Structure, Agency, and Intent: Preliminary Data Collection

Cory Siler[1], Stephen G. Ware[1]

[1]Narrative Intelligence Lab, Department of Computer Science, University of Kentucky, Lexington, KY, USA 40506

Abstract

We discuss the first phase of a multi-year effort to construct a dataset based on an interactive narrative exercise between pairs of humans. One human took the role of the player in a hypertext-based game, while the other took the role of a game master controlling the non-player characters. Alongside gameplay logs, we collected users' explanations for their own choices and for each other's choices, and their ratings for the quality of story structure and personal agency. Through repeated iterations of this process, we hope to support the design and testing of AI experience managers that maximize structure and agency by capturing a human game master's capability to reason about and influence the player's desires and expectations.

Keywords

experience management, narrative planning, player agency, player modeling

1. Introduction

A classic challenge in computer interactive narrative design [1, 2] is resolving conflicts between two desiderata: First, to be truly *interactive*, they should give the player an active role in the outcome of the story. Second, to be effective *narratives*, they should consist of meaningfully interconnected parts, not simply a meandering sequence of events. We refer to these criteria as *agency* and *structure* respectively.

In complex, procedural interactive narratives where it is not practical to anticipate all possible paths the player may take, an intelligent *experience manager* agent [3] may be designed to automatically decide how the story adapts to the player's choices.

How can an experience manager promote structure and agency? Because there can be conflict between these two criteria, one option is to prioritize one or the other. For instance, a well-structured story may be crafted through meticulous planning, but a player making unexpected choices could cause those plans to fail; to trade away agency for a stronger guarantee of structure, an experience manager could model its interaction with the player *adversarially* [4], planning to "defeat" a hypothetical player who derails the story whenever possible.

But in in human-led interactive narrative practices, such as improvisational theater [5] and tabletop role-playing games [6], participants often succeed in crafting high-structure, high-agency experiences together. A key factor is the expectation of cooperativity; by reasoning about others' intentions, signaling their own intentions, and assuming that others are doing the same, participants can coordinate to craft a story that fulfills their mutual goals.

There have been calls to model computer interactive narrative in the same way, where the system and player work together as storycrafting co-creators [7, 8, 9]. Our long-term objective is to build experience managers that operationalize this concept. But to do so, we must investigate: What makes a high-structure, high-agency story? How do interactive narrative participants coordinate with each other through their choices? And how does this coordination affect structure and agency in practice?

This paper discusses our first step in that investigation: collecting data from an interactive narrative hypertext game where the roles of both player and experience manager were filled by humans. Besides the logs from the gameplay itself, we elicited participants' explanations of *why* they made the choices they did, as well as why they believed their partner made choices. Through refinement and repetition of this process, we plan to amass a dataset where these explanations can be used to analyze how users reason about each other's intentions and expectations, and to evaluate how well a given experience management model can replicate that reasoning.

So far, we have conducted one iteration of our data collection effort, between students in a classroom exercise. We describe the initial design of the interactive narrative exercise and give our reflections on the design of future versions. We also present an observational study using the gameplay logs and user explanations combined with users' reported perceptions of the structure and agency afforded to them in the game. Although our data collection will need some adjustments to be as effective as possible for our original goal of helping to model implicit communication between participants, our data so far did provide the promising result of a positive correlation between a user's structure ratings and their agency ratings.

2. Related Work

For an overview of experience management in general, see Riedl and Bulitko [10]. Our work draws especially from planning-based experience management [11], where an author can exert control over the story structure through setting the goal of the planning problem. However, the player's actions may conflict with the experience manager's plans. *Narrative mediation* [12] aims to either preserve a plan by changing the outcome of the player's action, or find a new plan that is not threatened by the action. While the original approach is invoked at the moment of the player's action, a later iteration [13] incorporates player plan recognition to anticipate the action and adapt to it before it takes place. The tight control offered by mediation is important in certain applications, such as intelligent training systems where failure to meet the goal can result in a perverse lesson for the trainee [3, 14]. For applications that emphasize a personalized experience over a global objective, on the other hand, there has been research on experience managers that detect and fulfill player preferences [15, 16, 17].

