User evaluation of virtual reporting agents

Neil Madden
University of Nottingham
Nottingham NG8 1BB, UK
nem@cs.nott.ac.uk

Brian Logan
University of Nottingham
Nottingham NG8 1BB, UK
bsl@cs.nott.ac.uk

ABSTRACT

Many persistent online environments such as Massively-Multiplayer Online Games (MMOGs), feature weblogs and/or live reportage of participants' activities in the world. While such reports and commentary can enhance the user's enjoyment and increase their sense of shared experience, the demands of such large scale reporting on the participants can be considerable. To address this problem a number of in-game reporting and commentary systems have been proposed which use virtual "reporter" agents within the game to produce real-time and post-game commentary tailored to the interests of individual users. However, to date, there has been no evaluation of these systems from a user perspective. In this paper we present the results of a live evaluation study performed using an instance of the online role-playing game *Neverwinter Nights* augmented with *witness-narrator agents* to provide in-game and post-game reports. Our results indicate that reporting does increase enjoyment of the game, and that players play for longer when their activities are recorded on a community web page, suggesting that agent-based reporting is a promising approach to community building in online games and social environments.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

1. INTRODUCTION

A key motivation for participating in persistent virtual environments such as Massively-Multiplayer Online Games (MMOGs) is to be recognised by your peers for the characters or objects that you create, and the events that you are involved in [21, 22]. Users "... *enjoy the story-telling aspects of these worlds and enjoy creating avatars with histories that extend and tie in with the stories and lore of the world.*"[21] Many environments feature weblogs maintained by participants, which contain reports of events in the world or live reportage, such "Second Life TV" (now "Treet TV" [19]), which aims to broadcast video feeds back into the *Second Life* world and onto the web. These can enhance the enjoyment of the game, e.g., by making it easier to coordinate activities with other players, and, more generally, being talked about is a way of

building a reputation and progressing in the community of players. The possibility of appearing in reports (e.g., when doing well in a game) can help to motivate players, and the reported events can, in turn, influence the players' future activities thus helping to drive events in the environment. The latter may be particularly important in MMORPGs where the quests and challenges are periodically reset.

The resulting 'collaboratively authored' narrative which emerges from the interactions of the participants can strengthen the sense of participation and involvement in the environment. Some narrative strands, e.g., those relating to major conflicts involving many participants which form the 'backstory' of the environment may be widely shared, while others, such as an account of an individual quest, may be only of interest to a single user.

To support the creation of such narratives, a number of in-game reporting and commentary systems have been developed [8, 13] which can produce real-time and post-game commentary which is individually tailored to the interests of the spectators. These systems use the embodiment of "reporter" agents within the game as a metaphor for the system's activities and as a natural interface to the capabilities of the reporting system, allowing the conflicting demands of privacy and disclosure of information to be negotiated. However, while some preliminary results regarding system performance in terms of the number of events reported have been presented, to date there has been no evaluation of such systems from a user perspective.

In this paper, we present a detailed evaluation of an embodied reporting system from both a technical and user perspective. In particular, we present results of a live evaluation study that addresses the key questions as to whether such reporting increases players' enjoyment of the game, their sense of participation in the game (and whether they play more when reported on), and whether the agent technology is disruptive to the normal game mechanics. Finally, we go on to discuss the implications of our work for the design of agent-based commentary systems.

The remainder of the paper is organised as follows. In section 2 we outline the architecture of our agent-based commentary system, and the main agent roles within the framework. In section 3, we briefly describe the integration of the architecture with a commercial multiplayer role-playing game. In section 4 we summarise the results of the evaluation of the implemented system, from both a technical and user perspective, focussing on how the system, and the reports it generates, was perceived by players. In section 6 we compare this paper to related work, and in section 5 we highlight some implications of our work for the design of agent-based commentary systems, before concluding with some suggestions for future work in section 7.

Figure 1: Screenshot of implemented system, showing witness-narrator agent avatar.

2. SYSTEM ARCHITECTURE

The design of the reporting system attempts not only to provide automatic narrative generation for large numbers of simultaneous users, but also to allow those users to interact with the system and influence the storylines that are generated about them. For this reason, we refer to the approach as *collaborative* narrative generation: authoring of the narrative becomes a collaborative effort between the system itself and the participants in the game environment. In order to support this approach, the design of the system is based on embodied agents, each having a limited viewpoint of the game world, similar to the view afforded to players themselves. Agents are represented in the game world by custom avatars, which are designed to be easily distinguishable from other players and characters, while also being smaller than most other avatars to limit any disruption they might cause (see Figure 1). Players interact with these embodied narrator agents, either directly by giving them instructions ('follow me' or 'leave me alone'), or indirectly by adapting their behaviour when narrator agents are present, either to 'play up' for more interesting stories, or to avoid giving away their current game tactics. To ensure good coverage of the environment, and to develop a broader view of activities within the game, the agents are organised into a society, with different agents performing different roles. In the remainder of this section we briefly describe these agent roles and capabilities, and how the team as a whole is coordinated.

