both conditions represent a snapshot in the development timeline of a game. In study 2, we iterated on this closely-coupled aspect, further differentiating the interaction between conditions.

The design choice of representing the mechanics as simple text options is representative of how such a prototype would be created early in a game’s development, with no access to graphics or systems.

A second limitation of our work is that both games we used in the studies were already developed and the narrative prototypes were developed after the games, rather than before as would happen in a real-world use case.

A third limitation is that our games — although enjoyed by participants according to the means — were representative of what a small, indie studio would create as a demo and not what a triple-A studio would generate. Because the goal is for the fidelity of the narrative prototype to match the maturity of the game’s development, we don’t see this as a potential source of error; however, higher-fidelity narrative prototypes may yield different play experiences than their game-based counterparts of commercial quality.

Finally, participants were only exposed to the game for a short time (ten to fifteen minutes). Our results show no significant differences in play experience between the low-fidelity and high-fidelity versions, but we speculate that with a significantly longer play session, differences may emerge due to a lack of visual stimulation in the low-fidelity versions over a long period of time. Short sessions are ideal for playtesting, however, which fits the use case of our prototyping method.

Future Work

We would like to expand on the ideas in this paper and more deeply consider narrative prototyping. In both of our studies, we used games we had created. While this gave us an advantage in having control over every aspect of their design, the experiences were short and were not comparable to commercial games in terms of content and quality. The next step would be to replicate study 2 with an off-the-shelf narrative game, such as *The Stanley Parable* [52]. This would increase the generalizability of our results, showing that narrative prototypes can be effective on commercial games with complex content and interactions, build by professional writers. Second, we would like to document the processes we describe in the paper by iteratively developing a game using narrative prototyping, to test the feasibility of the method in practice. Finally, the narrative prototypes we used are focused on early-stage narrative evaluation and, to some extent, character dialogue. We would like to expand on this idea and look at creating different types of narrative prototypes, focusing on specific aspects of story development, such as world building, designing character personalities, and evaluating procedurally-generated narratives.

3.10 Conclusion

Iterative development and evaluation is important for all types of interactive systems, including interfaces and games. While iterative methods have become common practice for most game

designers, the integration of game narratives early in development is still new and unfamiliar to many developers. We proposed using text-based narrative prototypes as a means to iteratively test and develop game narratives, allowing writers to work closely and in parallel with the designers of the game systems. Through two studies, we compared high and low levels of graphic fidelity and player interaction with game mechanics and show that all experiential measures are not affected by playing the narrative prototype. Our findings provide evidence that narrative prototyping could feasibly be used by game studios to quickly and cheaply evaluate game ideas before committing development resources to them and to more closely integrate their narrative design into existing iterative development processes.

Table 3.1: Study 1 results - Means and SD for Low-Fidelity and High-Fidelity conditions, as well as F, p, and β values from a Multivariate Analysis of Variance (MANOVA) for Condition, Transportation, and the interaction.

	Low-Fidelity		High-Fidelity		Low-Trans		High-Trans		CONDITION			TRANS			COND * TRANS		
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	F	P	β	F	p	β	F	p	β
Enjoyment	4.88	1.35	4.30	1.73	4.00	1.65	5.28	1.08	2.82	0.10	0.60	19.26	0.00	0.01	1.59	0.21	0.75
Effort	5.44	1.27	5.53	1.27	4.99	1.33	5.97	0.93	0.23	0.64	0.92	13.36	0.00	0.04	0.04	0.84	0.94
Tension	3.14	1.52	2.81	1.21	2.79	1.33	3.29	1.42	0.95	0.33	0.83	3.97	0.05	0.47	0.95	0.33	0.83
Competence	5.09	1.19	4.90	1.41	4.67	1.18	5.37	1.29	0.33	0.57	0.91	6.72	0.01	0.25	0.50	0.48	0.89
Autonomy	3.93	1.44	3.40	1.46	3.28	1.47	4.20	1.33	2.55	0.12	0.63	11.87	0.00	0.06	2.01	0.16	0.69
Relatedness	4.07	1.40	3.93	1.34	3.79	1.28	4.27	1.40	0.14	0.71	0.93	3.04	0.09	0.57	0.51	0.48	0.89
Presence	4.48	1.39	3.99	1.66	3.50	1.36	5.05	1.23	2.08	0.15	0.68	28.85	0.00	0.00	0.00	0.96	0.95
Intuitive Controls	5.74	1.09	5.61	0.94	5.32	1.03	6.00	0.88	0.22	0.64	0.92	8.33	0.01	0.16	0.10	0.75	0.94
Similarity ID	3.31	1.56	3.73	1.56	2.85	1.30	4.21	1.52	2.43	0.12	0.65	19.61	0.00	0.01	0.21	0.65	0.92
Embodied Presence	4.02	1.67	3.47	1.77	2.85	1.42	4.69	1.52	1.88	0.14	0.72	30.33	0.00	0.00	0.03	0.85	0.95
Wishful ID	3.07	1.50	3.12	1.40	2.58	1.24	3.64	1.45	0.09	0.76	0.94	12.18	0.00	0.06	0.57	0.45	0.89
Motivation	0.95	0.13	0.94	0.18	0.91	0.21	0.97	0.16	0.10	0.75	0.94	2.48	0.12	0.64	0.25	0.62	0.92

Table 3.2: Study 2 results - Means and SD for Low-Fidelity and High-Fidelity conditions, as well as F, p, and β values from a Multivariate Analysis of Variance (MANOVA) for Condition, Transportation, and the interaction.

	Low-Fidelity		High-Fidelity		Low-Trans		High-Trans		CONDITION			TRANS			COND * TRANS		
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	F	P	β	F	p	β	F	p	β
Enjoyment	4.93	1.41	4.76	1.28	4.43	1.31	5.25	1.26	0.96	0.33	0.83	13.12	0.00	0.04	0.00	0.98	0.95
Effort	5.35	1.06	5.58	1.16	5.08	1.09	5.83	1.01	1.01	0.32	0.83	15.06	0.00	0.02	0.80	0.38	0.85
Tension	2.37	1.24	2.25	1.06	2.25	0.96	2.37	1.32	0.29	0.59	0.92	0.25	0.62	0.92	3.48	0.07	0.53
Competence	5.58	0.79	5.75	0.81	5.47	0.78	5.85	0.78	1.03	0.31	0.82	7.17	0.01	0.23	0.01	0.92	0.95
Similarity ID	3.48	1.42	3.06	1.37	2.84	1.24	3.71	1.44	4.00	0.05	0.48	13.92	0.00	0.04	0.33	0.57	0.91
Embodied Presence	3.83	1.60	3.64	1.65	3.02	1.35	4.43	1.56	1.12	0.29	0.82	28.42	0.00	0.00	1.21	0.27	0.80
Wishful ID	4.02	1.47	3.52	1.54	3.18	1.34	4.36	1.47	5.33	0.02	0.35	23.35	0.00	0.00	0.26	0.61	0.92
Transportation	4.59	0.88	4.37	0.89	4.05	0.84	4.91	0.72	4.28	0.04	0.46	40.47	0.00	0.04	0.00	0.96	0.95
Narrative Understanding	5.11	1.05	5.02	0.87	4.85	1.12	5.28	0.73	0.42	0.52	0.91	6.18	0.01	0.29	0.41	0.52	0.91
Attentional Focus	4.90	1.20	4.82	1.12	4.58	1.27	5.13	0.98	0.32	0.57	0.91	7.54	0.01	0.21	0.08	0.77	0.94
Narrative Presence	4.15	1.47	3.90	1.33	3.48	1.26	4.56	1.34	1.90	0.17	0.71	22.14	0.00	0.00	0.05	0.83	0.95
Emotional Engagement	3.75	1.49	3.50	1.51	3.01	1.25	4.23	1.48	1.83	0.18	0.72	25.04	0.00	0.00	0.68	.041	0.86

4 MANUSCRIPT C: RED OR BLUE PILL: FOSTERING IDENTIFICATION AND TRANSPORTATION IN RPGS

4.1 Introduction to Manuscript C

In Chapter 2, we showed that a belief alignment between the player and main playable character increased identification and, through identification, increased overall enjoyment. In that chapter the characters were designed a priori, and were identical for each player. Since we showed that player identification was strongly driven by whether or not the player shared the same beliefs exhibited by the main playable character, designing characters with strong beliefs may reduce the overall experience for many players with opposing beliefs. One solution to this problem is for all players to define the main playable character’s beliefs themselves by making dialogue and narrative choices to dictate how the game progresses.

In this chapter we will test this solution in the context of two-dimensional morality in a text-based narrative interface. In this Chapter we will address 4 main research questions:

1. Given free choice to make decisions, do players make choices in line with their personal beliefs?
2. How does forcing a player to act morally or immorally affect identification and transportation?
3. How does presenting a player with a subset of choices tailored to their play-style affect identification and transportation?
4. How is perceived effort and choice presentation driving the effects from the second and third research questions?

4.1.1 Publication and Individual Contribution

This work was published as the following: [41]. Although a multi-authored contribution, I was responsible for leading the research. My roles included: designing the experiments and constructing the interactive systems used in the experiments (i.e., constructing the text-based interfaces, coding the algorithms to dynamically present dialogue choices and construct player models based on

choices). Additionally, I was involved in the data analysis and contributed to the literature review and writing of the publication. I also presented the publication at the conference in which it was published. I contributed approximately 75% of the total work for this project and publication.

4.2 Abstract

Through free choice, individuals can exert control over the environment and experience agency. Research has suggested that tailoring aspects of choice to a player’s type can provide benefits; however, commercial role-playing games (RPGs) generally provide static opposing options from a spectrum (e.g., paragon versus renegade). To inform the design of choices in RPGs, we conducted three studies comparing configurations of dialogue choices across the morality spectrum in an RPG prototype. Study 1 compared two polar (opposing) choices with two moral or two immoral choices. Study 2 contrasted polar choices with similar choices tailored to the player’s morality, modelled from their initial 10 decisions. Study 3 compared polar choices with two tailored and one opposing choice. Our results consistently show that including the opposing choice is important to a player’s identification with their character, even when we can reliably predict the decision that they will make.

4.3 Introduction

Converging evidence from different fields of science agrees that having choices is valuable. Having the ability to choose autonomously allows for people to consciously exert control over the environment, resulting in a satisfied need for control and a feeling of agency [96, 141, 159, 161]. Humans [44, 242, 243] and animals [58, 243] alike prefer having choice over having no choice. Having more options available to choose from increases the probability of the preferred option being present [161, 197, 199]; however, having too many choices can overwhelm the chooser and decrease satisfaction with the decision [128, 172, 225, 251].

In games, choices are often implemented as a means of increasing interactivity [66, 67, 81, 111]. Role-playing games (RPGs) in particular, such as the Dragon Age [30] and Mass Effect [28] series, use choices as a tool to give the player control over their character and the narrative world [81]. It has been proposed that for in-game choices to have a positive effect on player experience, they need to give players a tangible outcome or consequence to having made the decision [126, 188]. Some games incorporate a tangible impact by maintaining a model of the player’s previous choices to use in tailoring future content (e.g., Mass Effect tracks ‘Paragon’ and ‘Renegade’ choices and some options are only available with a high enough score). On the other hand, it has also been proposed that the presentation of the choice itself is valuable, regardless of the outcome [159–161]. In addition to the importance of choice consequences, the meaningfulness of a presented choice has been proposed

as important to consider [126, 248, 253]. Previous approaches to create meaningful experiences in games include: tailoring the game narrative to match a player model [253], increasing feelings of social connections [196, 215], and enforcing a commitment to meaning when making decisions [248].

Even though choices are an important tool for enhancing play experience, little is known about the type of choices a player should be presented with, how choices are made by players, and what the resulting effects are on experience. In terms of choice types, many games that offer players a choice provide options that span a spectrum of experience (e.g., choosing between rude or polite dialogue). On one hand, giving players range in available choice should be good as it enhances their perceived freedom (i.e., objective agency) to make that choice. On the other hand, giving players a range in options may not result in as meaningful of an experience as a more carefully constructed subset of more similar options. Research on meaningfulness [253] suggests that tailoring choices to align with a model of previous choices, and therefore providing players with a smaller domain of possibilities from which to choose, can be beneficial.

This idea of providing players with more tailored choices in games is supported by player sentiments. In a review of Mass Effect 3 [28], the game is criticized for moving toward more polarized choices, as opposed to the more nuanced options in previous games [46]. Similarly, a blog post about Dragon Age: Inquisition [30] criticizes the game’s decisions for not considering previous choices made by the player or the character’s backstory [276].

Designers who wish to create meaningful, transportative, and enjoyable games that include choices have little guidance on the types of in-game choices that should be presented—*should choices include range or be more similar, but tailored to a player’s intended play style?*

In sum, there are two conflicting lines of reasoning. On one hand, psychological research on choice and agency teaches us how important variety in options is and that the feeling of a choice having a tangible impact is crucial. On the other hand, we have literature and player sentiment supporting the importance of a choice’s meaningfulness. How do these two perspectives interact? Is a choice as effective, or better, if we reduce the variety of options in exchange for making the choice more meaningful? Given the negative consequences of having too many options [128, 172, 225, 251], it is possible that making choices more meaningful has a higher growth potential for creating more effective decisions in games.

In a series of three studies, we evaluate how different types of choices affect player and narrative experience in the context of a role-playing game prototype. In particular, we compare choices with high objective agency (i.e., moral and immoral choices) with a more meaningful choice option that has less objective agency (i.e., similar morality, Study 1; and tailored to the player model, Studies 2 and 3). Our results suggest that giving the player a wider range of possibilities (i.e., including both moral and immoral choices compared to choices of similar moral valence) is important in fostering player identification and transportation into the narrative—even when the similar options

are tailored to the player’s intentions. Further, we show that providing meaningful choices with less objective agency (i.e., similar choices) does not harm transportation and identification when the opposing option is also included.

The contribution of our work is twofold: firstly, this paper can help designing decision options in games that positively contribute to player experience. Secondly, the results significantly contribute to the understanding of the underlying processes in a player’s experience of choice in games.

4.4 Related Work

4.4.1 Perceived Agency and Control

Converging evidence suggests that choice is inherently desirable, even when choice is not linked to the outcome [44, 159, 242, 243]. This is exemplified by the illusion of control, in which merely the possibility to choose results in a higher perception of control over the world [154, 155]. According to a recent review [161], there are three main benefits of choice and subsequently a positive perception of one’s control. Firstly, the expected value of having multiple choices is higher compared to having fewer choices, because the likelihood of being presented with a higher-value option increases with the number of available options [12, 263, 269]. Secondly, similar to the learned helplessness paradigm [9, 166], not being able to choose—and therefore not being in control—increases stress [63, 69, 104, 268]. Thirdly, choices tell us something about ourselves. According to the self-perception theory [24] and the dissonance-theory [85], individuals use their choices to infer their own attitude towards something or someone.

Sense of agency is linked to perceived control over one’s actions and their consequences in the world. Being an agent can also mean to be able to make a choice. If a positive sense of agency has been established, the outcomes of one’s actions are perceived as more positive compared to equal outcomes without the individual’s involvement. This phenomenon has been described in the framework of a self-serving bias [45, 112], and has also been reflected in findings showing reduced pain sensation after self-administration of pain or after predictable pain-related cues [20, 54, 68]. Furthermore, it has been suggested that humans select actions based not only on their outcome, but also on their control value (i.e., actions that result in higher control and therefore a higher sense of agency) [141, 221]. This suggests that choices indicating increased control are perceived as inherently rewarding [21, 23, 159]. Additionally, increased control has been associated with increased persistence after failures [10, 25, 87, 158], which is in line with research on the buffering effect of choice on negative emotion and losses [160, 161, 218].

In the context of games this means that aiming to elevate the sense of agency in players might not only enhance their enjoyment of the game, but is likely to increase their playtime and perseverance after in-game failures.

4.4.2 (Tailoring) Meaningful Choices in Games

Games, more than any other form of media, enable the player to choose and exert control over their experience, and although choices are a ubiquitous game dynamic, little research has been done on the effect of specific types of choices. Most recently, Iten et al [126] set out to investigate the effect of meaningful, moral choices on player experience, enjoyment and narrative engagement. They reported that for choices to be perceived as meaningful, the choice had to have tangible consequences (for similar arguments see [227, 260]).

This need for players to witness the consequences of their in-game choices aligns with other research into choices and agency in games. In particular, previous work has argued that choices in games should be designed with a commitment to meaning in mind. [248] Specifically, when players make a choice, they are creating a commitment on behalf of the player character. In order for this commitment to make sense, and to make the player feel as though their choices matter, the game system should adapt to hold the player accountable to the decisions they have made in a type of social contract.

Previous work has also shown that player experience can be enhanced by using player choices to build a computational model of players' preferred playstyle and tailor the game's narrative or decisions based on this model. [170, 213, 253]. In this system, individual play styles were modelled using five-dimensions: Fighter, Power Gamer, Tactician, Storyteller, and Method Actor. Player choices were tagged with values for each of these dimensions. When a player made a choice, their model was updated by adding the corresponding scores from the choice. Future content was then tailored based on the player's current model. This tailored storytelling is similar to a method employed in serious game design in which personalized feedback and information is used to promote behavioural change [151, 152].

4.4.3 Narrative Experience

Narrative experience has many facets but for the purpose of this paper, we discuss three major dimensions: transportation, narrative engagement and identification.

Transportation refers to the sensation of leaving your physical surroundings and being immersed in the narrative world [110]. The personality trait of transportability—i.e., the individual propensity to be transported—has also been shown to predict transportation [26, 61].

Narrative Engagement [49] is based on four different processes: narrative understanding (i.e., cognitive perspective taking and making sense of the story), attentional focus (i.e., undivided attention for the story), emotional engagement (i.e., empathy for the characters) and narrative presence (i.e., the sensation of having entered the narrative world).

Narrative engagement can be explained in the framework of mental models, which are cognitive

representations of aspects of the external world and influenced by an individual’s experiences [137, 138, 264]. For example, an individual might have a mental model of the earth containing shape and size. Similarly, a mental model of a story may contain, among other things, the chronological sequence of events and the spatial setting. Therefore, it was argued that narrative engagement is rooted in an individual’s mental model of a certain story. [26, 278].

Identification refers to the sensation that the player experiences the game from the character’s point-of-view [64]. Identification (with the character) has three main dimensions [258]: similarity identification (i.e., feeling to the character), embodied presence (i.e., taking the character’s perspective) and wishful identification (i.e., the character is a desired version of the player). Identification has been shown to foster intrinsic motivation [32, 33].

4.5 System Description

We present three studies in this paper. All three studies use an interactive dialogue system that maintains a morality model of the player based on their decisions, and tailors future choices based on this model. We briefly describe how the system and dynamic model are created and maintained.

4.5.1 Morality Model

In order to tailor the options presented to individual players, we first created a model to be able to predict a player’s choice. Given that the players would be making dialogue choices we elected to use a model of morality, based on Moral Foundation Theory [114] which has been previously used in games research [126, 150]. Our model consisted of the five primary axes of Moral Foundation Theory: care/harm, fairness, loyalty, respect, and purity. The player model consisted of a score for each of these five axes, as well as a combined score for overall morality.

Each dialogue option a player could choose was tagged with a score for each of the five axes. When the player made a choice, their model was updated by adding the vector from the option to their morality model.

After a short calibration phase of making decisions (we used the initial ten decisions), the player model at that time was saved and this was the model used by the system to tailor the decisions presented to the player.

4.5.2 Narrative System

We constructed an interactive dialogue system that allows players to play through a text-based adventure. The system can be configured with various rules and constraints to procedurally construct and present players with a set of options by combining predefined sentence fragments. Each fragment is tagged with the moral axes as described in the previous section. The system assigns

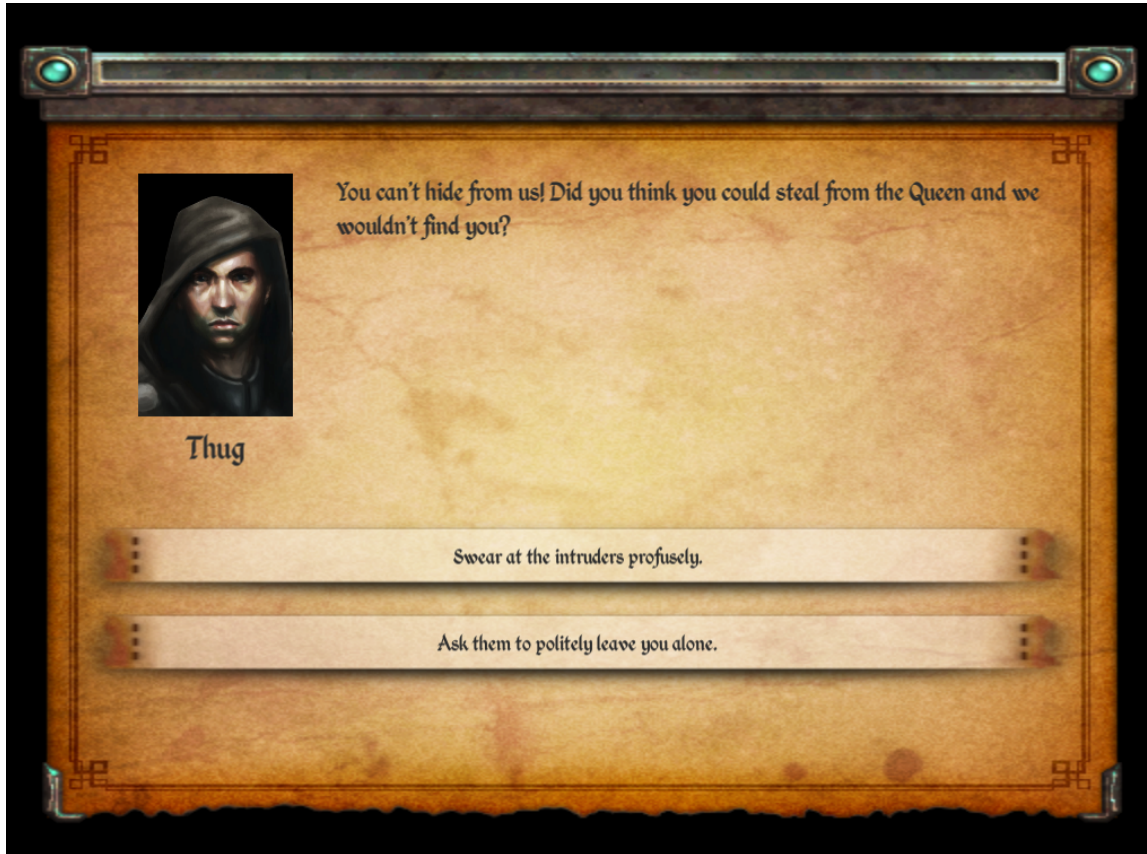


Figure 4.1: Screenshot of the interface design of dialogue system used in Manuscript C studies.

a morality score for each generated option, and automatically updates the player model when a choice is made by the player.