Although players in interactive narratives can find mean-

14th Intelligent Narrative Technologies Workshop at the 2024 Conference on Artificial Intelligence and Interactive Digital Entertainment, November 18, 2024, Lexington, KY

cory.siler@uky.edu (C. Siler); sgware@cs.uky.edu (S. G. Ware)

ingful ways to *defy* game expectations that are still done with a prosocial mentality [18], it has been proposed that players are typically accommodating to the story roles and opportunities offered to them by the game, and that leveraging this tendency can help an experience manager to achieve an effective balance of narrative goal achievement and player agency [2, 19]. In defining player agency as a harmony between what the player *can* do and what they *want* to do, Wardrip-Fruin et al. [20] emphasize that a game can guide the player's desires toward what is possible, rather than needing to make possible everything that could be desired. Player suggestibility has been explored in terms of visual [21] and natural-language [22, 23] cues in games.

Our in-progress dataset is unique in combining player intentions with structure and agency ratings in a symbolic planning-oriented interactive narrative environment, but there are existing datasets that have some of these features. A dataset by Zhu et al. [24] pairs players' natural-language descriptions of *Dungeons & Dragons* actions with corresponding formal commands for making a tool update a symbolic game state. Kreminski et al. [25] annotate players' natural-language stories about game experiences, using a qualitative coding that distinguishes elements such as the story point-of-view (player, character, or third-person perspective) and reasons for decisions (mechanical or character-centric).

3. Data Collection

We collected data from a classroom exercise with University of Kentucky students in an undergraduate game development class. The exercise consisted of pairs of students interacting with each other over the Web in a text adventure game with a Twine-like [26] hyperlink-based interface (Figure 1). One user took the role of the game master (GM), acting in an experience-manager-like capacity, and the other user took the role of the player.

Before the exercise, the class was given a brief presentation on the exercise and a basic description of the story characters. Then, to familiarize them with the game interface, users participated in a tutorial session where they were instructed to take specific actions. Users then played sessions in a freeform manner until the end of the class period, paired automatically by the server so that no user had the same role (GM or player) or partner twice in a row.

The game setting and mechanics were based on the *Save Gramma* adventure game [27]. The player controlled the protagonist of the story, who was simply referred to as "Player" and had the stated objective of bringing a potion to their cottage to cure their sick grandmother. The GM controlled the other characters: the Bandit, Guard, and Merchant. Available actions included traveling between four locations; picking up or attempting to trade or steal items; and attempting to attack characters.

One user was able to act at a time within a session, with control going to the GM by default until they opted to let the player take the next action. Furthermore, we modified some actions to allow for success or failure at the GM's discretion; for instance, when either user had one character attack another, the GM was prompted to choose whether the attack landed or missed.

At any time during gameplay, users were able to submit star ratings on a 5-point scale of their current evaluation of agency ("I can have meaningfully different experiences depending on my own choices") and structure ("These events feel like a story, rather than a random sequence"). After every three actions in the game besides walking, the interface gave users a visual reminder to update their ratings; however, at only one point in the game—at the end of the session—did the interface force users to rate.

At the end of each session, the interface gave each user a post-game interview (Figure 2): The user was incrementally shown a history of all the actions in the game session. After each action besides walking, the user was prompted to give text input to the question "Why did you choose this action?" for their own actions or "Why do you think *your partner* chose this action?" for the other user's actions.

Afterward, we asked students to voluntarily sign a consent form for the release of their data, and deleted the logs from each session that included a participant who did not sign a consent form.[1] The remaining data consisted of logs for 22 sessions among 18 participants.

4. Observations and Design Considerations

The game interface and the use of the *Save Gramma* story domain in a multiplayer text game were successful overall: Users quickly grasped most of the game elements after the tutorial, and a wide variety of stories emerged even with the simplicity of the domain. A point for improvement in later iterations is the set of goals available to the player: With the player having a single predefined game objective, there is not much room for investigating how the GM would reason about the player's long-term goals, so we are planning the next game to let the player choose from several objectives.

Along with the game itself, there are design decisions to iterate on about the form and timing of eliciting other information from the users: their evaluations of the story experience, and their explanations for their own and each other's actions.

