Applied human informatics, vol.4 No.1
Special Issue on Intelligent Environments for Health and Well-being

Examining the Impacts of Exaggerated Notification Texts on Smartphone Application Users

Shinnosuke DATE[1] Wenzhen XU[2] Masato TAYA[2, 3] Takeshi IWAMOTO[1]

[1] Toyama Prefectural University, Imizu, Toyama 939-0398, Japan
s-date@puc.pu-toyama.ac.jp
[2] KDDI Research, Inc., Fujimino, Saitama, 356-8502, Japan
[3] KDDI CORPORATION, Chiyoda, Tokyo 102-8460, Japan

Abstract. Exaggerations are widely used in creating attractive notifications, while overly exaggerated contexts may cause negative effects on users' engagements. To reduce the issues arising between the Click Through Rate (CTR) and user engagement with notifications, it is necessary to clarify the impacts of exaggeration on CTRs and user engagements. Since no comprehensive indicators have been successfully created yet, our first step is to propose an indicator to define and measure the strength of exaggeration of notification text. In addition, we created exaggerated notification text using these indicators and experimented with the application. In this paper, we have shown that exaggerated notifications may temporarily increase the CTR but not continuously. Although the results of the engagement study did not show any significant difference, the subjects who opened the exaggerated notifications showed a tendency to exit the application without performing any screen operations, indicating that exaggeration may reduce engagement.

Keywords. notification, application, exaggeration, CTR, engagement

Received: October 8, 2021; Revised: January 09, 2022; Accepted: February 14, 2022

1 Introduction

In recent years, smartphones have become an indispensable part of people's lives, and all different kinds of services and entertainment are being released as applications. According to the Mobile Market White Paper 2020 [1], while the average number of applications installed in a smartphone is 103.4, just 38.5 of these applications are used. This means that approximately 62.8% of the installed applications are left dormant. Therefore, in the smartphone application market, it is important to find a way to encourage users to continue using an application after it has been installed. The notification function of smartphones is often used as one such way to encourage users. Click Through Rate (CTR) [2] is known as a useful measure to indicate the percentage of users who react to the notification and studies have been conducted which aim to improve CTR [3][4][5].

These studies are expected to contribute to enhancing marketing through notifications in the future. On the other hand, as the market expands, exaggerated advertisements such as clickbait headlines have become problematic. Clickbait is defined as expressions using overemphasized language which, in many cases, is irrelevant to the actual content. Facebook took measures such as changing its algorithm to minimize hype and

misleading advertisements that fall under the category of clickbait in 2019[6]. Similarly, Google introduced a policy on clickbait advertisements in July 2020 to disallow the display of such advertisements [7]. Similar issues will arise as notifications become more important in digital marketing. However, unlike web advertisements, notifications are issued by pre-installed applications, and not only the CTR but also the engagement can be the most important indicators of usage. To increase engagement and encourage the continued use of the app, it is necessary to ensure its credibility. Exaggerated notifications, such as clickbait, may increase the CTR but cause distrust of such apps for users and engagement not only will not increase but may even decrease as a result. However, notifications are required to be attractive to get attention and increase the CTR of notifications as much as possible. In the English-speaking world, a survey by CleverTap [8] found that 250 power words can be used to make a notification more persuasive, but this kind of survey has not been conducted in Japan. Despite the use of these power words in the excessive texts, the credibility decreases, and this results in less engagement by users. In this dilemma between CTR and engagement in notifications, we need to know what type of exaggerated language undermines user engagement. The purpose of this study is to propose indices to indicate that the exaggerated notifications are exaggerated, and furthermore, to examine the relationship between the resulting CTR and engagement. We would like to show indices through the following steps.

1. Propose a hypothesis for an indicator of exaggeration.
2. Evaluate exaggerated notifications using the proposed indices and confirm the validity of the indices by conducting questionnaire surveys.
3. Verify the notifications using actual applications

The rest of the paper is organized as follows. We first describe the related work that improves the notification CTR in Section 2. Subsequently, we describe the definition of exaggeration in this study and the indices of exaggerated texts in Section 3. Then, the experiment to evaluate the validity of the indices is described in Section 4, and the results of the verification of the exaggerated notifications using actual applications are described in Section 5.

