

Effect of Humor on User Interest in a Recommendation Chatbot

Tomoya Asakura

Grad. Sch. of Systems Information Science
Future University Hakodate
Hakodate, JAPAN
g2121002@fun.ac.jp

Asuka Terai

Dept. of Complex and Intelligent Systems
Future University Hakodate
Hakodate, JAPAN
aterai@fun.ac.jp

Abstract—This study investigated the effects from including humor stimuli in an interaction with a recommendation chatbot on user interest in a recommended item. Three types of chatbots were developed with different frequencies of humor stimuli. A psychological experiment was conducted to investigate the differences in user interest in the recommended item. As a result, no significant direct effect on user interest was observed depending on the frequency of the humor stimuli. Nevertheless, the sense of humor, trust, and humanity had effects on the user interest. Moreover, the results suggested that humor stimuli had a positive effect on sense of humor and negative effect on trust.

Index Terms—humor, chatbot, recommend, human-computer interaction, persuade

I. INTRODUCTION

Chatbots have become popular in recent years, and certain chatbots such as “Siri” and “Google Home” incorporate humor to facilitate interactions with a person. However, chatbots for business, such as recommendation chatbots, generally do not incorporate humor. Nevertheless, salespeople often use humor when introducing their products to reduce the distance between them and the users and increase their trust, leading to increased interest in sales items [1]. Thus, it has been considered that humor might have an effect on user interest in items recommended by chatbots for business, as it does on those recommended by salespeople.

Previous research includes several examples of humor being applied to digital content. For example, when humor stimuli related to skiing were incorporated into a task of making ski trip reservations on a web page, the users’ satisfaction and enjoyment of the reservation process improved compared to the case without humor stimulation [2]. Furthermore, the effects of humor on both web pages and chatbots have been reported. In one study, when humor stimuli were applied as an element of “humanity” in a chatbot for a used clothing buy-back process, it was suggested that they were beneficial to transaction outcomes [3]. However, in this situation, users sold the clothes, and the effect of humor on the recommendation of the chatbot was not investigated.

In this study, aiming to investigate the effect of humor on user interest in the recommendation of a chatbot, three types of chatbots were developed with different frequencies of humor stimuli, and a psychological experiment was conducted

to investigate the differences among the user interest in the recommended item.

II. CHATBOT

We developed text-based chatbots as a web application using the chatbot creation tool “Dialogflow” [4]. These chatbots recommended computer game titles after communication.

A. Questions for Recommendation

The main dialogues of the chatbots consisted of questions concerning computer games, with the goal of recommending a game title. Four participants familiar with the games discussed the content and order of the questions, and the method of answering them. Then, ten questions were generated. Five other participants communicated with a chatbot providing the questions. Further adjustments to the questions were made based on their feedback. All of the questions are listed below. The order of the questions is from top to bottom.

- Which hardware do you prefer to play games on? PC/3DS/Switch/PS5/Smartphone
- Which do you prefer to play 2D games or 3D games?
- Which do you prefer single player games or multi-player games?
- What is your favorite game genre?
- Do you like to play against other players?
- Do you think “story” is important in games?
- Do you think “sound” is important in games?
- Do you think “character design” is important in games?
- Do you think game design is important? Game design is about game balance, gimmicks, structure, and direction.
- How difficult do you prefer games to be?

B. Humor Stimuli

For each of the 10 questions, three humor stimuli related to the computer game were generated. The humor stimuli were divided into three types: puns (Dajare, e.g., “Here’s a pun. Need inferior sounds? “(The question is pronounced as “ototteiru oto tte iru ?” in Japanese)), trivia (e.g., “By the way, it seems that the full name of Nintendo’s Luigi is Luigi Mario”), and human-like jokes (e.g., “Chatbot’s lacrimal gland is basically weak”). The puns were extracted from jokes posted on a Web site called “Dajare Station” [5]. The trivia

were based on game information, and the human-like jokes were generated by one of the authors. The five evaluators rated the humor stimuli using a questionnaire on a five-point scale concerning two evaluation items: “Is it funny?” (funniness) and “Is it appropriate as a response to a question?” (appropriateness). For each question, the humor stimulus with the highest funniness rating was selected from among the stimuli with appropriateness rating 3 or higher. The selected humor stimuli consisted of two puns, one trivia, and seven human-like jokes.