The questions about a user's intentions behind an action—"Why did you choose this action?" or "Why do you think *your partner* chose this action?"—had advantages and drawbacks for any choice of timing: Querying a user immediately after they took an action would capture the user's true expectations at the time better than querying at the end, which would be affected by the user's knowledge about how the story actually unfolded after the action. But immediate queries would also interrupt the flow of gameplay and risk affecting the user's decision-making process for later actions. Ultimately, our choice was informed by our intention to query about almost every action, in order to obtain enough data from a small number of playthroughs: Because so many mid-game queries would add up to a large distraction for the users, we chose to do all questioning in a post-game interview.

The responses to these questions took a wide variety of forms; we provide an informal catalog in Table 1. Although each kind provides valuable information in its own right, the diversity would also make it challenging to represent all of the explanations in a unified computational model. A collection of explanations like rows 1 through 3, for instance, could be formulated as goals for a narrative planning problem and be used to validate planning-based models of gameplay; meanwhile, rows 7 or 8 would be valuable for

[1]The process was approved by the university's IRB.

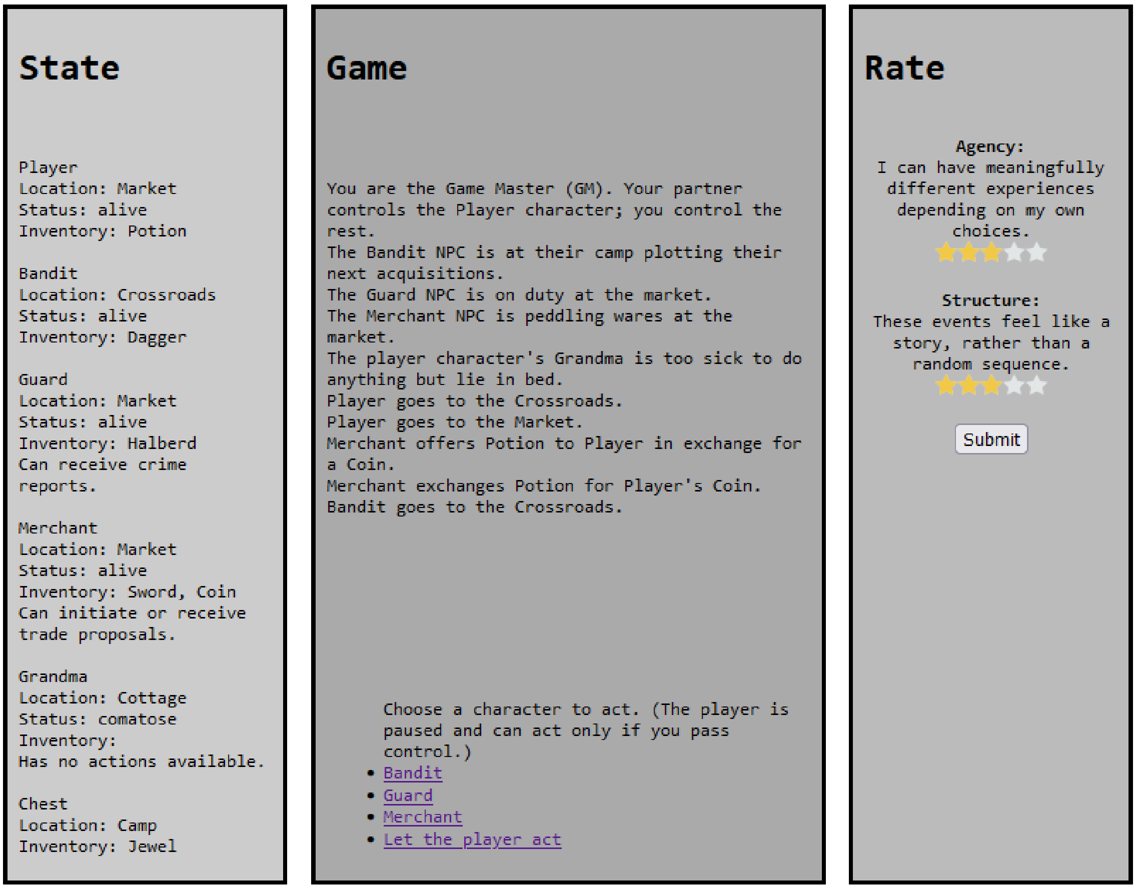

Figure 1: A mid-game view of the browser interface.