2 Related work

In this section, we will describe the related work that improves the notification CTR. Okoshi et al. conducted a study on detecting breakpoints, which is the timing at which notifications are likely to be opened [3]. In this research, the breakpoints were estimated by using machine learning. The big data of physical movements (driving, biking, walking, standing still, tilting the device, etc.) were obtained from smartphone devices. The authors reported that sending notifications at these breakpoints improved the CTR compared to normal conditions.

Fraser et al. reported the results of comparing the CTRs of attractive and unattractive notifications [4]. They defined attractive notifications as those with texts like clickbait headlines. Clickbait refers to incendiary headlines in advertisements and articles on the web designed to provoke users to click them [5]. In many cases, these headlines look interesting and appealing, but are less relevant to the actual product or contents. Fraser used the headlines of articles recognized as clickbait to create attractive notifications. The notifications were sent to 18 subjects, and their responses (open/close) were collected as data. By learning from this data, we created an agent that mimics the human

response to notifications. In the experiment, we sent attractive and unattractive notifications to the agent, and it reported that the CTR of the attractive notifications was higher than that of the unattractive ones.

3 Exaggeration

In this section, we describe the previous work on exaggeration and clarify the exaggeration specified in this study.

3.1 Exaggeration in Existing Research

The concept of exaggeration has long been the subject of various studies in the fields of psychology [9] and commerce [10, 11]. Although there are many interpretations of exaggerated advertisements, no clear definition has been established, nor have detailed studies have been conducted: some of them are completely irrelevant to the content [12] and others include unrealistic content [13]. Now, we would like to define the concept of exaggeration explicitly.

First, to define exaggeration, we focused on Kunieda's research on the impact of exaggerated expressions on consumer purchase intentions [14] and found two conclusions as the effects of exaggeration.

(A) First, exaggeration increases consumers' evaluation of the product. Exaggeration is effective for products with low prices and high purchase frequency. On the other hand, it is reported that exaggeration was not effective for products with high prices and low purchase frequency.

(B) Second, the distrust of exaggerated advertisements has a negative impact on consumers' purchase intentions. As mentioned in (A), even if the product evaluation is enhanced by exaggerated advertising, it is unlikely to lead to a purchase if the advertisement gives the consumers a negative impression or makes them distrust it. This is also consistent with the problem that exaggerated notifications, which we focus on in this study, can improve the CTR but may not lead to engagement due to the loss of trust in the application. Therefore, we consider the definition of exaggeration in this study with reference to these two effects of exaggerated advertising.

3.2 Definition of exaggeration in this study

(A) in the previous section, we mentioned that exaggeration can give users a positive impression. In the exaggeration of the notification text that is the subject of this study, giving a positive impression gives the user motivation to open the application. In general, it is said that catchphrases and headlines with a high degree of unexpectedness and attractiveness are effective. For example, the research on automatic generation of catchphrases [15] states that catchphrases should have "attractiveness" and "impact (unexpectedness)". Therefore, in this research, the elements that motivate users to open an application are "unexpectedness" and "attractiveness". "Unexpectedness" through exaggeration refers to unexpected texts that provide users with novelty and unexpected discoveries, and "attractiveness" through exaggeration refers to attractive texts that make users think the contents will be valuable. We think that the positive impression given by these two terms will increase the user expectations of the application, resulting in a higher notification CTR.

In contrast, in (B) described above, we mentioned that exaggeration can give a negative impression to users. In this study, exaggeration does give a negative impression to the user, which means that the user does not feel like opening the application due to the notification. In general, the factors that make users reluctant to start the application are the decrease in "credibility" and "usefulness". For example, in the international comparison survey of consumers' advertising avoidance [16], the low level of "credibility (believability)" and "usefulness" are cited as the factors which lead to avoiding advertisements. Therefore, in this study, we define the low level of "credibility" and "usefulness" as the main factors that make people reluctant to open applications. For notifications, exaggerated "credibility" is dubious text that makes people doubt its validity, and exaggerated "usefulness" is text that gives the impression that the information is useless. The negative impressions caused by these two factors are thought to reduce trust and engagement with the application. Table 1 summarizes the changes caused by exaggeration of the four indices described above. The increase/decrease of each element are shown in Table 1 by the impression of the exaggerated texts compared to the original texts. Based on the relationship between the increases/decreases in the four indices shown in Table 1, we define exaggeration in this study as "the expression of a text that contains at least one the following indices: an increase in unexpectedness, an increase in attractiveness, a decrease in credibility, and a decrease in usefulness compared to the original text". In this study, we propose the four relative exaggeration indices listed here, show the validity of evaluating exaggeration by using these indices, and verify the relationship between the CTR and engagement of the application.

