Importance of Well-Motivated Characters in Interactive Narratives: An Empirical Evaluation

Mei Si, Stacy Marsella, and David Pynadath

Institute for Creative Technologies
University of Southern California
Marina del Rey, CA 90292
{meisi,marsella,pynadath}@ict.usc.edu

Abstract. In order for the author to create his/her intended effects using interactive narratives, the user has to be able to understand his/her experience as designed by the author. In this paper, we argue that a key desideratum for interactive narrative frameworks is to model the characters' motivational consistency during the interaction. This work reports an empirical study for evaluating the importance of using well-motivated characters in interactive narratives. The results demonstrate that inconsistency in the characters' motivations can confuse the user and affect the user's expectations and interpretations of the events in the story.

1 Introduction

Narrative is a central part of the human experience. Its power to shape people's minds and affect people's behavior has been recognized throughout recorded history. With the rapid development of computer technology, a new form of media – computer aided interactive narrative has received increasing attention. It allows the user to take a role in a story and interact with other characters controlled by the system. The user's experience forms a unique story based on his/her interactions. The design of interactive narratives is often facilitated by authoring frameworks which can automatically generate the characters' behaviors during the interaction.

In this paper, we argue that a key desideratum for interactive narrative frameworks is to model the characters' motivational consistency during the interaction. In order for the author to create his/her intended effects, the user has to be able to understand his/her experience, i.e. what happened and why it happened, as designed by the author. The coherence of narrative, which refers to the sequence of events in the story having meaningful connections in both temporal and causal ways [11,1], has been identified as a crucial factor for ensuring that people understand their narrative experiences [17,2,8,10]. In the context of social interaction, a story being coherent usually requires the characters in it to be well-motivated. Thus, the characters' behaviors are interpretable to the audience/users.

Thespian is an interactive narrative framework designed with the characters' motivational consistency as a central concern. Thespian uses a two-layer system to drive the characters' interactions with the user. At the base is a multi-agent system comprised of goal-oriented autonomous agents that realize the characters in the story [13]. The
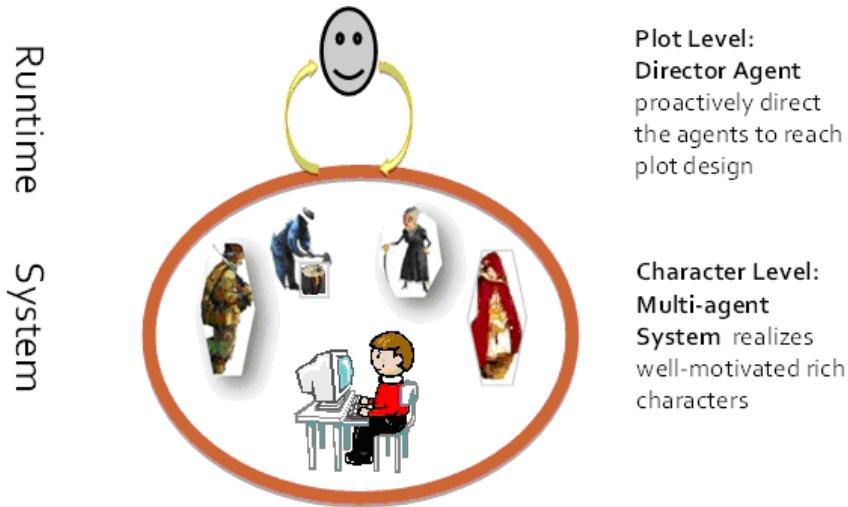
Fig. 1. Two-layer System for Interactive Narrative

characters' motivations are encoded as the agents' goals. Above this layer is a proactive director agent that continuously monitors the progress of the story and directs the characters toward the author's directorial goals [14]. Directorial goals are used by the author to indicate how he/she wants the story to progress, such as when an event should happen, or a character should change its belief about another character. Thespian supports directorial goals expressed as a combination of temporal and partial order constraints on the characters' actions and beliefs (including the user's).

A key aspect of Thespian's directorial control layer is that the director agent has access to models of the agents and the user. Automated approaches have been developed to model the characters' motivational consistency during the interaction and prevent broken characters from being created for reaching the directorial goals. Of course, the author can always specify the situations where the characters' motivations change.