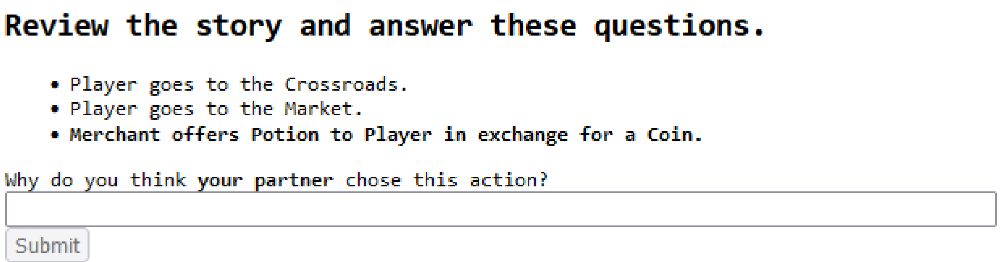

Figure 2: The post-game interview.

developing models of shared authorship but would require a notion of story-encompassing goals that does not fit as easily into a state-centric planning formalism.

When eliciting the users' perceptions of structure and agency, we chose a star rating system because, unlike the text-entry form for action explanations, it could easily be integrated into the game interface in a minimalist manner that allowed quick and (ideally) frequent feedback. However, some drawbacks became apparent later on: First, the ratings in our data turned out sparse, with many users providing only the mandatory final ratings and not the optional mid-session ratings. Second, the connotations of star ratings and the fact that structure-agency pairs were very often identical raises the possibility that users may have sometimes submitted ratings inattentive to the meaning—e.g., clicking a single rating in both star fields to represent the overall quality of the story.

We are exploring possible changes to the elicitation for future iterations. For the action explanations, we are considering how the way we ask questions to users could be changed to more precisely target styles of explanation that lend themselves to evaluating computational models. For the structure and agency evaluations, we are investigating ways to survey users that offer more psychometric validity than the star rating system. For instance, the IRIS project has constructed a set of scales to measure aspects of user experience in interactive narrative [28]; the *effectance* scale has been used to evaluate player agency in experience managers before [29, 30], but we are not aware of any equivalent scales for structure.

Table 1
Participant Action Explanations

#	Reason given for action	Example
1	As part of a plan to achieve a desired state of the world	The player took the potion. The player explained their choice as “to save grandma”.
2	To help prevent an unwanted state of the world	The player stole a weapon from the Guard, who had attacked them earlier. The player explained their choice as “self defense”.
3	To increase a metric considered valuable	The player looted a coin from the dead Merchant. The GM explained the player’s choice as “Because capitalism (who doesn’t want more money)”.
4	In response to a previous action	The Guard attacked the lawbreaking player. The player explained the GM’s choice as “I committed thievery right in front of him”.
5	Because the action is stereotypical for the character	The Bandit robbed the player. The player explained the GM’s choice as “an action a bandit would do”.
6	To establish a character trait	The Guard tried to rob the player. The GM explained their choice as “he decided on a career change, he’s now also a bandit”.
7	To create narrative conflict or game difficulty	The Bandit took the potion after the Guard died holding it. The GM explained their choice as “To keep the story going, it’d be too easy if the player got the potion immediately”.
8	To create a novel situation	The Merchant died offscreen at the beginning of the game, leaving their wares free to plunder. The GM explained their choice as “I thought it would be interesting for the player to not need money”.
9	To suggest a course of action to the other user	The Merchant offered to trade the player a sword in exchange for a jewel, which the player did not yet have. The GM explained their choice as “To try to get the player to go to the camp and steal the jewel”.
10	For amusement	The Guard looted the Bandit’s corpse. The GM explained their choice as “I think it’d be funny for the guard, after seeing the player rummage the corpse of the bandit, [to] also start rummaging through the belongings”.
11	To explore the game mechanics	The Bandit picked up the jewel. The GM explained their choice as “I didn’t know what this would do and thought it would be interesting”.