The society of agents is broken down into three main areas of responsibility, defined by the following roles. These roles are loosely based on the three agent types described in [8].

Reporters are responsible for discovering interesting events occuring within the environment and gathering as much information about them as possible as they are occuring.

Editors are responsible for aggregating reports from multiple reporters, checking them for consistency and completeness, and then combining them into a single coherent narrative.

Presenters are responsible for relating generated stories to an audience via some output medium, formatting the narrative appropriately, as well as collecting feedback from the audience to update the current focus areas of the system.

In [8], these roles were realised as concrete agent types. In the system described here, these capabilities are abstracted into roles, allowing a more flexible relationship between an agent and the roles it is capable of performing. In particular, a single agent may adopt one or more of these roles, and can adjust the roles it performs over time. Two primary classes of agents are defined in the existing implementation. *Witness-narrator agents* (WNAs) are embodied in the environment and combine each of the three roles, being capable of detecting and recognising events, aggregating events from other WNAs, and presenting simple summaries of recent events to participants in-game. External *commentator* agents are much simpler and combine only the editor and presenter roles. These agents collect event reports from WNAs, edit them, and then present more fully-formed stories to an external output medium—in this case, a community weblog. Example output from a presenter agent is shown in Figure 2.

Kale Veuthian has completed a quest!

Kale Veuthian completed a quest today in Etum Castle District! Congratulations!

It all started when Kale Veuthian spoke to Pata Brows, a female human, in Etum Castle District. Pata Brows asked Kale Veuthian to embark on a quest to fetch a magical fairy from Galum Forest. Kale Veuthian travelled to Galum Forest, encountering many difficulties on his travels. Kale Veuthian then captured a magical fairy in the forest. Finally, Kale Veuthian gave the magical fairy to Pata Brows, receiving Amulet of Intellect as a reward.

Figure 2: An example report.

In operation, the framework is designed to work as follows.

1. Firstly, embodied reporting agents (i.e., agents adopting the reporter role) observe activities of players in the game world, and classify these events against a formal event ontology.

2. Next, reporters identify particular observed events that are *interesting* (where interest is defined as matching one of a current set of *focus goals*, described below), and collect as much information about these events into formal report structures.

3. These formal reports are then sent to relevant editor agents.

4. Editors combine reports coming in from reporters, identifying reports that refer to the same event, and aggregating these into combined report structures.

5. The editor continues to expand the detail of reports as further information arrives, attempting to build up a higher-level description of the events using (simplified) techniques from the plan and activity recognition literature [5]. For example, a *battle* may be recognised from multiple reports of smaller scale combat events, or a *quest* may be recognised from a particular sequence of actions carried out by a single player.

6. Presenter agents periodically make requests to editors asking for any reports matching a particular focus goal. Any matching reports are then sent to the presenter, in their current form, and formatted for publication to a particular output medium (currently a weblog).

7. Additionally, players can directly query in-world witness-narrator agents to discover interesting events that have happened in this region recently. In this case, the W-N agent

generates an ad-hoc request to nearby agents to produce reports on any recent events, which are then summarised and told to the player.

The key coordinating concept used within this system is the notion of a *focus goal*, which represents the current priorities of the system as to what sorts of events should be reported on. These focus goals are initially set by the designer of the system, but can then be adapted according to audience and player interactions. For example, players within the game can directly interact with a narrator agent causing the generation of new focus goals related to their current activities. Each focus goal consists of a description of the type of events to report on (currently specified as a single class from the event ontology along with some constant values for particular properties, such as involving a particular character), together with a description of the location and times at which such reports are required. Locations are specified hierarchically, and so can range from the entire environment down to a single region (e.g., a building). Location and time information is used to direct reporter agents to appropriate areas in the environment. Each focus goal is associated with a team of agents that are actively working to maximise coverage of that goal. The system as a whole maintains a set of top-level focus goals, specified when the system is initially deployed. All agents in the society are members of the team for these focus goals. Agents can also dynamically generate new sub-focus-goals in response to events (e.g., to report on a *particular* battle occuring now), in which case a new team is formed from nearby agents to concentrate on that particular goal.

The system was implemented using the *Jason* agent-oriented programming language [4]. Each role was implemented as a separate *capability module* [14], and these were then combined, together with coordination and deliberation components, to create two types of agent: embedded witness-narrator agents, and external commentator agents. Commentator agents exists purely to provide a gateway for reports to be published to an external medium, such as a weblog, and to gather any feedback (such as ratings) from that medium to influence overall system focus goals. The implementation is described in more detail in [13].