Table. 1 Four relative exaggeration indices.

Indicator	Unexpectedness	Attractiveness	Credibility	Usefulness
+/-	+	+	-	-

4 Validity of the exaggeration indices

In the previous section, we discussed the definition of exaggeration and the four relative exaggeration indices. In this study, we conducted an experiment to evaluate validity of the exaggeration indices by producing a normal notification text and an exaggerated notification text (EN Text). Since the exaggerated text may not always apply to the exaggeration indices defined in this study, multiple texts were created and tested.

4.1 Purpose of this experiment

In the previous section, we defined exaggeration in this study as "the expression of a text that contains more than one of the following: an increase in unexpectedness, an increase in attractiveness, a decrease in credibility, and a decrease in usefulness compared to the original text". We also need to confirm that the four indices increase or decrease according to this definition by comparing the normal notification text to the EN Text. Therefore, the purpose of this experiment is to evaluate the validity of the four indices by conducting a subjective questionnaire survey on both the normal and EN Text of the smartphone application and confirming the increase or decrease of the indices.

4.2 Application

In order to create the normal text, this study uses a smartphone application called "Toyama Takeout Map" (hereinafter referred to as "the Application") as the experimental target [17]. Figure 1 shows an overview of this application. This smartphone application was developed to promote take-out food in Toyama Prefecture, as the restaurant industry was severely impacted by the COVID-19 pandemic. The Application has a function to introduce restaurants that offer take-out food. Figure 1 (1) shows the top screen when the application is opened, with a map showing restaurants and your current location at the top, and a list of restaurants at the bottom. The user can tap on the restaurant of interest from the map or list and view the information on the detail screen shown in (2). This application contains the information of 255 restaurants, and there are filters for categories and areas to help users find restaurants efficiently. In order not to affect actual user operation, this experiment will be conducted using the development version for experiments.

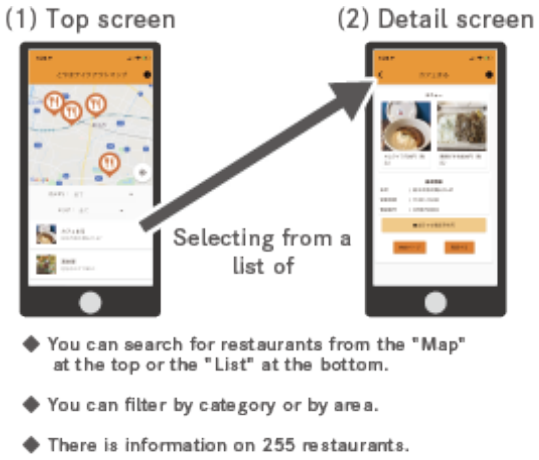

Fig. 1 Toyama Takeout Map App Outline.

4.3 Notification texts in this study

The purpose of this application is to "promote the use of take-out food to users". Because of this, the notification text in this study was created so that it would urge users to start this application. First, we created the normal notification text in accordance with the psychological effects of Cialdini's law [18] as an exaggeration-free text. There are six psychological effects in Cialdini's law: Liking, Reciprocation, Social proof, Authority and Scarcity (In this study, we excluded the Consistency of the sixth psychological effect.). Table 2 shows the definitions of the five psychological effects and an example of the notification texts. For each psychological effect shown in Table 2, four texts were created, for a total of 20 non-exaggerated notification texts. In addition, we paid attention not to include similar words or expressions in the texts. For each of these 20 texts, we added exaggerated expressions and created 20 exaggerated texts. We also clarified whether the EN text satisfies the exaggeration as defined in this study to conduct the experiment efficiently. Figure 2 shows a summary of the notification texts in this study.

