

Towards a Humorous Chat-bot Companion for Senior Citizens

Ria Mundhra¹, Ting Jen Lim², Hong Ngan Duong², Kheng Hui Yeo³, Andreea I. Niculescu³

¹Victoria Junior College, Singapore
ria.mundhra.2019@vjc.sg

²St. Andrews' Junior College, Singapore

³Institute for Infocomm Research, Singapore

Abstract

In the past decade, Singapore population has grown older with more elderly and fewer younger people. As such, government encourages research initiatives providing solutions to improve mental health and memory functions. In this paper, we discuss an within-subject experiment that we carried out to measure the impact of humour on information retention in senior citizens interacting with a chatbot. The recall and recognition rates of both humorous and non-humorous responses supported the hypothesis that humour aids recognition, but could not provide strong support for recall. We hope that studies going forward can assist the innovation of intelligent systems for other target groups and uses as well.

Keywords

Humour, chatbots, dialogue systems, wellness, elderly, memory

1. Introduction

It is expected that the proportion of the population who are elderly in Singapore will grow from 28% in 2030 to 47% in 2050 [1]. The growing elderly populace, complemented by changing family structures, rising divorce rate and estranged familial relations has resulted in the number of solitary elderly aged 65 and above rising from 14,500 in 2000 to 42,100 in 2014 [2]. As such, it is crucial to aid elderly in Singapore, especially solitary elderly, take care of their physical and mental health.

In spite of several attempts to solve the problem, it still persists not only in Singapore, but various other countries across the globe as well [3]. This could be because the current solutions to target loneliness and well-being of senior citizens, such as senior activity centres, therapy and support groups are often inaccessible [4]. Other solutions, such as robotic companions, may also be ineffective due to their high cost and unfamiliarity for digital immigrants who have had to adapt to the use of technology late into their lives [5]. Moreover, interactions with these companions often feel very mechanical, and the companion can often seem patronizing [6].

As such, a humorous personal companion chatbot, integrated with Facebook Messenger, was chosen to solve the problem. Humour has been shown to enhance both mental and physical health by moderating stress and elevating mood [7]. Specifically, the incorporation of relevant semantic humour, in the form of word play, into information has been proven to augment the memory retention of the information [8]. Moreover, research has proven that the incorporation of humour into artificial intelligence makes interaction seem more human-like [9].



Fig. 1 Health Advocado in the exercise advisor mode

The chatbot was deployed on Facebook Messenger (Fig. 1) as it is inexpensive to use and can be easily accessed anywhere, as long as the user has a Facebook account and an internet connection. Moreover, the platform is the most familiar social networking site among our target group of people aged 60 and above [10]. Facebook has also been employed by other humorous chatbots popular amongst Singaporeans, such as ‘The Bus Uncle’ [11].

We had two main objectives when developing the chatbot, to investigate the effect of semantic humour in artificial intelligence on memory, and to incorporate our findings into a humorous personal companion for senior citizens.

While there have been numerous studies conducted on the effect of humour on memory [8], we wished to see if this effect is replicated in artificial intelligence,

which is characterized by inflexibility of responses and lack of context [12]. Humour in conversation with chatbots is a relatively new topic on which little research has been done, hence we hope that our research paves the way for new studies to be conducted.

It was hypothesized that humour in an interactive system would improve retention and recognition of the information provided by the system.

2. Design and Implementation

The most significant problems faced by the target group included malnutrition, physical injury, decline in cognitive health and social isolation [13] [14]. To combat these problems, the following features were built into the personalized chatbot:

1. A conversational buddy able to sustain conversation with users over a range of topics and also able to provide emotional companionship and casual banter
2. A meal advisor that provides users with information on healthy ingredients, restaurants and recipes in a concise and funny way, to motivate them to take care of their nutrition
3. An exercise planner presenting users with personalized exercise regimes in a humorous way to help them recover from injuries as well as reduce the risk of injuring themselves again
4. A memory lane feature allowing users to reminisce the past Singapore via old photos, to use their cognition and reduce the rate of memory loss and cognitive decline [15]
5. A set of games helping users combat loneliness and enhance their memory, creativity and language skills.