The dialogue system leads players through an interactive narrative that follows a young roguish mage on an adventure that centered around performing a job for her estranged mother. The story dealt with complex relationships between the protagonist and her mother, as well as the protagonist and an old friend named Ember with whom things had ended badly. The story focused on themes of forgiveness, trust, and betrayal. Throughout the story there were 72 story fragments that players clicked through. Out of the 72 fragments, 35 were decisions that the player had to make, including both dialogue options and actions to move the story forward. See Figure 4.1 for a screenshot of the game.

4.5.3 Dialogue Choices

Though the system supports more complicated procedural construction of options, we elected to write the possible options by hand to limit the variance in experience between participants. For each decision, we created two versions of a moral option, as defined by our morality model, and

two versions of an immoral option. We ensured the two versions differed in content and the specific moral axes they were tagged with (e.g., one moral option may involve purity and the other may involve fairness), but the resulting options all had approximately the same morality score so the decisions were as balanced as possible.

We used the following guidelines when tagging the dialogue options with the moral axes:

Harm/Care : Options were tagged with this axis if the choice resulted in the character attempting to physically harm or protect another character. (e.g., throw a fireball at an enemy or stand in front of a character to protect them from danger).

Fairness/Reciprocity : Options were tagged with this axis if the choice involved the character acting fairly or unfairly to another character (e.g., performing a sneak attack or admitting that you are wrong).

Ingroup/Loyalty : Options were tagged with this axis if the choice involved the character acting loyally or disloyally towards their mother or friend, Ember (e.g., obeying her mother).

Authority/Respect : Options were tagged with this axis if the choice involved the character treating her mother or Ember with respect or disrespect. (e.g., insulting Ember).

Purity/Sanctity : Though this axis is traditionally associated with aspects such as chastity or hygiene, we used this axis to measure purity as was relevant to the narrative content, namely how the character feels about evil magic, which was fundamental to the plot. (e.g., whether or not the character uses an amulet full of dark magic or decides to destroy it.).

For the first ten story pieces, players were all presented with the same two options, in order to calibrate the individual player model. These decisions consisted of a moral option and an immoral option, and they evenly included a mix of all five moral axes as defined in our model.

After the calibration phase, when presenting a decision to the player, the system randomly selected from the pool of generated options to fulfill one of several rules:

POLAR : One option had positive morality (i.e., moral choice), one had negative (i.e., immoral choice).

MORAL : All of the options had a positive morality.

IMMORAL : All of the options had a negative morality.

TAILORED : All of the options had a morality that aligned with the calibrated morality model.

Once the options were chosen, they were presented to the player in a random order to eliminate sequence effects.

4.6 Recruitment, Outliers, and Bot Detection

For all studies presented in this paper, we recruited participants from Amazon’s Mechanical Turk, which has been shown to be a reliable research tool [169]. Due to recent concerns of data validity and an increase of bots on the platform [238], we employed strict filtering procedures to detect and remove outliers, inattentive participants, and bots from our data.

We removed all participants that displayed zero variance in their answers on any of the questionnaires to detect participants who just clicked the same answer on a Likert-scale on every single question. We also removed participants who spent more than three standard deviations longer or shorter than the mean on more than one questionnaire to remove participants that clicked through questionnaires too quickly and participants that were multitasking, or who left the study window open in their browser while doing something else. Finally, we asked several broad questions regarding the narrative to ensure participants had read and understood the story. Participants who didn’t answer the question we had asked (e.g., when asked about the character’s morality some bots replied by defining morality based on a web search) were removed. We describe how many participants were removed in each individual study description.

4.7 Study 1

RPGs generally present players with a spectrum of options (e.g., players can choose between a polite response and a rude response). In the case of the dialogue system we created, a presentation of one moral choice and one immoral choice reflects the way in which games currently present choices to players.

Previous work in the context of persuasive games has argued that a game is most persuasive when players have to invest mental effort into making a choice as opposed to picking options for the sake of picking options [272]. Applying this in the context of moral choices, a choice that requires more mental effort to make a selection should be more persuasive, which leads to positive outcomes [73, 201, 202].

We propose that a choice with both options on the same end of the moral spectrum would require players to exert more cognitive effort into making the choice. By making the choices similar, we reduce any preference biases so players are forced to choose between two things they feel similarly about.

The goal of Study 1 was to compare POLAR choices (one moral and one immoral option) with MORAL (two moral options) and IMMORAL (two immoral options) in a between-subjects study design with 90 participants, using the recruitment approach described in the previous section.

4.7.1 Measures

Here we present the constructs we measured, along with the number of items and Cronbach's alpha for each subscale.

Trait Measures

Transportability : Transportability (21 items, $\alpha = .903$), or an individual's trait susceptibility to be transported into a narrative world, was measured using the Transportability scale [26].

Sense of Agency (SONA, SOPA) : We measured an individual's sense of negative agency (SONA) (7 items, $\alpha = .925$), and sense of positive agency (SOPA) (5 items, $\alpha = .836$), using the Sense of Agency Scale [250].

Moral Foundations Questionnaire : An individual's affiliation with the five axes of Moral Foundation Theory [114] were measured using the short version of the Moral Foundation Questionnaire (MFQ) [108]. (care/harm (4 items, $\alpha = .657$), fairness (4 items, $\alpha = .696$), loyalty (4 items, $\alpha = .692$), respect (3 items, $\alpha = .827$), purity (4 items, $\alpha = .885$)).

State Measures

Transportation : The amount participants were immersed in the narrative world was measured using the transportation scale (11 items, $\alpha = .637$) [110].

Autonomy : Perceived autonomy (3 items, $\alpha = .886$) was measured using the autonomy subscale from the Player Experience of Needs Satisfaction Scale (PENS) [219].

Effort : We measured effort (4 items, $\alpha = .798$) using the effort-importance subscale of the Intrinsic Motivation Inventory (IMI) [174].

Enjoyment : We measured enjoyment (5 items, $\alpha = .839$) using the enjoyment/interest subscale from the IMI [174].

Identification : Player identification with the main character was measured using the Player Identification Scale [258], which measures three constructs: similarity identification (6 items, $\alpha = .927$), embodied presence (6 items, $\alpha = .930$), and wishful identification (5 items, $\alpha = .901$), and can be combined into a single measure of character identification [258].

Narrative Compliance Checks : To ensure participants read and understood the story, which was necessary to consider their responses, we asked several open questions regarding the narrative. Participants were asked to summarize the plot, list the name of the protagonist, describe whether they made decisions as themselves or as the character, and describe the protagonist's morality.

4.7.2 Data Analyses

Because we have less control over behaviour on MTurk, we need to remove participants who do not actually read the story or consider their responses (i.e., by just click through) and try to detect bot participants [238]. To do so, we calculated the variance in responses on each page of questions, and removed any participant who displayed a variance of 0 (e.g., clicked the middle bubble for each question). Additionally, we calculated the time spent on each questionnaire and removed any participant that was greater than 3 SD above or below the mean on more than one questionnaire (e.g., clicked through multiple questionnaires without reading or spent a large amount of time multitasking on another browser window).

In total we removed 11 participants, leaving 79 participants remaining in the dataset for analysis ($N_{polar} = 26$, $N_{moral} = 28$, $N_{immoral} = 25$).

We then conducted partial correlations of the trait variables (i.e., transportability, sense of positive agency, sense of negative agency, moral foundations) with the state variables (i.e., transportation, identification, effort, enjoyment, autonomy). Because the trait measures often correlated with the state measures, we included them as co-variates in our analyses.

We conducted a multivariate analysis of co-variance (MANCOVA) with choice (polar, moral, immoral) as a between-subjects factor, and transportability, sense of positive agency, and sense of negative agency as co-variates on the dependent measures of transportation, identification (similarity identification, embodied presence, wishful identification), effort, enjoyment, and autonomy. Mediation analyses were conducted with Process 3.00 in SPSS 25.

4.7.3 Results

The MANCOVA showed significant effects of choice on similarity identification ($F_{2,73} = 4.2, p = .020, \eta_p^2 = .11$), embodied presence ($F_{2,73} = 6.1, p = .004, \eta_p^2 = .15$), and wishful identification ($F_{2,73} = 6.5, p = .003, \eta_p^2 = .16$). There was also a significant effect of choice on transportation ($F_{2,73} = 4.1, p = .021, \eta_p^2 = .11$).

Bonferroni-Holm corrected pairwise comparisons showed that participants in the immoral condition experienced less similarity identification, embodied identification, wishful identification and transportation than participants in the polar or moral conditions (all $p < .05$).

In terms of experience, there were no main effects of choice on autonomy ($F_{2,73} = .06, p = .944$) or enjoyment ($F_{2,73} = 1.9, p = .165$), but was a significant effect on effort ($F_{2,73} = 5.2, p = .008, \eta_p^2 = .13$). Bonferroni-Holm corrected comparisons revealed that participants felt they invested less effort in the immoral condition than moral or polar (all $p < .02$).

Additionally, we calculated overall identification, as all identification measures showed the same trends. Results were in line with the subscale results for identification. We used this measure to

determine if invested effort mediated the effects of condition on identification (controlling for the same co-variates as in the MANCOVA). We found that choice predicted identification ($p=.003$), but that effort did not ($p=.057$); further, including effort in the model did not significantly reduce the direct effect of choice on identification ($p = .009$).

To determine whether the variance in identification and transportation could be explained by the differences in invested effort, we conducted mediation analyses, which investigates whether a latent mediating variable (M) explains the variance predicted between a predictor (X) and a dependent measure (Y) [38]. Because we must use dichotomous predictors [38] and we had three conditions, we just considered the polar and immoral conditions. We first modeled whether the prediction of choice-condition on identification was mediated by effort, and second whether the prediction of choice-condition on transportation was mediated by effort. Choice predicted identification ($\beta = .496, p = .005$) and remained a significant predictor, even when effort was included in the model ($\beta = .431, p = .017$). Further, the confidence interval for the indirect effect included zero (LLCI=-.226, ULCI=.012), suggesting that there was no mediation [38]. In terms of transportation, choice predicted transportation ($\beta = .172, p = .012$) but did not remain a significant predictor when effort was included in the model ($\beta = .118, p = .062$). Further, the confidence interval for the indirect effect did not include zero (LLCI=-.126, ULCI=-.001), suggesting that there was a partial mediation [38].

Finally, we were concerned that participants may have been randomly assigned to a choice condition that did not align with their preferred morality to play. Because the first 10 decisions were polar in each condition, we used those choices to calculate a predicted morality. In the moral condition, 11 participants were predicted to play morally and 17 were predicted to play immorally; in the immoral condition, 12 were predicted to play immorally, and 13 to play morally. The relatively similar distributions suggest that it is not a mismatch between intended playstyle and assigned condition that drives the lower transportation and identification in the immoral condition; however, that our experiment was not designed to test this hypothesis (and group sizes were too small to look for interaction effects) means that we can only postulate that mismatch is not responsible for the results.

4.7.4 Study 1 Discussion

Results of study 1 revealed that participants presented with only immoral options reported significantly lower identification on all subscales compared to the moral choice or polar conditions. Results were robust even when combining all identification subscales into one measure. Further, transportation—the metric that describes engagement from narratives—was also lower in the immoral condition. Finally, participants reported a lower subjective effort in the immoral condition. Also, we found that effort mediated the effect of choice on transportation, but not on identification.

Surprisingly, no differences in perceived autonomy were observed. We can explain this in that participants always decided between two options, and that in the specific case of our experiment design, experienced autonomy was not significantly affected by whether those choices were similar and moral, similar and immoral, or polar in their morality.

Unsurprisingly, there were no differences in overall enjoyment. We did not expect differences, as previous work showed that differences in identification and transportation do not translate into differences in enjoyment within narrative prototypes [43]. Although players may differently identify with the protagonist, and may experience greater transportation into the narrative, the enjoyment was measured as regarded the experience overall. It is not expected that differences in narrative transportation would result in differently-rated enjoyment of the experience in short-term engagement with a narrative prototype.

However, there are questions raised by our results. Specifically, participants were randomly assigned to conditions, thus the similarly-valent options (i.e., only moral or immoral options) were not matched to the player’s moral-compass. One could argue that the decision between two equally valenced options was not difficult because we did not control for the individual’s preferred decision-pattern.

4.8 Study 2

In Study 1 we showed that when presented with only immoral options, participants tended to identify less with their character and were less transported by the narrative than if given only moral options or one of each. Because participants were randomly assigned to the polar, moral, and immoral conditions, the presented choices may not have been meaningful if they did not align with the character’s personality, as played by the player. Our additional analyses showed the preferred morality as predicted by the initial 10 polar choices was not in line with the randomly-assigned condition; that is, approximately half of the participants in the moral and immoral conditions likely had a preferred morality that was incongruent with their assigned condition.

Our goal of Study 2 was to determine if polar or similar choices yielded a better experience, in a situation in which the similar choices were tailored to the participants’ preferred playstyle as predicted by their morality decisions.

4.8.1 Methods

We replicated the experimental design and analyses of Study 1, but instead of assigning participants to the moral and immoral options, we chose the moral or immoral path based on their initial 10 choices. As such, we randomly assigned participants to either the polar (presented polar options) or tailored (two moral options or two immoral options depending on their initial 10 choices) conditions.

To increase statistical power, we raised the number of participants in each condition from 30 to 50, with a total of 100 participants. We collected the same measures as in Study 1 and conducted the same data analyses. From the initial 100 participants, we removed 23 for non-compliance as described in Study 1, leaving 39 in the polar condition and 38 in the tailored condition. The reason there were significantly more participants removed in Study 2 compared to Study 1 was primarily due to an increase in bots on the platform the day we ran Study 2. As described in the earlier section, participants were asked questions relating to the content of the story, and participants that did not demonstrate a basic understanding of the questions being asked (e.g., when asked to define the plot of the narrative, some participants defined the terms *plot* or *narrative*) were removed. In total, we detected 8 participants who were likely bots using this method. The only difference in Study 2 is that we present the combined avatar identification measure for brevity, as the sub-scales all showed the same patterns [258].

4.8.2 Results

The MANCOVA (with transportability, sense of positive agency, sense of negative agency, and moral foundations included as co-variates) showed main effects of choice on identification ($F_{1,71} = 5.7, p = .019, \eta_p^2 = .08$), transportation ($F_{1,71} = 5.3, p = .024, \eta_p^2 = .07$), and effort ($F_{1,71} = 4.9, p = .030, \eta_p^2 = .06$). As Table 4.1 shows, polar choices yielded greater identification, transportation, and effort than tailored ones.

There were no effects of choice on autonomy ($F_{1,71} = .004, p = .953$) or enjoyment ($F_{1,71} = .04, p = .839$). We again conducted mediation analyses. We first modelled whether the prediction of choice-condition on identification was mediated by effort, and second whether the prediction of choice-condition on transportation was mediated by effort. Choice predicted identification ($\beta = .600, p = .019$) and remained a significant predictor, even when effort was included in the model ($\beta = .546, p = .039$). Further, the confidence interval for the indirect effect included zero (LLCI=-.401, ULCI=.086), suggesting that there was no mediation [38]. In terms of transportation, choice predicted transportation ($\beta = .335, p = .024$) but did not remain a significant predictor when effort was included in the model ($\beta = .191, p = .164$). Further, the confidence interval for the indirect effect did not include zero (LLCI=-.367, ULCI=-.014), suggesting that there was a partial mediation [38].

Based on the negative effects of the immoral condition in Study 1, we considered that the participants who were tailored to the immoral condition may have been driving down transportation. In the tailored condition, there were 21 participants assigned to the immoral condition and 17 assigned to the moral condition (based on their initial 10 decisions). Looking at the mean values of transportation though, participants experienced the greatest transportation in the polar condition (4.88), followed by the immoral tailored (4.66), and then moral tailored (4.40) conditions, suggesting

that it was not the immoral decisions harming experience in the tailored choice conditions.

4.8.3 Study 2 Discussion

Participants presented with polar decisions reported higher ratings of identification, transportation and effort. There was no significant effect of choice-condition on perceived autonomy or enjoyment. Further, mediation analysis showed that, in line with Study 1, effort mediated the influence of choice-condition on transportation, but did not predict identification or mediate that relationship—i.e., even when controlling for effort, the prediction of choice-condition on identification remained significant.

Interestingly, the tailored conditions again did not affect the perceived autonomy, which could be explained in that participants experienced the reduced effort instead of a lack of real choice. In study 3, we explore the effect of an option providing the best of both worlds: a meaningful choice between two options of the same morality, while also giving the option to choose the polar-opposite.

4.9 Study 3

Study 1 showed that polar and moral choices yielded higher transportation, identification, and invested effort than immoral ones. Study 2 showed that similar, but tailored, choices yielded lower transportation and identification than polar ones. Based on the results from the first two studies together, we hypothesized that the reason that we saw reduced experiences with tailored choices, and why we did not see differences in perceived autonomy, was that by tailoring choices to align with the participants' model, we took away the option that went against their model. As discussed in Study 2, there are reasons to expect that not having the choice that one wants to discard can harm experience.

To investigate whether the absence of the not-desired choice reduced identification and transportation, we adapted the tailored condition to present participants with three choices: two that aligned with their model and one that opposed. In this way, we maintain the tailored options and provided players with a meaningful choice, while also providing them with the option to deviate from the model, and not make the non-desired choice.

4.9.1 Methods

We replicated the experimental design and analyses of Study 2, presenting a tailored (two moral-one immoral or two immoral options-one moral depending on their initial 10 choices) condition. We gathered 50 participants. We collected the same measures as in Studies 1 and 2 and conducted the same data analyses. We compare the results to the polar condition from Study 2 as participants

were randomly recruited and Study 3 was gathered the day after Study 2. We removed 7 for non-compliance as described in Study 1, leaving 43 in the tailored 2.0 condition.

4.9.2 Results

The MANCOVA (with transportability, sense of positive agency, sense of negative agency, and moral foundations included as co-variates) showed no main effects of choice on identification ($F_{1,76} = .47, p = .494$), transportation ($F_{1,76} = .74, p = .392$), effort ($F_{1,76} < .001, p = .994$), autonomy ($F_{1,76} = .1.7, p = .194$), or enjoyment ($F_{1,76} = .026, p = .873$).

Table 4.1: Estimated Means \pm SE for Studies 1,2, & 3. Values of co-variates for each study are shown in bottom right.

	Study 1			Study 2		Study 3	
	Polar	Positive	Negative	Polar	Tailored	Polar	Tailored2.0
Transportation	4.89(.10)	4.87(.10)	4.58(.10)	4.88(.10)	4.54(.10)	4.80(.11)	4.67(.10)
Identification	4.68(.22)	4.73(.22)	3.69(.22)	4.45(.17)	3.85(.18)	4.34(.18)	4.18(.17)
Autonomy	5.37(.25)	5.46(.24)	5.27(.25)	5.13(.21)	5.14(.21)	5.03(.20)	5.39(.19)
Effort	5.83(.15)	5.86(.15)	5.41(.16)	5.83(.14)	5.37(.14)	5.74(.15)	5.73(.14)
Enjoyment	5.59(.20)	5.72(.19)	5.27(.20)	5.44(.17)	5.39(.17)	5.36(.20)	5.32(.19)
Similarity ID	4.68(.27)	4.56(.22)	3.62(.29)	S1. Trans=5.25, SOPA=5.43, SONA=2.50, MFQ=4.07			
Embodied Presence	5.17(.22)	5.31(.22)	4.26(.23)	S2. Trans=5.21, SOPA=5.49, SONA=2.33, MFQ=3.82			
Wishful ID	4.20(.26)	4.50(.26)	3.16(.27)	S3. Trans=5.12, SOPA=5.35, SONA=2.36, MFQ=3.80			

4.9.3 Study 3 Discussion

In Study 3, we adapted our tailored condition from study 2 into a new tailored 2.0 version to provide two similar options and one opposing option, to provide players with a meaningful model while still giving the option to customize the character’s personality. Results show that there was no significant difference between the conditions on any measure (i.e. identification, transportation, effort, autonomy or enjoyment). Thus, the tailored 2.0 decision condition in Study 3 significantly improved on the tailored condition of Study 2.

One limitation of study 3 lies in the fact, that in the tailored 2.0 condition has a confound. We cannot disentangle, in the current study design, whether the higher identification and effort were due to an increased number of options (i.e., from two to three) or the addition of the opposing choice. But participants again reported no difference in perceived autonomy, which might suggest that adding a third option did not significantly change their perception of their decision. Therefore, we propose that the difference found in Study 2 was due to the lack of alternatives and not the addition of one more option to pick from.

4.10 General Discussion

In the following sections, we summarize our results and discuss in detail their limitations and implications.

4.10.1 Summary of Results

Study 1

In Study 1, we showed that participants identified with their character less when only given the option to act immorally, compared to given only moral options or given a choice from both ends of the spectrum. Furthermore, transportation and effort were both also lower in the immoral condition. Invested effort mediated the effect of choice on transportation, but not on identification.

Study 2

In Study 2, we showed that even by tailoring the choices to align with the morality that participants established with their character in the initial 10 decisions, they still experienced lower identification, effort, and transportation than when given the options. Also, participants perceived effort mediated the influence of choice on transportation, but effort did not significantly mediate the prediction of identification through choice.