3. SYSTEM INTEGRATION

For the purposes of evaluation, the framework was integrated with an existing commercial role-playing game, *Neverwinter Nights* (NWN) [7]. NWN is a fairly typical role-playing game, in which player characters complete quests, battle monsters, and acquire treasure and experience to advance in the game. The game supports a multiplayer mode, with a single server capable of hosting up to around 60–70 players in a so-called 'persistent world', which can be kept running for months or even years. A number of development tools are supplied with the game to allow users (and researchers) to build their own persistent worlds, and to extend the capabilities of the system. The evaluation environment was developed by customising an existing popular persistent world implementation (known as *Rhun*), adding support for remote agents to create and manipulate avatars in the game, and to sense their immediate surroundings, with perception roughly equivalent to player characters.

The integration with NWN was achieved by intercepting events delivered to NPCs (non-player characters) in the game using the built-in scripting language, *NWScript*. The game was not designed to support this kind of third-party observation of events, and so instead we 'hook' into the events as they are delivered to each NPC and arrange for them to be also delivered to any nearby witness-narrator agents. Communications between the game server and the
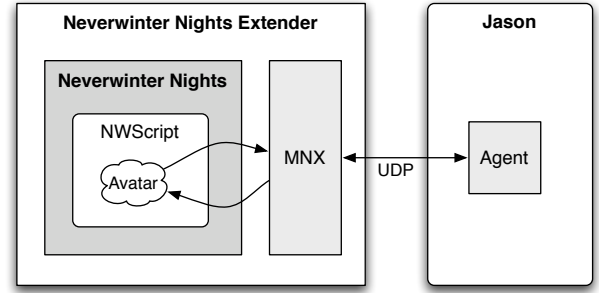
Figure 3: Integration of witness-narrator agents with *Neverwinter Nights*.

Jason agents are handled by the 'MNX' plugin for the *Neverwinter Nights Extender* (NWNX) project [17]. The MNX plugin allows scripts running on the server to communicate with the outside world using the UDP network protocol. The overall setup is shown in Figure 3. While event interception is possible for events involving NPCs, the same technique is not possible for player characters (PCs), as these characters are controlled directly by the player's inputs and not by scripts. In order to observe player events it is therefore necessary to also perform some polling. Each witness-narrator agent with an active reporter role periodically scans their surroundings to detect player characters, and then queries each such PC object to determine if they are currently performing an action. The NWN server limits the frequency of these polling requests for performance reasons to roughly one poll every 6 seconds, resulting in a significantly reduced rate of detection for events initiated by PCs compared to NPCs. Despite this, the detection rates were still high enough to make the study feasible. In addition to observing events, the NWN interface also periodically sends update messages to the witness-narrator agents detail the attributes of visible objects and characters (position, health, experience points etc.) as well as the agent's own current status.

Player characters can directly interact with witness-narrator agents via their avatars. The implementation of this uses NWN's built-in menu-driven conversation system. A player in the game can approach an embodied agent and engage them in conversation, bringing up a simple menu of options to choose from, much like interacting with any other NPC. The evaluated system allows players to choose one of four options: 'follow me', 'stop following me', 'who are you?' or 'what's going on around here?' They can also interact more bluntly with an agent by attacking it, in which case the agent will respond as if the player had told them to stop following them. Witness-narrator agents are marked as immortal in the game, and so cannot be killed. The reaction to damage instead provides a convenient way for a player to indicate that they do not wish to be reported on without having to engage the agent in conversation, for instance during a battle.

4. EVALUATION

The witness-narrator agent framework aims to increase the sense of participation in online role-playing games, and to foster a feeling amongst players that their participation contributes meaningfully to the evolution of an ongoing shared narrative experience. To evaluate the extent to which these aims have been achieved, we carried out an evaluation of the framework to determine how the players interacted with the witness-narrator agents, and whether the stories generated by the agents increased their enjoyment of the game. Specifically, we sought to answer four main questions:

- reports: do players find the reports accurate and interesting?

- impact: is the presence of the agents disruptive to normal gameplay or affect the way people play the game?

- participation: do players play the game more when witness-narrator agents are present?

- enjoyment: does appearing in a generated story make the game more enjoyable?

4.1 Methods

The evaluation was based on an extended live participation study. Given the relative novelty of the approach, this was primarily exploratory in nature, focussing on the players' perceptions of various aspects of the technology, rather than, e.g., a controlled study of individual features of the witness-narrator agent framework. The study was conducted by running a public *Neverwinter Nights* game server for an extended period and observing how players used the system. The data collected was mostly qualitative, consisting of comments and questionnaire responses from participants, and observations about their use of the technology. While player participation can be objectively measured by recording the amount of time players spend in the game, the questions about reports, impact and enjoyment cannot be answered using objective criteria. Therefore, although we collected data about the time the players spent playing the game, the actions they performed and whether these were reported, the focus of the evaluation was on the players' own assessment of their experiences with the witness-narrator framework as reported in a questionnaire completed by each participant at the completion of the study. The setup and results from this study are presented in section 4.3.