Some chatbot utterances included humour, mainly in the form of puns and wordplay, such as “corn is an a-maize-ing corn-lesterol corn-troller”. The puns made did not rely too heavily on culture or background, thus making the humour more universal. Puns and wordplay were also simpler to implement in a text-based system. The puns were augmented with relevant emojis as a visual form of humour and communication.

3. Methodology

To test the impact of humour on memory and recollection, a study was conducted with participants of age 60 and above with moderate to high English proficiency. The study was adapted from the experiment in “Memory for Information Paired with Humorous, Relevant Jokes” [8].

A pilot study with five participants was first conducted in order to detect and correct possible problems in the experimental set-up. Based on this pilot, we simplified our questionnaires and gave users more time to interact with the chatbot.

Participants were first given a brief introduction to the project, told that the study aimed to test the effectiveness of the chatbot, and then were guided on how to use the chatbot through a short video and explanation.

Participants were then asked to interact with the meal advisor portion of the chatbot for 20 minutes. It provided information about 18 different items from three categories; healthy ingredients, healthy recipes and healthy restaurants. Half of the information was presented in a humorous way, and the other half in a non-humorous way. Humorous and non-humorous statements were randomly spaced throughout the interaction. Participants were provided with scenarios to guide their interaction with the chatbot.

Participants were then given 10 minutes to complete a demographics and user experience questionnaire. This served to not only see their overall experience in interacting with the chatbot, but was also a filler task before the recall test, allowing us to more reliably distinguish between memory of humorous and non-humorous statements. Testing them immediately after the interaction could introduce test biases as they would still be able to recall the most recent replies regardless of the presence and quality of humour in the reply [16].

After the above questionnaires had been completed, we administered a recall test, where participants had 5 minutes to write down as many of the 18 presented facts as they could remember from their interaction with the chatbot. A recognition test then followed, where participants were presented with facts and tasked to recognize whether they have seen the fact in the chatbot earlier. 9 fake facts and 18 fact lures were added to the recognition test to increase the task difficulty. Fact lures refer to facts that are true, but were not presented by the chatbot during the interaction.

While the recognition test only tested participants' ability to recognize a piece of information as being familiar, the recall test measured their ability to retrieve related details about the information from their memory.

Lastly, participants were asked to rate how funny the jokes were. This served to verify that the humour was perceived as intended.

4. Results and Discussion

8 users participated in our final study. All participants were above the age of 60, from either Singapore or Malaysia, and rated themselves to have a relatively high level of English proficiency.

4.1. Evaluation of Humour

The mean humour rating given to each statement was calculated (*Table 1*), with 1 being not funny at all and 5 being very funny. The columns highlighted in yellow are non-humorous control statements, included for comparison.

Evaluation of Humour													
	Corn	Oats	Milk	Spinach	Kelp	Tuna	Sardine	Blueberry	Avocado	Nut	Dosirak	Wafuken	Marrow Noodles
Mean	4.25	2.50	3.75	3.63	3.00	2.63	3.63	4.13	2.75	3.87	4.00	3.88	3.75
No. of Samples	8	8	8	8	8	8	8	8	8	8	8	8	8
Std. Deviation	.707	1.414	1.389	1.408	1.414	1.506	.916	.641	1.488	.641	.926	1.126	.886

Table 1. Humour ratings for each statement

The non-humorous statements were consistently rated lower than humorous statements. This shows that the humorous statements were likely considered funnier by the target group. However, some humorous statements, such as those about spinach and sardines, were given comparatively lower scores. This could be because some puns are considered funnier than others.

4.2. Recall test

The Shapiro-Wilk test for normality was also conducted on the total number of humorous and non-humorous facts accurately recalled by participants, out of 9 humorous and 9 non-humorous facts. (*Table 2*).

Shapiro-Wilk Test for Normality

	Shapiro-Wilk		
	Statistic	df	Sig.
Humorous	.882	8	.197
Non-humorous	.724	8	.004

Table 2. Shapiro-Wilk test for normality for recall test

The data for non-humorous statements recalled is not normally distributed ($p=0.004 < \alpha=0.05$). Thus, the Wilcoxon signed rank test was used to compare between related samples instead. (*Table 3*).