Study 3

In Study 3, we updated the tailored option to also include the opposing (polar choice) and saw the effects from the first 2 studies go away, indicating that our updated tailored model was no worse than two polar options. Also, because we found no differences in Study 3, we have some confidence that the reason our tailored model produced lower identification in Study 2 was due to the absence of an alternative option that participants did not wish to choose.

4.10.2 Explanation of Results

Over three studies, we consistently show that the inclusion of the opposite choice is important and shows two robust effects. First of all, identification with the character increases. Secondly, effort increases and effort significantly mediated the corresponding increase in transportation.

With regard to identification, the decreased identification after being presented with only immoral options can be explained in the context of a self-serving bias [120] and a self-enhancement bias [228]. Both biases serve an individual for the purpose of maintaining self-esteem [35, 134]. In situations of self-threat [18]—a situation in which positive views about oneself are contradicted or otherwise threatened—preserving one’s self-esteem is especially important [53]. Transferred to

the context of games, participants may have rated their identification lower with a character that could only make immoral choices, because their view of themselves is generally positive; whereas identification was not impeded in either the polar, moral, or tailored 2.0 decision conditions, because they all provided a moral option to choose from. Following a similar line of reasoning, higher identification ratings in the polar and tailored 2.0 condition (and lower ratings in the tailored condition) could have resulted from a drop in self-efficacy [14, 17]. Self-efficacy refers to the general personal belief in one’s abilities to exert control over the environment [142]. Generally, a positive view of one’s self-efficacy is highly adaptive and crucial for well-being and psycho-social functioning [16, 218, 230]. Self-efficacy and the perception of control are linked to the perception of being able to choose freely [15, 161], which was only the case in the polar and tailored 2.0 conditions. When being presented with tailored options of the same morality, essentially only one choice is displayed, and this could lead to the perception that it is not possible to make the right choice given any situation, which in turn lowers self-efficacy.

Moreover, the significantly higher ratings of transportation in the polar and the tailored 2.0 conditions can be explained by its similarities to real-life decision making and to familiar game design. Realism has been reported as linked to increased transportation [27]. Further, transportation has been shown to be influenced by characteristics of the narrative that enhance suspense [245] and the perception of personal relevance [208]. Additionally, we found and replicated a significant mediation by perceived effort of choice on transportation but not on identification. This means that effort drives transportation but not identification. Or put differently, an increased effort leads to higher transportation—a measure for narrative enjoyment.

4.10.3 Reliability of the Prediction Model for Tailored Choices

We see that in the POLAR conditions for Study 1 and Study 2, the model generated by the system fairly reliably predicts the final morality model. For Study 1, the correlation between the actual and the predicted model was $r = .767, p < .01, N = 26$. For study 2, it was $r = .794, p < .01, N = 40$. Thus we assume that the model actually tailored the choices correctly based on the players’ decision used for calibration.

4.10.4 Autonomy vs. Agency & Trait vs. State

Autonomy and agency are terms that are often used interchangeably, but we argue in this section that autonomy and agency are two distinct constructs.

On one hand, autonomy as a state in games is usually measured using questions like “I experience a lot of freedom in the story.” or “The story provides me with interesting options and choices.” [219]. On the other hand, agency as a trait can be measured by high-level questions about the participant’s life or the story in general with questions like “I am in full control of what I do.”, “The things I do

are subject only to my free will” and “I am responsible for what I do.” [250].

Although both questionnaires contain choices, autonomy refers more to a perception of having options, while agency reflects the perception of having not only a choice but also that one has control over the decision and that choice has a consequence. We argue that our manipulation did not have an effect on autonomy because players were always free to make a choice, but their sense of agency may have been manipulated and an appropriate measurement of their state agency could have resulted in additional proof for our claims.

This might help explain why in our state measure of autonomy, we did not find an effect but an appropriate state measure of agency might have been affected by the different decision options. There are, to the best of our knowledge, currently no validated scales available that measure agency as a state, which might have been a crucial outcome variable to explain the experience of choices in the present study.

4.10.5 Game Design Considerations

Our results suggest that including a variety of choices (e.g., a mix of both moral and immoral) is important to a player’s identification with their character. Additionally, we see that having a variety of choices also led to higher transportation, but this was mediated by perceived effort. This implies several important considerations for game designers.

First, our results suggest that when presenting options to players, giving them varied options should lead to higher identification with the character (e.g., giving the option for bright pink hair in an avatar creator instead of limiting the choices to shades of brown).

Second, our results suggest that a variety of choices is helpful in fostering transportation, partially because they lead to a higher investment of effort in making the decision. This suggests that while a variety of customization options are important for fostering identification, the important design aspect for transportation is the perceived effort of the choice. This means that game designers should create decisions in such a way as to force the player to expend cognitive effort, rather than giving choices that the player will easily make without any consideration.

Finally, it is important to caution designers against giving players too many choices. As described earlier, previous work has shown that too many choices can lead to feelings of dissatisfaction with a choice [126,172,225,251]. Also, the addition of a third choice in Study 3 did not significantly change any of our measures compared to the polar condition, which suggests that it may not be the number of choices that are important, but instead the variety.

4.10.6 Limitations

There are four limitations to our study that have to be addressed. Firstly, we did not measure agency as a state. Although surprising, to our knowledge there are no validated measures available

to measure agency as a state, and given the background of the current literature, we opted to measure autonomy instead. Reformulating the trait questions or a similar questionnaire (e.g., self-care agency scale [143]) would be an option in future studies to explore effects on state agency as a result of in-game decisions.

Secondly, perceived effort showed a significant effect in Studies 1 and 2 but the construct is fairly non-specific and rather broad. Questions in the IMI effort subscale [216] include ones like “I put a lot of effort into this game.” and “It was important to me to do well at this game.” If we had measured cognitive workload using for example the NASA-TLX [115], we may have been able to unpack more specific effects. With that being said, the self-reported effort of participants did significantly mediate the relationship between choice condition and transportation in both Studies 1 and 2.

Thirdly, though previous work has shown that a low-fidelity narrative prototype can reasonably predict the narrative and play experience of a high-fidelity game using the same story [43], the prototype we used in the studies presented was only used by participants for a short amount of time and the conditions under which they played (i.e., under a controlled study as opposed to for enjoyment under their own volition) were different than that of a commercial game. As such, we can be reasonably confident that our results could be extended to a higher-fidelity interactive narrative, but the mechanisms of choices related to things besides dialogue choices (e.g., choices involving upgrades, choosing which quest to undertake, and customizing the appearance of the main avatar) are beyond the scope of this paper. We would expect similar results from studying other types of choice, but further research is needed to verify this.

Lastly, it is unclear from the design of Study 3 whether the results can be attributed to the presence of the alternative and opposing option or the fact that there was one additional option presented (i.e., 3 instead of 2 in the control condition). We considered this issue when designing the study; however, including 3 options in the POLAR condition would be difficult as there is no natural valence for the 3rd option, and it would not be comparable to the previous POLAR condition used. Given the results of the study-series together, and the non-significant difference in state autonomy when provided with the additional (i.e., 3rd) choice, we propose that only providing one additional option did not affect the player experience. Rather over the course of three studies, we reliably showed that having the polar choice was the driving force behind the effect.

4.10.7 Future Research Directions

The findings in the present study open new lines of potential research. Most importantly, the present study highlights the need for a state agency questionnaire that can be used in HCI and games research. Such a scale could be used to provide important information about player experience and how much players feel they are in control over their environment.

Further, several experimental manipulations should be considered by researchers to further explore the effects of different types of choices. One could argue that it would significantly alter the player experience if it was transparent that their choices are being modelled and predicted throughout the gameplay. Also moral dilemmas instead of polarized choices, could be used in order to provide more difficult choices. But the research of moral choices is not without its difficulties and there are several inter-individual differences that influence moral choice [60,139].

Finally, based on the present paper, three additional game design elements could be investigated. Firstly, researchers could evaluate the effects of a branching dialogue system on the sense of agency in the moment. A dialogue system like this might provide a stronger sense of agency by having natural consequences to individual choices participants have made. Participants could even be required to re-play the same game in order to notice the branching dialogue tree. Secondly, a colour-feedback system could be implemented to help offload cognitive load of the player in making decisions, similar to commercial games that employ polar choices. Previous work has shown that altering the colour of text can significantly alter player experience [39]. It is possible that highlighting moral and immoral choices could function similarly, providing players with visual feedback to confirm and strengthen the perceived morality of a choice. Thirdly, researchers might want to look into pacing by only giving players choices that align with their individual model, while occasionally providing players with opposing choices to manage their narrative experience curve. This would also require some form of repeated online measurement of the players' experience to in order to model their ideal moral curve for narrative experience.

4.11 Conclusion

Choice is inherently important to individuals and valued highly. Through choice, individuals can exert control over their environment and become autonomous agents. In games, especially in RPGs, the player is often confronted with plenty of decisions, through which they can shape their character and experience. In a series of three studies, we tested the effect of polar choices against various other choice-configurations (i.e., moral/ immoral in study 1, tailored in study 2 and tailored 2.0 in study 3). Our results across studies consistently showed that including the opposing choice is important to foster a player's identification with their character and their transportation into the narrative. These results still hold true even when a player's decision behaviour can be reliably predicted.

5 MANUSCRIPT D: PREDICTING BELIEFS FROM NPC DIALOGUES

5.1 Introduction to Manuscript D

In Manuscript A we showed that an alignment between the player’s real-world beliefs and those exhibited by the main playable character are beneficial for overall player experience. In practice, the results from the study are useful because they provide evidence to motivate game designers to tailor game content such that the main playable character’s beliefs align with those of the player in the real world.

In the study, however, we used external questionnaires to measure the player’s real-world beliefs. While this method was accurate and provided a useful model, the method lacks external validity, since most players would not fill out pages of questionnaires prior to playing the game in order to gather the data necessary to tailor the game’s content.

In the previous chapter, we proposed a method of implicitly measuring a player’s real-world beliefs through dialogue choices. We showed that we were able to create a consistent model of the player’s choices, which in turn predicted the moral direction they would take the main character through the narrative. We did not, however, see a clear alignment between the player’s real-world model of morality and that exhibited by the main player. We discussed potential reasons for this discrepancy in the previous chapter. In this chapter, we attempt to address this limitation by changing the underlying model from morality, which has a clear right and wrong answer, to benevolent sexism [106] and male role attitudes [62]. As we saw in the first study, benevolent sexism is a much more subtle construct in terms of morality. The beliefs are distributed evenly across genders and it is a construct that is perceived as moral at the individual level regardless of the individual alignment.

5.1.1 Publication and Individual Contribution

This work was published as the following: [42]. Although a multi-authored contribution, I was responsible for leading the research. My roles included: designing the experiments, providing the interactive dialogue system (reused from Manuscript C) and conducting a majority of the data analysis. Additionally, I was involved in the literature review, the creation of the dialogue, and

writing of the publication. I contributed approximately 70% of the total work for this project and publication.

5.2 Abstract

Game designers and developers benefit from gathering data from players; however, interrupting play with questionnaires can harm experience. Previous work has suggested that embedding questionnaires into games, such as through dialogue choices when interacting with non-player characters (NPCs) can help, but there is no evidence that dialogue choices can model the real-world beliefs of players. In this study we demonstrate two methods of successfully predicting responses to validated scales of sexist beliefs from NPC dialogues that do not differ in their resulting narrative engagement. Our findings open opportunities for better tailoring games and game experiences by modeling players through their in-game interactions.

5.3 Introduction

Data is important for game development and research, and self-report measures like questionnaires are essential tools for playtests, user studies, and adaptations in commercial games [91,93,179,180,184]. Previous work has suggested that player responses in dialogues with non-player characters (NPCs) might be used for data collection to avoid breaking the immersion or flow of players [91,93]. This approach is already being used in some commercial games, such as *Until Dawn* [93].

While earlier work has suggested the validity of NPC dialogues for player state measures (e.g., enjoyment, frustration), we do not yet know if they can be used to assess more stable constructs, such as the players' traits or beliefs. While there is some indication that players' attitudes (e.g., towards challenge) can be measured through dialogue responses [93], there might be barriers to measuring beliefs. Players may consider their beliefs in light of social desirability [157] and would respond in such a way that their dialogue responses are not predictive of their actual beliefs. Social desirability is particularly problematic when the player's beliefs are unpopular (e.g., racist or sexist beliefs) or are incongruent with the gaming context (e.g., pacifist beliefs when enjoying violent games). Further, when players enter the magic circle [122], they may answer dialogues as their character would respond, rather than as themselves. However, previous work has shown that the alignment between an individual's beliefs and the beliefs displayed by a game's protagonist is important for fostering player identification with the protagonist [40], suggesting that players may align character responses to their own as an unconscious means of fostering identification and enjoyment. As such, we require further research to evaluate if dialogue responses can predict the players' beliefs. Further, there are various ways in which such dialogues could be designed, including the framing of the dialogue writing, which could affect the players' behaviour or experience.

In this paper, we propose embedding the items of validated questionnaires that measure beliefs (usually through multiple items) into dialogue choices, and using the choices made by players to construct models that predict their beliefs. We further demonstrate two methods for converting validated questionnaires into interactive dialogue: Direct—prompting the player with the actual scale item and giving them choices to agree or disagree, and Indirect—prompting the player with a related statement and giving them a choices between the scale item or a statement opposite the scale item. We chose to model sexist beliefs, as an example of a held belief that may be challenging to assess within game dialogues due to social desirability and the context of gaming. Then, we conducted a user study in which 151 participants, who had previously completed the validated scales assessing their sexist beliefs, played through an interactive narrative using one of the two approaches (direct or indirect). We used the players’ dialogue choices to create a model for predicting their sexist beliefs and evaluated the narrative experience of both conditions.

Our results indicate that we could use dialogue choices to predict the players’ sexist beliefs, suggesting that player beliefs and attitudes are reflected in their dialogue choices. Further, the indirect and direct approaches were both useful for predicting beliefs and did not differ significantly in terms of the resulting narrative experience. Our findings shed further light on the validity of using in-game dialogues for assessing data about players, which has implications for data collection in game analytics, for games user researchers, and for adapting game features that potentially improve player experience by aligning game characters’ beliefs to the players’ beliefs.

5.4 Related Work

This research builds on earlier work on in-game assessments, dialogue choices, and gender and sexism in games.

5.4.1 In-game Assessments: Dialogue and Choices in Games

Earlier work has suggested that assessment can be embedded in dialogues with non-player characters [91]. By presenting dialogue options that are mapped to questionnaire items, it is possible to include assessment in an unobtrusive way [91], while problematic reducing interruptions in game-play [94]. While studies have not consistently showed that such an approach is beneficial in terms of player experience [93], it can be beneficial for data collection by increasing data quantity through number of answered prompts. Previous studies mostly examined the suitability of in-game assessments for measuring *state* variables, such as emotions [91, 94]. For example, it can be beneficial for playtesting to measure how a player experienced a game prototype, i.e., how a game affects the players’ *states*. For states, studies have shown that players’ post-game responses largely coincided with the states that they reported in dialogue choices [91, 93]. However, previous work has not

considered *traits* to a similar degree, i.e., whether in-game assessments can validly measure more constructs that are more stable than dynamically changing states. Such traits can be important for considering how players experience games, e.g., with the players' challenge orientation [254] as a trait that can help explain the players' enjoyment of in-game successes and failures [92]. Thus, it can be useful to measure traits through in-game assessments and there is some initial evidence from an earlier study, in which single item trait measures showed promise for valid measurements [93]. However, it is unclear if dialogues can be used to model more complex player traits, i.e., those that are generally assessed through multi-item scales. Further, we require further research that investigates *how* to design such dialogues, including different ways of framing dialogue prompts and choices.

5.4.2 Player Choice in Dialogues and Games

Using NPC dialogues for assessment assumes that players respond to dialogue options in a way that mirrors their values, i.e., that they answer in a way that reflects their traits. This assumption is in line with the *Media Equation* theory [211], a theory that suggests that humans interact with virtual entities similar to how they interact with other humans [135, 186].

However, choices in games are different in non-game contexts. Games enable players to make choices and control their experience. For players, games need to have tangible consequences to be perceived as meaningful [103, 261]. When players make a choice, they create a commitment on behalf of the player character [249], and for this commitment to make sense and feel like it matters, the game should hold the player accountable. Further, tailoring choices based on models of previous choices can be beneficial for player experience [171, 214, 252]. Dialogue choices that align with a player's playstyle have been shown to improve experience [41]. An easy way to make the player feel accountable is through moral choices. Choices in games are often presented as a moral conundrum between good or evil. Often a lot of work goes into making robust full experiences for players who choose to be good, bad, or in between. However, players do not usually make the morally bad choices, due to the significant contrasts in the choices.

In the Mass Effect series [29], players choose between binary choices of good or bad (paragon or renegade) each with content unique to those choices. Even with this choice system, 92% of players chose the paragon option when presented with a choice [200]. This is often due to certain choices being presented as the *right* choice both morally and within the gameplay itself, such as choosing to convince a squad mate to drop their gun or shooting them. This also likely lined up with their own real world beliefs [200]. Most people would not murder someone, especially if they had the option to talk themselves and the other person out of the situation. A moral choice becomes less of a significant choice, if it is presented with a solution that is regarded as the right choice by the game, even if it aligns with the players' personal beliefs. Offering a range of choices on a wide moral

spectrum, while enhancing players perceived freedom, may come at the cost of meaningful choices. If players already know the choices that they are going to make before a given scenario, the choice could be seen as meaningless.

Research suggests that dialogue choices in games can affect experience. Thus, the specific framing of dialogue options that can be used for assessment of traits might be limited due to considerations of player choice. If there is an obviously correct answer to a dialogue prompt, players might not respond with an answer that is reflective of their own traits. This is not an exclusive problem of games but also similar to social desirability biases that are a methodological concerns in studies and surveys [157]. However, problems might be exacerbated in a gaming context, in which choices have consequences for further play. Thus, we do not yet know how different dialogue options should be framed, e.g., in their specific wording, to allow the assessment of player traits without compromising the importance of player choice in games. Our study addresses this by investigating different styles of dialogue framing and their effects on response behaviour and player experience.

5.4.3 Gender and Sexism in Games

In this paper, we use sexist beliefs as a trait to assess through dialogue choice modeling. The representation of men in video games is more prevalent than women [78, 127, 271]. When women are portrayed in games, it is often in a sexualized way, with revealing clothing, unrealistic body images, or partial nudity [48, 78, 127]. This has led to strong criticism of the video game industry for catering to straight male players [88, 109, 224]. Research on sexualized imagery indicates men show a greater likelihood to harass women after play, usually with sexist remarks [75, 274]. Research on the female experience with sexualized imagery suggests they feel diminished self-efficacy [22] and increased self objectification after play [88, 90]. These effects have led to the conclusion that women may self select out of these environments [88] and reinforce the idea of games being dominated by masculine disclosure [220].

Sexism is typically conceptualized as hostility towards women, i.e. as hostile sexism [105]. However, there is also *benevolent sexism*, which are interrelated attitudes towards women that are stereotypical and restricted but are subjectively positive in tone or behaviours, e.g., suggesting to help a woman move a heavy object because she is seen as not strong enough to do it herself. Despite being well-intentioned, such attitudes and behaviour can reinforce traditional stereotypes and masculine dominance. Such views are thought to be reflected in a form of social ideology, in which there are protective views toward women, a reverence for roles as mothers and wives, and idealization as romantic love objects [105]. While these views of women are positive, they share common assumptions with hostile sexist beliefs, that women are restricted to domestic roles and are weaker. Hostile sexism characterizes women as unfit for certain roles while benevolent sexism rationalizes the confining to these roles [105].

In addition to players' beliefs about women, players might also differ in their beliefs about male roles. It is often assumed that attitudes of males roles are distinct from female roles, but are associated with attitudes towards gender roles and relationships. This means that an individual can have a progressive belief about female roles while adhering to a traditional view of male roles. Earlier research proposed scales to assess beliefs on men adhering to culturally defined standards of masculinity [206], suggesting that there is interest in assessing such traits. Beliefs about gender and sexism thus are important considerations for differentiating between players and there is evidence that those beliefs can be measured. Thus, these traits are highly relevant for study in the context of games, and therefore lend themselves for our study because they have been shown to be important for identification.

5.4.4 Gender and Sexism in Avatar Identification

Avatar identification facilitates games enjoyment, and it has been suggested that sexist characters can aid in avatar identification [119, 130]. Bowey et al. [40] found that the gender of the players does not predict avatar identification but their *beliefs* about gender roles and benevolent sexism do. They also found that visually sexist character stereotypes did not harm identification. However, sexist dialogue harmed the gameplay experience for players who did not hold sexist beliefs and did not improve the experience for those who did. This suggests that dialogue has a large impact on how players identify with avatars and how much they enjoy the game they are playing. Thus, beliefs about sexism are important considerations for understanding how players experience games, suggesting that they are ideal for this study because their assessment should not compromise moral choices but remain valid, e.g., circumventing problems of social desirability.

5.4.5 Summary

Earlier research suggests that assessments can be embedded in games and NPC dialogues. Yet, we require further research to evaluate if such an approach is suitable to measure player traits. We do not yet know how to frame dialogue options in a way that traits can be measured validly while still providing dialogue choices that are meaningful, e.g., through moral ambiguity. For this study, we aim to assess sexist beliefs in games because of their apparent importance in games and for the behaviour of players in interactions with avatars.

5.5 Methods: Dialogue Creation and User Study

We implemented two different approaches for dialogues that frame the dialogue somewhat differently. Then, we conducted a user study to evaluate those approaches and our main research

question, i.e., whether we can predict real-world sexist beliefs through player responses in dialogues.

5.5.1 Game/Narrative

We presented dialogue to our players using the same system used in a previous publication [41]. Though this system was constructed to tailor dialogue choices in real-time based on a dynamic player model, we configured the system to present two opposing choices for each dialogue choice. We also used the dynamic player model to keep track of the player’s choices and construct a final model at the end of the game.

