Mutual Implicit Question Answering for Shared Authorship: A Pilot Study on Player Expectations

Stephen G. Ware
sgware@uno.edu
Narrative Intelligence Lab, Department of Computer Science
University of New Orleans, New Orleans, LA, 70148, USA

Abstract

The typical goal of an experience manager in an interactive narrative is to create a sense of shared authorship that lends the player freedom to personalize the experience while still meeting the author's constraints on structure. This can be difficult when the player and author only communicate with one another through their actions. Each new action causes new questions to arise, assumptions to be made, and old questions to be answered. In this paper, I propose a technique called Mutual Implicit Question Answering, or MIQA, designed to allow an experience manager to both perceive and influence the momentum of an interactive story. It combines a generative model of narrative planning with analytical models of question answering and salience. I also present the results of a small, qualitative study of how people construct interactive narratives that lends insight for the eventual evaluation of a MIQA experience manager.

Introduction

Interactive narratives are often described in terms of player freedom and author structure. Players want the freedom to do many things. Static media like films offer little freedom, whereas sandbox games like Minecraft offer much. Authors structure narratives either for aesthetic reasons or because they face constraints on content and outcome. Highly structured narratives are coherent and compelling, whereas unstructured ones are chaotic and confusing. Intelligent narrative technologies often strive to provide *shared authorship*, a balance between the competing needs of freedom and structure that allows a human player and an artificial agent to meaningfully contribute to creating a narrative which is personal, meaningful, structured, and effective.

Despite the importance of shared authorship, relatively few attempts have been made to measure it directly. In this paper, I propose a technique called Mutual Implicit Question Answering, or MIQA (pronounced "Micha"), and explain how it could be used to measure and create shared authorship. I also describe a small, qualitative pilot study of human interactive storytelling that lends insight to the eventual evaluation of MIQA and motivates the importance of the author having an accurate perception of the player's expectations.

MIQA considers an interactive narrative as a multi-agent decision problem. The universe of possible stories is cast as a state-space graph, where nodes are world states and edges are actions that change the world state. Some actions can only be taken by the player(s), while others can only be taken by the author—typically each player controls a single character while the author controls all non-player characters, though notable exceptions exist (e.g. McCoy et al. 2014). Throughout this paper, I will refer to the second agent as the *author*, whether the author is a human or an artificial experience manager acting to carry out a human author's goals (Riedl et al. 2008; Roberts and Isbell 2008).

The virtual environment is a channel through which the player and author communicate with one another through their actions. This channel is often indirect and noisy because one side typically does not explain its beliefs, desires, and intentions directly to the other; rather, they must be inferred. The player is constantly asking questions of, assuming answers from, updating assumptions about, and making statements to the author through his or her actions. Human authors of interactive narratives (tutors, trainers, therapists, storytellers) pay attention to their audiences and do the same, but digital authors controlling virtual environments often fail at this task.

MIQA is an attempt to operationalize shared authorship by adapting explicit question answering processes long used for narrative analysis and making them an implicit part of the experience itself. It combines analytical and generative computational models of narrative. Together, I hope they will allow an artificial agent to both *perceive* and *influence* perceptions about the past and possible futures as the narrative is cocreated, enabling the author agent to better support both the player's freedom and its own structure.

Related Work

The challenge of balancing character and author goals was recognized as early as Meehan's Tale-Spin (Meehan 1977) and Dehn's response to it, Author (Dehn 1981). Tale-Spin simulated agents and then wove their plans together into a story, whereas Author planned for the author's goals and then found excuses for agents to act the way they did. Many interactive narrative systems are similar in spirit to one of these two non-interactive pioneers because they tend to first consider either player freedom or author control.

Approaches emphasizing player freedom often use a predict-and-support strategy (Thue et al. 2007; Sharma et al. 2007; Rowe et al. 2011). The system predicts the player's trajectory using goal recognition or a learned model of the player's preferences. It then places events important to the author along that trajectory. Problems arise when goal recognition fails, commonly because the system mistakes early exploratory behavior for an intentional expression of the player's preferences. By wrongly guessing the player's goal and acting to support it, the system risks creating a *self-fulfilling prophecy* or an *echo chamber* (e.g. the player fought a monster, so the game adds more monsters, so the player fights more monsters).