Median		
	Humorous	Non-humorous
Median	1.00	.00
Wilcoxon Signed Rank Test		
	Test	Sig.
1	Related-Samples Wilcoxon Signed Rank Test	.257

Asymptotic significances are displayed. The significance level is .050.

Table 3. Wilcoxon Signed Rank Test

The median for humorous statements recalled is higher than the median for non-humorous statements recalled. However, this difference is not significant ($p =$

0.257). This could be attributed to the floor effect, as many participants found the test very difficult, and hence most obtained very low scores. As a result, the test was unable to show a significant difference in the scores for humorous and non-humorous statements.

4.3. Recognition test

Data was collected on the mean number of humorous statements and non-humorous correctly identified, out of 9 each. The data was normally distributed ($p = 0.204$, Shapiro-Wilk test with $\alpha=0.05$) . (Table 4).

Shapiro-Wilk Test for Normality

	Statistic	df	Sig.
Humorous (True)	.884	8	.204
Non-humorous (True)	.952	8	.731

Table 4. Shapiro-Wilk test for normality for recognition test

Hence, the paired sample t-test for related groups was conducted on the data. (Table 5).

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Humorous (True)	6.63	8	1.923	.680
	Non-humorous (True)	5.88	8	1.553	.549

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Humorous (True) & Non-humorous (True)	8	.843	.008

Table 5. Paired sample t-test

From the table, the mean for humorous facts identified is higher than mean for non-humorous facts identified. This difference is statistically significant ($p = 0.008$).

4.4. User Experience

The mean response to each statement was calculated, with 1 being strongly disagree and 5 being strongly agree. (Table 6).

User Experience

	My experience with the chatbot is pleasant	The information on the chatbot is well organised	The terminology used is understandable	It is easy to understand the information provided by the chatbot	The chatbot is fast in responding to my messages	The response of the chatbot is accurate according to my query	It is easy to correct my mistakes in the chatbot
Mean	4.00	4.00	3.88	4.13	4.25	4.38	4.13
Std. Deviation	.756	.535	.835	.641	.886	.744	.641

Table 6. User experience ratings

On average, users were satisfied with our chatbot, however, some found the terminology used by the chatbot to be confusing. Specifically, users did not understand the meaning of words such as “antioxidants”. Some users also requested more graphics and more detailed information. We took this feedback into consideration.

4.5. Discussion

The results we obtained seem to support the hypothesis that humour, in the form of puns and wordplay, in an interactive system does improve recognition of previously-seen information. There is also some evidence to support the fact that this humour improves our ability to recall, although those results are not significant enough to draw any conclusions.

However, we are aware that our sample size is too small to draw indubitable conclusions. The study should be repeated with a larger sample size in order to smooth out biases such as the subjective nature of humour and difference in language proficiency of participants, which affects their comprehension of certain keywords within the chatbot’s responses. Participants could be recruited from senior activity centers, and incentives could be provided to participants to take part in our study.

5. Conclusion and Recommendations for Future Work

This research has numerous real-world applications due to the growing market for such products aimed to improve the health and well-being of senior citizens, especially as the world continues to face the problem of an ageing population. To optimize the usefulness of the chatbot as a personal companion and helper, semantic humour is incorporated into it. Currently, there is no comprehensive research done on the impact of humour in interactive systems on its effectiveness. As such, we hope our research on the impact of semantic humour in chatbots on recall and recognition paves the way to develop more intelligent systems for other target groups and purposes.

Furthermore, as humour is intrinsically related to language and culture, the language used by the chatbot should be periodically updated to reflect that of the older generation at the time. More work can also be done to programme the chatbot to auto-generate humour as well. The chatbot could also be enhanced to generate puns, that make use of phonetics and semantics, rather than sticking to a fixed set of humorous statements.

References

- [1] M. E. Siau, "Elderly to make up almost half of S'pore population by 2050," 13 August 2019. [Online]. Available:

<https://www.todayonline.com/singapore/elderly-make-almost-half-spore-population-2050-united-nations>.