While appropriate for the questions we wished to address, the use of a live participation study placed certain constraints on the experimental design that could be employed. In particular, it was difficult to perform a controlled study, e.g., with participants randomly assigned to one of two servers: one with agents enabled, and one without. Given the social nature of NWN and a player's investment in the development of character in a particular environment, there is little reason for a player to play in a new instance of NWN unless it has some significant new features (such as WNAs). It is therefore difficult to recruit players to form a control group. However without a large control group, it would have been difficult to ensure that the players assigned to each server were sufficiently similar to allow meaningful comparisons to be drawn. More importantly, the focus on the players' subjective evaluation of the technology meant that it was not clear how to formulate the majority of the research questions so as to allow the collection of meaningful data. In particular, the questions about the quality of the reports, the impact of the agents and whether players enjoyed having their activities reported on are only meaningful in an environment which incorporates the witness-narrator agent framework. Given the difficulties and the limited amount of additional data that would be produced, we decided not to pursue a controlled study.

In addition to the live evaluation study, we also performed a quantitative evaluation of the coverage achieved by the system. These coverage test were entirely automatic, using scripted bots to generate events which match as closely as possible those generated in an environment with human players.

Three machines were used to host the main servers for both studies: a public *Neverwinter Nights* server, an agents server running all agents in a single Jason instance and a MySQL database which which held the persistent beliefs of the agents (such as archives of previously produced reports and information about players who

have been encountered), and a webserver which hosted the public web pages produced by an AtomPub [10] commentator agent. The *Neverwinter Nights* server (a 3.2GHz Pentium 4 with 1GB of RAM) ran the standard Linux version of the NWN standalone server. Initial performance testing indicated that a single dedicated machine (a 3.2GHz P4 with 2GB of RAM) was capable of running up to 100 Jason agents (sufficient for the evaluation study), together with the associated MySQL database. The use of MySQL reduced the memory overhead required for the agents which store large numbers of beliefs related to previous events, and allowed recovery in the event of a system crash.

4.2 Coverage Tests

The purpose of the coverage tests was to determine the number of witness-narrator agents required to achieve a given level of coverage. The results were used to determine how many agents to use in the live evaluation study, and how best to distribute them over the large environment. Coverage was measured by using scripted participants ("bots") to simulate human players. Each bot followed a random walk from its spawning point, attacking any creature that it encountered (except the witness-narrator agents). All actions performed by the bots were recorded by the *Neverwinter Nights* server to create a log of all the low-level events that were generated. The witness-narrator agents then fed reports to a special log presenter agent that also created a log file in the same format. These two log files were then compared to evaluate coverage (how may events were actually reported).

The experiment used a fixed size environment (10 areas) and a fixed number of bots (5 per area). We measured the minimum number of agents required to achieve a certain level of coverage (events observed / events generated $\times$ 100). We took measurements at coverage levels from 10% to 90% in 10% increments. Measurements were taken by gradually increasing the number of agents in the environment until we would reliably achieve $x\%$ coverage over a 5 minute interval. A maximum of 100 agents was used.

Coverage (%)	# Agents
10	10
20	35
30	50
40	65
50	95
60	100
70–100	—

Table 1: Coverage Results

Table 1 show the results of these experiments. As can be seen, coverage increases more or less linearly with the number of agents up to about 40% before levelling off, and is limited to 60% even with the maximum number of agents. Investigation suggested that this was due at least in part to the failure of the *Neverwinter Nights* plugin to forward events to the agents when the number of events is high. Previous work [9] has shown that users rate reports as accurate and informative at significantly lower coverage levels and the performance of the system was deemed acceptable for the live evaluation.

4.3 Live User Evaluation

Once the initial technical tests had been carried out, the main live evaluation study was run in order to evaluate how players interact with the framework, and how well the technology performs in a real world situation.

The approach taken to the live evaluation study was to host a public *Neverwinter Nights* server and to then recruit players from the general *Neverwinter Nights* community to play on the server, interact with witness-narrator agents and an external commentator agent, and to collect data on the performance of the system through comments, feedback forms, and a final questionnaire. Participants were recruited from the existing population of *Neverwinter Nights* players by advertising on the official NWN message forums, NWN groups on Facebook etc, following the model suggested in [6]. These external participants were supplemented by local players (mostly undergraduate and graduate students). Participants were informed of the purpose of the study (necessary for informed consent) and were given brief instructions describing their rights and what data would be collected. They were asked to play the game as they normally would. The data collected from participants during the study included game logs detailing the times at which they logged in to the game, and their activities in the game world. Participants were also able to access the website with reports of their activities, and could read and rate reports for interest and accuracy as well as completing the online questionnaire. All data collected from participants was anonymised once we had been able to verify that they had fulfilled the requirements of the study.

In total, 11 participants completed the study and the questionnaire. 80% of the participants were aged 20-29 and the majority were male (73%). The sample population shows a clear bias towards young male adults, characteristic of computer game players (and the representative of the target audience for the witness narrator agents) [21]. Of these participants, most (73%) had some previous experience with NWN or similar role-playing games, although only relatively few respondents indicated that they played regularly (36%). The median value for the number of hours played per week was around one hour, which is significantly lower than result reported for MMORPGs in general of around 20 hours [21, 22].