Approaches emphasizing author control often use a plan-and-replan strategy (Cavazza, Charles, and Mead 2002; Riedl et al. 2008; Porteous, Cavazza, and Charles 2010; Ware and Young 2015; Robertson and Young 2015). They generate a story meeting the author's constraints and assume the player will act in the way most convenient to the author. When the player tries to take an unexpected action, the author either prevents it or replans the story. The high cost of replanning may result in a nearly identical story, only making significant changes when the original plan becomes impossible. When the story is inelastic, the communication that fosters shared authorship may break down with the player feeling *railroaded*—like the author is ignoring the player's attempts to personalize the experience.

The many predict-and-support and plan-and-replan systems developed to date are aware of these pitfalls and use various techniques to avoid them. This analysis is not meant as a criticism of those systems, but rather to point out the difficulties inherent in starting with either the player or author and then trying to account for the other. Through their actions, the player and the author both build momentum toward possible story futures. Predict-and-support may be too sensitive to the player's momentum, whereas plan-and-replan may not be sensitive enough. One of MIQA's goals is to provide a continuous evaluation of story momentum as affected by both the player and author.

MIQA owes much to communication-based and improvisation-based views of interactive narrative. Young (2002) and Cardona-Rivera and Young (2014) describe stories as a form of communication and mutual sense-making between the player and author. I adopt that paradigm and focus specifically on how questions are implicitly asked and answered. Like Magerko et al. (2010) and Samuel et al. (2016), I also borrow from improvisational theater. I consider an ideal shared authorship experience one where the player and author accept and build off of one another's momentum in the same way that improv actors should say "yes and."

MIQA is also similar to recent work by Davis et al. (2017) whose creative sense-making framework combines generative and analytical models. They quantify cocreativity as agents attempting to minimize surprise. Indeed, MIQA could be considered an application of their framework to the domain of interactive storytelling.

Mutual Implicit Question Answering

To participate in the cocreation of a narrative, an agent needs a model of the story. That model needs to track the temporal and causal relationships between events, as well as the beliefs, desires, and intentions of the author, audience, and virtual characters. It needs to support queries about the past and possible futures. I propose to accomplish this by combining generative plan-based models and analytical comprehension-based models of narrative. Together, these efforts will define what *can* and *should* happen.

Each action the author takes causes the player to ask new questions, make assumptions, and answer old questions about the author's intended trajectory for the story. Critics (Gulino 2004; Abbott 2008) and psychologists (Graesser, Lang, and Roberts 1991; Gerrig and Bernardo 1994) have found it useful to model the comprehension process as one where the audience asks questions and receives answers from the author. Regardless of whether this is actually happening at the neuronal level, this technique has proven a useful paradigm for formal criticism and empirical investigations. MIQA is an attempt to adapt this explicit question answering process to one that happens implicitly for both the player and author over the noisy channel of communication through action—Mutual Implicit Question Answering. To accomplish this, we must identify when an action causes someone to ask a question and how it is eventually answered by future actions. We also need to identify any assumptions made about the answer between asking and answering. I believe that one's accuracy at MIQA can serve as a measure of mutual understanding and shared authorship.

Generative Model: Narrative Planning

The event-driven nature of narratives has made AI research in automated planning a natural fit for computational models. A classical planning algorithm takes as input (1) the initial state, (2) a sequence of action templates with preconditions and effect, and (3) a goal. It outputs a plan, a sequence of executable ground actions that achieves the goal. Narrative planning expands on classical planning by adding additional constraints to ensure that agents act believably (Riedl and Young 2010). It must balance the author's goal with the goals of characters to ensure that every action a character takes is consistent with its beliefs, desires, and intentions. Young et al. (2013) provides a survey of various narrative planning systems.

Narrative planning is valuable to MIQA because it is generative. Given a partial story, a narrative planner can generate many possible, believable ways that it can end. In short, it defines the set of narratives that *can* happen. However, planning is computationally expensive (Helmert 2006), limiting its use for online systems. When planning for all possible futures is not feasible, a plan graph can be used. Plan graphs (Blum and Furst 1997) compress the exponentially-large space of possible plans into a polynomial-sized structure which can be used to estimate plans that resemble solutions (Hoffmann and Nebel 2001). When reasoning online, people rarely form exact plans for the future either.