The story we presented to players was framed as a prologue to the story from [40]. Players took on the role of the Knight prior to leaving on their adventure to rescue the kidnapped Princess. In this story, the player engages in a short conversation with the King, interviewing for the job of rescuing the Princess. The King asks the player many questions, all derived from existing validated questionnaires (see next section). Through the dialogue, the player has the opportunity to express their personally-held beliefs by either agreeing or disagreeing with the things the King is saying.

5.5.2 Dialogue Creation

We approached the problem of creating dialogue representative of the questionnaires from two perspectives, which we call *Direct* and *Indirect*. In the *Direct* approach, we present players with a statement similar to the actual items in the questionnaires, altered only to ground the statement in context to fit with the overarching story (e.g., “You seem well built. A young man should be physically tough, even if he’s not big.”). The player is presented with two possible responses: one statement that agrees with the prompt (e.g., “That’s my belief”), and one that disagrees with it (e.g., “You’re wrong”). See Table 5.1 for the complete direct dialogue derived from the Male Role Attitudes Scale [206]. In the *Indirect* approach we present the player with a statement related to the questionnaire item (e.g., “So many of my knights neglect their physical fitness.”). One response is the questionnaire item itself (e.g., “A young man should be physically tough, even if he’s not big”) and the alternative is a statement opposing the questionnaire item (e.g., “Even if a man is big, he does not need to be tough.”). See Table 5.2 for the complete indirect dialogue derived from the Male Role Attitudes Scale [206].

5.5.3 Prestudy

One week prior to gathering our data we posted a pre-screen study on Amazon’s Mechanical Turk (MTurk) to build our participant pool ($N = 300$). Participants filled out a basic demographics questionnaire, the Ambivalent Sexism Inventory (ASI, [105]), and the Male Role Attitudes Scale

(MRAS, [206]). The ASI measures an individual’s personal beliefs relating to benevolent sexism. The scale measures three subscales: protective paternalism (PP, the idea that women should be cherished and protected by men), complementary gender differentiation (CGD, the idea that women have moral superiority to men), and heterosexual intimacy (HI, the idea that every man needs a romantic relationship with a woman to be complete). In addition to these subscales, all of the items were also combined into a single measure of benevolent sexism. MRAS measures beliefs relating to masculine ideology and has only one scale within it. We gathered this data prior to the main experiment to prevent possible sequence effects participants may have had if they filled out both the questionnaires and interacted with the dialogue in the same session.

Out of the 300 participants, we removed participants based on the following criteria. First, we removed any participant who scored greater than three standard deviations from the mean on any of the calculated subscale constructs (i.e., male role attitudes, protective paternalism, complementary gender differentiation, and heterosexual intimacy, $N = 0$). Second, for each questionnaire we calculated the number of seconds per item (total time spent on the questionnaire divided by the number of individual items) and removed any participant who spent less than 1.5 seconds per item ($N = 45$). These criteria have been effectively used in previous studies with data from MTurk [40] to remove participants who simply clicked through the questions without reading them. Finally, at the time of running the main study, one participant from the pre-study was no longer a worker on MTurk so we were unable to invite them to participate in the main study. In total, we were left with 253 potential participants.

5.5.4 Main Study Design

Our main study followed a between-subjects experimental design. All participants played through one of two versions of the interactive dialogue (*direct* or *indirect*, as described above), and were assigned to their conditions at random.

Prior to playing through the game, participants completed the transportability questionnaire [70], which measures their susceptibility to be transported into a narrative world.

After playing through the game, participants completed the homophily questionnaire [176], which measures the perceived similarity in values to the main character, and the narrative engagement questionnaire [49], which divides narrative experience into narrative understanding, attentional focus, narrative presence, and emotional engagement.

Finally, participants were presented with a prompt to: *“Please indicate your approach to making decisions in the interactive narrative game by completing the following statement: I made decisions as:”*. Participants were given a slider that ranged from *“myself”* on the left to *“an external character”* on the right. We used this measure to know which participants made decisions that reflected their real-world beliefs, to more accurately compare the dialogue models to the real-world ques-

tionnaires.

5.5.5 Participants

One week after conducting the prestudy, we invited participants from our participant pool to complete the main study. We left the study open for five days to give participants a chance to complete the study. Out of the 253 participants invited, we gathered data from 157 (62% response rate). Of participants who responded, we removed a further 6 participants who were quicker than 1.2 seconds per item on either the transportability or narrative engagement questionnaire.

The final sample consisted of 151 participants (70 women, 79 men, 1 non-binary, and 1 undisclosed), with a mean age of 39 (SD=10.8). An overwhelming majority of players stated that they played games either every day (69, 45.7%) or a few times per week (68, 45%).

An ANOVA on transportability by condition was not significant ($F_{1,149} = 1.529, p = .218$), adding confidence that participants were randomly assigned to groups and a between-subjects comparison is justified. The means and standard deviations for all measures can be seen in Table 5.3.

5.6 Results

We analyze the data to answer our main research questions.

5.6.1 Did they play as themselves or an external character?

The first question we set out to answer was whether participants were making decisions on behalf of themselves, or playing the role of a separate protagonist. This question is important to answer first, because if the majority of participants were not making decisions on their own behalf then our initial premise of using in-game dialogue to predict a player’s real world beliefs would not be viable. There is evidence in previous work that players making in-game decisions tend to make decisions in alignment with their own personal beliefs [267], but we wanted to confirm this with our data.

The player’s perspective was measured using a slider, as discussed in the previous section. The slider had a total of 100 points, with 0 being all the way to the left (i.e., making decisions as themselves) and 100 being all the way to the right (i.e., making decisions as an externally). The mean average of this slider was 31.91, indicating that participants, on average, made decisions with their own beliefs in mind.

Additionally, we took this number and coded each participant as either *internal* (i.e., the slider was less or equal to 50) or as *external* (i.e, the slider was greater than 50).

Out of the 151 participants, 41 were *external* (27.15%) and 110 were *internal* (72.85%). We see that the overwhelming majority of participants made decisions on their own behalf as opposed to role-playing as a different character.

5.6.2 Do the dialogue models predict players' real-world beliefs?

Our next step was to test if the dialogue-generated models are useful for assessing the players' beliefs and whether there is a difference in the approach (direct vs indirect). For this, we calculated hierarchical moderated regressions predicting validated scale scores using the dialogue-generated model scores to test for the main effect at the first level. At the second level, we added dummy-coded condition and the interaction between condition and dialogue-model score as predictors to determine if beliefs differed between the groups and if the quality of prediction depended on condition. We correct for multiple tests by adjusting the significance threshold to $p < .001$ (divided by the number of tests).

The results (see Table 5.4) show that the dialogue-model scores were strong and significant predictors for the validated scale scores, which indicates that the modeling approach works for predicting the players' real-world beliefs.

Considering the condition and interaction, effects were not significant after familywise error correction. The lack of significant effects for condition show that the beliefs did not differ between the groups (direct and indirect), whereas the lack of significant differences in the interaction term suggest that we do not have evidence that either the direct or approach is better at creating predictive models.

5.6.3 Do participants prefer the Direct or Indirect methods for dialogue creation?

In addition to accuracy of prediction, we also looked at which condition resulted in a stronger narrative experience among participants. To test this, we conducted a Multivariate Analysis of Variance (MANOVA) using condition as a fixed factor, and the experience constructs as dependent variables, which included: narrative understanding, attentional focus, narrative presence, emotional engagement, and homophily.

The MANOVA revealed no significant effects for any dependent variable: narrative understanding ($F_{1,149} = 1.529, p = .218$), attentional focus ($F_{1,149} = 2.183e - 5, p = .996$), narrative presence ($F_{1,149} = 0.815, p = .368$), emotional engagement ($F_{1,149} = 0.129, p = .720$), and homophily ($F_{1,149} = 0.425, p = .516$). This suggests that neither approach resulted in a better narrative experience, and that neither group felt more similar in values to the main character.

5.7 Discussion

We summarize results, discuss applications and ethical issues, and present limitations and future work.

5.7.1 Summary of Results

Player Point of View

The majority of participants self-reported that they made choices in the interactive dialogue as themselves rather than an imaginary, external protagonist. This implies that when making the choices, they were applying their own beliefs and values, which gives confidence to our methodology that the models we construct based on dialogue choices will align closely with the external questionnaires. This is consistent with earlier research confirming that players tend to have aligned their out-of-game and in-game views, e.g., that identification with game characters represents a shift of self-perception [146,147] and involves a loss of self-awareness and its replacement with “emotional and cognitive connections with a character” [64]. This would suggest that players’ responses are suitable for predicting real-world beliefs.

Relationship between dialogue and real-world beliefs

To validate our proposed methodology, we needed to show that our models created based on interactive dialogue choices closely matched those created through validated, external questionnaires. To do this, we conducted hierarchical moderated regressions and showed that our models constructed from dialogue choices reliably predicted the models created from validated questionnaires. This demonstrates that our approach to using dialogue choices to implicitly construct models of a player’s real-world beliefs is valid and effective. Thus, our results confirm earlier work that dialogue responses can be used to assess information about players [91,93] and extend previous findings into the context of real-world sexist beliefs.

Effectiveness of dialogue creation methods

We looked at the effectiveness of our two approaches: *Direct* and *Indirect*. As the condition and interaction effects in the regressions were not significant, we do not have evidence that either approach resulted in a more accurate model, compared to the validated scale model. This suggests that there are different useful approaches that can be used without compromising the validity of data collection, which is important to inform designers who want to use such an approach but might not know if different dialogue design could affect data quality.

Preference of dialogue creation methods

Even though there were no substantial differences in the accuracy of either approach to creating dialogue, we wanted to check if either approach fostered a stronger experience for the player. For all of our dependent measures we did not find any significant main effects of condition, which means we do not have any evidence that either approach created a more engaging or enjoyable story. Again, this suggests that designers might use either approach, depending on their own preferences for dialogue style, without fear of negatively affecting player experience.

5.7.2 Game Design Applications

We describe three potential use cases for this work: (1) tailoring games to better align with the individual beliefs of players, (2) implicitly measuring players' beliefs for use in game analytics, and (3) providing researchers with an alternative to explicit self-report via questionnaires.

Our modeling approach could be used by game designers to dynamically tailor game content to better align with the beliefs and preferences of individual players. For example, consider games with well-defined protagonists, such as Joel and Ellie from *The Last of Us* [187] or Geralt from *The Witcher* series [210]. One of the features that make these protagonists compelling is their unique beliefs and values. While this makes the character complex and interesting, it also leads to the character taking positions on issues that differ from those of the real-world player. While this is not necessarily a bad thing, we know that similarity between the player and protagonist leads to higher perceived identification with the protagonist, and through that higher overall enjoyment of the game [40]. If game developers were able to subtly measure the player's real-world beliefs, the protagonist's personality and beliefs could be nudged slightly to better align with the player's beliefs, leading to the player perceiving a stronger bond with the protagonist and a higher overall enjoyment of the game.

Game designers could also use our method to gather player data for use in game analytics. In this use case, game designers could use our approach to construct player models through dialogue interaction. These models could then be used to better understand the demographics of their player base and gain richer insights from game analytics. Similarly, the methods described in this paper could be useful for game user researchers. Currently, researchers rely on questionnaires to construct subjective models of their players; using our method, researchers could construct these models through player interaction with the game instead. This would have two benefits. First, participants would spend more time in the game and less time filling out questionnaires, which has the potential to increase participant enjoyment. Second, the participant remains inside the game world while answering questionnaires, which may result in a more accurate reflection on their in-game experience.

5.7.3 Using Sexism as an Alignment Measure

In this work, we used measures of benevolent sexism and male role attitudes as our primary models for two main reasons: these measures have been used effectively for similar purposes in previous work [40], and they are measures that people tend to interpret more subjectively than measures such as morality, which have also been used in previous work [41].

That being said, there are issues with using measures of sexism in the way we used them in this paper that we need to address. While our results and related literature do suggest that an alignment with belief measures, such as benevolent sexism or male role attitudes, do increase avatar identification and enjoyment, this isn't the only thing designers should consider when creating game narratives and characters.

Because of the harm of sexism, intentionally creating a game with sexist content for the purpose of catering to an audience that identifies with those types of ideas is morally and socially questionable. Many popular games have been criticized for explicit hyper-masculine and hyper-feminine character designs (e.g., *Gears of War* [98], *Tomb Raider* [72]).

While our results and prior work suggest that these types of character designs may appeal to a small demographic that identifies with these characters, there is also a large demographic of potential players who would reject them. There is an argument to be had regarding whether designers should specifically tailor content to pander to their audience, or if they should create characters and ideas that challenge, and potentially influence or alter, the beliefs held by their audience members through empathizing with the characters' situations.

5.7.4 Ethics and Privacy

Collecting and analyzing player data is potentially problematic for the players' privacy. There are already issues with data privacy in tech and games [7, 37, 198]. Our approach has the potential to be used nefariously and it is important to consider that such data collection can be dangerous in terms of privacy [198]. This is particularly true for approaches that are unobtrusive. Thus, there is a tradeoff. While researchers and game developers can benefit from data collection that is not very obvious, if it does not interfere with the game experience [91, 94, 95, 231], it can also affect the players' perceptions of the game negatively. For instance, earlier work showed that data shared in NPC dialogues might be perceived as more intimate than data shared with a questionnaire [93]. Thus, such an unobtrusive approach should only be used with appropriate informed consent and sparingly, i.e., only when data collection is beneficial for players.

5.7.5 Limitations and Future Opportunities

In this work, we only tested our method with one type of belief, namely two measures of sexism: benevolent sexism and male role attitudes. While we are confident that our method for converting questionnaires into character dialogue would work for other beliefs and traits, further work should be conducted to verify the efficacy of this approach with different types of trait measures, such as personality types.

Also, our work only tested the dialogue in isolation, rather than in a full game setting. In the future, we would like to continue this work by embedding questionnaires into role-playing games and verifying our method still works when the player is immersed in a narrative world.

Finally, we would like to explore the possibility of creating player models through other types of choices players often make in games. While dialogue is the most obvious option, it is by no means the only one. For example, games often give players the opportunity to allocate skill points into character attributes (e.g., strength or intelligence) or special perks or abilities (e.g., increased critical hit damage or a fireball spell). While these types of choices are more closely tied to the efficacy of game mechanics than dialogue (e.g., increasing strength usually directly increases the amount of physical damage your character does), these types of choices could be presented in a way to subtly infer information about the player and build useful models. For example, allocating points in strength could mean the player sees themselves (or an ideal version of themselves) as strong. Designers could use this by altering how the character looks (making the character appear more muscular) or behaves (giving the character a rougher voice, making the character act more aggressive).

5.8 Conclusion

In this paper, we demonstrate two methods of successfully predicting responses to validated scales of sexist beliefs from interactions with NPCs. Our work can be used to inform the design of methods to assess player beliefs, values, and traits, without compromising engagement or immersion through typical explicit approaches to self-report.

Table 5.1: Interactive dialogue for the Direct condition, derived from the Male Role Attitudes Scale.

Original Item	Prompt	Response 1	Response 2
It is essential for a guy to get respect from others.	Why are you just standing there? It is respectful to bow to your King.	I agree, it is essential	I don't think it is essential.
A man always deserves the respect of his wife and children.	A man always deserves the respect of his wife and children.	Yes, very much so.	I don't think so.
I admire a guy who is totally sure of himself.	I hope you have more confidence than the other knights I have met with. I admire a guy who is totally sure of himself.	I think you are correct.	I don't think that's right.
A guy will lose respect if he talks about his problems.	When my Queen was still alive she always wanted to speak with me about her problems. It was maddening. A guy will lose respect if he talks about his problems.	I am on your side.	Not in my experience.
A young man should be physically tough, even if he's not big.	You seem well built. A young man should be physically tough, even if he's not big.	That's my belief.	You're wrong.
It bothers me when a guy acts like a girl.	Many of my younger knights are lacking manly qualities. Some even act more like princesses.	I'm with you on this one.	That's not my belief.
I don't think a husband should have to do housework.	You've been out in the world. I have heard from some of my advisors that there are households in which the husband is expected to cook and clean. I don't think a husband should have to do housework.	That is true.	I strongly disagree.
Men are always ready for sex.	A knight like you must have "saved" your fair share of maidens, if you know what I mean... I know what it's like, men are always ready for sex.	Hear hear.	I doubt that.

Table 5.2: Interactive dialogue for the Indirect condition, derived from the Male Role Attitudes Scale.

Original Item	Prompt	Response 1	Response 2
It is essential for a guy to get respect from others.	As the King, it is of the utmost importance that I am shown respect.	Well, it is essential for a guy to get respect from others.	Guys don't need to worry about gaining the respect of others.
A man always deserves the respect of his wife and children.	When my Queen was still alive, it was crucial that both she and my daughter respected me.	A man always deserves the respect of his wife and children.	Actually, the respect of family needs to be earned.
I admire a guy who is totally sure of himself.	I believe that confidence is one of the most important qualities in a man.	I admire a guy who is totally sure of himself.	I think that a little humility can go a long way.
A guy will lose respect if he talks about his problems.	My Queen always wanted to speak with me about her problems. It was maddening.	A guy will lose respect if he talks about his problems.	There is nothing shameful about talking through your problems.
A young man should be physically tough, even if he's not big.	So many of my knights neglect their physical fitness.	A young man should be physically tough, even if he's not big.	Even if a man is big, he does not need to be tough.
It bothers me when a guy acts like a girl.	Many of my knights are lacking manly qualities. Some even act more like princesses.	It bothers me when a guy acts like a girl.	Prince or princess... it's all the same to me. People can just be who they are.
I don't think a husband should have to do housework.	I heard from some of my advisors that there are households in which the husband is expected to cook and clean.	I don't think a husband should have to do housework.	I believe a husband should contribute to the household.
Men are always ready for sex.	A knight like you must have "saved" your fair share of maidens, if you know what I mean...	Men are always ready for sex.	Sometimes, men are uninterested in sex.

Table 5.3: Means and standard deviations for all measures used in the main study.

Construct	Mean	SD
Transportability	3.348	1.195
Male Role Attitudes	3.161	1.114
Benevolent Sexism	3.112	1.111
Protective Paternalism	3.348	1.195
Complementary Gender Differentiation	3.163	1.446
Heterosexual Intimacy	2.825	1.158
Narrative Understanding	5.024	0.816
Attentional Focus	5.159	1.722
Narrative Presence	4.898	1.566
Emotional Engagement	4.322	1.415
Homophily	4.097	1.516

Table 5.4: Regressions results with explained variance at the second level (R^2), standardized regression coefficients (β), and p values for regressions predicting questionnaire scores, using dialogue-model scores (DMS), dummy-coded condition (direct vs indirect, C), and the interaction ($DMS*C$).

	DMS		C		DMS*C		R^2
	β	p	β	p	β	p	
Protective Paternalism	.448	<.001	-.245	.081	.047	.737	.240
C. Gender Differentiation	.462	<.001	-.060	.594	-.012	.914	.218
Heterosexual Intimacy	.408	<.001	-.149	.127	.192	.049	.189
Benevolent Sexism	.658	<.001	-.169	.036	.068	.398	.452
Male Role Attitudes	.660	<.001	-.197	.035	.231	.013	.459

6 GENERAL DISCUSSION

6.1 Summary of Results

This section will provide a brief summary of the results from each manuscript.

6.1.1 Manuscript A

In Manuscript A we compared players' perceived experience between multiple versions of the same game: one version with gender neutral graphics and narrative (Gender Neutral), one version with gendered graphics and gender neutral narrative (Sexist Imagery), and one version with gendered graphics and gendered narrative (Sexist Imagery plus Dialogue). In this experiment we made the following findings:

- **Gender and Sexist Beliefs** - Looking at the relationship between gender and sexist beliefs, we found that gender predicted neither benevolent sexism nor male role attitudes.
- **Identification and Enjoyment** - Conversely, we found that a player's perceived identification with the protagonist positively predicted enjoyment.
- **Gender and Identification** - Despite common wisdom among gaming communities, we found that a player's self-identified gender did not predict perceived identification with the protagonist.
- **Sexist Beliefs and Identification** - Unlike with gender, we found that a player's sexist beliefs predict their identification with the protagonist.

The results from Manuscript A can be neatly summarized in Figure 6.1, which demonstrates the relationships between individual constructs, as determined through mediation models.

6.1.2 Manuscript B

Manuscript B includes 2 studies comparing experience between low fidelity (i.e., interactive text-based interfaces) and high fidelity (i.e., 3D graphics) games. The first study primarily compared measures of play experience (competence, autonomy, relatedness, presence, intuitive controls, enjoyment, effort, tension, similarity identification, embodied presence, and wishful identification).

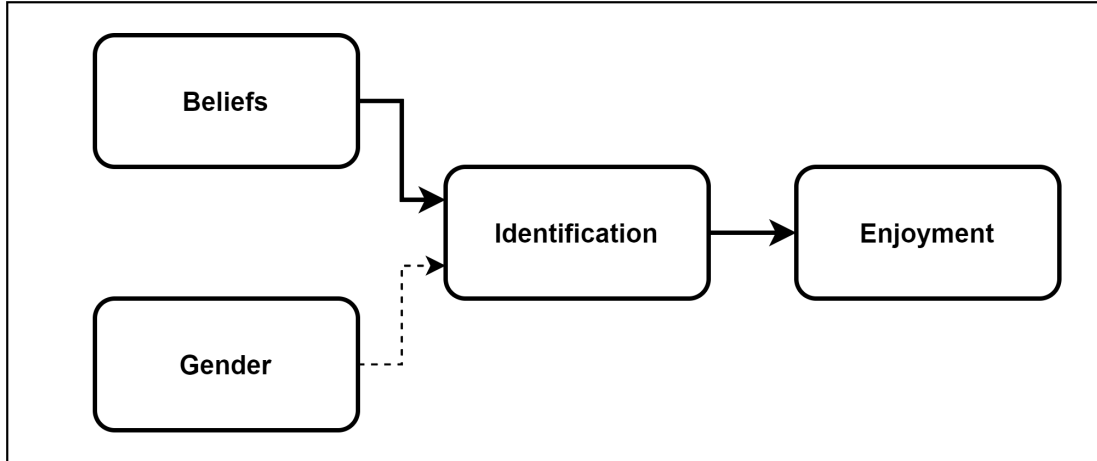


Figure 6.1: Mediation relationships as computed in Manuscript A. Arrows indicate a significant mediation and dotted lines indicate an insignificant mediation.