5. Analysis of the Initial Data

Issues discussed in Section 4, such as low rating frequency and nonuniformity of how users interpreted the post-game interview questions, meant that the initial data did not enable the experiments we originally planned such as using it to validate experience management techniques. In order to explore the data nonetheless, we conducted a series of *post hoc* analyses, positing and testing several hypotheses about factors that might correlate with structure and agency. Each hypothesis is described in a subsection later in this section.

For these analyses, we used Spearman’s rank correlation coefficient [31] to measure the correlations. Spearman coefficients range from -1 to 1 and measure how close two variables are to having a monotonic relationship. We chose Spearman because it supports ordinal data such as our ratings, rather than assuming continuous variables.

Because of the large number of hypotheses tested, we used false discovery rate control. This adjusts the p-values so the significance threshold (we use a standard $p < 0.05$) bounds the number of false positives relative to the total number of significant results. In particular, we chose the Benjamini-Yekutieli procedure [32] because it makes no assumptions about independence between hypothesis tests.

The numeric results are shown in Table 2: the Spearman correlation for each pairing, the original p-values (p_{orig}), and the adjusted p-values (p_{adj}). The rest of this section elaborates on each hypothesis and summarizes the results. Throughout the section, we use the terms “GM structure”, “GM agency”, “player structure”, and “player agency” as shorthand to refer to those participants’ star ratings of those qualities. Furthermore, we consider only the mandatory ratings from the end of the session due to the low frequency of optional mid-session ratings.

5.1. Structure and agency

There were statistically significant correlations between a user’s final structure ratings and their final agency ratings; this applied for both the GM and the player. These correlations were positive, supporting our claim that structure and agency are not necessarily in tension with each other.

On the other hand, although the correlation coefficients for all combinations of a GM rating and a player rating were positive, they were too small to be statistically significant.

5.2. Structure/agency and mutual understanding of intent

We hypothesized that higher agreement between a user’s stated intention for their action and their partner’s believed intention for their action would correlate with higher structure and agency ratings for the story. To test this, we labeled each pair of user explanations for the same action on

Table 2
Hypothesis Tests for Correlations

Section	Quantity 1	Quantity 2	Correlation	p_{orig}	p_{adj}
5.1	GM structure	GM agency	0.83	<0.01	$< \mathbf{0.01}$
5.1	GM structure	Player structure	0.27	0.23	1.00
5.1	GM structure	Player agency	0.10	0.67	1.00
5.1	GM agency	Player structure	0.15	0.53	1.00
5.1	GM agency	Player agency	0.07	0.76	1.00
5.1	Player structure	Player agency	0.72	<0.01	$< \mathbf{0.01}$
5.2	Explanation agreement	GM structure	0.07	0.27	1.00
5.2	Explanation agreement	GM agency	0.05	0.44	1.00
5.2	Explanation agreement	Player structure	-0.03	0.68	1.00
5.2	Explanation agreement	Player agency	-0.10	0.11	1.00
5.3	Denials of player actions	Player structure	0.28	0.23	1.00
5.3	Denials of player actions	Player agency	0.44	0.05	0.90
5.4	Player actions ÷ total	Player agency	0.06	0.80	1.00
5.5	Causal connections to ending	GM structure	0.18	0.43	1.00
5.5	Causal connections to ending	Player structure	-0.28	0.22	1.00

a four-point ordinal scale: A score of 1 indicated that the explanations directly contradicted each other (e.g., one user said the Guard's attack on the Bandit was unprovoked; the other said the attack was provoked by the Bandit's attack on the player), 2 indicated that the explanations were unrelated to each other but not contradictory (e.g., one user said the Merchant offered the potion for trade to avoid being killed by the player; the other said the Merchant made the offer because it was a profitable trade), 3 indicated that the explanations were plausibly related such as one being a subgoal of the other (e.g., one user said the player stole the potion to get the potion; the other said the player stole the potion to win the game), and 4 indicated that the explanations expressed the same idea (e.g., both users attributed the Merchant looting a corpse to the Merchant being greedy).