Analytical Models: QUEST and Indexter

Decades of research on narrative and media psychology have yielded valuable insights that have been translated into computational models of perception. Specifically, I focus on the QUEST model of question answering and the Indexter model of salience.

QUEST (Graesser, Lang, and Roberts 1991) models how people answer open-ended questions about narratives. A story is encoded as a directed graph called a QUEST Knowledge Structure (QKS). Nodes represent types of narrative content such as goals, events, and states. Edges represent relationships such as "event x was taken in service of goal y." QUEST also defines graph search procedures for answering questions like *why? how?* and *what enabled?* based on QKS structure. Graesser, Lang, and Roberts (1991) showed these procedures answer similarly to people. Cardona-Rivera et al. (2016) provide a mapping for translating narrative plans into QKS graphs, allowing us to ask these questions of a narrative plan. MIQA is based on question answering, and QUEST provides a basis for establishing when questions are asked and how they are answered.

Indexter models narrative salience—that is, how well the audience remembers a narrative's past events during online comprehension (Cardona-Rivera et al. 2012). It operationalizes the Event-Indexing Situation Model (Zwaan and Radvansky 1998) and states that the salience of a past event is correlated to recently narrated events based on how similar they are along five dimensions: *protagonist* (who), *time* (when), *location* (where), *causality* (how), and *intentionality* (why). Kives, Ware, and Baker (2015) showed Indexter reliably predicts which events are easier to remember. MIQA uses Indexter to quantify narrative momentum—where the story is going based on where it's been. Also, QUEST search procedures do not return a single answer to a question but a list of answers ranked by quality. Indexter can improve QUEST answer rankings by accounting for salience.

In general, QUEST and Indexter are analytical models. Given the set of stories the narrative planner says *can* happen, they rank them to determine which *should* happen.

The major limitation of these models is that analysis is typically done after the fact. When validating them, subjects were asked about past events (that is, the actual world). I intend to apply these models to possible future worlds. My previous work suggests this is possible. Farrell, Robertson, and Ware (2016) showed that, when faced with a hypothetical that makes the end of a story impossible, people answer QUEST questions as if they had rewritten the ending to imagine a non-narrated possible world. Farrell and Ware showed that Indexter salience can be used not only to measure (2016) but also influence (2017) expectations of an interactive narrative's future.

Experience Management with MIQA

The challenge of shared authorship lies in providing both player freedom and author structure. Structural constraints can be expressed in many ways, like what terminal states are acceptable, what actions should or should not occur, which states are desirable, etc. An ideal interactive environment permits many narratives that meet these constraints. So, given many possible futures with acceptable structure, a MIQA-based experience manager should choose actions for non-player characters that build on the player's momentum and shift it toward a future desirable for the author.

The improv metaphor is helpful here. As the player and author build the narrative action by action, they are causing new questions to be asked and old ones to be answered. Each action adds momentum to or away from possible futures. MIQA allows an author to both *detect* and *influence* which futures the player expects. When there are many ways to answer a question, the author should choose one that adds momentum to a future which meets the author's constraints. Likewise, when there are many questions the author can cause the audience to ask, it should choose one where the likely answers lie in futures which meet the author's constraints. The player and author should not fight one another's momentum as if in an adversarial game, but rather they should cooperate and build on one another's momentum, accepting it and shifting its direction to suit their needs, like a negotiation.

My eventual goal is to build a fully autonomous online narrative experience management agent that uses MIQA. This will require specifically identifying which questions are asked and when, as well as how they are answered. The accuracy of mutual implicit question answering can then be evaluated through explicit question answering by players and authors. I hope to demonstrate that MIQA accurately reflects each side's expectations of the other. Before beginning that undertaking, I sought to explore how one might elicit a player's expectations of the future and whether supporting them leads to better experiences of freedom and structure.

Pilot Study

I explored the viability of a MIQA-based approach by observing how humans perform mutual sense-making in a tabletop role playing game. I recruited an experienced game master (or GM) and two players who had all previously played *Dungeons and Dragons* together. I observed two interactive storytelling sessions, one representing effective shared authorship and one not, discretized these sessions, and analyzed how mutual understanding between the GM and players affected perceptions of player freedom and author structure. This simple, qualitative pilot study yielded valuable insights for the future development and evaluation of an artificial author agent.