C. Types of Chatbots

Three types of chatbots are developed: chatbots without a humor response, those with three or four humor responses, and those with ten humor responses. The humor responses were generated using the selected humor stimuli. The dialogues of the chatbot without a humor response consisted of only questions. The chatbot with three or four humor responses had a 1/3 chance of making a humor response after the user’s answer. The chatbot provided humor responses after the user answered questions 1, 4, 7, and 10, questions 2, 5, and 8, or questions 3, 6, and 9. The chatbot with ten humor responses provided a humor response every time the user provided an answer. The condition for the chatbot with a 1/3 chance of making a humor response was set according to a study on a previous chat system including puns [6], which reported that the inclusion of humor responses in one out of every three responses was increased the fun of the chat process. When the dialogues concerning the 10 questions were complete, the chatbots recommended a game title.

D. Game Title

A fictitious game title was used for the game title recommendation, so as to eliminate the possibility that a user’s knowledge of the recommended game title would make a difference in their interest in it. Six fictitious game titles were generated. Three of them were generated by one of the authors, and the others were generated by the fictitious English word generation Web site “This Word Does Not Exist” [7]. Five evaluators rated the six fictitious game titles using a questionnaire on a five-point scale concerning the evaluation item “Do you think it is a natural title for RPGs, action games, puzzle games, or games in other genres?” As a result, “Ephemeral Riot,” with the highest rating, was selected as the game title recommended by the chatbots.

III. PSYCHOLOGICAL EXPERIMENT

A. Participants

The participants were screened to identify those who was interested in computer games, as it could be expected that there would be relationships between their interest in such games and the evaluation of the recommended game title. As a screening question, three types of activities were selected from a list of 13 options as a hobby and an activity of interest, including “computational games.” The screening of the participants resulted in 296 participants for the experiment,

i.e., those who selected “computational game” as a hobby or an activity of interest.

B. procedure

The experimental procedure was as follows. 1. Communication with the chatbot. 2. Evaluation questionnaire.

C. Communication with the chatbot

The participants were divided into three groups: no humor response (101 participants), some humor response group (98 participants), and all humor response group (i.e., responded to all humor; 97 participants). People who answered for the screening question “computational games” as first, second, or third hobby or an activity of interest was randomly assigned to each group, so that the rates of users in each group were as equal as possible. The participants in each group used the chatbot for each condition. No humor response group used the chatbot without a humor response. Some humor response group used the chatbot with three or four humor responses. All humor response group used the chatbot with humor responses every time (ten humor responses).

D. Evaluation questionnaire

The items asked in the survey items were as follows. The user information items were “age,” “gender,” “weekly playing time of games,” and “number of games plays per year.” The participants were asked to select male or female for “gender” and to answer numerically to the other questions. The items concerning the recommended game title were “interest” and “willingness to investigate.” In addition, the items concerning chatbot were “likability,” “trust,” “appropriateness,” “enjoyment,” “humanity,” “willingness to reuse,” and “sense of humor.” The game title and chatbot items were rated on a five-point scale.

IV. RESULTS

A. Direct Effect of Humor on Interest

The average ratings of “sense of humor” are shown in Fig.1. A between-subjects one-way analysis of variance (ANOVA) was conducted to investigate whether sense of humor was appropriately controlled depending on the chatbot type. The main effect of chatbot type was observed ($F(2, 294) = 4.54, p < .05$). The post-hoc t-test revealed that the ratings under the no humor response condition were lower than those under the other conditions ($p < .05$). According to the results, the sense of humor was appropriately controlled depending on the chatbot type.