There was no statistical significance for the correlations (positive for GM, negative for player) of either user's ratings with the agreement scores. In later iterations, we plan to collect data that supports an analysis of whether users are *fulfilling* each other's intent, beyond simple awareness of that intent, and the relationship of between this fulfillment and structure or agency.

5.3. Player structure/agency and interventions

In the game, the GM was able to deny certain player choices; e.g., if the player character tried to rob an NPC, the GM could ensure the NPC escaped, or if an NPC tried to rob the player character and the player character tried to escape, the GM could ensure the escape attempt would fail. This is akin to what Riedl et al. [12] call the *intervention* style of narrative mediation, where an experience manager prevents a player from making a choice that threatens story goals.

Because interventions prevent a player from doing something they want to, but protect authorial intent, it is intuitive to expect them to improve structure at the cost of player agency. Based on this intuition, we investigated the relationship between interventions and the actual player ratings. We defined a metric of intervention frequency as the number of times in a session that the GM caused a player's action attempt to fail, divided by number of times in the session where that was possible. However, the (positive-coefficient) correlations between this metric and player structure or agency were not statistically significant.

5.4. Player agency and sharing of control

Recall that the GM chose when the player was allowed to act. We computed the ratio of player actions to total actions in each session. We posited that this ratio would have a relationship to player agency ratings, but the (positive) correlation coefficient was not statistically significant. A potential factor in this result is the lack of variation in this ratio; most sessions fell within a small range of player-to-GM action proportions.

5.5. Structure and causal connectedness

Narratologists have proposed *causality* as a critical tool for story coherence [33]. At a high level, two story events A and B are said to be causally connected if B would not have happened without A; we discuss a specific computational formulation later in this section.

Among other factors, causal connection of an event to the ending of a text story has been shown to predict readers' perceived importance of the event [34]. This notion has been influential to the field of planning-based story generation; for instance, partial-order narrative planners [35] are built to ensure that all actions are causal ancestors of a goal specified as the endpoint of the story. In this part of the analysis, we examined whether a causality-based model of story coherence matched how users rated structure in our interactive narrative.

We define a measure of causal connectedness in stories as follows. The planning domain counterpart to our game [36] specifies the preconditions needed to be fulfilled before an action is available, and the effects that taking the action will have on the game state. A *causal link* between two actions specifies a precondition of the later action that was fulfilled by the earlier action. A *causal chain* is a sequence of causal links where no precondition appears twice and each later action of one causal link is the earlier action of the next causal link. The measure we used is the proportion of story actions that are on a causal chain to the story ending (resulting from the death or success of the player character), as a ratio to the total number of story actions.

We computed our causal connectedness measure for each story by using Sabre [37] to detect the enablement relationships between actions in the gameplay log. The (positive for GM, negative for player) correlation coefficient between this measure and structure rating was not statistically significant for either user.

A limitation is that we defined causality only in terms of action preconditions and effects, e.g., the Bandit must possess the potion for the player to steal it from the Bandit, so the Bandit picking up the potion enables the player's theft. Causality defined more broadly would incorporate how action choices affect each other via motivational changes or information flow [38], e.g., the Guard chases the Bandit only because the player has reported a crime by the Bandit. Another limitation is that we considered causal connectedness only with respect to the ending of the story. A fuller account of causal connectedness would also include character goals and goal failures [33]. However, analyzing these relationships would require ground-truth information about the characters' beliefs and intentions to exist; in our human exercise such information only existed subjectively in the minds of users as they roleplayed, revealed only in small fragments during the post-game interviews.

6. Conclusions

This paper presented the beginnings of a dataset we will construct over multiple years. The data from each iteration, as well as more detailed descriptions of the data collection process and related publications, will be made available online.[2] A distinct feature of our data collection is the supplementing of game logs with users' explanations of both their own and their partners' choices. We hope to use this data to develop better models of how users reason about each other in human interactive narratives and how this can translate to computer interactive narratives where an experience manager agent reasons about the implicit communications between itself and the player through gameplay choices.