Data Collection

Both role-playing sessions were performed on Roll20.net and Discord, applications for dice-based role-playing games and multi-party voice communication respectively. The GM and players had used these to play previous games together and were comfortable with this medium. I observed and recorded the audio of both sessions for future discretization.

The game system chosen by the GM was a highly simplified version of Dungeons and Dragons, in which a character is described by distributing a limited pool of points into six attributes: strength, constitution, dexterity, charisma, intelligence, and wisdom. These points are added to dice rolls

when a character attempts an action that might fail, so the distribution of points determines a character's strengths and weaknesses. Each player controls one character while the GM controls all other characters and the environment. The many other features of the system were discarded for simplicity and to keep the sessions focused on the story, rather than on other aspects of these games, like combat.

The GM was asked, without the knowledge of the players, to ensure that the first session offered little freedom while the second offered significant freedom. Based on player self-reports of freedom, this was successful.

Throughout each session, the GM and players were interrupted and asked to answer *why* questions about important events in the game using a typed short-answer free response. They were also asked to rate their perceptions of freedom and structure using the 5-point scales given in Figure 1. Their answers were recorded privately so they did not affect one another's perceptions of the story. Interruptions occurred about 140 seconds apart on average, providing relatively fine-grained measurements of how freedom and structure changed over time.

I then listened to the recorded sessions and coded the narratives into text as a series of simple events. An event represents a single action where one or several subjects performs a single verb on zero or more objects. Examples include walking from one place to another, buying or selling an object, asking or answering a question, attacking a target, etc. Each event was also labeled as a GM action (if it was part of the environment or required the consent of a non-player character) or a player action (if it required the consent of a player) or both if both had to consent, such as moving as a group, buying, and selling. This level of granularity reflects the kinds of actions that are commonly available in digital interactive games, including my previous work (Ware and Young 2015). The guiding principle used for coding events was to imagine the session had taken place in one of the many D&D video games and that each event should represent a single entry in the game's log as it is played. When a fully digital MIQA agent is deployed, it will be in this kind of environment and use this level of granularity to represent the narrative, so this coding scheme is designed to generate the kind of data on which a MIQA agent will eventually be evaluated. To maintain consistency, the same person coded both stories, all events in a story were coded in the same sitting, and both stories were coded on the same day.

After coding, both players and the GM were shown the text events one at a time in an online survey. After each event, subjects were individually asked to describe their expectations for the future of the story at that moment using a typed short-answer free response. These responses were then coded as above. Given that these are predictions about the future, some events were allowed to be ambiguous, such as "Jort will go somewhere." Subjects were asked to describe the future as they imagined it at that time, disregarding knowledge of the actual outcome. Accurately recalling these perceptions can be difficult, which is why I relied on veteran role-players with experience separating in-game from out-of-game knowledge. The data collected reflects numerous wrong beliefs and plans that were formed but never carried out, suggesting that subjects did recall their perceptions with at least some accuracy.

Scale used to measure freedom:

0. **None:** There is only one thing I'm allowed to do, and I cannot affect the course of the story.
1. **Little:** My actions only affect small, mostly meaningless elements of the story.
2. **Fair:** My actions affect some important elements of the story.
3. **Much:** My actions affect most of the important elements of the story.
4. **All:** I can do anything I want, and the story will change to suit my actions.

Scale used to measure structure:

0. **None:** The story is a random sequence of events that don't make sense.
1. **Little:** The main plot is unclear or keeps changing, but some things make sense.
2. **Fair:** The main plot is clear, with a few plot holes and loose ends.
3. **Much:** Most events make sense, and I see how they fit together to make a story.
4. **All:** Every element is carefully crafted and meaningful.

Figure 1: The survey used by GM and players to report perceptions of freedom and structure.

Story Summaries

Before the first session, players created their characters. The first player emphasized strength and constitution and played the role of a courageous but gullible human warrior named Jort Doogan. The second emphasized intelligence and wisdom and played the role of a meek, bookish elf scholar named Splunt Junkman.

In the first session, Jort and Splunt are recruited by a dark elf pirate named Talasanor who sails to an encampment of hostile orks to retrieve one of a pair of magical swords. After slaying the orks and obtaining the sword, Talasanor reveals that he is the true king of their nation and the only one who can wield the sword. After the group returns home, a war breaks out between angels and demons whom Talasanor must vanquish using the magical sword. In both adventures, Talasanor does most of the work, while Jort and Splunt are merely along for the ride, unable to contribute in any meaningful way and usually with little understanding of why they are there or what they are supposed to be doing. This session lasted 61 minutes and was coded into 49 events. Subjects were interrupted to answer a question and report freedom and structure 23 times.