In contrast, much of the computational approaches to interactive narratives does not directly address the characters' motivational consistency during the interaction [7,3,9,12,16]. As a result, the author has to either manually exam the huge amount of paths through the story to ensure that the characters' actions are consistent with their motivations, or bear the risk of having inconsistent characters interacting with the user. However, there is no trade off between using well-motivated characters and achieving event design goals – forcing a character to act inconsistently with its motivations to create events in the story will not lead to the author's desired effects. In this work, we report an empirical evaluation on the importance of well-motivated characters in interactive narratives and demonstrate this effect. The stories used in this study were prepared using the Thespian framework. The details of the study and its results are presented in this paper, followed by discussion and future work.

2   Empirical Evaluation

This study is designed to exam the effect of having inconsistent characters in interactive narratives. In particular, it evaluates how inconsistency in a main character's motivations affects the user's experience and understanding of the story.

The Little Red Riding Hood story is used as the example domain. The story is implemented using the Thespian framework. The user's role is the wolf.

This section starts by giving a brief description of the event design goals supported by Thespian, which is necessary for understanding the design of the stores used in the study. The rest of the section presents the study's design, procedure, hypotheses, and how the materials for this study are prepared.

2.1   Thespian's Support for Directorial Goals

Currently, six different types of goals are supported as listed in Table 1. The author can combine any number of goals defined using this syntax to specify his/her designs of the story. The events in the syntax can be either an action, e.g. "wolf-eat-Granny" or a character's belief, e.g. "wolf: wolf's hunger = 0 (the wolf believes that the value of the wolf's state feature hunger is 0) ". "anybody" can be used in defining actions in directorial goals. It indicates that the corresponding field of the action can be filled with any character, e.g. "anybody-kill-wolf". The approach used by Thespian for achieving directorial goals are discussed in [14], and [15] presents an evaluation of the effectiveness of the approach.

2.2   Experimental Design

The study utilized a 2 by 2 between-group design. Two factors were varied: what the author wants the user to experience and whether all characters have consistent motivations

Table 1. Syntax for Specifying Directorial Goals

orders =	$[event1, event2]$
	$event2$ should happen after $event1$
earlierThan =	$[event, step]$
	$event$ should happen before $step$ steps of interaction
laterThan =	$[event, step]$
	$event$ should happen after $step$ steps of interaction
earlierThan2 =	$[event1, event2, step]$
	$event2$ should happen within $step$ steps after $event1$ happened
laterThan2 =	$[event1, event2, step]$
	$event2$ should happen after $step$ steps after $event1$ happened
NoObjIfLater =	$[event, step]$
	if there is a constraint that requires $event$ to happen, and $event$ hasn't happen after $step$ steps of interaction, the constraint is not valid any more

during the interaction. Each of the independent variables had two variations. The virtual characters' motivations were either kept consistent or allowed to be inconsistent during the interaction. The two sets of directorial goals listed in Tables 2 and 3 were used for defining two different target effects of the story, as illustrated in Figures 2a and 2b respectively. The first set of directorial goals ideally will create stories in which the user's dramatic experience contains multiple exciting moments, and the climax is reached at the end of the story. The second set of directorial goals tries to restrict the user's experience of dramatic moments to two instances – a smaller spike at the beginning and the climax at the end of the story.

Table 2. Directorial Goals I

orders =	[[wolf-eat-Granny, anybody-kill-wolf],
	[Red-giveCake-Granny, wolf-eat-Red]
	[Red-giveCake-Granny, wolf-eat-Granny]]
earlierThan =	[60: [anybody-talkAboutGranny-wolf], 90: [wolf-eat-Red], 120: [wolf-eat-Granny]]
earlierThan2 =	[(wolf-eat-Granny, 30, [anybody-kill-wolf])]
NoObjIfLater =	[95: [wolf-eat-Granny]
laterThan =	[wolf-eat-Granny: 90, wolf-eat-Red: 60]
laterThan2 =	[(wolf-eat-Red, 10, wolf-eat-Granny)]