The second study primarily compared measures of narrative experience (enjoyment, effort, tension, competence, similarity identification, embodied presence, wishful identification, transportation into narrative worlds, narrative understanding, attentional focus, narrative presence, emotional engagement).

In the first study, we found no significant differences between conditions on any experiential measures. Additionally, our statistical tests had enough power to assert the null hypothesis on all measures except for autonomy, presence, enjoyment, similarity identification, and embodied presence.

In the second study, we found significant difference between conditions on the measures of transportation, similarity identification, and wishful identification, though these differences were extremely small. Additionally, we had enough statistical power to assert the null hypothesis on all other measures except for narrative presence and emotional engagement.

These two studies demonstrate that even though we were comparing games that were very different in terms of graphic fidelity, we found very small (if any) differences between all experiential measures.

6.1.3 Manuscript C

Manuscript C includes three studies exploring the mechanisms of how players make dialogue choices in video games. In all three studies, players were presented with a dialogue prompt or question, and were provided two or three possible responses. The number of responses and the content of each response (i.e., moral or immoral) was dependent on the particular study and experimental condition.

In the first study, participants were assigned to one of three conditions: Polar (one moral choice and one immoral choice), Moral (two moral choices), or Immoral (two immoral choices). In this study, participants reported lower levels of similarity identification, embodied presence, wishful identification, and transportation in the Immoral condition, opposed to the two other conditions. Similarly, participant reported investing less effort in the immoral condition, when compared to the other two conditions.

In the second study, participant were assigned to one of two conditions: Polar (one moral choice and one immoral choice) or Tailored (two choices of similar moral valence, in alignment with the player's previous choices). In this study, the Polar condition yielded greater identification, transportation, and effort than the tailored condition.

In the third study, we added a third condition to the two conditions in the second study. We called this condition Tailored 2.0, since it took the basic structure of the Tailored condition (i.e., two choices of moral valence aligning with the player's previous choices) but added a third, opposing choice. The main motivation for this study was to better understand the results from the second study. In that study, we found that opposing choices (Polar condition) fostered stronger identification, transportation, and effort than choices that aligned with the player's playstyle (Tailored condition). In this third study, we saw all of the significant effects from the second study go away, indicating that, while it could potentially be useful to model the player's choices, using the models to alter which choices players were presented did not seem to be better than the current industry standard (Polar choices).

In addition to looking at main effects, we also constructed mediation models looking at the relationships between Choice (the condition), identification, transportation, and effort. We saw the same relationships in all of the studies. In the first model (Figure 6.2), the type of choice presented to player (i.e., Polar or some version of tailored/aligned choices) predicted Identification directly. For transportation, the same relationship exists, but the relationship between choice and transportation is mediated by effort (Figure 6.3).

These models seem to indicate that having more diverse choices (i.e., Polar choices) results in stronger identification with the protagonist. This seems to make sense, as more diverse choices could be seen as giving the player more freedom in constructing the personality of their character, which could be argued is analogous to physically designing the character, which has been shown to increase identification [255].

In a similar vein, we see a relationship between Choice and transportation. Unlike with identification, however, this relationship is mediated by effort. This indicates that the reason Polar choices caused stronger levels of transportation was not necessarily because of the diversity of choices, but instead due to the diversity of choices causing participants to exert more effort, which was the actual driving factor on transportation. Intuitively this makes sense; investing more effort (and

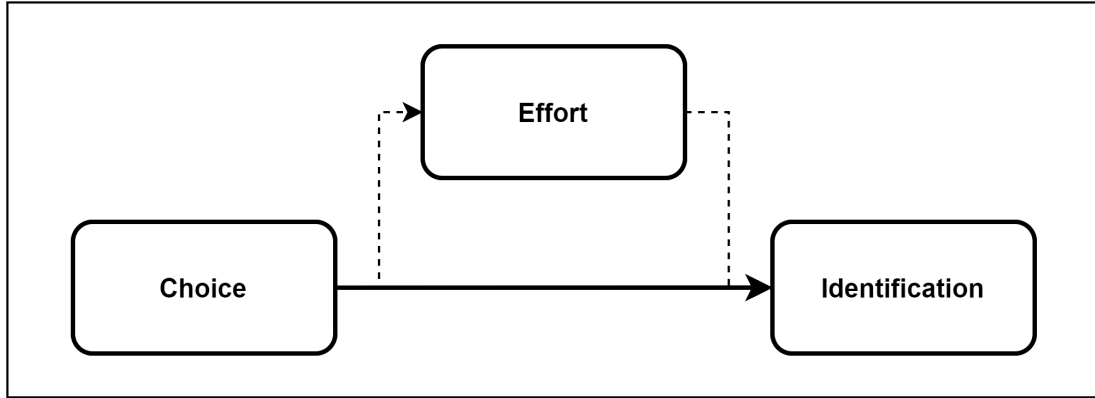


Figure 6.2: Mediation relationship between Choice, Effort, and Identification as computed in Manuscript C. Arrows indicate a significant mediation and dotted lines indicate an insignificant mediation.

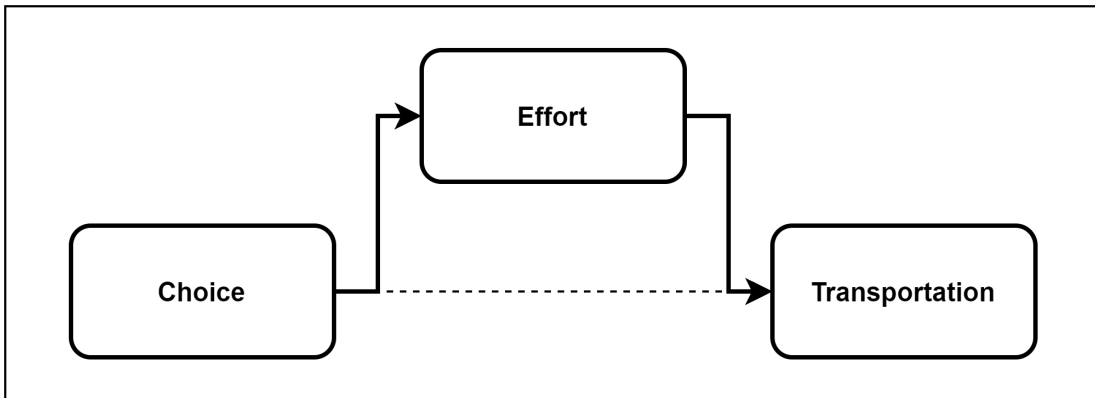


Figure 6.3: Mediation relationship between Choice, Effort, and Transportation as computed in Manuscript C. Arrows indicate a significant mediation and dotted lines indicate an insignificant mediation.

therefore more time) in making a choice makes one more invested in the outcome of the choice. It would make sense that an increase in investment could lead to a stronger immersion (transportation). Additionally, we see similar effects in other fields. Previous work has shown that when exerting more effort when reading financial statements, participants experienced a greater effect on judgements [189].

6.1.4 Manuscript D

In Manuscript D we conducted one user study which ties together all of the previous manuscripts. In Manuscript C we constructed models to better understand how players make choices and the driving factors for identification and transportation. In this study, we used the dialogue systems (based on the text prototypes evaluated in Manuscript B) and modelling approaches from Manuscript C to

evaluate a better research approach the main research question addressed in Manuscript A.

In Manuscript A, we looked at the relationships between a game’s protagonist and the real-world player controlling them. In particular, we looked at the sexist beliefs being exhibited by the protagonist and how the alignment (or opposition) with the player’s real world beliefs drove identification and enjoyment.

In this study, we only had one validated approach for measuring the player’s sexist beliefs: validated questionnaires, namely the Ambivalent Sexism Inventory (ASI, [105]) and Male Role Attitudes Scale (MRAS, [206]). While these two scales are validated and shown to work for research purposes, these scales wouldn’t work in a real game setting. If, for example, developers wanted to better understand the beliefs and values of their players, it would not be practical to have players fill out pages of questionnaires prior to playing the game. Players would get bored and possibly not answer honestly. If a developer intended to use the player’s responses to tailor game content in some way to increase the player’s experience, inaccurate measures would not only prevent any possible boosts to experience, but could actually create a game that is in opposition to the player’s beliefs, negatively affecting experience.

One solution to this problem, which is addressed in Manuscript D, is to embed the questionnaires inside of the game’s dialogue, making the questions a part of the gameplay, instead of making players answer them in isolation.

In this study we defined two methods of converting questionnaire items into dialogue: Direct (giving the player the prompt from the questionnaire and giving them two dialogue options, one in agreement and one in disagreement), and Indirect (giving the player a prompt related to the topic of the questionnaire and giving them two dialogue options, one the actual prompt from the questionnaire and the other an opposing statement).

We tested these two methods by comparing the models created from the dialogue responses to those created by the actual questionnaires. We first confirmed that the majority of participants made choices from their own perspective, as opposed to from the perspective of a separate character. Using hierarchical linear regressions, we showed that our dialogue-generated models predicted the scores calculated by the corresponding validated questionnaire. Additionally, we found no significant differences in experience between the two methods of dialogue creation (Direct and Indirect), indicating that neither was preferred by players.

6.2 Contributions

This section will briefly describe the main contributions this dissertation will provide to both industry and research methodologies. Some contributions have an effect on both, and thus will be briefly discussed in the appropriate context in both sections.

6.2.1 Industry Contributions

This section will summarize the main contributions of this dissertation as they relate to use by industry game developers and designers.

The main contribution of Manuscript A is the mediation models that show significant relationships between players' beliefs, identification with the protagonist, and overall enjoyment of the game. In particular, we showed that when a player's personal beliefs are in opposition with those demonstrated by the game's protagonist, the player identifies with the protagonist less. Additionally, we saw that the amount the player identified with the protagonist predicted the amount they perceived enjoying the game (See Figure 6.1).

A secondary finding was that an individual's demographic membership (i.e., self-identified gender) did not play a significant role in identifying with the hyper-masculine protagonist, but an alignment of beliefs did.

In isolation, this work suggests that game designers need to understand their audience beyond easy to gather demographic data (such as gender). It isn't enough for game designers to put their players into simple boxes, they need to understand the beliefs and values of their audiences and be sensitive to that when creating characters.

Manuscript B provides a brief framework for creating a text-based prototype as well as evidence that it can be used effectively to make early design decisions without needing a fully fleshed out game. Designers could use our methodology to test multiple competing game ideas both cheaply and quickly, allowing senior designers to make key decisions earlier in the development process. This work also provides a suggestion for how this methodology could be integrated into existing workflows, to allow designers to create deeper, more well-integrated relationships between gameplay and narrative.

Manuscript C provides several mediation models that help designers better understand how players perceive and make narrative choices, (see Figure 6.2 and 6.3). The main takeaways from these models are that designers should focus on providing players with diverse choices (e.g., both moral and immoral) to increase player identification, as well as choices that require players to exert effort in making the decision (e.g, difficult choices with no clear right answer, choices with major consequences) to increase player transportation.

Finally, Manuscript D provides game designers with a framework for embedding validated questionnaires into interactive dialogue choices. Game designers could use our approaches to gather data about their players for the purposes of tailoring game content and increasing the accuracy of gameplay data by better understanding the demographics of their players.

6.2.2 Methodological Contributions

This section will summarize the main contributions of this dissertation as they relate to providing methodological contributions to game user researchers.

The main methodological contribution from Manuscript B was the design and application of narrative prototypes for use in research. A major limitation we experienced in Manuscript A was the long development time required to create a full 3D game for the purpose of testing a single set of research questions.

This limitation could be at least partially remedied by researchers using our text-prototypes methodology. This methodology can be used to more quickly test research questions and hypotheses. While it is true that many elements of games require a complete game to accurately evaluate a research question, there is a great deal of research that focuses on extremely specific game elements (e.g., dialogue choices in games) that don't require a fully fleshed out game to test. For these types of research questions, a great deal of time and money is wasted on unnecessary development. This methodology allows researchers to focus primarily on the research questions in isolation rather than wasting time on unnecessary development. Additionally, this approach could encourage more cross-discipline research into games, by making the field more accessible to less technical researchers.

The studies in Manuscript C provide a practical research use-case for the narrative prototypes evaluated in Manuscript B. These studies show how researchers can study complex research questions relating to games (i.e., the mechanics of how players make dialogue choices) without needing the complete game to gather useful data.

Manuscript D gives researchers an alternate method of gathering questionnaire data from participants. Instead of asking participants to fill out pages of questionnaires, researchers could carefully integrate the questions into their games through dialogue choices. This would allow participants to be more immersed in the game when they answer the questions, and also potentially increase participant's overall experience throughout their study participation.

6.3 Ethics

6.3.1 Ethics of Data Collection

Collecting and analyzing player data is potentially problematic for players' privacy. While our suggested methods are, for the most part, benevolent in terms of application, by gathering data as described in our methods, game companies have the responsibility to treat the gathered data responsibly. In order for our methods to be used in commercial games, players should expect game developers and publishers to be transparent with what data they are gathering and how it is being used. Companies should ensure that players give informed consent prior to gathering any data from

the methods described in this dissertation.

Furthermore, not every player will consent to their data being used in the ways described in this dissertation. As such, it is crucial for game developers to treat our methods as a way of augmenting experience (e.g., tailoring protagonist behaviours to better align with the player’s personally held beliefs to increase identification) and not as a game mechanic in and of itself. By treating our methods in this way, players can all play the game, and simply opt in or out of having their data used in this way (e.g., through an option in the game settings, similar to how players can toggle on and off sensitive content such as blood in some games).

6.4 Reflections on Manuscripts

The research presented in this dissertation was conducted across a span of 5 years (2016-2021). As such, my personal views and beliefs have evolved and changed a great deal over that time, which I believe is reflected in the evolution of the research projects presented in this dissertation. In this section, I will go through each research project and reflect on how it aligned with my beliefs then compared to now, as well as highlight any missed opportunities or limitations that weren’t considered at the time of the initial publication.

6.4.1 Manuscript A

As the first project chronologically, my views on research methods and approaches have deviated the most since the time of writing Manuscript A. Of particular consequence is the treatment of gender as a binary variable that primarily drove the research questions, experiment design, and data analysis.

The initial motivations behind this project was to examine the effects of sexist content on measures of play experience (e.g., enjoyment), as well as investigating whether there were short- or long-term effects on individually held sexist beliefs when exposed to sexist content. As a result, the game content was designed around a gender-binary framework (i.e., characters modelled around traditional binary gender assumptions such as muscular knights and petite, delicate princesses). Similarly, many aspects of the study design were constructed under assumptions of gender-binary ideas, such as the measures of sexism themselves (these measures have built-in assumptions of the gender-binary which are immediately exclusive of players who identify as non-binary), to the way self-identified gender was measured (i.e., choosing between ‘Male’, ‘Female’, or ‘Other’).

Beyond the study design, once I started gathering data it became clear that the more interesting story was the relationship between self-identified gender and beliefs in sexist ideas (i.e., benevolent sexism and male role attitudes). As discussed in the Results and Discussion sections for Manuscript A, participants’ self-identified binary gender did not play a significant role in any of the statistical

models. This was interpreted as an argument for researchers and game designers to move away from sexist content, and questioned the importance of researchers utilizing gender binary as a fundamental experimental variable.

That being said, the design of the research methods themselves can be viewed as exclusive to individuals who identify as non-binary, despite the intention of making an argument in favour of gender non-binary (i.e., the driving force of the data was individually-held beliefs instead of gender-binary categorization), the study itself was based around assumptions of gender-binary, as described above.

Given a chance to design this experiment again, given what I now know and believe, I would do several things differently. First, the method of gathering and reporting participant gender would be more inclusive. Instead of prompting participants with choices of {‘Male’, ‘Female’ or ‘Other’}, I would present the following choices: {‘woman’, ‘man’, ‘non-binary’, ‘prefer not to disclose’, ‘prefer to self-describe’ }. The former was the method used in Manuscripts A, B, and C, and the latter, more inclusive method, was used in Manuscript D.

In addition to how gender is gathered, I would avoid approaching gender as an independent variable in the analysis. As demonstrated in the results from Manuscript A, gender didn’t play a significant role in the statistical models, and it is difficult to be both statistically rigorous (i.e., having participants evenly distributed across the factor) and be inclusive for all participants.

6.4.2 Manuscript B

The experimental design of the two studies in Manuscript B had similar limitations to Manuscript A in terms of how participant gender was measured and reported in the paper. If I had the opportunity to do those studies again, I would make the same changes as suggested in the previous section.

Additionally, the design choice of colour-coding in the mini-map of the high-fidelity version of the Space Adventure game (Figure 3.1) presents an issue for participants with colour-blindness. This limitation was not considered at the time of creating the systems used in the study, and should definitely be addressed if the game was to be used in future research projects. The map was used as a way to visualize to the player where they currently were in the spaceship, highlighting which rooms were currently inaccessible, the current goal, etc. The simplest solution would be to add icons to the minimap to indicate these states, as opposed to relying on the colour of the rooms.

Another limitation in this Manuscript was the use of blog posts to motivate the main problem and motivation of the research. The reason for including these blog posts was to ground the research in industry practices. Unfortunately, it is rare for many of these practices to be published in academic venues, since the details of these practices are often closely guarded secrets. As such, in order to examine some of these practices it was necessary to piece together various comments

and statements from blog posts to get a feel for the bigger picture. While this was the best solution I could come up with at the time, using blog posts as a reference in this way is limited because there is no guarantee the original blog posts will still be accessible in the future.

6.4.3 Manuscript C

Clarifying the Title

There are two main issues to raise with respect to Manuscript C. The first issue is the title of the paper. The reasoning behind this title was a reference to the film *The Matrix*¹. In this film, the protagonist is living in a computer simulation and is offered two pills: a red pill and a blue pill. The blue pill would make the character go back to their life of blissful ignorance inside the simulation and the red pill causes the character to awaken from the simulation and accept the truth of their reality.

There are two reasons for this reference in the title of the paper. First, the studies presented in this Manuscript deal with binary dialogue choices, and the Matrix is one of the most famous forms of media in which a binary choice is used in such a fundamental way to the plot. The second reason for this reference was to the way major dialogue choices are presented in the *Mass Effect* series²: colour-coded as red (renegade choices) or blue (paragon choices).

References to the red pill are also used in far-right conspiracy communities [185] as a metaphor for becoming awakened to the truth of the world by accepting the content of a particular conspiracy. Though this metaphor was being used in these communities at the time this Manuscript was published, I was unaware of this use until recently. Because of the use of this metaphor, it is important to explain the reason for the title, and to clarify that the work is in no way affiliated with, or endorsing, the conspiracies associated with the metaphor.

Limitations with Moral Foundation Theory

The dialogue choices used in the main stimuli of this study were framed around morality as measured using the Moral Foundations Questionnaire [114]. The reason for using this particular scale was to more seamlessly build on a related study published around the time of this research [126]. The scale itself is problematic, particularly the Purity subscale. Many of the questions refer to outdated ideas based on sexism, racism, and nationalism. Given the opportunity to do this research over again, I would remove this scale entirely and consider other ways of framing and contextualizing morality. One way would be to use dialogue and decisions from existing role playing games as the structure for a morality model.

¹The Matrix, Warner Bros. Pictures, 1999

²Mass Effect, Bioware, 2007

Also, in the original published version of this manuscript, it is stated that due to the problematic purity axis, the dialogue used the purity access to refer to interactions with ‘dark magic’. In this version of the manuscript, references to ‘dark magic’ have been replaced with ‘evil magic’ to avoid historically racist references.

6.4.4 Manuscript D

Manuscript D was written and published in parallel to the writing of this dissertation, so there are few things to mention in terms of reflection. That being said, I will mention a few things of interest relating to this manuscript.

The main belief models used in this research were benevolent sexism [105] and male role attitudes [206], the same measures as in Manuscript A. As with my reflections on Manuscript A, I would like to point out the gender-binary assumptions built into measures of sexism and the exclusion this gives to participants who identify as non-binary. The main reason for the use of these scales was to connect this final piece of research back to the first (Manuscript A), and one way I chose to do this was to use the same scales to remain consistent. That being said, there are a number of other scales that could have been used in this study to be more inclusive of the participants. A questionnaire like transportability [61], for example, would measure a trait, similar to sexism, without some of the issues that come with a gender-binary assumption.

7 CONCLUSION

Researchers and game designers have long known that when players identify with protagonists in games they enjoy the game more. Much of the research into this topic, however, focuses on the character's physical appearance, instead of deeper traits such as personally-held beliefs and personality. In terms of these types of traits, each individual person possesses a unique combination of beliefs, values, and personality elements. Based on what researchers know about character identification, it is likely beneficial for game protagonists to reflect the player's own traits through elements of their in-game behaviour such as dialogue.

Given the diversity of individuals, however, designing a single protagonist that adequately reflects each individual player is impossible. The solution to this problem is to tailor the protagonist's in-game behaviour to more closely align with the beliefs and values of the player. To do this, however, game designers need to gather information about each individual player (i.e., measure their real-world beliefs and values). The current methods for gathering this type of information is through validated, subjective questionnaires. While this approach works in a formal research setting, making players fill out pages of questionnaires in a commercial game would likely be poorly received by players, even if they end up creating a better overall experience.

This dissertation addresses the problem of how game designers can effectively measure players' real-world beliefs inside of a game setting, and how tailoring characters to player's beliefs affects game experience. Specifically, this work makes four main contributions:

1. Understanding that an alignment of beliefs between the player and protagonist leads to stronger measures of Identification and Enjoyment.
2. Demonstrating the effectiveness of using low-fidelity, text-based prototypes as a proxy for high-fidelity graphical games for the purpose of evaluating basic player experience.
3. Demonstrating a method for using a player's in-game dialogue choices to construct a dynamic model of the player, and through this system better understand the psychological mechanisms of choice. Specifically, 1)that giving a player more varied choices increases their perceived identification with the protagonist, and 2)Giving the player more varied choices increases the amount of effort they put in to making the choice, and through that effort players are more transported into the narrative world.