4.4 Experiments

In this section, we describe an experiment to compare normal notification texts and EN Text. First, in this experiment, we created a questionnaire for the 40 notification texts. In the questionnaire, the four indices described in Section 4.2 were evaluated quantitatively on a scale of 1 to 7 for each text. An example of the questionnaire is shown in Figure 3. The questionnaire was filled out by 13 subjects who are males in their twenties from the laboratory in Toyama Prefectural University and the response data was collected.

Table. 2 Psychological effects and examples of notification texts.

Psychological effects	Definition	Examples
Liking	The psychology of wanting to respond aggressively to requests from a person you have feelings for.	No more making three meals! Sometimes takeout is just the thing to change things up!
Reciprocation	The psychological belief that one must make concessions or compromises to a favored party.	Don't let COVID-19 beat you! Support your favorite restaurants by ordering take-out food!
Social proof	The psychology of wanting to be in tune with what many people are choosing.	Now, more and more people are opting for takeout! Would you like to try take-out food?
Authority	The psychology of trusting and being influenced by someone with authority, such as a title.	Top restaurants ranked by the TV station are all here! Open the app and check out the restaurants!
Scarcity	The psychology of feeling that the more limited something is the more valuable.	Only now! Takeout a limited time menu for your convenience!

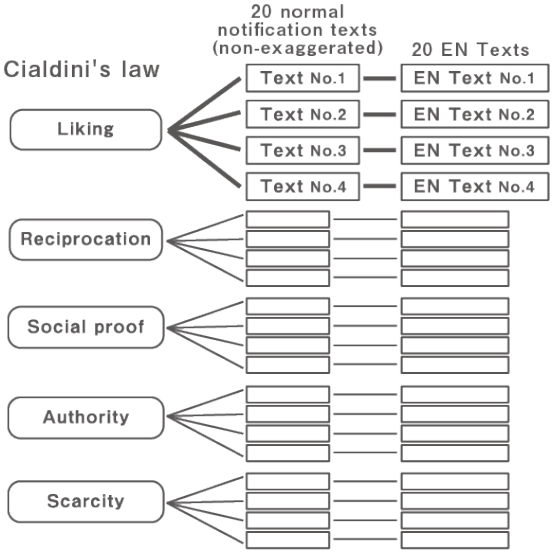

Fig. 2 Summary of notification texts in this study.

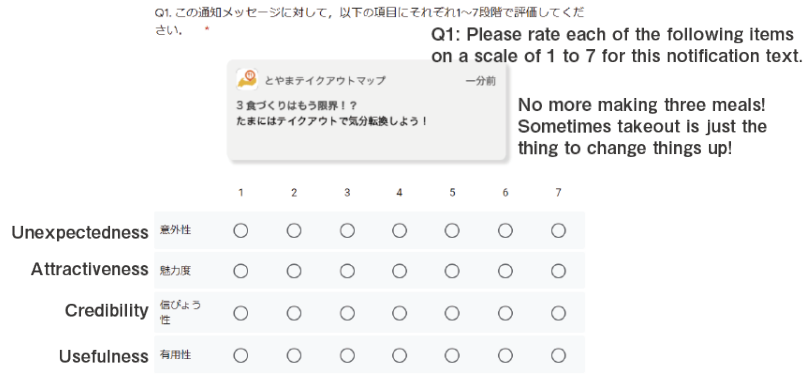

Fig. 3 Examples of questionnaire.

