Clareus Scientific Science and Engineering
Volume 2 Issue 6 July 2025
ISSN: 3065-1182

Clareus Scientific

Persuasion Process from the Perspective of Decision-Making Model in Werewolf Games

Citation: Asuka Nakai., et al. "Persuasion Process from the Perspective of Decision-Making Model in Werewolf Games". Clareus Scientific Science and Engineering 2.6 (2025): 05-13.

Article Type: Research Article

Received: May 26, 2025

Published: July 12, 2025

Asuka Nakai* and Takeshi Ito

The University of Electro-Communications, Chofu Tokyo, 1828585, Japan

***Corresponding Author:** Asuka Nakai, The University of Electro-Communications, Chofu Tokyo, 1828585, Japan.

Abstract

This study analyzed players' decision-making process in Werewolf and the success factors of persuasion in a 5-player Werewolf. In Experiment 1, we focused on the villager player and analyzed the process of deciding who to vote for based on what other players say. We proposed a decision-making model in which logical and economic rationality are prioritized, and when they still cannot decide, "empathy" becomes the deciding factor. Experiment 2 focused on persuasion situations and examined how successful persuasion is. The results confirmed that persuasion using logical persuasion materials is successful. This supported the model's validity in Experiment 1 that logical rationality is the overriding criterion for decision-making. However, we could not observe the success factors when both persuaders had logical material. Future research will investigate the aspects of successful persuasion more deeply through complex persuasion situations and games with more proficient players.

Keywords: Werewolf Game; Decision-making; Persuasion; Logical Rationality

Introduction

Werewolf games (hereafter referred to as "Werewolf") are classified as identity-hiding, multiplayer, and incomplete information communication games. They are one of the most popular party games among young people in Japan. Players are dealt their position cards at the start of play. They are informed whether they are in the werewolf camp (werewolves and maniacs) or the villager camp (others) but are not informed of players' positions other than themselves. The objective is to lead one's camp to victory by estimating the positions of others and falsifying one's position. This requires not only advanced communication skills but also logical reasoning. The "Artificial Intelligence based Werewolf Project" researches and develops AI and robots that play werewolves to realize artificial agents that can play with humans without discomfort [1]. To achieve this goal, it is essential to elucidate the thinking mechanism of human players, and research from a cognitive science perspective is also expected.

In this study, we will conduct a cognitive experiment to investigate the detailed thought process through interviews with players, and propose a cognitive model of the decision-making process of players. We focus on “villagers” and “maniacs” who cannot know the positions of others due to their positions in a 5-player Werewolf game (from now on referred to as “5-player Werewolf”).

Furthermore, focusing on persuasion behavior, which is often observed when players discuss with each other, we will try to decipher from the above decision-making model what factors are involved in the success or failure of persuasion.

Related works

Fujii et al.’s research can be cited as analyzing the human decision-making process [2]. To verify the reliability of the CVM (Contingent Valuation Method), which measures the value of the natural environment, Fujii et al. conducted a study to examine subjects’ decision-making process using verbal protocol analysis, which has been used in cognitive science. Specifically, subjects were given a situation, asked to read a question aloud, and then asked to utter everything that came to their minds up to the answer. The data from the start of the utterance to the point where the decision was made was used as the utterance data, and the contents were categorized and analyzed by item. The results suggest that “bias” cannot be adequately eliminated even by NOAA (National Oceanic and Atmospheric Administration) guidelines. Thus, verbal protocol analysis has the potential to reveal rigorous decision-making behavior. In this study, we want to continue clarifying the werewolves’ decision-making process by conducting a verbal protocol analysis.

Crano et al.’s research shows the study of attitudes and persuasion [3]. This research summarizes findings on the formation of people’s attitudes through persuasion. Crano et al. mentioned dual-process models such as the elaboration likelihood model (ELM) and the heuristic/systematic model (HSM) as thought-process models of the recipients of persuasion. These models comprise two different processes: conducting based on careful consideration of information and performing based on personal associations with the recipient or simple reasoning. Therefore, persuasive messages are classified according to two elements: logical elements, such as well-reasoned and data-based, and emotional elements, such as peripheral cues and heuristics.