4. Using the methods from the previous contribution, this dissertation demonstrates a method for embedding validated questionnaires into in-game dialogue, and shows that this method creates player models that are comparable to those created through the standard practices of Likert questionnaires.

Together, these contributions provide several novel methodologies for use in both academia and industry, as well as contributes a greater understanding of the psychological mechanisms of player choice and identification between player and protagonist.

REFERENCES

- [1] One-way ancova in spss statistics.
- [2] Part 4: “adopt, adapt and improve” - agile qa bioware style.
- [3] Half-life review, Nov 1998.
- [4] The last of us review, June 2013.
- [5] Jedi: Fallen order review, Nov 2019.
- [6] Knight archetype, Jun 2020.
- [7] A. Boutilier [The Star]. Video game companies are collecting massive amounts of data about you. <https://www.thestar.com/news/canada/2015/12/29/how-much-data-are-video-games-collecting-about-you.html>, 2019. Accessed: 2020-07-28.
- [8] Espen Aarseth. A narrative theory of games. In *Proceedings of the international conference on the foundations of digital Games*, pages 129–133, 2012.
- [9] Lyn Y Abramson, Martin E Seligman, and John D Teasdale. Learned helplessness in humans: Critique and reformulation. *Journal of abnormal psychology*, 87(1):49, 1978.
- [10] Gregory R Andrews and Ray L Debus. Persistence and the causal perception of failure: Modifying cognitive attributions. *Journal of educational psychology*, 70(2):154, 1978.
- [11] Jonathan Arnowitz, Michael Arent, and Nevin Berger. *Effective prototyping for software makers*. Elsevier, 2010.
- [12] John William Atkinson. An introduction to motivation. 1964.
- [13] Rachel Bailey, Kevin Wise, and Paul Bolls. How avatar customizability affects children’s arousal and subjective presence during junk food–sponsored online video games. *CyberPsychology & Behavior*, 12(3):277–283, 2009.
- [14] Albert Bandura. Self-efficacy: toward a unifying theory of behavioral change. *Psychological review*, 84(2):191, 1977.
- [15] Albert Bandura, Gian Vittorio Caprara, Claudio Barbaranelli, Maria Gerbino, and Concetta Pastorelli. Role of affective self-regulatory efficacy in diverse spheres of psychosocial functioning. *Child development*, 74(3):769–782, 2003.
- [16] Albert Bandura and Edwin A Locke. Negative self-efficacy and goal effects revisited. *Journal of applied psychology*, 88(1):87, 2003.
- [17] Albert Bandura and Sebastian Wessels. *Self-efficacy*. W.H. Freeman & Company, 1997.
- [18] Roy F Baumeister, Laura Smart, and Joseph M Boden. Relation of threatened egotism to violence and aggression: The dark side of high self-esteem. *Psychological review*, 103(1):5, 1996.
- [19] BBC. Blog death threats spark debate, 2007.

- [20] Brianna Beck, Steven Di Costa, and Patrick Haggard. Having control over the external world increases the implicit sense of agency. *Cognition*, 162:54–60, 2017.
- [21] Jeffery G Bednark and Elizabeth A Franz. Agency attribution: Event-related potentials and outcome monitoring. *Experimental Brain Research*, 232(4):1117–1126, 2014.
- [22] Elizabeth Behm-Morawitz and Dana Mastro. The effects of the sexualization of female video game characters on gender stereotyping and female self-concept. *Sex roles*, 61(11-12):808–823, 2009.
- [23] Nicole Behne, Henning Scheich, and André Brechmann. The left dorsal striatum is involved in the processing of neutral feedback. *Neuroreport*, 19(15):1497–1500, 2008.
- [24] Daryl J Bem, Michael A Wallach, and Nathan Kogan. Group decision making under risk of aversive consequences. *Journal of Personality and Social Psychology*, 1(5):453, 1965.
- [25] Jamil P Bhanji and Mauricio R Delgado. The social brain and reward: social information processing in the human striatum. *Wiley Interdisciplinary Reviews: Cognitive Science*, 5(1):61–73, 2014.
- [26] Helena Bilandzic and Rick W Busselle. Transportation and transportability in the cultivation of genre-consistent attitudes and estimates. *Journal of Communication*, 58(3):508–529, 2008.
- [27] Helena Bilandzic and Rick W Busselle. Enjoyment of films as a function of narrative experience, perceived realism and transportability. *Communications*, 36(1):29–50, 2011.
- [28] Bioware. *Mass Effect*. Electronic Arts, 2007. Video Game.
- [29] Bioware. *Mass Effect*. Bioware, 2007. Video Game.
- [30] Bioware. *Dragon Age: Inquisition*. Electronic Arts, 2014. Video Game.
- [31] Max Birk and Regan L Mandryk. Control your game-self: effects of controller type on enjoyment, motivation, and personality in game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 685–694, 2013.
- [32] Max V Birk, Cheralyn Atkins, Jason T Bowey, and Regan L Mandryk. Fostering intrinsic motivation through avatar identification in digital games. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 2982–2995, 2016.
- [33] Max V Birk and Regan L Mandryk. Combating attrition in digital self-improvement programs using avatar customization. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 660. ACM, 2018.
- [34] Max V Birk, Dereck Toker, Regan L Mandryk, and Cristina Conati. Modeling motivation in a social network game using player-centric traits and personality traits. In *International Conference on User Modeling, Adaptation, and Personalization*, pages 18–30. Springer, 2015.
- [35] Bruce Blaine and Jennifer Crocker. Self-esteem and self-serving biases in reactions to positive and negative events: An integrative review. In *Self-esteem*, pages 55–85. Springer, 1993.
- [36] Janet Bloomfield. #gamergate is fundamentally a fight for men, for masculinity and for male spaces., 2015.
- [37] BN. Meeks [NBCNews.com]. Is privacy possible in the digital age? <http://www.nbcnews.com/id/3078854/t/privacy-possible-digital-age/>, 2000. Accessed: 2020-09-09.
- [38] Jocelyn H Bolin. Hayes, andrew f.(2013). introduction to mediation, moderation, and conditional process analysis: A regression-based approach. new york, ny: The guilford press. *Journal of Educational Measurement*, 51(3):335–337, 2014.

- [39] Jason T Bowey, Max V Birk, and Regan L Mandryk. Manipulating leaderboards to induce player experience. In *Proceedings of the 2015 annual symposium on computer-human interaction in play*, pages 115–120, 2015.
- [40] Jason T Bowey, Ansgar E Depping, and Regan L Mandryk. Don’t talk dirty to me: How sexist beliefs affect experience in sexist games. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 1530–1543, 2017.
- [41] Jason T Bowey, Maximilian A Friehs, and Regan L Mandryk. Red or blue pill: fostering identification and transportation through dialogue choices in rpgs. In *Proceedings of the 14th International Conference on the Foundations of Digital Games*, pages 1–11, 2019.
- [42] Jason T Bowey, Julian Frommel, Brandon Piller, and Regan L Mandryk. Predicting beliefs from npc dialogues. In *Proceedings of the 3rd IEEE Conference on Games*, 2021.
- [43] Jason T Bowey and Regan L Mandryk. Those are not the stories you are looking for: Using text prototypes to evaluate game narratives early. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, pages 265–276. ACM, 2017.
- [44] Nicola J Bawn, Daniel Read, and Barbara Summers. The lure of choice. *Journal of Behavioral Decision Making*, 16(4):297–308, 2003.
- [45] Gifford W Bradley. Self-serving biases in the attribution process: A reexamination of the fact or fiction question. *Journal of personality and social psychology*, 36(1):56, 1978.
- [46] Brett Bunge. ”mass effect” choices: The power of a personalized journey, 2017.
- [47] Bungie. *Destiny*. Activision, 2014. Video Game.
- [48] Melinda CR Burgess, Steven Paul Stermer, and Stephen R Burgess. Sex, lies, and video games: The portrayal of male and female characters on video game covers. *Sex roles*, 57(5-6):419–433, 2007.
- [49] Rick Busselle and Helena Bilandzic. Measuring narrative engagement. *Media Psychology*, 12(4):321–347, 2009.
- [50] Judith Butler. *Gender trouble: Feminism and the subversion of identity*. routledge, 2011.
- [51] Bill Buxton. *Sketching user experiences: getting the design right and the right design*. Morgan kaufmann, 2010.
- [52] Galactic Cafe. *The Stanley Parable*. Galactic Cafe, 2013. Video Game.
- [53] W Keith Campbell and Constantine Sedikides. Self-threat magnifies the self-serving bias: A meta-analytic integration. *Review of general Psychology*, 3(1):23, 1999.
- [54] Katrina Carlsson, Jesper Andersson, Predrag Petrovic, Karl Magnus Petersson, Arne Öhman, and Martin Ingvar. Predictability modulates the affective and sensory-discriminative neural processing of pain. *Neuroimage*, 32(4):1804–1814, 2006.
- [55] Justine Cassell. Chess for girls?: Feminism and computer games. 1998.
- [56] Justine Cassell and Henry Jenkins. *From Barbie to Mortal Kombat: gender and computer games*. MIT press, 2000.
- [57] Michael B Catani and David W Biers. Usability evaluation and prototype fidelity: Users and usability professionals. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 42, pages 1331–1335. SAGE Publications Sage CA: Los Angeles, CA, 1998.
- [58] A Charles Catania and Terje Sagvolden. Preference for free choice over forced choice in pigeons. *Journal of the experimental analysis of behavior*, 34(1):77–86, 1980.

- [59] Shira Chess and Adrienne Shaw. A conspiracy of fishes, or, how we learned to stop worrying about # gamergate and embrace hegemonic masculinity. *Journal of Broadcasting & Electronic Media*, 59(1):208–220, 2015.
- [60] Julia F Christensen and Antoni Gomila. Moral dilemmas in cognitive neuroscience of moral decision-making: A principled review. *Neuroscience & Biobehavioral Reviews*, 36(4):1249–1264, 2012.
- [61] Katheryn R Christy and Jesse Fox. Transportability and presence as predictors of avatar identification within narrative video games. *Cyberpsychology, Behavior, and Social Networking*, 19(4):283–287, 2016.
- [62] Judy Y Chu, Michelle V Porche, and Deborah L Tolman. The adolescent masculinity ideology in relationships scale: Development and validation of a new measure for boys. *Men and Masculinities*, 8(1):93–115, 2005.
- [63] Ros Clubb and Georgia Mason. Animal welfare: captivity effects on wide-ranging carnivores. *Nature*, 425(6957):473, 2003.
- [64] Jonathan Cohen. Defining identification: A theoretical look at the identification of audiences with media characters. *Mass communication & society*, 4(3):245–264, 2001.
- [65] Joanne Cohoon and William Aspray. *Lost in translation: Gender and high school computer science*. Mit Press, 2008.
- [66] Greg Costkyan. I have no words & i must design: toward a critical vocabulary for games. In *Proceedings of the computer games and digital cultures conference, Finland*, 2002.
- [67] Chris Crawford. *Chris Crawford on interactive storytelling*. New Riders, 2012.
- [68] G Crombez, Frank Baeyens, and Paul Eelen. Sensory and temporal information about impending pain: the influence of predictability on pain. *Behaviour research and therapy*, 32(6):611–622, 1994.
- [69] Geert Crombez, Christopher Eccleston, Petra De Vlieger, Stefaan Van Damme, and Armand De Clercq. Is it better to have controlled and lost than never to have controlled at all? an experimental investigation of control over pain. *PAIN®*, 137(3):631–639, 2008.
- [70] Sonya Dal Cin, Mark P Zanna, and Geoffrey T Fong. Narrative persuasion and overcoming resistance. *Resistance and persuasion*, 2:175–191, 2004.
- [71] Suzanne De Castell and Mary Bryson. Retooling play: Dystopia, dysphoria, and difference. *From Barbie to Mortal Kombat: gender and computer games*, pages 232–261, 1998.
- [72] Core Design. *Tomb Raider*. Edits Intractive, 1996. Video Game.
- [73] Ralph J DiClemente, Richard A Crosby, and Michelle C Kegler. *Emerging theories in health promotion practice and research*. John Wiley & Sons, 2009.
- [74] Rockstar San Diego. *Red Dead Redemption*. Rockstar Games, 2010. Video Game.
- [75] Karen E Dill, Brian P Brown, and Michael A Collins. Effects of exposure to sex-stereotyped video game characters on tolerance of sexual harassment. *Journal of Experimental Social Psychology*, 44(5):1402–1408, 2008.
- [76] Igor Dolgov, William J Graves, Matthew R Nearents, Jeremy D Schwark, and C Brooks Volkman. Effects of cooperative gaming and avatar customization on subsequent spontaneous helping behavior. *Computers in human behavior*, 33:49–55, 2014.

- [77] Jianming Dong. Book review: Paper prototyping-the fast and easy way to design and refine user interfaces, by carolyn snyder. *International Journal of Human-Computer Interaction*, 17(2):287–288, 2004.
- [78] Edward Downs and Stacy L Smith. Keeping abreast of hypersexuality: A video game character content analysis. *Sex roles*, 62(11-12):721–733, 2010.
- [79] Christopher Dring. More money is spent on games than movies and music combined, says ihs. *MCV*. Retrieved June, 30:2017, 2015.
- [80] Magy Seif El-Nasr, Anders Drachen, and Alessandro Canossa. *Game analytics*. Springer, 2016.
- [81] Malte Elson, Johannes Breuer, James D Ivory, and Thorsten Quandt. More than stories with buttons: Narrative, mechanics, and context as determinants of player experience in digital games. *Journal of Communication*, 64(3):521–542, 2014.
- [82] Daniel Engelberg and Ahmed Seffah. A framework for rapid mid-fidelity prototyping of web sites. In *IFIP World Computer Congress, TC 13*, pages 203–215. Springer, 2002.
- [83] Franz Faul, Edgar Erdfelder, Axel Buchner, and Albert-Georg Lang. Statistical power analyses using g* power 3.1: Tests for correlation and regression analyses. *Behavior research methods*, 41(4):1149–1160, 2009.
- [84] Olga Eizner Favreau. Sex and gender comparisons: Does null hypothesis testing create a false dichotomy? *Feminism & Psychology*, 7(1):63–81, 1997.
- [85] Leon Festinger. *A theory of cognitive dissonance*, volume 2. Stanford university press, 1962.
- [86] Andy Field. *Discovering statistics using IBM SPSS statistics*. sage, 2013.
- [87] Susan Folkman and Richard S Lazarus. Coping as a mediator of emotion. *Journal of personality and social psychology*, 54(3):466, 1988.
- [88] Jesse Fox, Jeremy N Bailenson, and Liz Tricase. The embodiment of sexualized virtual selves: The proteus effect and experiences of self-objectification via avatars. *Computers in Human Behavior*, 29(3):930–938, 2013.
- [89] Jesse Fox, Rachel A Ralston, Cody K Cooper, and Kaitlyn A Jones. Sexualized avatars lead to women’s self-objectification and acceptance of rape myths. *Psychology of Women Quarterly*, 39(3):349–362, 2015.
- [90] Jesse Fox and Wai Yen Tang. Sexism in online video games: The role of conformity to masculine norms and social dominance orientation. *Computers in Human Behavior*, 33:314–320, 2014.
- [91] Julian Frommel, Fabian Fischbach, Katja Rogers, and Michael Weber. Emotion-based dynamic difficulty adjustment using parameterized difficulty and self-reports of emotion. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play, CHI PLAY ’18*, page 163–171, New York, NY, USA, 2018. Association for Computing Machinery.
- [92] Julian Frommel, Madison Klarkowski, and Regan L Mandryk. The Struggle is Spiel: On Failure and Success in Games. In *Proceedings of the 16th International Conference on the Foundations of Digital Games (FDG)*, FDG ’21, Virtual, 2021. ACM.
- [93] Julian Frommel, Cody Phillips, and Regan L Mandryk. Gathering Self-Report Data in Games through NPC Dialogues: Effects on Data Quality, Data Quantity, Player Experience, and Information Intimacy. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, CHI ’21*, Virtual, 2021. ACM.

- [94] Julian Frommel, Katja Rogers, Julia Brich, Daniel Besserer, Leonard Bradatsch, Isabel Ortinou, Ramona Schabenberger, Valentin Riemer, Claudia Schrader, and Michael Weber. Integrated questionnaires: Maintaining presence in game environments for self-reported data acquisition. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY '15, page 359–368, New York, NY, USA, 2015. Association for Computing Machinery.
- [95] Julian Frommel, Valentin Sagl, Ansgar E Depping, Colby Johanson, Matthew K Miller, and Regan L Mandryk. Recognizing affiliation: Using behavioural traces to predict the quality of social interactions in online games. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–16, 2020.
- [96] Juri Fujiwara, Nobuo Usui, Soyoung Q Park, Tony Williams, Toshio Iijima, Masato Taira, Ken-Ichiro Tsutsui, and Philippe N Tobler. Value of freedom to choose encoded by the human brain. *Journal of neurophysiology*, 110(8):1915–1929, 2013.
- [97] Tracy Fullerton. *Game design workshop: a playcentric approach to creating innovative games*. CRC press, 2014.
- [98] Epic Games. *Gears of War*. Microsoft Game Studios, 2006. Video Game.
- [99] Platinum Games. *Bayonetta*. Nintendo, Sega, 2009. Video Game.
- [100] Kathrin M Gerling, Max Birk, Regan L Mandryk, and Andre Doucette. The effects of graphical fidelity on player experience. In *Proceedings of international conference on Making Sense of Converging Media*, pages 229–236, 2013.
- [101] Kathrin Maria Gerling, Matthew Miller, Regan L Mandryk, Max Valentin Birk, and Jan David Smeddinck. Effects of balancing for physical abilities on player performance, experience and self-esteem in exergames. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2201–2210, 2014.
- [102] Richard J Gerrig. *Experiencing narrative worlds: On the psychological activities of reading*. Yale University Press, 1993.
- [103] David Gibson and Karen Schrier. *Ethics and Game Design*. 2010.
- [104] Gary B Glavin, William P Paré, Tone Sandbak, Hans-Kristian Bakke, and Robert Murison. Restraint stress in biomedical research: an update. *Neuroscience & Biobehavioral Reviews*, 18(2):223–249, 1994.
- [105] Peter Glick and Susan T Fiske. The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of personality and social psychology*, 70(3):491, 1996.
- [106] Peter Glick and Susan T Fiske. The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. In *Social Cognition*, pages 116–160. Routledge, 2018.
- [107] John D Gould and Clayton Lewis. Designing for usability: key principles and what designers think. *Communications of the ACM*, 28(3):300–311, 1985.
- [108] Jesse Graham, Brian A Nosek, Jonathan Haidt, Ravi Iyer, Spassena Koleva, and Peter H Ditto. Mapping the moral domain. *Journal of personality and social psychology*, 101(2):366, 2011.
- [109] Kishonna L Gray. Deviant bodies, stigmatized identities, and racist acts: Examining the experiences of african-american gamers in xbox live. *New Review of Hypermedia and Multimedia*, 18(4):261–276, 2012.
- [110] Melanie C Green and Timothy C Brock. The role of transportation in the persuasiveness of public narratives. *Journal of personality and social psychology*, 79(5):701, 2000.

- [111] Melanie C Green and Keenan M Jenkins. Interactive narratives: Processes and outcomes in user-directed stories. *Journal of Communication*, 64(3):479–500, 2014.
- [112] Jeff Greenberg, Tom Pyszczynski, John Burling, and Karyn Tibbs. Depression, self-focused attention, and the self-serving attributional bias. *Personality and Individual Differences*, 13(9):959–965, 1992.
- [113] Adams Greenwood-Ericksen and Juliana Goriounov. Assessment of responses to narrative in games. 2016.
- [114] Jonathan Haidt, Craig Joseph, et al. The moral mind: How five sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules. *The innate mind*, 3:367–391, 2007.
- [115] Sandra G Hart and Lowell E Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier, 1988.
- [116] Tilo Hartmann and Christoph Klimmt. Gender and computer games: Exploring females’ dislikes. *Journal of Computer-Mediated Communication*, 11(4):910–931, 2006.
- [117] Andrew F Hayes. Process: A versatile computational tool for observed variable mediation, moderation, and conditional process modeling, 2012.
- [118] Andrew F Hayes. *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford publications, 2017.
- [119] Dorothée Hefner, Christoph Klimmt, and Peter Vorderer. Identification with the player character as determinant of video game enjoyment. In *International conference on entertainment computing*, pages 39–48. Springer, 2007.
- [120] Fritz Helder. The psychology of interpersonal relations. *New York*, 1958.
- [121] Andrew Hogue, Sunbir Gill, and Michael Jenkin. Automated avatar creation for 3d games. In *Proceedings of the 2007 conference on Future Play*, pages 174–180, 2007.
- [122] Johan Huizinga. *Homo ludens*. Editora Perspectiva SA, 2020.
- [123] Robin Hunicke, Marc LeBlanc, and Robert Zubek. Mda: A formal approach to game design and game research. In *Proceedings of the AAAI Workshop on Challenges in Game AI*, volume 4, page 1722, 2004.
- [124] IO Interactive. *Hitman:Absolution*. Square Enix, 2012. Video Game.
- [125] Katherine Isbister and Noah Schaffer. *Game usability: Advancing the player experience*. CRC press, 2008.
- [126] Glena H Iten, Sharon T Steinemann, and Klaus Opwis. Choosing to help monsters: A mixed-method examination of meaningful choices in narrative-rich games and interactive narratives. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 341. ACM, 2018.
- [127] James D Ivory. Still a man’s game: Gender representation in online reviews of video games. *Mass Communication & Society*, 9(1):103–114, 2006.
- [128] Sheena S Iyengar and Mark R Lepper. When choice is demotivating: Can one desire too much of a good thing? *Journal of personality and social psychology*, 79(6):995, 2000.
- [129] Mary R Jackman. *The velvet glove: Paternalism and conflict in gender, class, and race relations*. Univ of California Press, 1994.