The study was carried out over approximately 2 weeks, with participants mostly playing for an hour or two on a single day. The resulting pattern of play limited the opportunities for interaction between players and the activities they could perform in the game. A newly created player is initially quite weak as a character, which limits their ability to undertake quests. As a result, the majority of reports produced were accounts of battles fought between players and various creatures in the environment. The only exception to this was a single occassion in which a player managed to 'level-up' (i.e., gain an experience level).

4.3.1 Reports

An example report of a quest from the live evaluation is shown in Figure 2. However, as noted above, due to the scope of the study most of the reports were battle reports (see Figure 4). The battle reports are perhaps the least interesting of the reports produced, often consisting of little more than a blow-by-blow account of the action, with little effort put in to making the report interesting from a narrative point of view. Despite these drawbacks, the majority of respondents (90%) rated the reports as "interesting", with only a single participant rating the reports as "not interesting" (see Figure 5).

The individual responses to this question are quite revealing. A selection of some of the responses are as follows:

- *"It's interesting to get some updates on what is going on, but the reports themselves are a bit dull."*

- *"I liked to read about my victories."*

- *"They reported on my interactions in the game which was*

Skirmish in Galum Hills!

There was a battle in Galum Hills today, involving Kale Veuthian, a Goblin and a Hobgoblin.

It all started when a Goblin attacked Kale Veuthian with their light mace. Then, Kale Veuthian attacked Goblin with their shortsword. The Goblin attacked Kale Veuthian with their light mace. Kale Veuthian attacked the Goblin with their shortsword. The Goblin attacked Kale Veuthian with their morningstar. Kale Veuthian attacked the Goblin with their shortsword. The Goblin was slain by Kale Veuthian. Kale Veuthian attacked a Hobgoblin with their shortsword. The Hobgoblin attacked Kale Veuthian with their longsword. Kale Veuthian attacked the Hobgoblin with their shortsword. The Hobgoblin attacked Kale Veuthian with their longsword. Finally, the Hobgoblin was slain by Kale Veuthian.

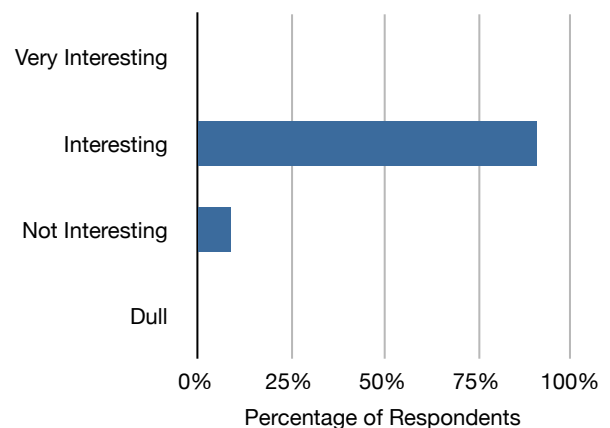**Figure 4: An example battle report.**

Figure 5: How interesting were the reports?

usefull [sic] in backtracking over events and detailed my killinmg [sic] spree."

From these responses, we can see that the main interest for respondents lay in the chronicling of their individual accomplishments in battle. However, several respondents noted the repetitive nature of the reports.

A corollary of the relatively detailed nature of the reports generated during the live study was that the resulting reports were rated as being quite accurate. All respondents felt that the reports were accurate or very accurate, with almost half (45%) rating the reports as "very accurate". Sample responses to this question included:

- *"I didn't see any discrepancies with the reports and what actually happened as far as I can tell."*

- *"[T]hey reported on everything I did."*

- *"[The agents] seem to leave out some events and the order is not always accurate."*

While the participants were appeared to generally happy with the web-based reports, their evaluation of the in-game reporting is less clear. One participant commented that *"nothing ever seems to be going on when I ask them."* This latter complaint is likely

due to the relatively small scale nature of the study. It is expected that with a larger group of participants it would be more likely that the witness-narrator agents would have something to report when asked to do so in-game.