Another analysis of persuasion content in werewolves is that of Zaitsu et al., who classified persuasion in werewolves into three types: “logical persuasion,” in which the player tells the opponent his/her viewpoint and the reasons for coming to that viewpoint, “emotional persuasion,” in which the player emphasizes the correctness of his/her argument, and “persuasion using meta-information,” in which the player uses information that cannot be known within the current game [4]. They also defined successful persuasion in werewolves as “not being selected as the enemy team at the end.” They analyzed the effects of the different persuasion contents on the persuaders by examining whether the experimental participants who received the three types of persuasion by the dialogue system trusted the persuasion contents. In this study, we would like to clarify what kind of persuasion content can win trust concerning the three types of persuasion described above.

5-player Werewolf

Werewolf are generally party games played with many players (9 or more). Still, while Werewolf with many players are as exciting as party games with many roles, many players need roles such as villagers. The game becomes slow in the early stages due to a lack of information, or some players are eliminated without much success or clear reasons, depending on their roles. Therefore, in this study, we will focus on 5-player Werewolf, which are said to be the smallest unit of werewolves.

In 5-player Werewolf, there is a minimum number of roles to be considered in a werewolf game: the werewolf camp of one werewolf and one maniac and the villager camp of one seer and two villagers. The game also retains the essential strategic elements of a werewolf, such as acquiring information through roles and deception. At the beginning of the game, the five players are assigned roles, and Day 1 begins with the seer divining one person. If the werewolf survives until the night of Day 2, the werewolf camp wins. Unlike the multiplayer werewolf game, there is a strategy of deception in all roles. Although the game is settled in a short period, it is a strategically deep game.

Experiment 1

Purpose

Focusing on the villagers who cannot know the roles of others due to their role abilities in 5-player Werewolf, we analyze how they decide who to vote for based on their utterances during the game and interviews immediately after the game to clarify their decision-making process.

Method

Twelve students (eleven males and one female), well-trained in 5-player Werewolf, were the experiment subjects. To ensure that the participants fully understood the basics of the 5-player Werewolf, they were asked to have played the game at least ten times.

The experiment used the following procedure: Participants were divided into three overlapping groups (each participant could participate in at most two groups), and the following procedure was repeated.

1. The participants were asked to play 5-player Werewolf, and a video of the werewolves was taken. In this experiment, the duration of one discussion at noon was set to 5 minutes.
2. Immediately after the end of the game, we showed the game video to one of the villagers who had survived until the Day 2 and asked him/her in an interview how he had decided who to vote for.
3. Villagers' statements in the game conversation and the interview were analyzed as follows.
4. The conversations of the game players were transcribed for each player over time and encoded with tags that matched the content of their statements. The tags used were those from other studies, plus those deemed necessary based on the results of preliminary experiments [5].
5. Interview statements were transcribed in the text next to the player's tagged conversation in the game. The starting point for describing these statements was aligned with the in-game video played during the interview.

In the game of Werewolf, players mainly engage in conversation during the game. Therefore, it is difficult to obtain the thoughts of players during the game using online speech protocols. Therefore, in this method, we decided to ask about the thought process in the form of an interview conducted immediately after However, we conducted the interviews carefully, being careful about this point, so we think that the interview contents are generally reliable. As this is a study that examines thought processes, it is an experiment that is difficult to analyze quantitatively due to its nature. We will attempt a qualitative analysis focusing on some typical interview results.