The average ratings concerning “interest” are shown in Fig.2. A between-subjects one-way ANOVA was conducted to investigate the direct effect of the chatbot type on the interest. There was no main effect from the chatbot type ($F(2, 294) = 0.437, n.s.$).

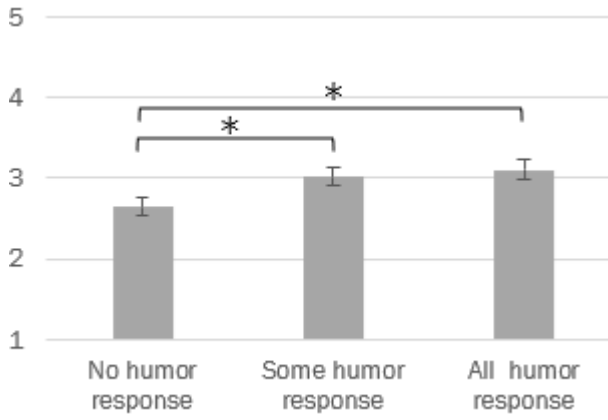


Fig. 1. Average ratings for "sense of humor"

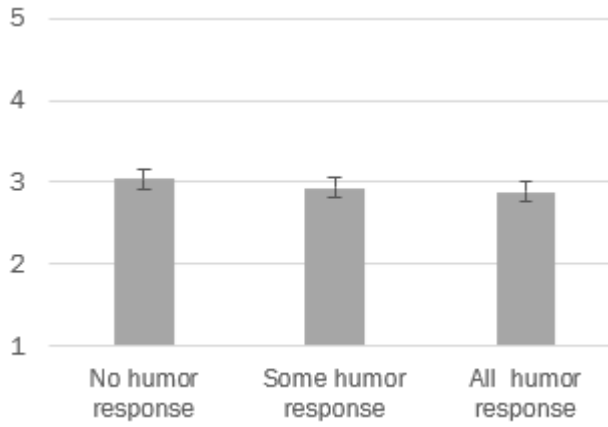


Fig. 2. Average ratings for "interest"

B. Analysis of Factors Influencing Interest

A multiple regression analysis was conducted to investigate the factors influencing the interest. In the analysis, the dependent variable was "interest," and the independent variables were "age", "gender", "likability", "trust", "appropriateness", "enjoyment", "humanity", "willingness to reuse" and "sense of humor". If users are interested in a recommended game title, they aim to investigate what is the title. Therefore, "willingness to investigate" was not used as an independent variable. The results (adjusted $R^2 = 0.567$) show in Table I that the "trust" and "sense of humor" factors have significant effects on user interest. In addition, the "humanity" factor has a significant trend effect on interest.

Therefore, the effects of the type of chatbot on "trust" and "humanity" were investigated. The average ratings of "trust" are shown in Fig.3. A between-subjects ANOVA was conducted, and the main effect of the chatbot type was observed ($F(2, 294) = 7.15, p < .05$). A post-hoc t-test revealed that the ratings under the "no humor" response condition were higher than those under the other conditions ($p < .05$). According to the results, the humor responses reduced trust in the chatbot. The average ratings of "humanity" are shown

TABLE I
MULTIPLE REGRESSION ANALYSIS RESULTS

	Estimate	Std.Error	t-value	Pr(> t)
(Intercept)	-0.000	0.0382	0.000	1.000
age	-0.042	0.039	-1.067	0.287
gender	0.039	0.040	-1.067	0.328
likeability	0.060	0.085	0.704	0.482
trust	0.460	0.058	7.948	0.000
appropriateness	0.043	0.050	0.861	0.390
enjoyment	0.033	0.070	0.479	0.632
humanity	0.112	0.059	1.892	0.060
willingness to reuse	0.056	0.066	0.859	0.391
sense of humor	0.153	0.055	2.763	0.006

in Fig.4 Similarly, a between-subjects ANOVA was conducted; however, no main effect was found ($F(2, 294) = 1.14, n.s.$).