Table 3. Directorial Goals II

orders =	[[wolf-eat-Granny, anybody-kill-wolf],
	[Red-giveCake-Granny, wolf-eat-Red]
	[Red-giveCake-Granny, wolf-eat-Granny]]
earlierThan =	[30: [anybody-talkAboutGranny-wolf], 120: [wolf-eat-Red], 120: [wolf-eat-Granny]]
earlierThan2 =	[(wolf-eat-Granny, 30, [anybody-kill-wolf])]
laterThan =	[wolf-eat-Granny:90, wolf-eat-Red:90]
laterThan2 =	[(anybody-talkAboutGranny-wolf, 50, wolf-eat-Granny)]

The dependent variables are the type of the story experienced by the subjects and their understandings of the characters' motivations and relationships. To simplify the data collection process, the subjects did not interact directly with the interactive narrative system. Instead, they watched animated interaction histories, and were instructed to imagine that he/she is the user. The subjects' answers of what they think are the user's experience is used as the estimation of the real user's experience. The data were collected through a post-test survey. χ^2 tests were used for examining whether the subjects' answers in different conditions are statistically different[1].

---

[1] Alternatively, the subjects could directly interact with the interactive narrative system. However, in that case a subject's data would only be valid when the interaction is consistent with the directorial goals; but directorial control does not always succeed (see [15] for details).

Table 4. Conditions

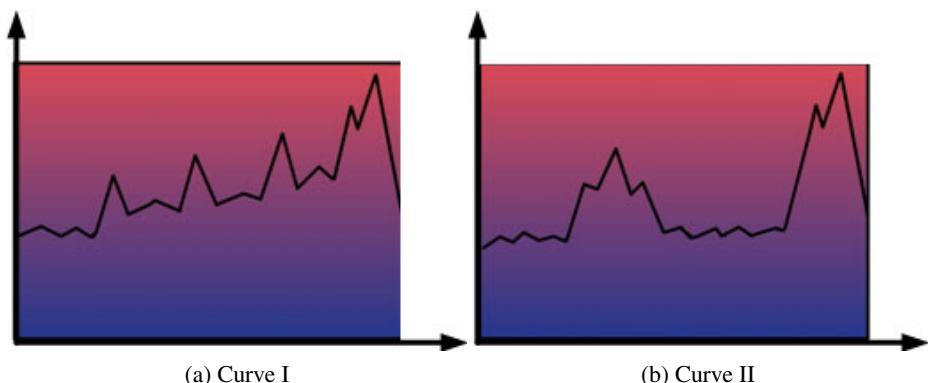|	Directorial Goal I	Directorial Goal II
Consistent Motivations	Condition I (76)	Condition II (69)
Inconsistent Motivations	Condition III (67)	Condition IV (74)

2.3   Procedure

This study was conducted online. Subjects between the ages of 18 and 40 were re-cruited via the Internet. The recruiting ad was posted at Craglist.com and other similar advertising websites. The subjects were informed that one winner will be automatically selected from every 40 subjects. The winner was awarded a $50 gift card.

The whole study takes around 15 minutes to complete. Subjects were randomly as-signed to one of the four conditions, as listed in Table 4. In all the conditions, the subjects first read a background story, then read/watched an animated story, and finally filled out a questionnaire. The subjects were instructed to imagine that he/she is the user, who plays the character wolf, while reading/watching the story. The background story was the same for all the conditions, and the animated story varied for each condition.

2.4   Materials

The background story provides the subjects basic information about the characters' motivations and abilities. In particular, it informs the subject that the wolf wants to eat Red and Granny, both the hunter and the woodcutters are capable of killing the wolf. Further, the woodcutter will only kill the wolf if he sees the wolf eating people.

The stories used in the consistent character motivations conditions (Conditions I & II) are recorded histories of human users interesting with Thespian. The director

(a) Curve I   (b) Curve II

Fig. 2. Event Design of the Story

agent was given the goals listed in Tables 2 and 3 respectively for generating the stories needed for Condition I and Condition II. One interaction history, where directorial control succeeds, is used for each condition. The stories used in the inconsistent character motivations conditions (Conditions III & IV) are replications of the corresponding stories used in the consistent character motivations conditions, with the actor of one key event replaced with another character whose motivation is broken by conducting the action. More specifically, in this experiment the event of Red telling the wolf about Granny's location is replaced with the hunter telling the wolf the same information.