4.5 Results

In this section, we describe the data obtained as a result of the experiment. As an example, Table 3 shows the results of "Unexpectedness" for No.1 and EN Texts among the results of this experiment. For Table 3, the 13 subjects are represented by p1-13. The difference in Table 3 is the result of subtracting the "Unexpectedness" rating of the normal notification text from the "Unexpectedness" rating of the EN Text. A difference of 1 or more indicates an increase, a difference of -1 or less indicates a decrease, and a difference of 0 indicates no change in "Unexpectedness". As shown in Table 3, 7 respondents chose "increase," 3 chose "decrease," and 3 chose "no change" for "Unexpectedness" of No. 1. These results were obtained for each of the 20 pairs of notification texts, and the numbers of "increased," "no change," and "decreased" are summarized in Table 4. In Table 4, the numbers in parentheses indicate the number of "increased", "no change", and "decreased" from the left. The n in Table 4 indicates the number of subjects, and in this experiment, n=13. Table 5 shows the total number of respondents for each indicator (increase, no increase, decrease). Table 5 shows that the number of respondents who selected "increased" for "Unexpectedness" was high, while the number of respondents who selected "decreased" for "Credibility" and "Usefulness" was high. On the other hand, there was no difference in the number of respondents who selected "Attractiveness". Also, sub-group dependency was not found in this experiment. In the next section, we will test whether the increase or decrease of the indicators is significant for each notification.

Table. 3 Results of "Unexpectedness" in No.1 notification text.

Subjects	Normal Notification Texts	EN Texts	Difference	+/-
p1	3	6	+3	+
p2	4	6	+2	+
p3	5	6	+1	+
p4	2	6	+4	+
p5	1	4	+3	+
p6	3	1	-2	-
p7	3	1	-2	-
p8	1	7	+6	+
p9	2	3	+1	+
p10	6	5	-1	-

p11	2	2	±0	±
p12	4	4	±0	±
p13	4	4	±0	±

4.6 The sign test

In this section, we describe a method for statistically testing the results of an experiment. In this study, we used the sign test [19] as a method to investigate if there is a significant difference in each of the four indicators between the normal notification text and EN text. In this experiment, we used the two-tailed test. As an example, for "unexpectedness" in the text of No. 2, the hypothesis H_0 and the alternative hypothesis H_1 are as follows.

$$\begin{cases} Hypothesis\ H_0: No.2\ has\ no\ difference\ in\ unexpectedness\ before\ and\ after\ exaggeration. \\ Alternative\ hypothesis\ H_1: No.2\ has\ a\ difference\ in\ unexpectedness\ before\ and\ after\ exaggeration. \end{cases}$$

First, we compute the test statistic S for this hypothesis, which is obtained by summing the pairs of corresponding data x_{1i} and x_{2i} such that $x_{1i} > x_{2i}$. Here, i is the sample number and represents each subject, x_{1i} is the number of subjects i evaluated the "Unexpectedness" of the text of No. 2, and x_{2i} is the number of subjects i evaluated the "Unexpectedness" of the EN Text of No. 2. Therefore, S is the sum of the number of people whose evaluation of the "Unexpectedness" of No. 2 increased after the exaggeration. From Table 4 in the previous section, the number of people who increased their evaluation of the "Unexpectedness" of No. 2 is 11, so S=11. Next, we find the rejection limit point r_a, which is given by the sign test table and is determined by the significance level a (5%) and N. Here N is the sample size, which in this study refers to the number of subjects. When the relationship between S and ra is $r_a \leq S$ or N-$r_a \leq S$, we reject the hypothesis H_0. However, x_{1i}=x_{2i} is not included in N. In No. 2, "Unexpectedness," the number of subjects (13) is subtracted from the number of subjects (2) in the "None" category to get N = 11, and r_a =2 from the sign test table. Since N-$r_a \leq S$, the hypothesis H_0 is rejected, indicating that there is a significant difference in the number of increases in "Unexpectedness" in No. 2. Using this sign test, each notification text was tested for each of the four indicators.

Table. 4 The number of people after exaggeration (increase, none, decrease) for each notification.

n=13	Unexpectedness	Attractiveness	Credibility	Usefulness
No.1	(7, 3, 3)	(3, 4, 6)	(1, 0, 12)	(1, 2, 10)
No.2	(11, 2, 0)	(7, 4, 2)	(7, 3, 3)	(6, 5, 2)
No.3	(9, 3, 1)	(5, 2, 6)	(4, 3, 6)	(6, 3, 4)
No.4	(8, 4, 1)	(7, 2, 4)	(4, 3, 6)	(5, 0, 8)
No.5	(3, 4, 6)	(7, 4, 2)	(6, 5, 2)	(3, 6, 4)
No.6	(7, 4, 2)	(7, 2, 4)	(5, 4, 4)	(4, 7, 2)
No.7	(5, 5, 3)	(3, 5, 5)	(3, 2, 8)	(1, 4, 8)
No.8	(1, 7, 5)	(2, 9, 2)	(4, 3, 6)	(2, 8, 3)
No.9	(7, 5, 1)	(3, 2, 8)	(3, 1, 9)	(2, 2, 9)
No.10	(6, 4, 3)	(4, 5, 4)	(2, 6, 5)	(4, 3, 6)
No.11	(7, 5, 1)	(3, 5, 5)	(2, 6, 5)	(1, 7, 5)

