Predicting Beliefs from NPC Dialogues

Jason T Bowey
Interaction Lab
University of Saskatchewan
Saskatoon, Canada
jason.bowey@usask.ca

Julian Frommel
Interaction Lab
University of Saskatchewan
Saskatoon, Canada
julian.frommel@usask.ca

Brandon Piller
Interaction Lab
University of Saskatchewan
Saskatoon, Canada
brandon.piller@usask.ca

Regan L Mandryk
Interaction Lab
University of Saskatchewan
Saskatoon, Canada
regan.mandryk@usask.ca

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.

Index Terms—questionnaires, surveys, GUR, games user research, dialogues, modeling, beliefs

I. 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 [1]–[5]. 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 [2], [3]. This approach is already being used in some commercial games, such as *Until Dawn* [3].

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 [3], there might be barriers to measuring beliefs. Players may consider their beliefs in light of social desirability [6] 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 [7], 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 [8], 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 choice 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.

II. Related Work

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

A. In-game Assessments: Dialogue and Choices in Games

Earlier work has suggested that assessment can be embedded in dialogues with non-player characters [2]. By presenting dialogue options that are mapped to questionnaire items, it is possible to include assessment in an unobtrusive way [2], while problematic reducing interruptions in gameplay [9]. While studies have not consistently showed that such an approach is beneficial in terms of player experience [3], 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 [2], [9]. 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 [2], [3]. 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 [10] as a trait that can help explain the players' enjoyment of in-game successes and failures [11]. 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 [3]. 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.

B. 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 [12], a theory that suggests that humans interact with virtual entities similar to how they interact with other humans [13], [14].

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 [15], [16]. When players make a choice, they create a commitment on behalf of the player character [17], 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 [18]–[20]. Dialogue choices that align with a player's playstyle have been shown to improve experience [21]. 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 [22], 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 [23]. 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 [23]. 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 [6]. 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.

C. 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 [24]–[26]. When women are portrayed in games, it is often in a sexualized way, with revealing clothing, unrealistic body images, or partial nudity [24], [25], [27]. This has led to strong criticism of the video game industry for catering to straight male players [28]–[30]. Research on sexualized imagery indicates men show a greater likelihood to harass women after play, usually with sexist remarks [31], [32]. Research on the female experience with sexualized imagery suggests they feel diminished self-efficacy [33] and increased self objectification after play [28], [34]. These effects have led to the conclusion that women may self select out of these environments [28] and reinforce the idea of games being dominated by masculine disclosure [35].

Sexism is typically conceptualized as hostility towards women, i.e. as hostile sexism [36]. 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 [36]. 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 [36].

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 [37], 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.

D. Gender and Sexism in Avatar Identification

Avatar identification facilitates games enjoyment, and it has been suggested that sexist characters can aid in avatar identification [38], [39]. Bowey et al. [8] 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.

E. 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.

III. 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.

A. Game/Narrative

We presented dialogue to our players using the same system used in a previous publication [21]. 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 [8]. 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.

B. 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 I for the complete direct dialogue derived from the Male Role Attitudes Scale [37]. 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 II for the complete indirect dialogue derived from the Male Role Attitudes Scale [37].

C. 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

TABLE I

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 II

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.

demographics questionnaire, the Ambivalent Sexism Inventory (ASI, [36]), and the Male Role Attitudes Scale (MRAS, [37]). 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., MRAS, PP, CGD, and HI, $N = 0$). 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 [8] 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 participant in the main study. In total, we were left with 253 potential participants.

D. 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 [40], which measures their susceptibility to be transported into a narrative world.

After playing through the game, participants completed the Homophily questionnaire [41], which measures the perceived similarity in values to the main character, and the Narrative Engagement questionnaire [42], which divides narrative experience into narrative understanding, attentional focus, narrative presence, and emotional engagement.

TABLE III
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 IV
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		
	β	p	β	p	β	p	R^2
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

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 questionnaires.

E. 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 been seen in Table III.

IV. RESULTS

We analyze the data to answer our main research questions.

A. 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 [43], 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.

B. 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 IV) 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.

C. 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.

V. DISCUSSION

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

A. Summary of Results

1) 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 [44], [45] and involves a loss of self-awareness and its replacement with "emotional and cognitive connections with a character" [46]. This would suggest that players' responses are suitable for predicting real-world beliefs.

2) 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 [2], [3] and extend previous findings into the context of real-world sexist beliefs.

3) 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.

4) 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.

B. 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 [47] or Geralt from The Witcher series [48]. 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 [8]. 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.