Outputs from Thespian are text-based story composed of dialogue acts. To make it more natural for the user to read, a Java applet was developed to automatically convert dialogue acts to surface sentences, show the story, and illustrate the story with pictures, i.e. show texts with their corresponding pictures.

The post-test questionnaire contains two parts. In the first part, the subjects were asked to choose from the two curves (Figures 2a and 2b) the one that better describes the protagonist's – the wolf's – dramatic experience during the story.

In the second part, the subjects need to fill out a short survey consisting of eight questions based on what they have been shown so far – the background story and the animated story. These questions are designed to collect information on the subjects' comprehension of the story – their beliefs of the motivations of the characters, the relationships among the characters, and the characters' predictions about other characters' behaviors which inform the subjects' understanding of the characters' relationships. Each question is a statement, such as "the hunter will kill the wolf whenever he gets a chance." The subjects need to indicate whether the statement is true, false, or they cannot decide based on what they know.

2.5   Hypotheses

The main hypothesis of this study is that the users' experiences and comprehensions of the story are affected by the consistency of the characters' motivations. More specifically, it is hypothesized that regardless of the intended design of the story, in the inconsistent conditions (Conditions III & IV) more subjects will choose the two-spike curve (Figure 2b) to describe the wolf's experience. This is because when the virtual characters' motivations are inconsistent, the event that reveals the inconsistency becomes a significant event of the story. Together with the final scene, in which a character dies, this leads to two major dramatic moments in the story.

The above hypothesis is made based on the assumption that the subjects will notice the inconsistency in the hunter character's motivations. However, it is not uncommon for the audience to be tolerant to broken characters [6]. Therefore, the post-test survey tests whether the inconsistency is detected by the subjects.

The survey contains eight questions, which can be divided into two groups. The baseline questions (questions 2, 4, 5, 6, 8) ask about the relationships between the wolf, the woodcutter, Red and Granny. Their answers should be the same across all conditions. The rest of the questions (questions 1, 3, 7) are the "experimental" questions. They ask about the relationships between the wolf, the hunter and Granny. In the inconsistent conditions (Conditions III & IV), the hunter voluntarily provides Granny's location to the wolf. If the subjects paid attention to this unusual event, they are likely to feel

confused about the characters' relationships, and believe that the hunter has a wrong expectation about the wolf or the hunter has a bad relationship with Granny. Therefore, the subjects' answers to the "experimental" questions inform whether the inconsistency in the hunter's motivations is noticed, which would affect the prediction of the main hypothesis.

Finally, no hypothesis is made about how the design of the story (the directorial goal) affects the subjects' understanding of the characters' motivations and relationships.

3   Results

The data for this study were collected from 286 internet users in the United States over a 4-week period. Table 4 lists the conditions in this study and the number of subjects assigned to each condition.

The results of this evaluation confirm the hypotheses. The subjects did notice the inconsistency in the hunter's motivations. Further, when the characters' motivations are inconsistent, more subjects chose the two-spike curve (Figure 2b) to describe the wolf's experience regardless of the design of the story. The details of the results are presented below.

3.1   Subjects' Experiences of the Story

Table 5 summarizes the subjects' choices of dramatic experience in each condition. It can be observed that inconsistent character motivations lead to more subjects choosing the two-spike curve (Figure 2b) regardless of the design of the story. The subjects' choices in Condition I and Condition III are significantly different ($\chi^2 = 4.445$, p = .04). Similarly, their choices in Condition II and Condition IV are significantly different ($\chi^2 = 10.285$, p = .00).

Table 5. Subjects' Choices of the Wolf's Experience

Condition	Choose Curve I	Choose Curve II
I	66%	34%
II	55%	45%
III	54%	46%
IV	36%	64%

The results in Table 5 also show that when the virtual characters' motivations are kept consistent during the interaction, the first set of directorial goals realized the author's design of the story better than the second set of directorial goals. In Condition I, significantly more subjects chose Figure 2a ($\chi^2 = 7.579$, p = .01, compared to a 50% - 50% distribution). In Condition II, the subjects' choices were rather random ($\chi^2 = 0.71$, p = .40, compared to a 50% - 50% distribution). Both sets of the directorial goals were designed to create the corresponding dramatic experiences. Multiple reasons may contribute to this result. One possibility is the second set of goals is simply not designed

well for creating that experience. It is also possible that because the curve in Figure 2a is a common view of story – Aristotelian tension curve, people may tend to believe or expect that most stories have that structure.