No.12	(3, 4, 6)	(7, 2, 4)	(4, 4, 5)	(6, 2, 5)
No.13	(3, 7, 3)	(2, 3, 8)	(3, 4, 6)	(1, 8, 4)
No.14	(4, 2, 7)	(5, 2, 6)	(7, 4, 2)	(4, 6, 3)
No.15	(6, 4, 3)	(4, 6, 3)	(5, 6, 2)	(4, 5, 4)
No.16	(6, 3, 4)	(5, 3, 5)	(3, 5, 5)	(1, 4, 8)
No.17	(3, 4, 6)	(4, 4, 5)	(4, 2, 7)	(4, 4, 5)
No.18	(5, 5, 3)	(3, 6, 4)	(2, 7, 4)	(4, 5, 4)
No.19	(6, 3, 4)	(5, 5, 3)	(4, 3, 6)	(4, 4, 5)
No.20	(7, 2, 4)	(2, 6, 5)	(2, 3, 8)	(1, 6, 6)

4.7 Result of the sign test

First, there were six notification texts in which none of the indicators were significantly different. For the remaining 14 notification texts, exaggeration significantly changed one to three indicators. Of the 14 notifications, only 3 did not satisfy the hypothesis of this study that "Unexpectedness" and "Attractiveness" would increase, and "Credibility" and "Usefulness" would decrease, while the other 11 all satisfied the hypothesis. These results are summarized in Table 6. The definition of exaggeration in this study is "the expression of a text that contains more than one of the following: an increase in unexpectedness, an increase in attractiveness, a decrease in credibility, and a decrease in usefulness compared to the original text". We consider that of the 20 EN texts created in this study, 11 of the 14 that showed a significant change in some indicators showed the characteristics of hyperbole as defined. Therefore, it is reasonable to evaluate exaggeration by using the four indices proposed in this study. The effects of 11 EN texts were four for "Liking", three for "Reciprocation", two for "Social proof", and one each for "Authority" and "Scarcity". For the texts based on Cialdini's law used in this study, "authority" and "scarcity" were rarely exaggerated. We thought that these effects already included the possibility that the texts were exaggerated.

The next step is to send these notifications in an actual application and to verify whether they affect the CTR of notification texts and engagement.

Table. 6 The number of notification text with significant differences.

Number of indicators that showed significant differences	Number of the texts	The text that did not satisfy the hypothesis.
0	6	NA
1	9	3
2	4	0
3	1	0

5 Experiments using the Application

We conducted an experiment to investigate CTR and engagement using the EN Texts presented in the previous section. For the experiments, we used five notification texts that showed significant differences in two or more indices that were the result of strong exaggeration. Table 7 shows the increase/decrease of the four indicators in the notification letters used. In the table, "+" refers to those that were rated higher after the exaggeration, and "-" refers to those that were rated lower after the exaggeration. Blank columns indicate that no significant difference was found.

Table. 7 Results of the sign test (with two or more significant differences).

No.	Unexpectedness	Attractiveness	Credibility	Usefulness
No.1			-	-
No.2	+	+		
No.9	+		-	-
No.11	+			-
No.20			-	-

5.1 Method of experiments

We asked the subjects to install the "Toyama Takeout Map" described in the previous section and conducted an experiment to send notifications to them. In order not to interfere with this experiment, we informed the subjects of the original purpose of the system after the experiment was completed.

Next, the notification texts which were to be sent are shown in Table 8. No. in Table 8 is the number assigned to the normal notification text and the corresponding EN Text pair. These normal notification texts and EN Texts were sent to the subjects, and their responses were logged. Specifically, we obtained the opening of the notification, the launching of the application, and the screen transition operation data.