Results

Thirteen games were played, and interviews were conducted for six games that could be analyzed. Among them, we conducted a tagging analysis of the in-game conversation, focusing on games that fit the purpose of this study, i.e., games in which the villagers' decision-making decides the game-winner on the Day 2 of the game. Table 1 is one of the game records of the target game. T.1

First, explain the terms and symbols used in the game record. The game record is a table summarizing the game's flow, including each player's position, main statements, actions during the night, and where he/she voted. Table 1 shows the actions and statements from the night of Day 0 to the vote on Day 2, in chronological order. The squares of players eliminated from the game are blacked out to indicate they are no longer involved. Table 2 shows the meaning of abbreviations and symbols used in the game record. T.2

Let us look at the progression of this game in Table 1. In this game, on Day 1, a werewolf and a maniac come out as seers in addition to the seer. The timing of the maniac's coming out is a little late, and he is then suspected. Looking at the divination details, the seer in Player 2 suddenly finds the werewolf in Player 1, the werewolf claims the villager in Player 3 as a villager, and the maniac claims Player 3 as a werewolf. The two players, the maniac and the werewolf, disagree about their claims. Villager Player 4 will have to decide who to vote for on Day 1 under these circumstances. Also, on Day 2, Player 1 divines Player 2 to be a werewolf, and since Player 2 has already guessed werewolf, he still needs to mention the other decision results. From Player 4's point of view, both of them have succeeded

logically. This situation becomes a situation where one has to decide who to vote for. Player 4 is the player who has to decide based solely on what others have said so far and is the typical research subject of this study. Therefore, we focused our analysis on this game.

Player	1	2	3	4	5
Roles	Werewolf	Seer	Villager	Villager	Maniac
Day 0, Night		Pre>1			
Day 1, Daytime	CO:S	CO:S			CO:S
	Pre>3V	Pre>1W			
					Pre>3W
		Sus>5			
			Sus>5		
	Sus>2				
			Sus>1,2,5	Sus>1,2,5	
		Vote>1			
			Vote>5		
				Sus>1,2,5	
Day 1, Vote	5	1	5	5	3
Day 1, Night	At>3	Pre>3			
Day 2, Daytime	Pre>2W				
Day 2, Vote	2	1		1	

Table 1: A game record to be analyzed.

Terms and Symbols	meanings
CO:	Coming out one's role
COX	Withdraw the role coming out
Pre>	Informing the results of the prediction
PD>	Role estimation from the player's view
Po>	Possibility of other player roles
Sus>	Suspect a specific player
Obj>	Objections to specific players
Cons>	Consent to specific players
Qes>	Question to specific players
Vote>	Intention to vote for specific players
Pr_Vote>	Call for a vote for specific players
At>	Who the werewolf attacks

Table 2: Terms and symbols used in the game records and their meanings.

Interviews with Player 4 revealed the flow of this game, as seen by Player 4. This is outlined below. (● denotes objective observation, ○ denotes subjective content).

Day 1

- Player 2 claims to vote for the Player 1 based on his own divination result.
- Player 3 insists on voting for Player 5, which is doubtful from any player's point of view.
- Player 1 claims that Player 2 is suspicious because he insists that Player 1 is a werewolf.
  - Player 2's explanation is sympathetic.
  - Player 3's explanation seems to make sense.

Day 2

- Player 1 judges Player 2 as a werewolf.
- Player 2 claims that he is still alive so that Player 1 can say that Player 2 is a werewolf and get Player 4 to vote for him.
  - Player 2's explanation matches Player 3's explanation.
  - Lack of sympathy for Player 1 throughout.

Player	1	2	3	4	5
Roles	Werewolf	Seer	Villager	Villager	Maniac
Total number of statements	78	91	65	45	13
Illogical statement	0	5	1	2	0
Concurrence	15	4	5	6	1
Refutation	4	4	1	0	0
Statements directed at others	7	17	10	0	0
Statements from the self-perspective	6	12	5	0	0

Table 3: Statement aggregation results for each player.

Table 3 shows the total number of statements made by the players. Player 2 has more statements than Player 1. Looking at the content, neither Player 1 nor 2 showed any logical breakdown.