4.3.2 Impact

Figure 6 shows the responses to whether the witness-narrator agents were disruptive to the game-play. Despite their small size, the agents were found to be "slightly disruptive" by the majority of respondents, with only 3 people stating that the agents were not at all disruptive to the gameplay. However, no respondents felt that the agents represented a major source of disruption to their usual gameplay. From the free-text responses to this question, the disruption is due to the agents sometimes getting in the way of the players, rather than, e.g., issues of privacy. From observing players during the game, it appears that this is partially the result of teamwork: when a battle breaks out the scene can quickly become flooded as a number of nearby witness-narrator agents immediately move into the vicinity in order to ensure good coverage of the battle. Another source of complaint was that when an agent is following a player, the agent will sometimes 'jump' directly into the path of the player. This seems to be an artefact of the implementation of the follow behaviour in *Neverwinter Nights*: if a player moves too far away from the agent, the game will simply 'teleport' the agent back to a position near the player, which sometimes happens to be directly in the way. Some sample comments from this question include:

- *"Sometimes there would be multiple agents and they would block my way, but this did not happen very often."*

- *"They're not intrusive byt [sic] they sometimes get in the way. They sometimes jump right in front of you when you are moving."*

However the intrusiveness of the agents was felt to be engaging by some players, who liked the "paparazzi scrum" (see Section 4.3.4 below).
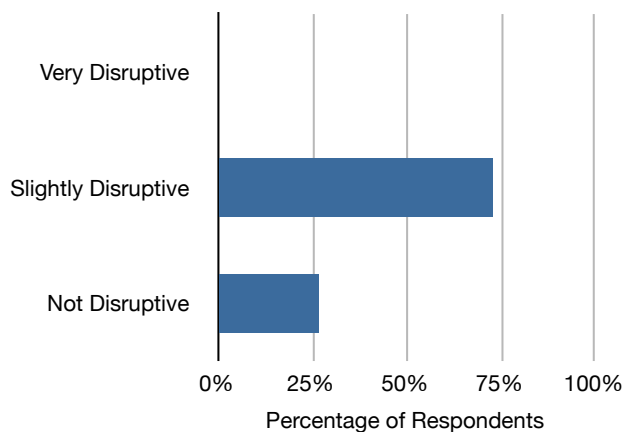

Figure 6: Were the agents disruptive in any way?

4.3.3 Participation

Figure 7 shows the response of the participants to whether they felt they played the game more or less when their actions were being reported on by the witness-narrator agents. The results show a clearly expressed preference for when the agents were present, with 7 respondents preferring to play with the agents present, and only 4

being indifferent to the agents. No respondents said that they would play less when the agents are present. The main reason given for increased participation was that players enjoyed being able to read reports of what they had done after the fact. One respondent wrote *"I liked to be able to look back at what I had done"*.
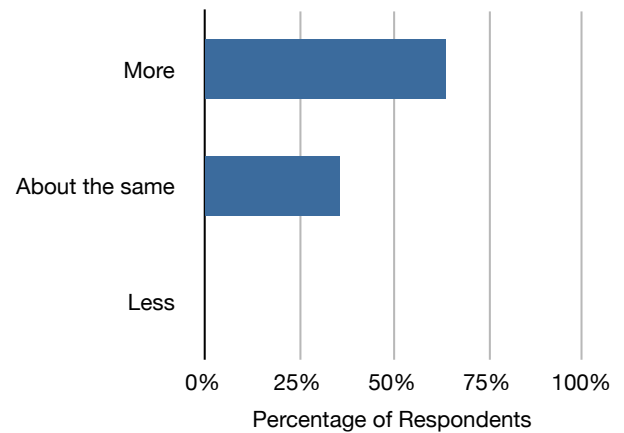

Figure 7: Did you play more or less when the agents were present?

4.3.4 Enjoyment

The results of the most important question—whether the presence of the agents actually increased participants' enjoyment of the game—are shown in Figure 8. These results are very promising, with all respondents (11) indicating that they enjoyed the game more because of the witness-narrator agents and having their actions reported on. The responses to this question reveal some of the reasons for this:

- *"It was cool to feel like what I did in the game was recorded and didn't end when I logged off."*

- *"[B]ecause I like to read about winning battles after I played."*

- *"Yes the reporters reacted to my actions much like a paparazzi scrum which was quite fun when on a killing spree."*
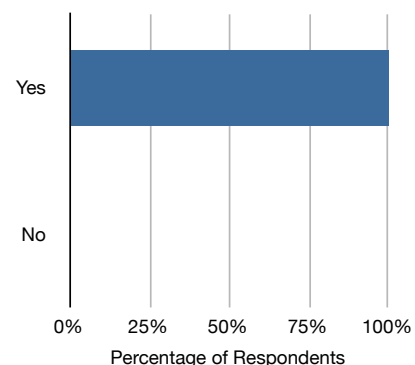

Figure 8: Did the agents increase your enjoyment of the game?

5. DISCUSSION

In this section we draw out some general lessons for the use of embodied agents in commentary systems, and on the use of games as a platform for MAS and AI research.

From the responses, we can see that the main interest for respondents lay in the chronicling of their individual accomplishments in battle. As stated above, the prose generation aspect of the presenter capability was not developed beyond basic text templating. The resulting lack of variation in the generated reports is clearly a limitation and several respondents commented on the repetitive nature of the reports. Although most users were satisfied with the textual reports, one participant suggested that screenshots would make the reports more interesting. The format and content of the reports is clearly an area that requires more work.

