

Evaluating Group Decision Making Systems Using Role-Playing Games

Takeshi Nishida¹

Yuu Shibata¹

Takeo Igarashi^{1,2}

¹Department of Computer Science, The University of Tokyo / ²PRESTO, JST

7-3-1 Hongo, Bunkyo-ku, Tokyo JAPAN, 113-0033

tnishida@ui.is.s.u-tokyo.ac.jp

yuu@is.s.u-tokyo.ac.jp

takeo@acm.org

ABSTRACT

This paper proposes the use of role-playing games (RPGs) to evaluate group decision making systems, and introduces a game scenario, *Lupus in Tabula*, which we recommend for this purpose. We briefly describe the game scenario and the reasons why we recommend it. Finally, we report the results of a case study in which it was applied to a two-dimensional avatar chat. The results suggest that RPG-based evaluation can rapidly reveal the central issues associated with the success or failure of the system.

Categories and Subject Descriptors

H.5.3 [Information interfaces and presentation] Group and Organization Interfaces – *Evaluation/methodology*

General Terms

Design, Experimentation, Human Factors.

Keywords

Evaluation methods and strategies, Group decision making, Role-playing games, *Lupus in Tabula*.

1. INTRODUCTION

Evaluating CSCW systems [3, 6] is more difficult than evaluating single-user applications for several reasons. First, if we wish to run an evaluation in a real context using real tasks, it takes a long time to collect data. Even if we manage to run the system on real tasks, it is still difficult to obtain quantitative metrics for CSCW systems. In addition, it is nearly impossible to run multiple evaluations under identical conditions, making it difficult to compare multiple systems on an equal footing.

Replacing real tasks with virtual ones can solve many of these problems, simplifying the evaluation process and speeding up data collection [4, 5]. However, using virtual tasks increases the risk of missing the important issues related to the real usage, and quantitative metrics are still elusive. Moreover, it is difficult to recruit a large number of participants, especially if the task is boring.

We suggest using activities from role-playing games as a set of tasks for evaluating CSCW systems. We believe that RPGs are an appropriate abstraction for real tasks: detailed enough to offer real

challenges, but simple enough to be used in practical evaluations.

Our first target applications are group discussion and decision making systems: teleconference systems, BBS, chat, etc. After considering a variety of game scenarios, we chose the RPG *Lupus in Tabula* for our preliminary experiment. We will begin by describing the game scenario and the reasons why we recommend it for this purpose. We will then report the current status of our project and conclude with a discussion of future challenges.

2. LUPUS IN TABULA

2.1 Overview

Lupus in Tabula (Are you a werewolf?) is a RPG in which a team of villagers cooperates to defend their village from a team of werewolves. The players in this game are divided into two groups: humans and werewolves. The game continues until one group wins by eliminating all of its opponents.

The game progresses as follows. Werewolves attack a human every night, but it is difficult for the humans to catch them because they mingle with the villagers during the day. As a countermeasure against the werewolves, survivors gather every day to discuss the problem. At the end of the discussion, they lynch one person among them, whom they suspect of being a werewolf. The players repeat this day and night cycle until the massacre ends. Naturally, the length of each day and night can proceed at a rate faster than real time.

More detailed rules and instructions can be found on the official *Lupus in Tabula* website [2]. Note that local rules sometimes vary.

2.2 Playing Styles

Originally, *Lupus in Tabula* was played face-to-face; all players met in the same place and had an oral discussion. Werewolves had their own period of communication, called the *night phase*, during which they gestured silently to one another while the villagers kept their eyes closed.

Its recent implementation as a network game, via BBS and online chat, impacted the game in several ways. First, any non-verbal information, such as gestures and ambient atmosphere, was lost in the online version. Second, the werewolves could communicate with each other in more detail at any time, using the independent communication channel opened for them. The night phase was removed in these versions; the werewolves communicate and decide the victim during the day in the background.

2.3 *Lupus in Tabula* for Evaluations

RPGs have a rich potential for providing tasks for CSCW evaluations. Because RPGs are exciting, we can easily encourage

many people to repeatedly participate in the experiment. In addition, the task time is considerably shorter than that of a real task, enabling quick cycles of evaluation and improvement.