In the second session, which was recorded one day later, Jort and Splunt are called to a small town to hunt a werewolf who is troubling the townsfolk. The previous were-

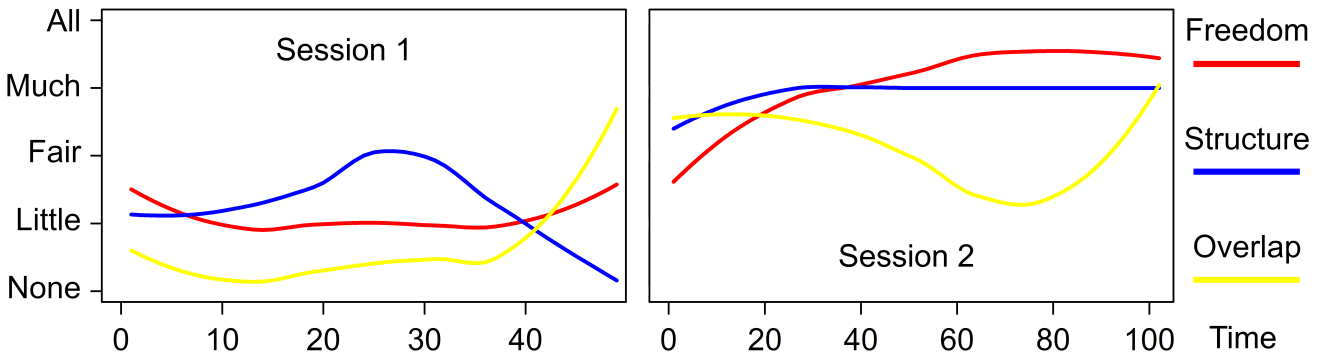

Figure 2: Average player self-report of freedom and structure, and the degree to which the GM's expectations of the future overlapped with the players.

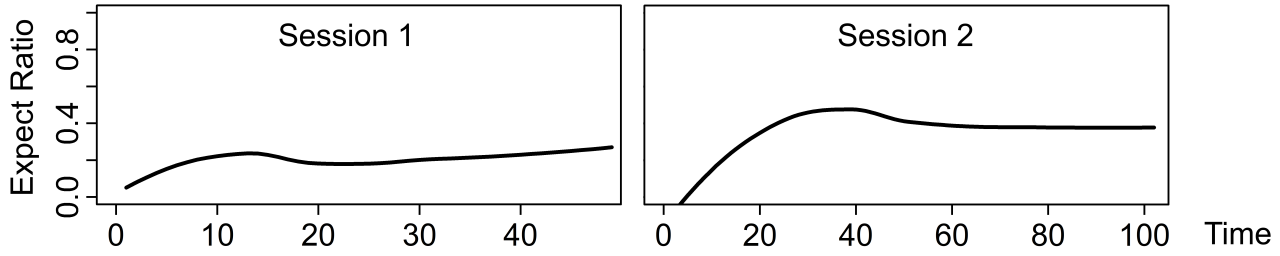

Figure 3: Expect ratios over time.

wolf hunter, now dead, left behind a list of four suspects whom the players investigate one by one. After failing to find evidence or trap the werewolf, the players return to town to find another victim, this time with telltale evidence that leads them to arrest the second suspect, Old Man Jenkins. This session lasted 116 minutes and was discretized into 102 events. Subjects were interrupted 58 times.

Insights for Question Answering

One straight-forward way to validate a model of implicit questions answering is through explicit question answering. One might validate a MIQA agent by interrupting players, asking them to answer open questions about the narrative, and comparing those answers to MIQA's. The first valuable insight gained from this experiment was that this approach may not be viable.

All questions asked to subjects during interruptions were *why* questions about an event that had just occurred. For example, in the first session, after "Jort attacks the ork," subjects were interrupted and asked "Why did Jort attack the ork?" QUEST answers *why* questions about actions in terms of goals, so these questions were designed to elicit subjects' expectations about the future. However, subjects tended to answer in terms of motivations rather than goals. Presumably the goal of attacking the ork is something like "to kill the ork," but both players and the GM answered this question as "because the ork attacked Jort." The ork's attack is the event that *motivated* the goal of killing the ork, but it is not the goal itself. This is indicative of how subjects answered questions in general, referring to a past events that motivated an action rather than a future event or goal that action was in service of.