- [2] J. Tai, "Number of seniors aged 65 and above who live by themselves has tripled since 2000," 17 August 2015. [Online]. Available: <https://www.straitstimes.com/singapore/old-and-home-alone-in-singapore>.
- [3] United Nations, Department of Economic and Social Affairs, Population Division, "World Population Ageing 2015," 2015.
- [4] M. Pardasani, "Senior Centers: Characteristics of Participants and Nonparticipants.," *Activities, Adaptation & Aging*, pp. 48-70, 2010.
- [5] D. Gkouskos and J. Burgos, "I'm in! Towards participatory healthcare of elderly through IOT," in *Proceedings of the 7th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare*, Lund, Sweden, 2017.
- [6] K. Binsted, "Using humour to make natural language interfaces more friendly," in *The 1995 International Joint Conference on AI workshop on AI and Entertainment*, Montreal, 1995.
- [7] T. R. Herzog and S. J. Strevey, "Contact With Nature, Sense of Humor, and Psychological Well-Being," 2008. [Online]. Available: https://scholarworks.gvsu.edu/psy_articles/47.
- [8] S. Atir, "Memory for Information Paired with Humorous, Relevant Jokes.," Yale University, 2010.
- [9] P. Siarri, "Can you teach humor to an AI?," 3 April 2019. [Online]. Available: <https://medium.com/futuresin/can-you-teach-humor-to-an-ai-13ef4cff6cac>.
- [10] I. Griniute, "Social Media and Older People: Promising Future?," 2019. [Online]. Available: <https://all-digital.org/social-media-older-people-promising-future/>.
- [11] A. Murthy, "Why I created Bus Uncle," 17 February 2017. [Online]. Available: <https://medium.com/@abhilashmurthy/why-i-created-bus-uncle-40beb665c197>.
- [12] M. Ryan, "Are AI 'Thinking Machines' Really Thinking?," 1 November 2019. [Online]. Available: <https://towardsdatascience.com/are-ai-thinking-machines-really-thinking-1fa9d758de6d>.
- [13] S. Smith, "10 Common Elderly Health Issues," 11 February 2016. [Online]. Available: <https://vitalrecord.tamhsc.edu/10-common-elderly-health-issues/>.
- [14] C. Pardue-Spears, "Common Problems Faced by the Elderly in the US," 5 February 2018. [Online]. Available: <https://www.familymattershc.com/common-problems-for-elderly/>.

- [15] B. Woods, A. E. Spector, C. A. Jones, M. Orrell and S. P. Davies, "Reminiscence therapy for dementia.," *Cochrane Database of Systematic Reviews*, no. 2, 2005.
- [16] D. Kowalczyk, "Recency Effect in Psychology: Definition & Example," 24 July 2015. [Online]. Available: <https://study.com/academy/lesson/recency-effect-in-psychology-definition-example-quiz.html>.
- [17] H. Chi, E. Agama and Z. G. Prodanoff, "Developing serious games to promote cognitive abilities for the elderly," in *2017 IEEE 5th International Conference on Serious Games and Applications for Health (SeGAH)*, Perth, 2017.
- [18] A. J. Chapman and P. Crompton, "Humorous presentations of material and presentations of humorous material: A review of the humor and memory literature and two experimental studies," in *Practical Aspects of Memory*, London, Academic Press, 1978, pp. 84-92.
- [19] S. Bertsch, B. J. Pesta, R. Wiscott and M. A. McDaniel, "The generation effect: A meta-analytic review," *Memory and Cognition*, vol. 35, no. 2, pp. 201-210, 2007.
- [20] M. Berger, T. H. Wagner and L. C. Baker, "Internet use and stigmatized illness," *Soc Sci Med*, vol. 61, pp. 1821-1827, 2005.
- [21] E. M. Berg and L. G. Lippman, "Does humor in radio advertising affect recognition of novel product brand names?," *The Journal of General Psychology*, vol. 128, pp. 194-205, 2001.
- [22] A. Augello, G. Saccone, S. Gaglio and G. Pilato, "Humorist Bot: Bringing Computational Humour in a Chat-Bot System.," in *CISIS 2008: 2nd International Conference on Complex, Intelligent and Software Intensive Systems*, 2008.