Discussion

The change in villager's thinking

Looking at the thought process of Player 4 (the villager) on Day 1, he compares the claims of Players 2 and 3, who were sympathetic, claiming who to vote for. Player 2 claimed that Player 1 was suspicious, while Player 3 claimed that Player 5 was suspicious. Although there is no difference between these two arguments regarding logical rationality, comparing their arguments, Player 2 and Player 3 decided that the latter's argument was better regarding the villagers' probability of winning. As a result, he decided to vote for Player 5, in agreement with Player 3's opinion.

On the Day 2, he reported that they did not sympathize with Player 1 compared to Player 2. Player 2 was suspicious of Player 1, and Player 1 was suspicious of Player 2. In terms of logical rationality, there is no difference between their claims. There is also no difference in the odds of winning other than the logical consequence that one of them is a werewolf. Under these circumstances, the final criterion for the judgment was "empathy." It was impossible to determine from the speech tags alone where the difference in sympathy for these two people came from.

Comparison of the number of statements

Here, we focused on the number of statements made by the players. Of players 1, 2, and 3, there was a difference in the amount of "attempts to execute other players" and "explaining the situation from one's point of view." These statements were more frequent in Players 2 and 3 than in Player 1, which may have given Player 4 sympathy during the play. This suggests that more statements about the execution target and explanations of the situation from the player's perspective may evoke sympathy for the villager players.

Conclusion

Based on the above, we can consider the decision-making process of the villager players in the order shown in Fig. 1 below.

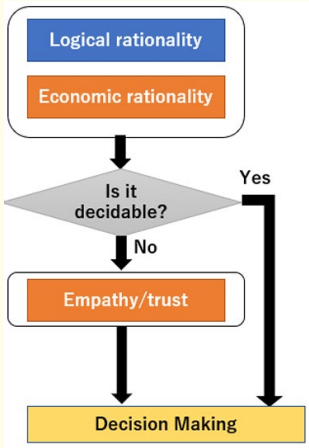

Figure 1: Decision-making process of players (villagers).

First, we focus on "logical rationality" to see if there are any logical contradictions. If there is no contradiction, they consider "economic rationality" based on their camp's winning rate. If no difference is found in either of these cases, the two opposing hypotheses are examined based on the credibility of the previous statements, such as the amount of speech from one's perspective.

Figure 1 shows the hypothesized decision-making model for this experiment. In this decision-making situation, each player is expected to try to persuade the other player about the two conflicting events. Therefore, in the next experiment, we would like to test the validity of this decision-making model by focusing on the persuasion situation and identifying what influences the success or failure of the persuasion.

Experiment 2

Purpose

This experiment focuses on persuasion situations and examines how persuasion is used in actual in-game discussions. Furthermore, we will discuss the success or failure of persuasion and determine what influences its success or failure.

Method

This experiment was conducted with 16 male students aged 20 or older who needed more experience with 5-player Werewolf. At first, the rules of 5-player Werewolf were explained to them so they could understand the game's content. Then, the participants were asked to play 5-player Werewolf, and the speech data for all 35 games were obtained by capturing the scene with a video camera. Since the participants took turns participating in the games, each participant played about 16 games.

Games to be analyzed and definition of successful persuasion

To make it easier to understand the persuaders and the persuaded, the situation to be analyzed was limited to "one person unsure about voting in the Day 2 discussion, while two others are trying to convince him or her. The number of games that met the above conditions was 10 out of 35. Furthermore, among these ten games, we conducted a detailed analysis of the speech of the successful and unsuccessful persuaders and their game development in two games in which the positions of the two persuaders were equal from the viewpoint of the persuaders. These games were relatively complex in determining who to vote for, with only the information given on Day 1 (divination results, Day 1 voting destination, etc.).