I speculate this occurred because subjects find it easier to answer in terms of past events that have definitely happened rather than future events which may not happen. QUEST questions were designed to be asked after the story has been read, and this pilot study suggests subjects answer differently when asked while the story is still unfolding. This does not necessarily indicate subjects don't reason about about the future, since subjects reported specific beliefs or plan for the future, only that a better way of eliciting those beliefs is needed. Ultimately, for this pilot study, I simply asked subjects to describe their expectations for the future at each moment of the story in a post-survey, but this method will likely be unreliable for inexperienced role-players.

Freedom, Structure, and Agency

The first interesting result I observed is that player reports of freedom and structure were both lower in the first session and higher in the second, as seen in Figure 1. Indeed, freedom and structure were strongly positively correlated (Pearson's product-moment correlation, $\rho = 0.83$, $p < 0.001$). This is surprising if one considers player freedom and authorial structure as opposing ends of the same spectrum, where increasing one requires decreasing the other. These results support the definition of agency advanced by Wardrip-Fruin et al. (2009), who claim agency is not simply the ability to do anything, but rather a balance between what the system invites the player to do and what they can actually do. High freedom *and* high structure can both be achieved in the ideal shared authorship experience.

Overlapping Visions of the Future

The first hypothesis I explored was that freedom and structure are higher when the GM's expectations for the future overlap more with those of the players. To evaluate the degree of overlap, three raters (the author and 2 research assistants) were independently shown the coded stories one step at a time alongside the GM's and players' expectations of the future. They rated the degree of overlap on a 5 point scale (from "totally different" to "totally the same") for the GM and player 1 (P1), GM and player 2 (P2), and P1 and P2. They achieved an inter-rater reliability of $\alpha = 0.706$ (Krippendorff 2004). Before performing the analysis, I removed the first and last event from each story as outliers because, as one might expect, they demonstrate very low and very high overlap respectively.

Like freedom and structure, the overlap between the GM and players' vision of the future was lower in the first session and higher in the second. However, regression analyses of freedom and structure as a function of both overlap and session number revealed a negative correlation between overlap and freedom ($F_{(2,144)} = 294.9$, $p < 0.001$, $R^2 = 0.80$) and overlap and structure ($F_{(2,144)} = 335.9$, $p < 0.001$, $R^2 = 0.82$). This held when comparing GM to P1, GM to P2, and GM to the averages for P1 and P2. Thus, I cannot support this hypothesis, except to say that the story with higher freedom and structure overall also showed higher overlap overall.

Fulfilling Player Expectations

Overlapping visions of the future are not necessarily the best indicator of mutual sense-making. GMs frequently conceal plans or mislead the player for the sake of the story. The players need not know that Old Man Jenkins is the werewolf to share authorship with the GM. It is more important, then, for the GM to have an accurate model of what the players expect. This can measured indirectly by observing how often the GM took actions that the players expected.

Every event in both sessions was marked as either expected or unexpected by the players based on whether it ever appeared in a player's vision of the future before it happened. Ground events were expected by ambiguous events; for example, "Jort and Splunt discover Old Man Jenkins is the werewolf" was considered expected if "Jort and Splunt will discover that someone is the werewolf" appeared in Jort's or Spunt's vision of the future. Given this definition of an expected event, after each event I calculated the ratio of expected GM events to total GM events—i.e. the proportion of things the GM has done up until that moment that the players expected the GM to do. I dub this the *expect ratio*, and visualize it over time for both session in Figure 3.

Regression analyses of freedom and structure as a function of both expect ratio and session number found a significant positive correlation between expect ratio and freedom ($F_{(2,144)} = 403.6$, $p < 0.001$, $R^2 = 0.85$) and expect ratio and structure ($F_{(2,144)} = 336.2$, $p < 0.001$, $R^2 = 0.82$). Though this analysis includes session number as a predictor variable, I also repeated it for each session individually. I found that, in the first session, expect ratio and freedom have a significant negative correlation ($F_{(2,47)} = 8.6$, $p = 0.005$, $R^2 = 0.14$). This may actually be additional support for my hypothesis, because in that session the GM was intentionally trying to create a low freedom story, so it makes sense that a skilled GM would use an accurate perception of the players' expectations to limit player freedom.