We believe that *Lupus in Tabula* is particularly well-suited for evaluating group decision making systems in three reasons. First, the discussion in *Lupus in Tabula* has a clear goal that is shared within the group, so that participants are never unsure what to do next. Since the two groups have directly opposing objectives, it is also possible to examine the performance of the design against nuisances.

Second, the task includes notification issues that arise from the day acycle. For example, it is important for the players to be aware of the time left for the discussion at any moment, and displaying a countdown clock is not always the best solution. This provides a chance to explore different methods of notifying players of events that are appropriate for various styles of collaboration.

Third, there are several measurable variables which may reflect the usability issues. We are especially interested in the question; does victory depend on the UI? We can answer this question by playing the game many times using different UIs. Our assumption is that the better UI produces better discussion, resulting in a higher winning percentage of the villagers.

2.4 Discussions

There are other games and scenarios appropriate for evaluation purposes. For example, social dilemma games are suited for evaluating trust in communication [4], and the Shape Factory game is a good model for distributed groups working together over the network [5]. Similarly, we believe that *Lupus in Tabula* is suited for evaluating group decision making systems.

We are not using different UIs for each group in a single game. Varying the UIs for each group means that two variables are varied at once. It would make analyzing the results more difficult. In addition, it might spoil the game by giving unnecessary information to the players (they might make inferences based on the fact that different UIs are used).

3. MESSYWOLF

We implemented a two-dimensional avatar chat system for playing *Lupus in Tabula* and used it to run experiments. We call our system MessyWolf, since like the MessyBoard system [1], it uses a layout in which messages move freely in two-dimensional space.

We offered an additional communication channel for the werewolves as a separate 2-D layer. By using this function properly, the werewolves could communicate with messages interspersed with the discussions of the villagers.

3.1 Selected Design Implications

We received numerous requests from users for the ability to emphasize certain messages. Since the system did not have such a mechanism, many participants simulated animation effects by repeatedly moving a message. In addition, some clever werewolves tried to divert suspicions by moving accusatory messages away from their avatars. This tactic might be prevented by allowing users to emphasize messages.



Figure 1 Arrow characters were preferred to indicate (a) future and (b) past topics (messages are in Japanese): (a) write ↓, if you have any idea, (b) ← ↑ one or other of two.

Players preferred using arrow characters to indicate past and future topics (Figure 1). However, these arrows became confusing when messages were later moved.

4. FUTURE WORK

We face a variety of future challenges. First and foremost, we must examine more carefully our evaluation strategy and the scope of its application. In particular, we are interested in examining additional measurable variables, such as winning percentages and utterance frequencies, since these variables may provide insight into usability issues.

Finally, we are planning to perform evaluation experiments that include a wide variety of collaboration styles, such as face-to-face, chat, video, and other techniques.

5. REFERENCES

- [1] Adam M. Fass, Jodi Forlizzi and Randy Pausch. MessyDesk and MessyBoard: Two Designs Inspired By the Goal of Improving Human Memory. In *Proc. of Designing Interactive Systems (DIS)*, 2002, pp. 303-311.
- [2] *Lupus in Tabula* (from the daVinci Games web site): http://www.davincigames.com/page_eng.cfm?sez=01&gioco=lit.
- [3] Dennis, C. Neale. John, M. Carroll. and Mary, Beth Rosson. Evaluating computer-supported cooperative work: models and frameworks. In *Proc. of the CSCW2004*, 2004, pp. 112-121.
- [4] Jens Riegelsberger, M. Angela Sasse and John D. McCarthy. The researcher's dilemma: evaluating trust in computer-mediated communication. In the *International Journal of Human-Computer Studies*, 58(6), 2003, pp. 759-781.
- [5] Bos, N., Olson, J., Nan, N., Shami, N., Hoch, S., and Johnston, E. Collocation bindness in partially distributed groups: is there a downside to being collocated? In *Proc. of CHI 06*, 2006, pp. 1313-1321.
- [6] Steven, R. Haynes. Sandeep, Puro. and Amie, L. Skattebo. Situating evaluation in scenarios of use. In *Proc. of the CSCW2004*, 2004, pp. 92-101.