Results

The first of the two games analyzed is shown in Table 4. In this game, the maniac, the seer, and the villager came out as seers on Day 1, and the villager withdrew his coming out shortly after that. The divination details are as follows: the maniac in Player 2 claims that the werewolf in Player 5 is not a werewolf, and the actual seer in Player 3 claims that the villager in Player 1 is not a werewolf. As for Player 4, who withdrew his coming out, there were both speculations that he was a villager and speculations that he was in the werewolf camp, so only this player was not suspected. Player 5, who thought Player 4 was a villager, then suggested voting for either Player 1 or Player 5, who was likelier to be a werewolf. However, it was pointed out that at this stage, there was no proof that Player 4 was a villager and that there was a possibility that neither Player 1 nor Player 5 was in the werewolf camp, so it was suggested again to vote for either Player 2 or Player 3, one of which was definitely in the werewolf camp. The result was that Player 2, the maniac, was executed.

Player	1	2	3	4	5
Roles	Villager	Maniac	Seer	Villager	Werewolf
Day 0, Night			Pre>1		
Day 1, Daytime				CO:S	
		CO:S			
			CO:S		
				COX	
		Pre>5V	Pre>1V		
					PD>4V
			Sus>4W		
					Pr_Vote>1,5
	Sus>4				
					PD>4V
					Pr_Vote>2,3
			Pr_Vote>2,3		
					Qes>4
				Pr_Vote>2	
Day 1, Vote	2	3	2	2	3
Day 1, Night			Pre>4		At>3
Day 2, Daytime					PD>4V Sus>1W
				Sus>5W	
					Sus>4W
				Qes>1	
				PD>1V	PD>1V
					PD>5V
				PD>4V	
					Obj>4
Day 2, Vote	4			5	4

Table 4: First game record in the games analyzed.

On the Day 2, the discussion began with the two players who had come out as seers on the Day 1 gone. Assuming that Player 1 is a werewolf here, Player 3, who divined that Player 1 is not a werewolf, is lying, i.e., a maniac, from the standpoint of Player 1. So, Player 1 doesn't need to attack Player 3. Because if Player 3 remains, a Power Play can be made on Day 2. "Power Play" refers to a play in which a werewolf comes forward and votes against a villager in collusion with a maniac. However, Player 3 was attacked by the werewolf on the night of Day 1. This shows that Player 1 is doubtful to be a werewolf. In this situation, Player 4 and Player 5 must convince Player 1 that they are not werewolves.

Player 5 claimed he was not a werewolf because he had suggested on the 1st day that he be included in the voting list. In response, Player 4 countered but did not offer a logical explanation like Player 5. As a result, Player 1 did not support Player 4's counterargument and ultimately voted for Player 4, resulting in a victory for the werewolf camp.

Next, the game record of the second game among the games analyzed is shown in Table 5.

In this game, the werewolf, the maniac, and the villager came out to be the seer on Day 1. The werewolf and the villager withdrew their coming out immediately after that, and then the actual seer came out late. The divination details are that the maniac in Player 2 claims that the villager in Player 3 is not a werewolf, and the actual seer in Player 4 claims that the maniac in Player 2 is not a werewolf. Player 4 then proposed to vote for Player 3 based on his perspective, but the other players became suspicious of him because of his late coming out, and Player 4 was executed as a result.

Player	1	2	3	4	5
Roles	Werewolf	Maniac	Villager	Seer	Villager
Day 0, Night				Pre>2	
Day 1, Daytime	CO:S	CO:S			CO:S
	COX				COX
				CO:S	
		Pre>3V		Pre>2V	
				Sus>2M	
					Sus>4M
	Pr_Vote>1,4,5				
					Cons>1
				PD>1V, 5V Su>3W	
					Su>4M
				Pr_Vote>3	
		PD>3V			
			Po>3W, 2M		
				Pr_Vote>3	
		Su>4W, 3M			
					Pr_Vote>4
Day 1, Vote	4	3	4	3	4
Day 1, Night	At>2				
Day 2, Daytime					PD>3V Sus>1W
	Sus>5W				
					Sus>2M
					PD>5V
Day 2, Vote	5		1		1

Table 5: First game record in the games analyzed.