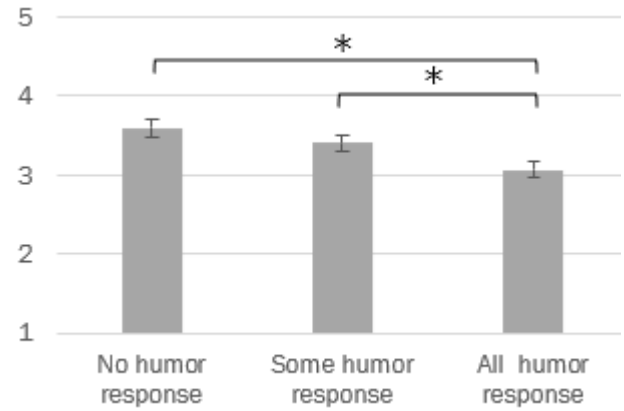


Fig. 3. Average ratings for "trust"

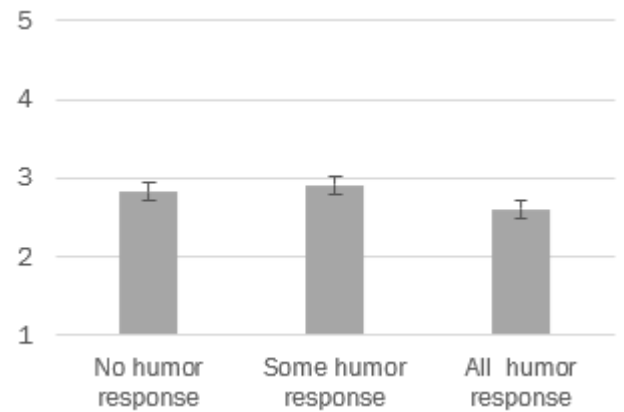


Fig. 4. Average ratings for "humanity"

C. path analysis

A path analysis was conducted to examine the relationships between "chatbot types," "interest," "sense of humor," "trust," and "humanity." A path analysis is a method that aims to graphically present the complex causal relationships among a number of variables as a statistically inferred model. In this

path analysis, “the number of humor responses” in the chatbots were used instead of “chatbot types.” As a path analysis requires assumptions regarding the relationships among the variables, such assumptions were made based on the above results: the effects of the number of humor responses (chatbot types) effects on “sense of humor” and “trust,” and the effects of “sense of humor,” “trust,” and “humanity” on “interest.” The results from the path analysis are shown in Fig.5.

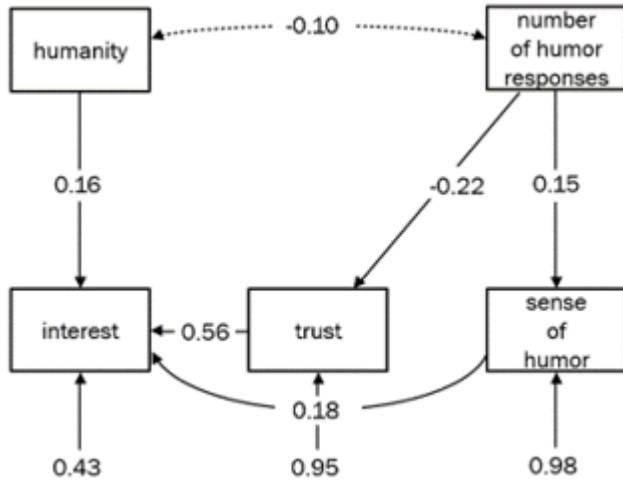


Fig. 5. Results of path analysis

However, the path analysis model did not fit the data well ($p < .001$, $gfi = 0.636$, $cfi = 0.496$, $rmsea = 0.475$, $aic = 2264.8$). Therefore, certain models were tested, and the model with the best fit (Fig.6) was found ($p = 0.167$, $gfi = 0.995$, $cfi = 0.997$, $rmsea = 0.052$, $aic = 2000.7$). The results suggest that the coefficient of the influence from “trust” on “interest” (0.56) is larger than those from the other factors on “interest” (0.18, 0.16). In addition, the coefficient of the influence from “number of humor responses” on “trust” is negative, (-0.17) and the coefficient of the influence from “number of humor responses” to “sense of humor” is positive (0.21).