C. 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 [8], and they are measures that people tend to interpret more subjectively than measures such as morality, which have also been used in previous work [21].

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 [49], Tomb Raider [50]).

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.

D. 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 [51]–[53]. 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 [52]. 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 [2], [9], [54], [55], 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 [3]. 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.

E. 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).

VI. 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.

References

[1] E. D. Mekler, J. A. Bopp, A. N. Tuch, and K. 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*, 2014, pp. 927–936.

[2] J. Frommel, F. Fischbach, K. Rogers, and M. 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*, ser. CHI PLAY '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 163–171. [Online]. Available: https://doi.org/10.1145/3242671.3242682

[3] J. Frommel, C. Phillips, and R. 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*, ser. CHI '21. Virtual: ACM, 2021.

[4] G. Mountain, "Psychology profiling in Silent Hill: Shattered Memories," 2010, video Presented at the Paris Game/AI Conference.

[5] S. H. Memories, "Silent hill: Shattered memories - behind the psych profile," https://www.youtube.com/watch?v=OszjLIxbI9c, 2013, accessed: 2020-07-13.

[6] P. J. Lavrakas, "Social desirability," in *Encyclopedia of Survey Research Methods*, P. J. Lavrakas, Ed. Thousand Oaks, California, USA: Sage Publications, 2008, pp. 825–826.

[7] J. Huizinga, *Homo ludens*. Editora Perspectiva SA, 2020.

[8] J. T. Bowey, A. E. Depping, and R. 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*, 2017, pp. 1530–1543.

[9] J. Frommel, K. Rogers, J. Brich, D. Besserer, L. Bradatsch, I. Ortinau, R. Schabenberger, V. Riemer, C. Schrader, and M. 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*, ser. CHI PLAY '15.   New York, NY, USA: Association for Computing Machinery, 2015, p. 359–368. [Online]. Available: https://doi.org/10.1145/2793107.2793130

[10] G. F. Tondello, K. Arrambide, G. Ribeiro, A. J.-l. Cen, and L. 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*.   Springer, 2019, pp. 375–395.

[11] J. Frommel, M. Klarkowski, and R. 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)*, ser. FDG '21.   Virtual: ACM, 2021.

[12] B. Reeves and C. I. Nass, *The media equation: How people treat computers, television, and new media like real people and places.*   Cambridge: Cambridge university press, 1996.

[13] C. Nass and Y. Moon, "Machines and mindlessness: Social responses to computers," *Journal of social issues*, vol. 56, no. 1, pp. 81–103, 2000.

[14] D. Johnson and J. Gardner, "Effects of team-based computer interaction: The media equation and game design considerations," in *Entertainment Computing - ICEC 2005*, F. Kishino, Y. Kitamura, H. Kato, and N. Nagata, Eds.   Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 468–479.

[15] D. Gibson and K. Schrier, *Ethics and Game Design*, 2010.

[16] L. Vikaros and D. Degand, "Moral Development through Social Narratives and Game Design," 2010.

[17] K. Tanenbaum and T. 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*, vol. 3, no. 2, 2009.

[18] D. Thue, V. Bulitko, M. Spetch, and T. 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.

[19] M. Mateas and A. 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.

[20] M. Riedl, D. Thue, and V. Bulitko, "Game AI as storytelling," in *Artificial Intelligence for Computer Games*, 2011.

[21] J. T. Bowey, M. A. Friehs, and R. 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*, 2019, pp. 1–11.

[22] Bioware, *Mass Effect*.   Bioware, 2007, video Game.

[23] Paul Tassi, "You'll Be Surprised What Percent Of 'Mass Effect' Players Chose Paragon," feb 2020.

[24] E. Downs and S. L. Smith, "Keeping abreast of hypersexuality: A video game character content analysis," *Sex roles*, vol. 62, no. 11-12, pp. 721–733, 2010.

[25] J. D. Ivory, "Still a man's game: Gender representation in online reviews of video games," *Mass Communication & Society*, vol. 9, no. 1, pp. 103–114, 2006.

[26] D. Williams, N. Martins, M. Consalvo, and J. D. Ivory, "The virtual census: Representations of gender, race and age in video games," *New Media & Society*, vol. 11, no. 5, pp. 815–834, 2009.

[27] M. C. Burgess, S. P. Stermer, and S. R. Burgess, "Sex, lies, and video games: The portrayal of male and female characters on video game covers," *Sex roles*, vol. 57, no. 5-6, pp. 419–433, 2007.