Assuming that Player 3 is a werewolf here, Player 2, who divined that Player 3 is not a werewolf, is lying, i.e., a maniac, from the standpoint of Player 3. So, Player 3 doesn't need to attack Player 2. Because if Player 3 remains, a Power Play can be made on Day 2.

However, Player 2 was attacked by the werewolf on the night of Day 1. This shows that Player 3 is doubtful to be a werewolf. In this situation, Player 1 and Player 5 must convince Player 3 that they are not werewolves. This development is similar to the first game.

In this game, there was no significant difference in behavior between Player 1 and Player 5 during the Day 1 discussion. However, Player 2 had voted for Player 3, who was clearly not a werewolf, when voting on Day 1. This indicates that Player 2 may be a maniac. Player 5 mentioned this point and insisted that he would not attack Player 2 if he were a werewolf, being aware of this. This assertion won Player 5 the trust of Player 3, and Player 3 voted for Player 1, resulting in a win for the villagers' camp.

Discussion

In both of the two games analyzed in this study, the players moved to the Day 2 with only one of the persuaders having persuasive material from the discussion on the Day 1, and the player who used that persuasive material to persuade logically was successful.

We speculate that the persuaders referred to "logical rationality," which considers logical contradictions in the werewolf player's decision-making process model proposed in Experiment 1. In this respect, the validity of this decision-making model was indicated. However, the examples of these two games showed only arguments that did not fall outside the "logical rationality" category in the decision-making process model.

In general, as players become more proficient at playing, they play various games on Day 1 to accumulate information that might be used to persuade them on Day 2. In Experiment 2, we did not initially see any plays where the villagers played the seers. Still, we gradually began to see plays in which they played the seers. This play is intended to avoid a Power Play situation on the Day 2 if the villagers do nothing. To prevent this, villagers will actively play coming out to complicate matters, gain more information, and leave more persuasive material on the Day 2.

How are decisions made by non-persuaders when players have evenly logical persuasive material on Day 2? To test the validity of the decision-making model in Figure 1, we need to investigate decision-making in more complex phases.

Conclusion

This study conducted two experiments on 5-player Werewolf to investigate the decision-making process and the success factors of persuasion in wargames. In Experiment 1, the model showed that logical and economic rationality are essential factors in villagers' decision-making and that empathy can ultimately be the deciding factor. Experiment 2 focused on the persuasion scene and examined the triumphant persuasion phase. As a result, having logical, persuasive materials and persuading based on them is adequate. In this respect, the validity of the proposed model, in which logical rationality is the preferential decision-making criterion in Experiment 1, was supported. However, we could not investigate how successful persuasion is when both persuaders have logical persuasion material and need help deciding which is more reasonable.

Therefore, we plan to conduct future experiments with more proficient players and investigate more complex persuasion situations to examine economic rationality and how empathy is achieved.

References

1. Artificial Intelligence based Werewolf. https://aiwolf.org/en/
2. Fujii S., et al. "An Analysis of Decision-Making Processes under CVM: A Cognitive-Psychological Test of NOAA's Guidelines". Journal of Japan Society of Civil Engineers, Ser. D3 (Infrastructure Planning and Management) 19 (2002): 91-98. In Japanese
3. Crano WD and Prislin R. "Attitude and persuasion". Annual Review of Psychology 57.1 (2006): 345-374.
4. Zaitsu T and Katagami D. "Persuasion and effects to human players by the AI-based werewolf game player agent". The 36th Annual Conference of the Japanese Society for Artificial Intelligence 4I3-OS-26b-03 (2022): 1-4. in Japanese
5. Sugimoto M and Ito T. "How do villagers make decision in the game of 5-players' werewolf?". The 34th Annual Meeting of Japanese Cognitive Science Society P1-26F (2017): 826-832. in Japanese