Some participants noted that some events were missed and/or reported out of sequence. It is not clear if this is attributable to limitations of coverage discussed in section 3, or whether the detailed nature of the reports ecourages the players to expect comprehensive reporting. It would be interesting to know if participants would be more tolerant of missing detail in more complex reports of, e.g., a quest.

One participant commented on the limited in-game presenting capability of the witness-narrator agents, stating *"nothing ever seems to be going on when I ask them."* This may be due to the limited scope of the study. With a larger group of participants it is more likely that the witness-narrator agents would have something to report when asked to do so in-game.

While the responses to the participation question are encouraging, with a majority preferring to play when the agents were present, the results need to be interpreted with caution as the actual amount of time spent in the game world during the study by individual particpants was rather low. The greatest time spent in the game was a single participant who played on several consecutive days and for a total of around 5 hours. It therefore difficult to draw conclusions regarding the participants' likely responses to the agents over a longer period.

Lastly, we make some general observations regarding the use of games as a platform for AI. As argued in [12], games have many advantages for AI (and particularly agent) research. They provide a dynamic environment, with sufficient complexity to be challenging without being overwhelming, while at the same time allowing more or less natural interaction with human users. However, while many games engines provide a relatively clean API, this is often not designed with agent interaction in mind. Integration exercises such as that described here are facilitated by rich, sciptable APIs, with good documentation (and preferably access to source code). The games which are good candidates as research platforms are often fairly mature, and hence have a relatively small population of active players. The consequent difficulties in recruiting participants for studies such as that described here can, to some extent at least, negate the advantages of using games as a platform for research.

6. RELATED WORK

Game commentary systems are a well established area of research and development, and a number of existing technologies support the concept. Games routinely provide rudimentary commentaries in the form of high-score tables and in-game notifications of other players' achievements. Recently there has been some effort to publish such data more widely, such as the 'XBox Live gamerscore' system [20] which collects data on how many 'achievements' each player has accumulated in the games they play and displays this information to other players.

A number of systems have been developed using agents to present commentaries in an engaging way. The Virtual Human Presenter [16] uses an agent to present pre-scripted material, while assistant agents point to written notes on a whiteboard. The Life-Like Presentation Agent [2] adopts a similar approach, but with more support for creating presentations using different media types. The best known commercial example of an animated news presenter is Ananova [1].

The RoboCup football competition for robots [11] has become a focus for automated game commentary systems [3]. The ROCCO system implements a commentator system for RoboCup by analysing a visual scene graph to generate reports of matches [3]. Two animated agents have been created to present commentary from ROCCO, introducing personality and an element of bias into their presentation [2]. The MIKE commentary system uses distributed agents to analyse and commentate on RoboCup in real-time [3].

The closest work to the witness-narrator agents is the reporting agents framework developed by Fielding et al. and applied to the task of reporting on events occurring in the *Unreal Tournament* fast-paced action game [8]. The framework consists of dedicated reporter, editor and presenter agents that cooperate to produce commentary in a number of different formats, such as a live IRC chat channel or a post-match website summary. The framework evaluated in the current paper builds on the notions from this earlier work, expanding the generality and capabilities of the system to scale up to massively multiplayer online games. While this previous work was evaluated in terms of the level of coverage that could be achieved in various configurations, little evaluation was performed from a user perspective.

However while there has been considerable work on techniques for producing commentary from computer games, there has been relatively little work on evaluating the impact that such systems have on players. In particular, the questions of whether these commentary systems increase enjoyment of the game, or whether they are disruptive, have not been addressed.

7. CONCLUSIONS AND FUTURE WORK

In this paper, we presented a detailed evaluation of an agent-based commentary system for computer games from both a technical and user perspective. A live evaluation study was used to investigate four key questions: as to whether such reporting increases players' enjoyment of the game, their sense of participation in the game (and whether they play more when reported on), whether the agent technology is disruptive to the normal game mechanics, and the quality of the reports produced. The study was performed using an instance of the online role-playing game *Neverwinter Nights* augmented with *witness-narrator agents* to provide in-game and post-game reports. Our results indicate that reporting does increase enjoyment of the game, and that players play for longer when their activities are recorded on a community web page, suggesting that agent-based reporting is a promising approach to community building in online games and social environments.

The evaluation also revealed a number of problems with the current system. Perhaps the biggest current drawback is the comparatively poor narrative prose generation that is performed. Improvement of this capability requires work in at least two areas:

1. development of a more comprehensive description of a story, incorporating ideas of dramatic arc in order to produce more compelling content;

2. incorporation of more sophisticated natural language generation (NLG) in order to create more variation and conciseness in the generated prose.

The current approach relies on simple text templates and some very rudimentary rules for generating some basic variation in the text produced. These techniques, while adequate, limit the quality of the produced output. A clear area for future work on the framework would therefore have to begin with improving the text output capabilities, perhaps incorporating a complete NLG component. This suggests the introduction of a new *narrator* role into the agent framework, such as that described in [18]. We are also interested in integrating methods to improve the variation in generated narratives, using techniques such as those proposed in [15].