[28] J. Fox, J. N. Bailenson, and L. Tricase, "The embodiment of sexualized virtual selves: The proteus effect and experiences of self-objectification via avatars," *Computers in Human Behavior*, vol. 29, no. 3, pp. 930–938, 2013.

[29] K. 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*, vol. 18, no. 4, pp. 261–276, 2012.

[30] A. Sarkeesian, "Women as background decoration (part 1)," *Feminist Frequency*, 2011.

[31] K. E. Dill, B. P. Brown, and M. A. Collins, "Effects of exposure to sex-stereotyped video game characters on tolerance of sexual harassment," *Journal of Experimental Social Psychology*, vol. 44, no. 5, pp. 1402–1408, 2008.

[32] M. Z. Yao, C. Mahood, and D. Linz, "Sexual priming, gender stereotyping, and likelihood to sexually harass: Examining the cognitive effects of playing a sexually-explicit video game," *Sex roles*, vol. 62, no. 1-2, pp. 77–88, 2010.

[33] E. Behm-Morawitz and D. Mastro, "The effects of the sexualization of female video game characters on gender stereotyping and female self-concept," *Sex roles*, vol. 61, no. 11-12, pp. 808–823, 2009.

[34] J. Fox and W. Y. Tang, "Sexism in online video games: The role of conformity to masculine norms and social dominance orientation," *Computers in Human Behavior*, vol. 33, pp. 314–320, 2014.

[35] A. Salter and B. Blodgett, "Hypermasculinity & dickwolves: The contentious role of women in the new gaming public," *Journal of broadcasting & electronic media*, vol. 56, no. 3, pp. 401–416, 2012.

[36] P. Glick and S. T. Fiske, "The ambivalent sexism inventory: Differentiating hostile and benevolent sexism." *Journal of personality and social psychology*, vol. 70, no. 3, p. 491, 1996.

[37] J. H. Pleck, F. L. Sonenstein, and L. C. Ku, "Attitudes toward male roles among adolescent males: A discriminant validity analysis," *Sex roles*, vol. 30, no. 7-8, pp. 481–501, 1994.

[38] D. Hefner, C. Klimmt, and P. Vorderer, "Identification with the player character as determinant of video game enjoyment," in *International conference on entertainment computing*.   Springer, 2007, pp. 39–48.

[39] J. Jansz, "The emotional appeal of violent video games for adolescent males," *Communication theory*, vol. 15, no. 3, pp. 219–241, 2005.

[40] S. Dal Cin, M. P. Zanna, and G. T. Fong, "Narrative persuasion and overcoming resistance," *Resistance and persuasion*, vol. 2, pp. 175–191, 2004.

[41] L. L. McCroskey, J. C. McCroskey, and V. P. Richmond, "Analysis and improvement of the measurement of interpersonal attraction and homophily," *Communication Quarterly*, vol. 54, no. 1, pp. 1–31, 2006.

[42] R. Busselle and H. Bilandzic, "Measuring narrative engagement," *Media Psychology*, vol. 12, no. 4, pp. 321–347, 2009.

[43] A. J. Weaver and N. Lewis, "Mirrored morality: An exploration of moral choice in video games," *Cyberpsychology, Behavior, and Social Networking*, vol. 15, no. 11, pp. 610–614, 2012.

[44] C. Klimmt, D. Hefner, and P. Vorderer, "The video game experience as "true" identification: A theory of enjoyable alterations of players' self-perception," *Communication theory*, vol. 19, no. 4, pp. 351–373, 2009.

[45] C. Klimmt, D. Hefner, P. Vorderer, C. Roth, and C. Blake, "Identification with video game characters as automatic shift of self-perceptions," *Media Psychology*, vol. 13, no. 4, pp. 323–338, 2010.

[46] J. Cohen, "Defining identification: A theoretical look at the identification of audiences with media characters," *Mass communication & society*, vol. 4, no. 3, pp. 245–264, 2001.

[47] Naughty Dog, *The Last of Us*.   Sony Computer Entertainment, 2013, video Game.

[48] CD Projekt RED, *The Witcher*.   Atari, 2007, video Game.

[49] Epic Games, *Gears of War*.   Microsoft Game Studios, 2006, video Game.

[50] Core Design, *Tomb Raider*.   Edits Intractive, 1996, video Game.

[51] 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.

[52] 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.

[53] 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.

[54] V. Shute and M. Ventura, *Stealth assessment: Measuring and supporting learning in video games*.   Cambridge, Massachusetts, USA: The MIT Press, 2013.

[55] J. Frommel, V. Sagl, A. E. Depping, C. Johanson, M. K. Miller, and R. 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*, 2020, pp. 1–16.