- [130] Jeroen Jansz. The emotional appeal of violent video games for adolescent males. *Communication theory*, 15(3):219–241, 2005.
- [131] Charlene Jennett, Anna L Cox, Paul Cairns, Samira Dhoparee, Andrew Epps, Tim Tijs, and Alison Walton. Measuring and defining the experience of immersion in games. *International journal of human-computer studies*, 66(9):641–661, 2008.
- [132] Jennifer Jenson and Suzanne De Castell. Theorizing gender and digital gameplay: Oversights, accidents and surprises. *Eludamos. Journal for Computer Game Culture*, 2(1):15–25, 2008.
- [133] Jennifer Jenson and Suzanne de Castell. Girls@ play: An ethnographic study of gender and digital gameplay. *Feminist Media Studies*, 11(2):167–179, 2011.
- [134] Oliver P John and Richard W Robins. Accuracy and bias in self-perception: individual differences in self-enhancement and the role of narcissism. *Journal of personality and social psychology*, 66(1):206, 1994.
- [135] Daniel Johnson and John Gardner. Effects of team-based computer interaction: The media equation and game design considerations. In Fumio Kishino, Yoshifumi Kitamura, Hirokazu Kato, and Noriko Nagata, editors, *Entertainment Computing - ICEC 2005*, pages 468–479, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg.
- [136] Daniel Johnson and John Gardner. Personality, motivation and video games. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*, pages 276–279, 2010.
- [137] Philip N Johnson-Laird and Ruth MJ Byrne. Modal reasoning, models, and manktelow and over. *Cognition*, 43(2):173–182, 1992.
- [138] PN Johnson-Laird and RMJ Byrne. *Essays in cognitive psychology. deduction.* hillsdale, nj, us, 1991.
- [139] Guy Kahane, Jim AC Everett, Brian D Earp, Miguel Farias, and Julian Savulescu. ‘utilitarian’ judgments in sacrificial moral dilemmas do not reflect impartial concern for the greater good. *Cognition*, 134:193–209, 2015.
- [140] Michael J Kahn and Amanda Prail. Formal usability inspections. In *Usability inspection methods*, pages 141–171. 1994.
- [141] N Karsh and B Eitam. I control therefore i do: Judgments of agency influence action selection. *Cognition*, 138:122–131, 2015.
- [142] Alan E Kazdin. Imagery elaboration and self-efficacy in the covert modeling treatment of unassertive behavior. *Journal of Consulting and Clinical Psychology*, 47(4):725, 1979.
- [143] Barbara Y Kearney and Barbara J Fleischer. Development of an instrument to measure exercise of self-care agency. *Research in nursing & health*, 2(1):25–34, 1979.
- [144] Herbert C Kelman. Processes of opinion change. *Public opinion quarterly*, 25(1):57–78, 1961.
- [145] Marleen JE Klaassen and Jochen Peter. Gender (in) equality in internet pornography: A content analysis of popular pornographic internet videos. *The Journal of Sex Research*, 52(7):721–735, 2015.
- [146] Christoph Klimmt, Dorothée Hefner, and Peter Vorderer. The video game experience as “true” identification: A theory of enjoyable alterations of players’ self-perception. *Communication theory*, 19(4):351–373, 2009.
- [147] Christoph Klimmt, Dorothée Hefner, Peter Vorderer, Christian Roth, and Christopher Blake. Identification with video game characters as automatic shift of self-perceptions. *Media Psychology*, 13(4):323–338, 2010.

- [148] Christian Knutt. Agdc: Bioware charts writing for mass effect, 2007.
- [149] Barbara Köhler, Juan Haladjian, Blagina Simeonova, and Damir Ismailović. Feedback in low vs. high fidelity visuals for game prototypes. In *2012 Second International Workshop on Games and Software Engineering: Realizing User Engagement with Game Engineering Techniques (GAS)*, pages 42–47. IEEE, 2012.
- [150] Marina Krcmar and Drew P Cingel. Moral foundations theory and moral reasoning in video game play: using real-life morality in a game context. *Journal of Broadcasting & Electronic Media*, 60(1):87–103, 2016.
- [151] Matthew W Kreuter, Victor J Strecher, and Bernard Glassman. One size does not fit all: the case for tailoring print materials. *Annals of behavioral medicine*, 21(4):276, 1999.
- [152] Willemieke Kroeze, Andrea Werkman, and Johannes Brug. A systematic review of randomized trials on the effectiveness of computer-tailored education on physical activity and dietary behaviors. *Annals of behavioral medicine*, 31(3):205–223, 2006.
- [153] Lenelis Kruse, Ernst Weimer, and Franc Wagner. What men and women are said to be: Social representation and language. *Journal of Language and Social Psychology*, 7(3-4):243–262, 1988.
- [154] Ellen J Langer and Judith Rodin. The effects of choice and enhanced personal responsibility for the aged: A field experiment in an institutional setting. *Journal of personality and social psychology*, 34(2):191, 1976.
- [155] Ellen J Langer and Jane Roth. Heads i win, tails it’s chance: The illusion of control as a function of the sequence of outcomes in a purely chance task. *Journal of personality and social psychology*, 32(6):951, 1975.
- [156] Craig Larman. *Agile and iterative development: a manager’s guide*. Addison-Wesley Professional, 2004.
- [157] Paul J Lavrakas. Social desirability. In Paul J Lavrakas, editor, *Encyclopedia of Survey Research Methods*, pages 825–826. Sage Publications, Thousand Oaks, California, USA, 2008.
- [158] Joseph E LeDoux and Jack M Gorman. A call to action: overcoming anxiety through active coping. *American Journal of Psychiatry*, 158(12):1953–1955, 2001.
- [159] Lauren A Leotti and Mauricio R Delgado. The inherent reward of choice. *Psychological science*, 22(10):1310–1318, 2011.
- [160] Lauren A Leotti and Mauricio R Delgado. The value of exercising control over monetary gains and losses. *Psychological science*, 25(2):596–604, 2014.
- [161] Lauren A Leotti, Sheena S Iyengar, and Kevin N Ochsner. Born to choose: The origins and value of the need for control. *Trends in cognitive sciences*, 14(10):457–463, 2010.
- [162] B Lewis-Evans. finding out what they think: A rough primer to user research, part 1. gamasutra, 2012.
- [163] Dong Dong Li, Albert Kien Liau, and Angeline Khoo. Player–avatar identification in video gaming: Concept and measurement. *Computers in Human Behavior*, 29(1):257–263, 2013.
- [164] Matthew Lombard, Robert D Reich, Maria E Grabe, Cheryl C Bracken, and Theresa B Ditton. Presence and television. the role of screen size. *Human communication research*, 26(1):75–98, 2000.
- [165] C Neil Macrae, Charles Stangor, and Miles Hewstone. *Stereotypes and stereotyping*. Guilford Press, 1996.

- [166] Steven F Maier and Martin E Seligman. Learned helplessness: Theory and evidence. *Journal of experimental psychology: general*, 105(1):3, 1976.
- [167] Regan L Mandryk and M Stella Atkins. A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. *International journal of human-computer studies*, 65(4):329–347, 2007.
- [168] Regan L Mandryk, Kori M Inkpen, and Thomas W Calvert. Using psychophysiological techniques to measure user experience with entertainment technologies. *Behaviour & information technology*, 25(2):141–158, 2006.
- [169] Winter Mason and Siddharth Suri. Conducting behavioral research on amazon’s mechanical Turk. *Behavior research methods*, 44(1):1–23, 2012.
- [170] Michael Mateas and Andrew Stern. Integrating plot, character and natural language processing in the interactive drama façade. In *Proceedings of the 1st International Conference on Technologies for Interactive Digital Storytelling and Entertainment (TIDSE-03)*, volume 2, 2003.
- [171] Michael Mateas and Andrew Stern. Integrating Plot, Character and Natural Language Processing in the Interactive Drama Façade. In *Proceedings of the Technologies for Interactive Digital Storytelling and Entertainment (TIDSE) Conference*, 2003.
- [172] Kurt Matzler, Martin Waiguny, and Johann Fuller. Spoiled for choice: consumer confusion in internet-based mass customization. *Innovative Marketing*, 3(3):7–18, 2007.
- [173] Victoria McArthur. The ux of avatar customization. In *Proceedings of the 2017 CHI conference on human factors in computing systems*, pages 5029–5033, 2017.
- [174] Edward McAuley, Terry Duncan, and Vance V Tammen. Psychometric properties of the intrinsic motivation inventory in a competitive sport setting: A confirmatory factor analysis. *Research quarterly for exercise and sport*, 60(1):48–58, 1989.
- [175] Allen R McConnell and Russell H Fazio. Women as men and people: Effects of gender-marked language. *Personality and Social Psychology Bulletin*, 22(10):1004–1013, 1996.
- [176] Linda L McCroskey, James C McCroskey, and Virginia P Richmond. Analysis and improvement of the measurement of interpersonal attraction and homophily. *Communication Quarterly*, 54(1):1–31, 2006.
- [177] Tulay Tetiker McNally. Part 1: The end of the dark ages for qa in game development, 2013.
- [178] Adam W Meade and S Bartholomew Craig. Identifying careless responses in survey data. *Psychological methods*, 17(3):437, 2012.
- [179] Elisa D Mekler, Julia Ayumi Bopp, Alexandre N Tuch, and Klaus Opwis. A systematic review of quantitative studies on the enjoyment of digital entertainment games. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 927–936, 2014.
- [180] Silent Hill Memories. Silent hill: Shattered memories - behind the psych profile. <https://www.youtube.com/watch?v=OszjLlxBI9c>, 2013. Accessed: 2020-07-13.
- [181] Pejman Mirza-Babaei, Lennart E Nacke, John Gregory, Nick Collins, and Geraldine Fitzpatrick. How does it play better? exploring user testing and biometric storyboards in games user research. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1499–1508, 2013.
- [182] Oliver Moore. Woman’s call to end video game misogyny sparks vicious online attacks. *The Globe and Mail*, 2012.

- [183] David Morley. *Television, audiences and cultural studies*. Routledge, 2003.
- [184] Gwaredd Mountain. Psychology profiling in Silent Hill: Shattered Memories, 2010. Video Presented at the Paris Game/AI Conference.
- [185] JB Mountford. Topic modeling the red pill. *Social Sciences*, 7(3):42, 2018.
- [186] Clifford Nass and Youngme Moon. Machines and mindlessness: Social responses to computers. *Journal of social issues*, 56(1):81–103, 2000.
- [187] Naughty Dog. *The Last of Us*. Sony Computer Entertainment, 2013. Video Game.
- [188] Jeff L Nay and José P Zagal. Meaning without consequence: virtue ethics and inconsequential choices in games. In *Proceedings of the 12th International Conference on the Foundations of Digital Games*, page 14. ACM, 2017.
- [189] Mark W Nelson and William B Tayler. Information pursuit in financial statement analysis: Effects of choice, effort, and reconciliation. *The Accounting Review*, 82(3):731–758, 2007.
- [190] Team Ninja. *Dead or Alive Xtreme Beach Volleyball*. Tecmo, 2003. Video Game.
- [191] Nintendo. *Super Mario Bros*. Nintendo, 1985. Video Game.
- [192] Nintendo. *The Legend of Zelda: Twilight Princess*. Nintendo, 2006. Video Game.
- [193] Kamala O Norris. Gender stereotypes, aggression, and computer games: An online survey of women. *Cyberpsychology & Behavior*, 7(6):714–727, 2004.
- [194] Rockstar North. *Grand Theft Auto V*. Rockstar Games, 2013. Video Game.
- [195] Martha C Nussbaum. Objectification. *Philosophy & Public Affairs*, 24(4):249–291, 1995.
- [196] Mary Beth Oliver, Nicholas David Bowman, Julia K Woolley, Ryan Rogers, Brett I Sherrick, and Mun-Young Chung. Video games as meaningful entertainment experiences. *Psychology of Popular Media Culture*, 5(4):390, 2016.
- [197] Elinor Ostrom. A behavioral approach to the rational choice theory of collective action: Presidential address, american political science association, 1997. *American political science review*, 92(1):1–22, 1998.
- [198] P. Stafford [Polygon]. The dangers of in-game data collection. <https://www.polygon.com/features/2019/5/9/18522937/video-game-privacy-player-data-collection>, 2019. Accessed: 2020-07-28.
- [199] Prasanta K Pattanaik and Yongsheng Xu. On preference and freedom. *Theory and Decision*, 44(2):173–198, 1998.
- [200] Paul Tassi. You’ll Be Surprised What Percent Of ‘Mass Effect’ Players Chose Paragon, feb 2020.
- [201] Richard E Petty, Jamie Barden, and S Christian Wheeler. The elaboration likelihood model of persuasion: Developing health promotions for sustained behavioral change. 2009.
- [202] Richard E Petty and John T Cacioppo. The elaboration likelihood model of persuasion. In *Communication and persuasion*, pages 1–24. Springer, 1986.
- [203] Susan P Phillips. Defining and measuring gender: a social determinant of health whose time has come. *International Journal for Equity in Health*, 4(1):1–4, 2005.
- [204] David Pinelle, Nelson Wong, and Tadeusz Stach. Heuristic evaluation for games: usability principles for video game design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1453–1462, 2008.

- [205] Pixel Crushers. Dialogue system for unity.
- [206] Joseph H Pleck, Freya L Sonenstein, and Leighton C Ku. Attitudes toward male roles among adolescent males: A discriminant validity analysis. *Sex roles*, 30(7-8):481–501, 1994.
- [207] Andrew K Przybylski, Netta Weinstein, Richard M Ryan, and C Scott Rigby. Having to versus wanting to play: Background and consequences of harmonious versus obsessive engagement in video games. *CyberPsychology & Behavior*, 12(5):485–492, 2009.
- [208] Jessie M Quintero Johnson and Angeline Sangalang. Testing the explanatory power of two measures of narrative involvement: An investigation of the influence of transportation and narrative engagement on the process of narrative persuasion. *Media Psychology*, 20(1):144–173, 2017.
- [209] Sheri Graner Ray. *Gender Inclusive Game Design: Expanding the Market (Advances in Computer Graphics and Game Development Series)*. Charles River Media, Inc., 2003.
- [210] CD Projekt RED. *The Witcher*. Atari, 2007. Video Game.
- [211] Byron Reeves and Clifford Ivar Nass. *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge university press, Cambridge, 1996.
- [212] Cecilia L Ridgeway. *Gender, interaction, and inequality*. Springer, 1992.
- [213] Mark Riedl, David Thue, and Vadim Bulitko. Game ai as storytelling. In *Artificial intelligence for computer games*, pages 125–150. Springer, 2011.
- [214] Mark Riedl, David Thue, and Vadim Bulitko. Game AI as storytelling. In *Artificial Intelligence for Computer Games*. 2011.
- [215] Ryan Rogers, Julia Woolley, Brett Sherrick, Nicholas David Bowman, and Mary Beth Oliver. Fun versus meaningful video game experiences: A qualitative analysis of user responses. *The Computer Games Journal*, 6(1-2):63–79, 2017.
- [216] Richard M Ryan and Edward L Deci. Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology*, 25(1):54–67, 2000.
- [217] Richard M Ryan and Edward L Deci. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist*, 55(1):68, 2000.
- [218] Richard M Ryan and Edward L Deci. Self-regulation and the problem of human autonomy: Does psychology need choice, self-determination, and will? *Journal of personality*, 74(6):1557–1586, 2006.
- [219] Richard M Ryan, C Scott Rigby, and Andrew Przybylski. The motivational pull of video games: A self-determination theory approach. *Motivation and emotion*, 30(4):344–360, 2006.
- [220] Anastasia Salter and Bridget Blodgett. Hypermasculinity & dickwolves: The contentious role of women in the new gaming public. *Journal of broadcasting & electronic media*, 56(3):401–416, 2012.
- [221] Kazuyuki Samejima, Yasumasa Ueda, Kenji Doya, and Minoru Kimura. Representation of action-specific reward values in the striatum. *Science*, 310(5752):1337–1340, 2005.
- [222] Jose Luis González Sánchez, Natalia Padilla Zea, and Francisco L Gutiérrez. From usability to playability: Introduction to player-centred video game development process. In *International Conference on Human Centered Design*, pages 65–74. Springer, 2009.
- [223] Anita Sarkeesian. Tropes vs. women: # 2 women in refrigerators. *Feminist Frequency*, 2011.

- [224] Anita Sarkeesian. Women as background decoration (part 1). *Feminist Frequency*, 2011.
- [225] Benjamin Scheibehenne, Rainer Greifeneder, and Peter M Todd. Can there ever be too many options? a meta-analytic review of choice overload. *Journal of Consumer Research*, 37(3):409–425, 2010.
- [226] Jason Schreler. The messy, true story behind the making of destiny, 2015.
- [227] Karen Schrier. *Ethics and game design: teaching values through play: teaching values through play*. IGI Global, 2010.
- [228] Constantine Sedikides and Michael J Strube. The multiply motivated self. *Personality and Social Psychology Bulletin*, 21(12):1330–1335, 1995.
- [229] Reinhard Sefelin, Manfred Tscheligi, and Verena Giller. Paper prototyping-what is it good for? a comparison of paper-and computer-based low-fidelity prototyping. In *CHI'03 extended abstracts on Human factors in computing systems*, pages 778–779, 2003.
- [230] Deane H Shapiro Jr, Carolyn E Schwartz, and John A Astin. Controlling ourselves, controlling our world: Psychology’s role in understanding positive and negative consequences of seeking and gaining control. *American psychologist*, 51(12):1213, 1996.
- [231] Valerie Shute and Matthew Ventura. *Stealth assessment: Measuring and supporting learning in video games*. The MIT Press, Cambridge, Massachusetts, USA, 2013.
- [232] Paul Skalski, Ron Tamborini, Ashleigh Shelton, Michael Buncher, and Pete Lindmark. Mapping the road to fun: Natural video game controllers, presence, and game enjoyment. *New Media & Society*, 13(2):224–242, 2011.
- [233] Gearbox Software. *Duke Nukem Forever*. 2K Games, 2011. Video Game.
- [234] Christina Hoff Sommers. What critics of gamergate get wrong. *Youtube*, 2014.
- [235] Christina Hoff Sommers. The war on gamers continues. *Youtube*, 2015.
- [236] Janet T Spence and Robert L Helmreich. *Masculinity and femininity: Their psychological dimensions, correlates, and antecedents*. University of Texas Press, 1979.
- [237] Gamasutra Staff. Gdc: See how bioware melded story with systems to make dragon age, 2015.
- [238] Chris Stokel-Walker. Bots on amazon’s mechanical turk are ruining psychology studies, 2018.
- [239] Santa Monica Studio. *God of War*. Sony Computer Entertainment, 2005. Video Game.
- [240] Lionhead Studios. *Fable II*. Microsoft Game Studios, 2008. Video Game.
- [241] PBS Digital Studios. Thoughts on #gamergate. *Youtube*, 2014.
- [242] Shuji Suzuki. Effects of number of alternatives on choice in humans. *Behavioural Processes*, 39(2):205–214, 1997.
- [243] Shuji Suzuki. Selection of forced-and free-choice by monkeys (macaca fascicularis). *Perceptual and motor skills*, 88(1):242–250, 1999.
- [244] Janet K Swim, Robyn Mallett, and Charles Stangor. Understanding subtle sexism: Detection and use of sexist language. *Sex roles*, 51(3-4):117–128, 2004.
- [245] Nurit Tal-Or and Jonathan Cohen. Unpacking engagement: convergence and divergence in transportation and identification. *Annals of the International Communication Association*, 40(1):33–66, 2016.

- [246] Ron Tamborini, Nicholas David Bowman, Allison Eden, Matthew Grizzard, and Ashley Organ. Defining media enjoyment as the satisfaction of intrinsic needs. *Journal of communication*, 60(4):758–777, 2010.
- [247] Chek Tien Tan, Sander Bakkes, and Yusuf Pisan. Inferring player experiences using facial expressions analysis. In *Proceedings of the 2014 Conference on Interactive Entertainment*, pages 1–8, 2014.
- [248] Karen Tanenbaum and Joshua Tanenbaum. Commitment to meaning: a reframing of agency in games. 2009.
- [249] Karen Tanenbaum and Tess Tanenbaum. Commitment to Meaning. A Reframing of Agency in Games. *After Media — Embodiment and Context, Digital Arts and Culture 2009, Arts Computation Engineering, UC Irvine*, 3(2), 2009.
- [250] Adam Tapal, Ela Oren, Reuven Dar, and Baruch Eitam. The sense of agency scale: A measure of consciously perceived control over one’s mind, body, and the immediate environment. *Frontiers in psychology*, 8:1552, 2017.
- [251] Nguyen T Thai and Ulku Yuksel. Too many destinations to visit: Tourists’ dilemma? *Annals of Tourism Research*, 62:38–53, 2017.
- [252] David Thue, Vadim Bulitko, Marcia Spetch, and Trevon Romanuik. A computational model of perceived agency in video games. In *Proceedings of the 7th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, AIIDE 2011*, 2011.
- [253] David Thue, Vadim Bulitko, Marcia Spetch, and Trevon Romanuik. A computational model of perceived agency in video games. In *AIIDE*, 2011.
- [254] Gustavo F Tondello, Karina Arrambide, Giovanni Ribeiro, Andrew Jian-lan Cen, and Lennart E Nacke. “i don’t fit into a single type”: A trait model and scale of game playing preferences. In *IFIP Conference on Human-Computer Interaction*, pages 375–395. Springer, 2019.
- [255] Sabine Trepte and Leonard Reinecke. Avatar creation and video game enjoyment. *Journal of Media Psychology*, 2010.
- [256] Selen Turkyay and Charles K Kinzer. The effects of avatar-based customization on player identification. *International Journal of Gaming and Computer-Mediated Simulations (IJGCMS)*, 6(1):1–25, 2014.
- [257] Maarten van de Kant, Stephanie Wilson, Mathilde Bekker, Hilary Johnson, and Peter Johnson. Patchwork: a software tool for early design. In *CHI 98 Conference Summary on Human Factors in Computing Systems*, pages 221–222, 1998.
- [258] Jan Van Looy, Cédric Courtois, Melanie De Vocht, and Lieven De Marez. Player identification in online games: Validation of a scale for measuring identification in mmogs. *Media Psychology*, 15(2):197–221, 2012.
- [259] Eva A Van Reijmersdal, Jeroen Jansz, Oscar Peters, and Guda Van Noort. Why girls go pink: Game character identification and game-players’ motivations. *Computers in human behavior*, 29(6):2640–2649, 2013.
- [260] Lance Vikaros and Darnel Degand. Moral development through social narratives and game design. In *Ethics and game design: Teaching values through play*, pages 197–215. IGI Global, 2010.
- [261] Lance Vikaros and Darnel Degand. Moral Development through Social Narratives and Game Design. 2010.