This analysis is helpful even if the GM took expected actions by accident because it is based solely on the perception of the players. It demonstrates that fulfilling player expectations, whether on accident or on purpose, plays a role in creating high freedom and high structure stories.

As the number of events in the story increases, each new expected GM action has less impact on the expect ratio. For longer role-playing sessions, it may be necessary calculate this ratio so that it gives more weight to recent events.

The same relationship holds when the GM's perceptions of agency and structure are modeled as a function of the proportion of actions players took that the GM expected, but the relationships were not significant ($p = 0.148$ and $p = 0.199$ respectively) given the low power of the study. I also calculated the proportion of actions (1) that players took that players expected and (2) that the GM took that the GM expected, but no significant positive correlations between these values and their respective perceptions of agency or structure were revealed. This is interesting because it suggests that one does not necessarily need to do what one expects to do to experience high agency and structure.

Conclusion

Interactive narratives are a form of mutual sense-making across a noisy channel of communication through action. I have proposed a technique call Mutual Implicit Question Answering for allowing an artificial agent to participate in sense-making by combing a generative narrative planning model of what *can* happen with analytical models of what *should* happen based on question answering and salience.

Our eventual goal is to build a virtual experience manager that uses MIQA to both perceive and influence the player's expectations for the future to create narratives which offer both player freedom and authorial structure. A simple, qualitative pilot study of how humans perform this process tentatively supports our hypothesis that an accurate model of player expectations for the future can aid in experience management. Of course, given the low power of this study and the amount of arbitrary coding involved in processing, all conclusions should be treated as exploratory. Still, this study provided valuable insights about experimental designs for the eventual evaluation of a MIQA-based agent, chief among them the need for an accurate, unobtrusive means of eliciting a player's expectations. In conclusion, I hope this work lays the foundation for a new generation of shared authorship in interactive narratives that are personal, meaningful, structured, and effective.

Acknowledgments

This research was supported by NSF award IIS-1464127.

References

Abbott, H. P. 2008. *The Cambridge introduction to narrative*. Cambridge University Press.

Blum, A. L., and Furst, M. L. 1997. Fast planning through planning graph analysis. *AI* 90(1):281–300.

Cardona-Rivera, R. E., and Young, R. M. 2014. Games as conversation. In *Proceedings of the Games and Natural Language Processing workshop at AIIDE*, 2–8.

Cardona-Rivera, R. E.; Cassell, B. A.; Ware, S. G.; and Young, R. M. 2012. Indexter: a computational model of the event-indexing situation model for characterizing narratives. In *Proceedings of the 3rd Workshop on Computational Models of Narrative*, 34–43.

Cardona-Rivera, R. E.; Price, T.; Winer, D.; and Young, R. M. 2016. Question answering in the context of stories generated by computers. *ACS* 4:227–245.

Cavazza, M.; Charles, F.; and Mead, S. J. 2002. Character-based interactive storytelling. *IEEE Intelligent Systems, special issue on AI in Interactive Entertainment* 17(4):17–24.

Davis, N.; Hsiao, C.-P.; Singh, K. Y.; Lin, B.; and Magerko, B. 2017. Creative sense-making: Quantifying interaction dynamics in co-creation. In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition*, 356–366.

Dehn, N. 1981. Story generation after TALE-SPIN. In *Proceedings of IJCAI*, volume 81, 16–18.

Farrell, R., and Ware, S. G. 2016. Predicting user choices in interactive narratives using indexter's pairwise event salience hypothesis. In *Proceedings of ICIDS*, 147–155.

Farrell, R., and Ware, S. G. 2017. Influencing user choices in interactive narratives using indexter's pairwise event salience hypothesis. In *Proceedings of AIIDE*.

Farrell, R.; Robertson, S.; and Ware, S. G. 2016. Asking hypothetical questions about stories using QUEST. In *Proceedings of ICIDS*, 136–146.

Gerrig, R. J., and Bernardo, A. B. I. 1994. Readers as problem-solvers in the experience of suspense. *Poetics* 22(6):459–472.