The other major area for future research is in the development of the idea of *motive recognition* for inferring the deeper reasons behind players' actions. We believe that a significant aspect of what constitutes an interesting story relates to the motives, schemes, and tactics that underly character interactions. These aspects are currently woefully underexplored in this work. Future work will look at exploring more comprehensive plan recognition, as well as looking at deontological and social aspects of characters to reveal conflicts of interest, double-crossing, revenge, betrayal, and other motivations that could potentially significantly improve the narratives that are produced.

The multi-agent teamwork approach taken has been largely successful, however more research is required to better understand how to coordinate teams in larger scale game environments capable of supporting thousands rather than tens of players. The use of focus goals is also a promising direction for further work. For example, the rules for forming and disbanding teams in response to new and ongoing focus goals could be further investigated, to develop a clear approach for when a focus goal can be considered achieved or otherwise able to be dropped.

8. REFERENCES

[1] Ananova. http://www.ananova.com/video. (Accessed 13th October 2009).

[2] E. André and T. Rist. Presenting through performing: On the use of multiple animated characters in knowledge-based presentation systems. In *Proceedings of the International Conference on Intelligent User Interfaces*, pages 1–8. ACM Press, 2000.

[3] Elisabeth André, Kim Binsted, Kumiko Tanaka-Ishii, Sean Luke, Gerf Herzog, and Thomas Rist. Three RoboCup simulation league commentator systems. *AI Magazine*, 21(1):57–66, March 2000.

[4] Rafael H. Bordini, Jomi Fred Hübner, and Michael Wooldridge. *Programming Multi-Agent Systems in AgentSpeak using Jason*. John Wiley & Sons Ltd, 2007.

[5] Sandra Carberry. Techniques for plan recognition. *User Modeling and User-Adapted Interaction*, 11(1–2):31–48, 2001.

[6] R. Coomber. Using the Internet for survey research. *Sociological Research Online*, 2(2), 1997. http://www.socresonline.org.uk/2/2/2.html.

[7] Bioware Corp. Neverwinter Nights. http://nwn.bioware.com, 2002. (Accessed 13th October 2009).

[8] Dan Fielding, Mike Fraser, Brian Logan, and Steve Benford. Extending game participation with embodied reporting agents. In *Proceedings of the ACM SIGCHI International Conference on Advances in Computer Entertainment Technology (ACE 2004)*, Singapore, June 2004. ACM Press.

[9] Daniel Fielding, Brian Logan, and Steve Benford. Balancing the needs of players and spectators in agent-based commentary systems. In *Proceedings of the Fifth International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2006)*, pages 996–998, Hakodate, Hokkaido, Japan, May 2006.

[10] J. Gregorio and B. de Hora. The Atom Publishing Protocol. Technical Report RFC 5023, The Internet Engineering Task Force (IETF), October 2007.

[11] H. Kitano, M. Asada, Y. Kuniyoshi, I. Noda, and E. Osawa. Robocup: The robot world cup initiative. In *Proceedings Autonomous Agents 1997*, pages 340–347. ACM, 1997.

[12] John E. Laird and Michael van Lent. Human-level AI's killer application: Interactive computer games. *AI Magazine*, 22(2):15–26, Summer 2001.

[13] Neil Madden and Brian Logan. Collaborative narrative generation in persistent virtual environments. In *Proceedings of the AAAI Fall Symposium on Intelligent Narrative Technologies*, Arlington, Virginia, USA, November 2007.

[14] Neil Madden and Brian Logan. Modularity and compositionality in Jason. In *Proceedings of ProMAS 2009*, Budapest, Hungary, May 2009.

[15] Nick Montfort. Ordering events in interactive fiction narratives. In *Proceedings of the AAAI Fall Symposium on Intelligent Narrative Technologies*, 2007.

[16] T. Noma and N. Badler. A virtual human presenter. In *Proc. IJCAI Workshop on Animated Interface Agents: Making Them Intelligent*, pages 45–51, 1997.

[17] Neverwinter nights extender (NWNX). http://www.nwnx.org. (Accessed 13th October 2009).

[18] Ivo Swartjes and Mariët Theune. The virtual storyteller: Story generation by simulation. In *Proc. 20th Belgian-Netherlands Conference on Artificial Intelligence (BNAIC)*, pages 257–264, October 2008.

[19] Treet TV. http://treet.tv. (Accessed 13th October 2009).

[20] Description of Xbox 360 gamer profiles. http://support.microsoft.com/kb/905882/?sd=xbox. (Accessed 13th October 2009).

[21] Nick Yee. The demographics, motivations, and derived experiences of users of massively multi-user online graphical environments. *PRESENCE: Teleoperators and Virtual Environments*, 15(3):309–329, 2006.

[22] Nick Yee. Motivations of play in online games. *CyberPsychology and Behavior*, 9:772–775, 2006.