3.2   Subjects' Comprehension of the Story

In reporting the results, the data from the two consistent character motivations conditions (Conditions I & II) are merged into one group, and the data from the two inconsistent character motivations conditions (Conditions III & IV) are merged into another group.

Experimental Questions. The results of the "experimental" questions are exactly as expected. In the inconsistent character motivations conditions, the hunter not only did not kill the wolf, but also informed the wolf of Granny's location. As a result, more subjects were confused about the relationships between the hunter, the wolf and Granny.

Question 1: "The hunter will kill the wolf whenever he gets a chance."
    In the inconsistent conditions, more subjects chose "cannot decide", and fewer subjects chose true. The difference is significant ($\chi^2 = 371.47$, p = .00).

Question 3: "The hunter and Granny don't get along well."
    In the inconsistent conditions, more subjects chose true, and fewer subjects chose false or "cannot decide". The difference is significant ($\chi^2 = 18.37$, p = .00).

Question 7: "The hunter knows the wolf will always eat people whenever it gets a chance."
    In the inconsistent conditions, more subjects chose "cannot decide", and fewer subjects chose true. The difference is significant ($\chi^2 = 26.05$, p = .00).

Baseline Questions. It has been hypothesized that the subjects' answers to the baseline questions are the same regardless of the experimental conditions. This hypothesis is confirmed in the subjects' answers to two questions: questions 2 and 5.

Question 2: "The wolf and the woodcutter are friends."
    Overall, the subjects' choices in the consistent conditions do not differ from those in the inconsistent conditions. Most of the subjects chose "cannot decide" ($\chi^2 = 0.05$, p = .98).

Question 5: "Red didn't expect the wolf to eat people."
    Overall, the subjects' choices in the consistent conditions do not differ from those in the inconsistent conditions ($\chi^2 = 3.17$, p = .21). The majority of the subjects chose true.

For questions 4, 6 and 8, the subjects' answers in the inconsistent conditions are in fact more consistent with Thespian's model of the story. These results could simply be an artifact. It is also possible that a more deliberate decision-making process was involved in the inconsistent conditions because the subjects felt confused [5,4]. None of these questions directly asks about information provided in the background story or in the animated story. To answer the questions, the subjects need to make inferences based on the information they know. A more deliberate decision-making process can help the subjects to understand the story better and therefore, make more correct choices. Following are the details of the results.

Question 4: "Red doesn't like Granny. She went to visit Granny just because her mum asked her to. She would rather Granny die."

In this story, Red is not modeled as disliking Granny. In the inconsistent conditions, more subjects chose false and fewer subjects chose "cannot decide". The difference is significant (χ^2 = 10.36, p = .01). In addition to the two possible explanations listed above, the fact that Red told the wolf where Granny lives in the consistent conditions may also cause this effect.

Question 6: "The woodcutter didn't expect the wolf to eat people."

In the inconsistent conditions, more subjects chose true, and fewer subjects chose false or "cannot decide". The difference is significant (χ^2 = 17.83, p = .00).

Question 8: "The wolf didn't eat Red at the first place because the woodcutter is close by."

In the inconsistent conditions, more subjects chose true, and fewer subjects chose false or "cannot decide". The difference is significant (χ^2 = 6.63, p = .04).

4   Discussion and Future Work

In this study, we used a simple example to demonstrate how broken characters can hurt the achievement of the author's desired effects. The broken character is easy to fix in this case. The author can simply add a special rule in the directorial goals to prevent the hunter from telling the wolf Granny's location. However, to detect all of such broken cases is a difficult problem. Because of the support of user interactivity, there are a huge amount of paths through the story. It is impossible for the author to follow each path and check whether the characters behave appropriately in it. Further, as more behavior rules are added by the author, they may start to conflict with each other. As a result, the author has to either sacrifice the richness of interaction or spend extensive effort to define the characters' behaviors. Therefore, it is a better design of the authoring framework if it contains sophisticated character models that can automatically reason about the characters' motivations while generating their behaviors.