Graesser, A. C.; Lang, K. L.; and Roberts, R. M. 1991. Question answering in the context of stories. *Journal of Experimental Psychology: General* 120(3):254.

Gulino, P. 2004. *Screenwriting: the sequence approach*. A&C Black.

Helmert, M. 2006. New complexity results for classical planning benchmarks. In *Proceedings of ICAPS*, 52–62.

Hoffmann, J., and Nebel, B. 2001. The FF planning system: fast plan generation through heuristic search. *Journal of Artificial Intelligence Research* 14:253–302.

Kives, C.; Ware, S. G.; and Baker, L. J. 2015. Evaluating the pairwise event salience hypothesis in Indexter. In *Proceedings of AIIDE*, 30–36.

Krippendorff, K. 2004. *Content analysis: an introduction to its methodology*. Sage.

Magerko, B.; Fiesler, C.; Baumer, A.; and Fuller, D. 2010. Bottoms up: improvisational micro-agents. In *Proceedings of the 3rd Intelligent Narrative Technologies workshop*.

McCoy, J.; Treanor, M.; Samuel, B.; Reed, A. A.; Mateas, M.; and Wardrip-Fruin, N. 2014. Social story worlds with Comme il Faut. *IEEE Transactions on Computational Intelligence and Artificial Intelligence in Games* 6(2):97–112.

Meehan, J. R. 1977. TALE-SPIN, an interactive program that writes stories. In *Proceedings of IJCAI*, 91–98.

Porteous, J.; Cavazza, M.; and Charles, F. 2010. Applying planning to interactive storytelling: Narrative control using state constraints. *ACM Transactions on Intelligent Systems and Technology* 1(2):1–21.

Riedl, M. O., and Young, R. M. 2010. Narrative planning: balancing plot and character. *JAIR* 39(1):217–268.

Riedl, M. O.; Stern, A.; Dini, D.; and Alderman, J. 2008. Dynamic experience management in virtual worlds for entertainment, education, and training. *International Transactions on Systems Science and Applications, special issue on Agent Based Systems for Human Learning* 4(2):23–42.

Roberts, D. L., and Isbell, C. L. 2008. A survey and qualitative analysis of recent advances in drama management. *Int. Trans. on Systems Science and Applications, special issue on Agent Based Systems for Human Learning* 4(2):61–75.

Robertson, J., and Young, R. M. 2015. Automated gameplay generation from declarative world representations. In *Proceedings of AIIDE*, 72–78.

Rowe, J. P.; Shores, L. R.; Mott, B. W.; and Lester, J. C. 2011. Integrating learning, problem solving, and engagement in narrative-centered learning environments. *International Journal of AI in Education* 21(1-2):115–133.

Samuel, B.; Ryan, J.; Summerville, A. J.; Mateas, M.; and Wardrip-Fruin, N. 2016. Bad News: an experiment in computationally assisted performance. In *Proceedings of the ICIDS*, 108–120.

Sharma, M.; Ontanón, S.; Strong, C. R.; Mehta, M.; and Ram, A. 2007. Towards player preference modeling for drama management in interactive stories. In *Proceedings of FLAIRS*, 571–576.

Thue, D.; Bulitko, V.; Spetch, M.; and Wasylishen, E. 2007. Interactive storytelling: a player modelling approach. In *Proceedings of AIIDE*, 43–48.

Wardrip-Fruin, N.; Mateas, M.; Dow, S.; and Sali, S. 2009. Agency reconsidered. In *Proceedings of DIGRA*.

Ware, S. G., and Young, R. M. 2015. Intentionality and conflict in The Best Laid Plans interactive narrative virtual environment. *IEEE TCIAIG* 8(4):402–411.

Young, R. M.; Ware, S. G.; Cassell, B. A.; and Robertson, J. 2013. Plans and planning in narrative generation: a review of plan-based approaches to the generation of story, discourse and interactivity in narratives. *Sprache und Datenverarbeitung, Special Issue on Formal and Computational Models of Narrative* 37(1-2):41–64.

Young, R. M. 2002. The cooperative contract in interactive entertainment. In *Socially Intelligent Agents*. chapter 28, 229–234.

Zwaan, R. A., and Radvansky, G. A. 1998. Situation models in language comprehension and memory. *Psychological Bulletin* 123(2):162.