V. DISCUSSION

In this study, a psychological experiment using three types of recommendation chatbots (with and without humor responses) was conducted to investigate the differences between user interest in the recommended item. The results indicate that no significant direct effect on user interest was observed depending on the frequency of the humor stimuli. The results of path analysis revealed that the user interest in the recommended item was affected by sense of humor, trust, and humanity, and that trust had the strongest effect on interest. The inclusion of humor stimuli increases sense of humor; however, it lowers trust, which is important in generating interest. Therefore, the effect of the humor on the interest may be negated.

Although the three types of humor stimuli were used in this study, it could be considered that there were differences

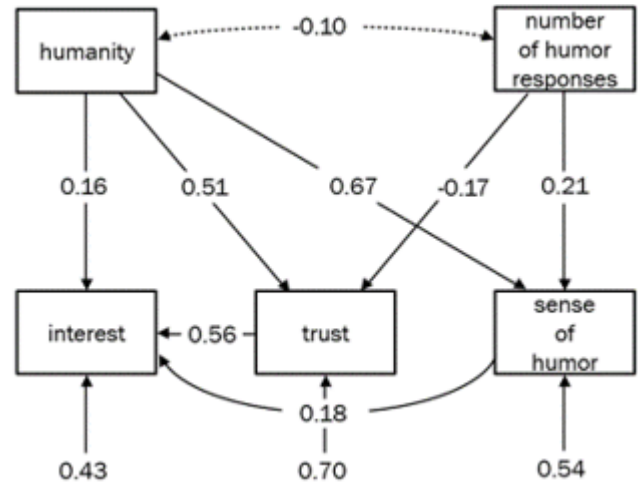


Fig. 6. Improved results of path analysis

in their effects on the trust in the chatbot. It is necessary to examine types of humor stimuli able to elicit humor without lowering trust in the chatbot. In addition, although we used a path analysis, it created a model based on an assumption of the causal relationships among the variables. It is necessary to investigate the causal relationships among the variables and the effects of other factors.

REFERENCES

- [1] B. Lussier, Y. Grégoire, and M. A. Vachon, “The role of humor usage on creativity, trust and performance in business relationships: An analysis of the salesperson-customer dyad,” *Industrial Marketing Management*, vol.65, pp. 168-181, 2017.
- [2] S. Schanke, G. Burtch, and G. Ray, “Estimating the impact of “humanizing” customer service chatbots,” *Information Systems Research*, vol. 32, No. 3, pp. 736-751, 2021.
- [3] W. M. van Dolen, K. de Ruyter, and S. Streukens, “The effect of humor in electronic service encounters,” *Journal of Economic Psychology*, vol. 29, No. 2, pp. 160-179, 2008.
- [4] Google. “Dialogflow.” Dialogflow. <https://dialogflow.cloud.google.com>.
- [5] Dajare Station. “Dajare Station.” Dajare Station. <https://dajare.jp>
- [6] P. Dybala, M. Ptaszynski, S. Higuchi, R. Rzepka and K. Araki, “Humor prevails!-implementing a joke generator into a conversational system.” in *Proc. Australasian Joint Conference on Artificial Intelligence*. Springer, Berlin, Heidelberg, 2008, pp. 214-225.
- [7] D. Thomas. “This Word Does Not Exist.” This Word Does Not Exist. <https://www.thisworddoesnotexist.com>.