On the other hand, people are also known to be tolerant to broken characters [6]. People may either ignore the inconsistency in the characters or use their own imagination to bring the gap. The latter case may or may not be what the author wants since once the user's interpretation of the story deviates from the author's design, the user's experience is less controlled by the author.

Thespian's default design enforces the characters to always behave consistently with their motivations. The author is allowed to set a threshold for the degree of inconsistency allowed in each character (see [14] for details). In the extreme case, the author can let the director agent ignore the constraint of maintaining consistent character motivations completely. However, currently there is no metrics for computationally assessing what degree of inconsistency in the characters can be allowed for a story. The author has to set the threshold based on experience or trial and error. In our future work, we plan to develop automated approaches and heuristics for helping the author set this threshold.

5   Conclusion

This work provides an empirical evaluation on the importance of using well-motivated characters in interactive narratives. In particular, it demonstrates that inconsistency in a main character's motivations affects the user's experience and understanding of the story, and therefore hurts the achievement of the author's desired effects. This result supports the design decision we made in Thespian: modeling well-motivated characters as the basis for generating the interaction and constraining the characters' behaviors with the event design of the story.

References

1. Bordwell, D., Thompson, K.: Film Art: an introduction, 5th edn. McGraw-Hill, New York (1997)
2. Bruner, J.: Acts of Meaning. Harvard University Press, Cambridge (1990)
3. Cavazza, M., Charles, F., Mead, S.J.: Agents' Interaction in Virtual Storytelling. In: de Antonio, A., Aylett, R.S., Ballin, D. (eds.) IVA 2001. LNCS (LNAI), vol. 2190, pp. 156–170. Springer, Heidelberg (2001)
4. Elsbach, K.D., Barr, P.S.: The effects of mood on individual's use of structured decision protocols. Organization Science 10, 181–198 (1999)
5. Forgas, J.P.: Mood and Judgment: The Affect Infusion Model (AIM). Psychological Bulletin 117, 39–66 (1995)
6. Kelso, M.T., Weyhrauch, P.W., Bates, J.: Dramatic Presence. Presence: Teleoperators and Virtual Environments 2, 1–15 (1993)
7. Mateas, M., Stern, A.: Integrating Plot, Character and Natural Language Processing in the Interactive Drama Façade. In: The International Conference on Technologies for Interactive Digital Storytelling and Entertainment (2003)
8. Nelson, K.: Narratives from the Crib. Harvard University Press, Cambridge (1989)
9. Nelson, M.J., Roberts, D.L., Isbell, C.L., Mateas, M.: Reinforcement Learning for Declarative OptimizationBased Drama Management. In: AAMAS (2006)
10. Ochs, E., Capps, L.: Living Narrative: Creating Lives in Everyday Storytelling. Harvard University Press, Cambridge (2001)
11. Onega, S., Landa, J.A.G.: Narratology: An Introduction. Longman, London (1996)
12. Riedl, M.O., Saretto, C.J., Young, R.M.: Managing Interaction Between Users and Agents in a Multi-agent Storytelling Environment. In: AAMAS, pp. 741–748 (2003)
13. Si, M., Marsella, S.C., Pynadath, D.V.: Thespian: Using Multi-Agent Fitting to Craft Interactive Drama. In: AAMAS, pp. 21–28 (2005)
14. Si, M., Marsella, S.C., Pynadath, D.V.: Directorial Control in a Decision-Theoretic Framework for Interactive Narrative. In: Proceedings of International Conference on Interactive Digital Storytelling, Guimarães, Portugal, pp. 221–233 (2009)
15. Si, M., Marsella, S.C., Pynadath, D.V.: Evaluating Directorial Control in a Character-Centric Interactive Narrative Framework. In: AAMAS, Toronto, Canada (2010)
16. Szilas, N.: IDtension: a narrative engine for interactive drama. In: The 1st International Conference on Technologies for Interactive Digital Storytelling and Entertainment, pp. 14–25 (March 2003)
17. Wilkens, T., Hughes, A., Wildemuth, B.M., Marchionini, G.: The Role of Narrative in Understanding Digital Video: An Exploratory Analysis. In: The Annual Meeting of the American Society for Information Science, pp. 323–329 (2003)