- [262] Robert A Virzi, Jeffrey L Sokolov, and Demetrios Karis. Usability problem identification using both low-and high-fidelity prototypes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 236–243, 1996.
- [263] John Von Neumann and Oskar Morgenstern. *Theory of games and economic behavior*, 2nd rev. 1947.
- [264] Stella Vosniadou and William F Brewer. Mental models of the earth: A study of conceptual change in childhood. *Cognitive psychology*, 24(4):535–585, 1992.
- [265] P Warman. Global games market will reach \$102.9 billion in 2017. *Newzoo-games market research*, 2015.
- [266] Diane Watson, Mark Hancock, Regan L Mandryk, and Max Birk. Deconstructing the touch experience. In *Proceedings of the 2013 ACM international conference on Interactive tabletops and surfaces*, pages 199–208, 2013.
- [267] Andrew J Weaver and Nicky Lewis. Mirrored morality: An exploration of moral choice in video games. *Cyberpsychology, Behavior, and Social Networking*, 15(11):610–614, 2012.
- [268] Gerdi Weidner, Ronald Friend, Thomas J Ficarrotto, and Nancy R Mendell. Hostility and cardiovascular reactivity to stress in women and men. *Psychosomatic Medicine*, 1989.
- [269] Allan Wigfield and Jacquelynne S Eccles. Expectancy–value theory of achievement motivation. *Contemporary educational psychology*, 25(1):68–81, 2000.
- [270] Michael E Wiklund, Christopher Thurrott, and Joseph S Dumas. Does the fidelity of software prototypes affect the perception of usability? In *Proceedings of the Human Factors Society Annual Meeting*, volume 36, pages 399–403. SAGE Publications Sage CA: Los Angeles, CA, 1992.
- [271] Dmitri Williams, Nicole Martins, Mia Consalvo, and James D Ivory. The virtual census: Representations of gender, race and age in video games. *New Media & Society*, 11(5):815–834, 2009.
- [272] Will Wright and Ian Bogost. *Persuasive games: The expressive power of videogames*. Mit Press, 2007.
- [273] Brianna Wu. No skin thick enough: The daily harassment of women in the game industry. *Polygon. Np, Web*, 16, 2014.
- [274] Mike Z Yao, Chad Mahood, and Daniel Linz. Sexual priming, gender stereotyping, and likelihood to sexually harass: Examining the cognitive effects of playing a sexually-explicit video game. *Sex roles*, 62(1-2):77–88, 2010.
- [275] Nick Yee, Nicolas Ducheneaut, Han-Tai Shiao, and Les Nelson. Through the azerothian looking glass: Mapping in-game preferences to real world demographics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2811–2814, 2012.
- [276] Yummylee. I have some criticisms for dragon age: Inquisition., 2014.
- [277] Andreas Zecher. Open letter to the gaming community. *Medium*, September, 1, 2014.
- [278] Rolf A Zwaan, Mark C Langston, and Arthur C Graesser. The construction of situation models in narrative comprehension: An event-indexing model. *Psychological science*, 6(5):292–297, 1995.

APPENDIX A

MANUSCRIPT A CONSENT FORM

Title: Game Evaluation

Researcher(s): Jason Bowey, Dr. Regan Mandryk, Department of Computer Science, University of Saskatchewan, 306-966-2327, jason.bowey@usask.ca

Purpose(s) and Objective(s) of the Research: The purpose of this project is to evaluate a game currently in development.

Procedures:

- In this study, you will be playing through a short game. After, you will be asked to answer a few questions about yourself and your experience playing the game.
- This study will take approximately 30 minutes to complete.

Funded by: The Natural Sciences and Engineering Research Council of Canada (NSERC).

Potential Risks and Benefits: There are no known or anticipated risks to you by participating in this research.

Confidentiality:

- Confidentiality will be maintained throughout the study. The entire process and data will be anonymized. Data will only be presented in the aggregate and any individual user comments will be anonymized prior to presentation in academic venues.
- Only the principle researcher and her research assistants will have access to the data to ensure that your confidentiality is protected.
- Storage of Data:
 - Data (including survey and interview responses, logs of computer use, and videos of interaction) will be stored on a secure password-protected server for 7 years after data collection.
 - After 7 years, the data will be destroyed. Paper data will be shredded and digital data will be wiped from hard disks beyond any possibility for data recovery.

Right to Withdraw:

- Your participation is voluntary. You may withdraw from the research project for any reason, at any time without explanation.
- Should you wish to withdraw, you may do so at any point, and we will not use your data; we will destroy all records of your data.
- Your right to withdraw data from the study will apply until the data have been aggregated (one week after study completion). After this date, it is possible that some form of research dissemination will have already occurred and it may not be possible to withdraw your data.

Follow up: To obtain results from the study, please contact Jason Bowey (jason.bowey@usask.ca)

Questions or Concerns:

- Contact the researcher(s) using the information at the top.
- This research project has been approved on ethical grounds by the University of Saskatchewan Research Ethics Board. Any questions regarding your rights as a participant may be addressed to that committee through the Research Ethics Office ethics.office@usask.ca (306) 966-2975. Out of town participants may call toll free (888) 966-2975.

APPENDIX B

MANUSCRIPT B CONSENT FORM

Title: Early Game Evaluation

Researcher(s): Jason Bowey, Dr. Regan Mandryk, Department of Computer Science, University of Saskatchewan, 306-966-2327, jason.bowey@usask.ca

Purpose(s) and Objective(s) of the Research: The purpose of this project is to get early feedback and suggestions for the narrative of a game currently under development.

Procedures:

- In this study you will play an early version of a game narrative and answer some questions about your overall experience.
- this study will take approximately 30 minutes to complete.

Funded by: The Natural Sciences and Engineering Research Council of Canada (NSERC).

Potential Risks and Benefits: There are no known or anticipated risks to you by participating in this research.

Confidentiality:

- Confidentiality will be maintained throughout the study. The entire process and data will be anonymized. Data will only be presented in the aggregate and any individual user comments will be anonymized prior to presentation in academic venues.
- Only the principle researcher and her research assistants will have access to the data to ensure that your confidentiality is protected.
- Storage of Data:
 - Data (including survey and interview responses, logs of computer use, and videos of interaction) will be stored on a secure password-protected server for 7 years after data collection.
 - After 7 years, the data will be destroyed. Paper data will be shredded and digital data will be wiped from hard disks beyond any possibility for data recovery.

Right to Withdraw:

- Your participation is voluntary. You may withdraw from the research project for any reason, at any time without explanation.
- Should you wish to withdraw, you may do so at any point, and we will not use your data; we will destroy all records of your data.
- Your right to withdraw data from the study will apply until the data have been aggregated (one week after study completion). After this date, it is possible that some form of research dissemination will have already occurred and it may not be possible to withdraw your data.

Follow up: To obtain results from the study, please contact Jason Bowey (jason.bowey@usask.ca)

Questions or Concerns:

- Contact the researcher(s) using the information at the top.
- This research project has been approved on ethical grounds by the University of Saskatchewan Research Ethics Board. Any questions regarding your rights as a participant may be addressed to that committee through the Research Ethics Office ethics.office@usask.ca (306) 966-2975. Out of town participants may call toll free (888) 966-2975.

APPENDIX C

MANUSCRIPT C CONSENT FORM

Title: Early Game Evaluation

Researcher(s): Jason Bowey, Dr. Regan Mandryk, Department of Computer Science, University of Saskatchewan, 306-966-2327, jason.bowey@usask.ca

Purpose(s) and Objective(s) of the Research: The purpose of this project is to get early feedback and suggestions for the narrative of a game currently under development.

Procedures:

- In this study you will play an early version of a game narrative and answer some questions about your overall experience.
- this study will take approximately 30 minutes to complete.

Funded by: The Natural Sciences and Engineering Research Council of Canada (NSERC).

Potential Risks and Benefits: There are no known or anticipated risks to you by participating in this research.

Confidentiality:

- Confidentiality will be maintained throughout the study. The entire process and data will be anonymized. Data will only be presented in the aggregate and any individual user comments will be anonymized prior to presentation in academic venues.
- Only the principle researcher and her research assistants will have access to the data to ensure that your confidentiality is protected.
- Storage of Data:
 - Data (including survey and interview responses, logs of computer use, and videos of interaction) will be stored on a secure password-protected server for 7 years after data collection.
 - After 7 years, the data will be destroyed. Paper data will be shredded and digital data will be wiped from hard disks beyond any possibility for data recovery.

Right to Withdraw:

- Your participation is voluntary. You may withdraw from the research project for any reason, at any time without explanation.
- Should you wish to withdraw, you may do so at any point, and we will not use your data; we will destroy all records of your data.
- Your right to withdraw data from the study will apply until the data have been aggregated (one week after study completion). After this date, it is possible that some form of research dissemination will have already occurred and it may not be possible to withdraw your data.

Follow up: To obtain results from the study, please contact Jason Bowey (jason.bowey@usask.ca)

Questions or Concerns:

- Contact the researcher(s) using the information at the top.
- This research project has been approved on ethical grounds by the University of Saskatchewan Research Ethics Board. Any questions regarding your rights as a participant may be addressed to that committee through the Research Ethics Office ethics.office@usask.ca (306) 966-2975. Out of town participants may call toll free (888) 966-2975.

APPENDIX D

MANUSCRIPT D CONSENT FORM

Title: Interactive Narrative

Researcher(s): Jason Bowey, Dr. Regan Mandryk, Department of Computer Science, University of Saskatchewan, 306-966-2327, jason.bowey@usask.ca

Purpose(s) and Objective(s) of the Research: The purpose of this study is to better understand how players make decisions in interactive narratives.

Procedures:

- In this study you will play through an interactive narrative and answer questions about your experience.
- this study will take approximately 10 minutes to complete.

Funded by: The Natural Sciences and Engineering Research Council of Canada (NSERC).

Potential Risks and Benefits: There are no known or anticipated risks to you by participating in this research.

Confidentiality:

- Confidentiality will be maintained throughout the study. The entire process and data will be anonymized. Data will only be presented in the aggregate and any individual user comments will be anonymized prior to presentation in academic venues.
- Only the principle researcher and her research assistants will have access to the data to ensure that your confidentiality is protected.
- Storage of Data:
 - Data (including survey and interview responses, logs of computer use, and videos of interaction) will be stored on a secure password-protected server for 7 years after data collection.
 - After 7 years, the data will be destroyed. Paper data will be shredded and digital data will be wiped from hard disks beyond any possibility for data recovery.

Right to Withdraw:

- Your participation is voluntary. You may withdraw from the research project for any reason, at any time without explanation.
- Should you wish to withdraw, you may do so at any point, and we will not use your data; we will destroy all records of your data.
- Your right to withdraw data from the study will apply until the data have been aggregated (one week after study completion). After this date, it is possible that some form of research dissemination will have already occurred and it may not be possible to withdraw your data.

Follow up: To obtain results from the study, please contact Jason Bowey (jason.bowey@usask.ca)

Questions or Concerns:

- Contact the researcher(s) using the information at the top.

- This research project has been approved on ethical grounds by the University of Saskatchewan Research Ethics Board. Any questions regarding your rights as a participant may be addressed to that committee through the Research Ethics Office ethics.office@usask.ca (306) 966-2975. Out of town participants may call toll free (888) 966-2975.

APPENDIX E

AMBIVALENT SEXISM INVENTORY QUESTIONNAIRE ITEMS

Protective Paternalism
A good woman should be set on a pedestal.
Women should be cherished and protected by men.
Men should sacrifice to provide for women.
In a disaster, women need not be rescued first. (-)
Complementary Gender Differentiation
Women have a superior moral sensibility.
Women have a quality of purity few men possess.
Women have a more refined sense of culture, taste.
Heterosexual Intimacy
Every man ought to have a woman he adores.
Men are complete without women. (-)
Despite accomplishment, men are incomplete without women.
People are often happy without heterosexual romance. (-)

Table E.1: Individual questionnaire items from the Ambivalent Sexism Inventory. Questionnaire presented in a Likert scale ranging from *Strongly Disagree* to *Strongly Agree*. Reverse-coded questionnaire items indicated by (-).

APPENDIX F

HOMOPHILY QUESTIONNAIRE ITEMS

Homophily

This character thinks like me.
This character behaves like me.
This character has thoughts and ideas that are similar to mine.
This character is like me.
This character shares my values.
This character has a lot in common with me.
This character is similar to me.

Table F.1: Individual questionnaire items from the Homophily questionnaire. Questionnaire presented in a Likert scale ranging from *Strongly Disagree* to *Strongly Disagree*.

APPENDIX G

INTRINSIC MOTIVATION INVENTORY QUESTIONNAIRE ITEMS

Enjoyment
I enjoyed the game very much.
Playing the game was fun.
I would describe this game as very interesting.
While playing the game, I was thinking about how much I enjoyed it.
This game did not hold my attention. (-)
Effort
I put a lot off effort into this game.
It was important to me to do well at this game.
I tried very hard at playing the game.
I didn't try very hard at playing the game. (-)
Tension
I felt tense while playing the game.
I felt pressured while playing the game.
I was anxious while playing the game.
I was very relaxed while playing the game. (-)
Competence
I think I am pretty good at the game.
I am satisfied with my performance at this game.
After playing the game for a while, I felt pretty competent.
I am pretty skilled at the game.
I couldn't play this game very well. (-)

Table G.1: Individual questionnaire items from the Intrinsic Motivation Inventory questionnaire. Questionnaire presented in a Likert scale ranging from *Strongly Disagree* to *Strongly Agree*. Reverse-coded items indicated by (-).

APPENDIX H

MALE ROLE ATTITUDES QUESTIONNAIRE ITEMS

Male Role Attitudes

It is essential for a guy to get respect from others.
A man always deserves the respect of his wife and children.
I admire a guy who is totally sure of himself.
A guy will lose respect if he talks about his problems.
A young man should be physically tough, even if he's not big.
It bothers me when a guy acts like a girl.
I don't think a husband should have to do housework.
Men are always ready for sex.

Table H.1: Individual questionnaire items from the Male Role Attitudes questionnaire. Questionnaire items presented in a Likert scale ranging from *Strongly Disagree* to *Strongly Agree*.

APPENDIX I

MORAL FOUNDATIONS QUESTIONNAIRE

Harm/Care
Whether or not someone suffered emotionally.*
Whether or not someone cared for someone weak or vulnerable.*
Compassion for those who are suffering is the most crucial virtue.
One of the worst things a person could do is hurt a defenseless animal.
Fairness
Whether or not people were treated differently from others.*
Whether or not someone acted unfairly.*
When the government makes laws, the number one principle should be ensuring that everyone is treated fairly.
Justice is the most important requirement for a society.
Loyalty
Whether or not someone's actions showed love for his or her country.*
Whether or not someone did something to betray his or her group.*
I am proud of my country's history.
People should be loyal to their family members, even when they have done something wrong.
Respect
Whether or not someone showed a lack of respect for authority.*
Whether or not someone conformed to the traditions of society.*
Respect for authority is something all children need to learn.
Purity
Whether or not someone violated standards of purity and decency.*
Whether or not someone did something disgusting.*
People should not do things that are disgusting, even if no one is harmed.
I would call some acts wrong on the grounds that they are unnatural.

Table I.1: Individual questionnaire items from the Moral Foundations Questionnaire. Questionnaire items marked with * were presented in a Likert scale ranging from *Not at all Relevant* to *Extremely Relevant*. Otherwise, items presented on Likert scale ranging from *Strongly Disagree* to *Strongly Agree*.

APPENDIX J

NARRATIVE ENGAGEMENT QUESTIONNAIRE ITEMS

Narrative Understanding
At points, I had a hard time making sense of what was going on in the game. (-)
My understanding of the characters is unclear. (-)
I had a hard time recognizing the thread of the story. (-)
Attentional Focus
I found my mind wandering while playing the game. (-)
While playing the game I found myself thinking about other things. (-)
I had a hard time keeping my mind on the game. (-)
Narrative Presence
During the game, my body was in the room, but my mind was inside the world created by the game.
The game created a new world, and then that world suddenly disappeared when the game ended.
At times during the program, the story was closer to me than the real world.
Emotional Engagement
The story affected me emotionally.
During the game, when a main character succeeded, I felt happy, and when they suffered in some way I felt sad.
I felt sorry for some of the characters in the game.

Table J.1: Individual questionnaire items from the Narrative Engagement questionnaire. Questionnaire items presented on Likert scale ranging from *Strongly Disagree* to *Strongly Agree*. Reverse-coded items indicated by (-).

APPENDIX K

PLAYER EXPERIENCE OF NEEDS SATISFACTION QUESTIONNAIRE ITEMS

Competence
I feel competent at the game.
I feel very capable and effective when playing.
My ability to play the game is well matched with the game’s challenges.
Autonomy
The game provides me with interesting options and choices.
The game lets you do interesting things.
I experienced a lot of freedom in the game.
Relatedness
I find the relationships I form in this game fulfilling.
I find the relationships I form in this game important.
I don’t feel close to other players. (-)
Presence
When playing the game, I feel transported to another time and place.
When moving through the game world I feel as if I am actually there.
I am not impacted emotionally by events in the game. (-)
The game was emotionally engaging.
I experience feelings as deeply in the game as I have in real life.
When playing the game I feel as if I was part of the story.
When I accomplish something in the game I feel as if I was part of the story.
I had reactions to events and characters in the game as if they were real.
Intuitive Controls
Learning the game controls was easy.
The game controls are intuitive.
When I wanted to do something in the game, it was easy to remember the corresponding control.

Table K.1: Individual questionnaire items from the Player Experience of Needs Satisfaction questionnaire. Questionnaire items presented on Likert scale ranging from *Strongly Disagree* to *Strongly Agree*. Reverse-coded items indicated by (-).

APPENDIX L

PLAYER IDENTIFICATION SCALE QUESTIONNAIRE ITEMS

Similarity Identification

- My character was like me in many ways.
- My character resembled me.
- I identified with my character.
- My character was an extension of myself.
- My character was similar to me.
- I resembled my character.

Embodied Presence

- When I was playing, it felt as if I was my character.
- I felt like I was inside my character when playing.
- In the game, it was as if I became one with my character.
- When I was playing I was transported into my character.
- When playing, it felt as if my character's body became my own.
- In the game, it was as if I acted directly through my character.

Wishful Identification

- If I could become my character, I would.
 - I would like to be more like my character.
 - My character is an example to me.
 - My character is a better me.
 - My character has characteristics that I would like to have.
-

Table L.1: Individual questionnaire items from the Player Identification questionnaire. Questionnaire items presented on Likert scale ranging from *Strongly Disagree* to *Strongly Agree*.

APPENDIX M

SENSE OF AGENCY QUESTIONNAIRE ITEMS

Sense of Positive Agency

I am in full control of what I do.
I am the author of my actions.
Things I do are subject only to my free will.
The decision whether and when to act is within my hands.
My behaviour is planned by me from the very beginning to the very end.
I am completely responsible for everything that results from my actions.

Sense of Negative Agency

I am just an instrument in the hands of somebody or something else.
My actions just happen without my intention.
The consequences of my actions feel like they don't logically follow my actions.
My movements are automatic - my body simply makes them.
The outcomes of my actions generally surprise me.
Nothing I do is actually voluntary.
While I am in action, I feel like I am a remote controlled robot.

Table M.1: Individual questionnaire items from the Sense of Agency questionnaire. Questionnaire items presented on Likert scale ranging from *Strongly Disagree* to *Strongly Agree*.

APPENDIX N

TRANSPORTABILITY QUESTIONNAIRE ITEMS

Transportability

- I can easily envision the events in the story.
 - I find I can easily lose myself in the story.
 - I find it difficult to tune out activity around me. (-)
 - I can easily envision myself in the events described in a story.
 - I get mentally involved in the story.
 - I can easily put stories out of my mind after I've finished reading them. (-)
 - I sometimes feel as if I am part of the story.
 - I am often impatient to find out how the story ends.
 - I find that I can easily take the perspective of the character(s) in the story.
 - I am often emotionally affected by what I've read.
 - I have vivid images of the characters.
 - I find myself accepting events that I might have otherwise considered unrealistic.
 - I find myself thinking what the characters may be thinking.
 - I find myself thinking of other ways the story could have ended.
 - My mind often wanders. (-)
 - I find myself feeling what the characters may feel.
 - I find that events in the story are relevant to my everyday life.
 - I often find that reading stories has an impact on the way I see things.
 - I easily identify with characters in the story.
 - I have vivid images of the events in the story.
-

Table N.1: Individual questionnaire items from the Transportability questionnaire. Questionnaire items presented on Likert scale ranging from *Strongly Disagree* to *Strongly Agree*. Reverse-coded items indicated by (-).

APPENDIX O

TRANSPORTATION QUESTIONNAIRE ITEMS

Transportation

While I was interacting with the system, I could easily picture the events taking place.

While I was interacting with the system, activity going on in the room around me was on my mind. (-)

I could picture myself in the scene of the events described in the narrative.

I was mentally involved in the narrative while playing it.

After finishing the narrative, I found it easy to put it out of my mind. (-)

I wanted to learn how to narrative ended.

The narrative affected me emotionally.

I found myself thinking of ways the narrative could have turned out differently.

I found my mind wandering while reading the narrative. (-)

The events in the narrative are relevant to my everyday life.

The events in the narrative have changed my life.

Table O.1: Individual questionnaire items from the Transportation questionnaire. Questionnaire items presented on Likert scale rangind from *Strongly Disagree* to *Strongly Agree*. Reverse-coded items indicated by (-).