Our analysis of the initial data is exploratory in nature rather than drawing definitive conclusions, and it is an observational study rather than a formal controlled experiment that establishes causation. However, the analysis provides preliminary support for one of the premises of our broader project: the possibility for user's senses of cohesive story structure and personal agency to exist interdependently rather than in conflict.

Beyond data collection, our project will eventually grow to allow for testing community-designed experience managers with a larger and more diverse pool of human players in a variety of story domains, using the same platform that we are developing for human-to-human gameplay. We are interested in feedback from the community about how to make both the platform and our eventual dataset as useful as possible for other experience management researchers.

Acknowledgments

We thank the other researchers from our lab who helped us test and improve the interactive narrative game used in the data collection: Alex Barrera, Rachelyn Farrell, Mira Fisher, and Lasantha Senanayake. We also thank the reviewers, including those who provided very detailed feedback on a previous submission.

This material is based upon work supported by the National Science Foundation under Grant No. IIS-2145153. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

[2] http://cs.uky.edu/~sgware/projects/pairedstories/

References

[1] J. Bates, Virtual reality, art, and entertainment, Presence: Teleoperators & Virtual Environments 1 (1992) 133–138.

[2] E. W. Adams, Interactive narratives revisited: Ten years of research, in: Games Developers Conference, 2005.

[3] M. O. Riedl, A. Stern, D. Dini, J. Alderman, Dynamic experience management in virtual worlds for entertainment, education, and training, International Transactions on Systems Science and Applications 4 (2008) 23–42.

[4] J. Robertson, R. M. Young, A model of superposed states, in: Experimental AI in Games Workshop, volume 10, 2016, pp. 65–71.

[5] B. Magerko, W. Manzoul, M. Riedl, A. Baumer, D. Fuller, K. Luther, C. Pearce, An empirical study of cognition and theatrical improvisation, in: ACM Conference on Creativity and Cognition, volume 7, 2009, pp. 117–126.

[6] G. A. Fine, Shared Fantasy: Role Playing Games as Social Worlds, University of Chicago Press, 2002.

[7] R. M. Young, The cooperative contract in interactive entertainment, in: K. Dautenhahn, A. Bond, L. Cañamero, B. Edmonds (Eds.), Socially Intelligent Agents: Creating Relationships with Computers and Robots, Springer, 2002, pp. 229–234.

[8] R. E. Cardona-Rivera, R. M. Young, Games as conversation, in: Games and Natural Language Processing Workshop, volume 3, 2014, pp. 2–8.

[9] S. G. Ware, Mutual implicit question answering for shared authorship: A pilot study on player expectations, in: Intelligent Narrative Technologies Workshop, volume 10, 2017, pp. 259–265.

[10] M. O. Riedl, V. Bulitko, Interactive narrative: an intelligent systems approach., AI Magazine 34 (2013) 67–77.

[11] R. M. Young, Notes on the use of plan structures in the creation of interactive plot, in: AAAI Fall Symposium on Narrative Intelligence, 1999, pp. 164–167.

[12] M. Riedl, C. J. Saretto, R. M. Young, Managing interaction between users and agents in a multi-agent storytelling environment, in: International Conference on Autonomous Agents and Multiagent Systems, volume 2, 2003, pp. 741–748.

[13] J. Harris, R. M. Young, Proactive mediation in plan-based narrative environments, IEEE Transactions on Computational Intelligence and AI in Games 1 (2009) 233–244.

[14] M. Fisher, C. Siler, S. G. Ware, Intelligent de-escalation training via emotion-inspired narrative planning, in: Intelligent Narrative Technologies Workshop, volume 13, 2022.

[15] D. Thue, V. Bulitko, M. Spetch, E. Wasylishen, Interactive storytelling: A player modelling approach, in: AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, volume 3, 2007, pp. 43–48.

[16] M. Sharma, S. Ontañón, M. Mehta, A. Ram, Drama management and player modeling for interactive fiction games, Computational Intelligence 26 (2010) 183–211.

[17] K. K. Yu, M. Guzdial, N. R. Sturtevant, M. Cselinacz, C. Corfe, I. H. Lyall, C. Smith, Adventures of AI directors early in the development of Nightingale, in: AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, volume 18, 2022, pp. 70–77.

[18] P. Sidhu, M. Carter, Benevolent transgressive play in *Dungeons & Dragons*, Simulation & Gaming 1 (2023) 1–22.

[19] R. Cardona-Rivera, R. Young, Symbolic plan recognition in interactive narrative environments, in: AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, volume 11, 2015, pp. 16–22.

[20] N. Wardrip-Fruin, M. Mateas, S. Dow, S. Sali, Agency reconsidered, in: Digital Games Research Association Conference, 2009.

[21] R. E. Cardona-Rivera, R. M. Young, A cognitivist theory of affordances for games, in: Digital Games Research Association Conference, 2013.

[22] R. Figueiredo, A. Paiva, "I want to slay that dragon!" - influencing choice in interactive storytelling, in: International Conference on Interactive Digital Storytelling, volume 3, 2010, pp. 26–37.

[23] D. L. Roberts, M. L. Furst, B. Dorn, C. L. Isbell, Using influence and persuasion to shape player experiences, in: ACM SIGGRAPH Symposium on Video Games, 2009, pp. 23–30.

[24] A. Zhu, K. Aggarwal, A. Feng, L. J. Martin, C. Callison-Burch, FIREBALL: A dataset of *Dungeons and Dragons* actual-play with structured game state information, in: Annual Meeting of the Association for Computational Linguistics, volume 61, 2023, pp. 4171–4193.

[25] M. Kreminski, B. Samuel, E. Melcer, N. Wardrip-Fruin, Evaluating AI-based games through retellings, in: AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, volume 15, 2019, pp. 45–51.

[26] C. Klimas, Twine / an open-source tool for telling interactive, nonlinear stories, http://twinery.org/, 2009.

[27] S. G. Ware, E. T. Garcia, M. Fisher, A. Shirvani, R. Farrell, Multi-agent narrative experience management as story graph pruning, IEEE Transactions on Games 15 (2022) 378–387.

[28] C. Klimmt, C. Roth, I. Vermeulen, P. Vorderer, The empirical assessment of the user experience in interactive storytelling: Construct validation of candidate evaluation measures, Technical Report, European Commission, 2010.

[29] D. Thue, V. Bulitko, M. Spetch, T. Romanuik, A computational model of perceived agency in video games, in: AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, volume 7, 2011, pp. 91–96.

[30] A. Ramirez, V. Bulitko, M. Spetch, Evaluating planning-based experience managers for agency and fun in text-based interactive narrative, in: AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, volume 9, 2013, pp. 65–71.

[31] P. Schober, C. Boer, L. A. Schwarte, Correlation coefficients: appropriate use and interpretation, Anesthesia & Analgesia 126 (2018) 1763–1768.

[32] Y. Benjamini, D. Yekutieli, The control of the false discovery rate in multiple testing under dependency, Annals of Statistics 29 (2001) 1165–1188.

[33] T. Trabasso, T. Secco, P. van den Broek, Causal cohesion and story coherence, in: H. Mandl, N. L. Stein, T. Trabasso (Eds.), Learning and Comprehension of Text, Lawrence Erlbaum Associates, Inc., 1984, pp. 83–111.

[34] T. Trabasso, L. L. Sperry, Causal relatedness and importance of story events, Journal of Memory and Language 24 (1985) 595–611.

[35] M. O. Riedl, R. M. Young, Narrative planning: balancing plot and character, Journal of Artificial Intelligence Research 39 (2010) 217–268.

[36] S. G. Ware, R. Farrell, A collection of benchmark problems for the Sabre narrative planner, Technical Report, Narrative Intelligence Lab, University of Kentucky, 2023.

[37] S. G. Ware, C. Siler, Sabre: a narrative planner supporting intention and deep theory of mind, in: AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, volume 17, 2021, pp. 99–106.

[38] S. M. Khan, M. Rostamigiv, On explaining agent behaviour via root cause analysis: A formal account grounded in theory of mind, in: European Conference on Artificial Intelligence, volume 26, 2023, pp. 1239–1247.