The subjects were 22 students from laboratories, and they did not join this experiment stated in Section 4. The experiment was conducted for 10 weekdays from April 19, 2021, to May 7, 2021 (Period 1), and for another 10 weekdays from May 10, 2021 to May 21, 2021 (Period 2). Notifications were sent twice a day over a total of 20 weekdays. Notifications were sent at 11:00 a.m. and 5:00 p.m., which are the times when users are most active in the released version of the application. In addition, we thought it would interfere with the experiment if the same notification was displayed to the subjects on the same day or the next day. Therefore, we scheduled the same notification and the same pair of notifications at least two days apart. In addition, each notification was sent to each subject twice in period 1 and twice in period 2, for a total of four times.

Table. 8 Normal notification texts and EN texts sent in this experiment.

No	Normal notification texts	EN Texts
No.1	No more making three meals! Sometimes takeout is just the thing to change things up!	No more making three meals! A change of pace with takeout makes you feel 10 years younger!
No.2	Want a good meal? Come on, let's get take-out!	Want to eat a good meal? Take-out food is better than home cooking!
No.9	The number of restaurants that offer takeout food is rapidly increasing! That restaurant near your house is also offering this service!	93 out of 100 people eat take-out food! That restaurant near your house is also offering take-out food!
No.11	Now, more and more people are opting for takeout! Takeout your meals too!	More and more people are opting for takeout! Don't you want to be part of this big wave...?
No.20	We also offer a variety menu especially for take-out! Don't miss it! Try take-out food now!	Did you know that we offer special menus only for takeout-...

5.2 CTR and engagement

The CTR in this experiment is derived by dividing the number of people who tapped the pop-up window on the screen when receiving a notification by the total number of notifications. Here, the number of notifications, which is the parameter, is 44 times for one notification, since 22 subjects x 2 notifications were sent in period 1, and 88 times for the entire period, since 22 subjects x 4 notifications were sent. Next, we define engagement in this study. Engagement typically refers to the actual use of an application. However, the definition and the indices to measure it differ depending on the application itself. In general, social networking applications use the number of accesses to the app and the usage time as engagement. In contrast, accommodation booking applications use the number of reservations made as engagement, and do not focus on the number of times the app is opened or the usage time of the application itself. In this way, engagement needs to be defined according to the application. In our experiment, we used a simple application that allows users to view information on multiple restaurants offering take-out food. Therefore, in this study, we defined the use of the application as when users view restaurants' information. Thus, engagement in this study is defined as the action of a screen transition to browse such information. In other words, if there was no screen transition, we did not count it as an engagement regardless of the activation of the application. Figure 4 shows the relationship between engagement and screen transitions of the application in this study. Using this verification, the number of people who made these screen transitions was obtained, and the screen transition rate was defined as the number of opened notifications.

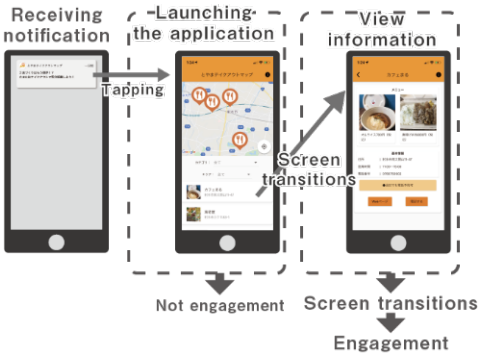

Fig. 4 Relationship between engagement and screen transition in this study.

5.3 Results

Table 9 shows the results of the CTR of the normal notification texts and EN Texts for each period. The CTRs in Table 9 were calculated based on the sum of the five notifications. In addition, the number in parentheses next to the CTR in Table 9 indicates the number of people who opened the notification. In this study, there are five normal notification texts and five EN Texts, so the number of to send a notification was sent in period 1 is 44 times five, or 220 times, and the number of times in the entire period is 88 times five, or 440 times. The results in Table 9 show that the CTR of EN Texts is higher in period 1 and lower in period 2. In addition, the CTR did not change between periods 1 and 2.

Next, we calculated the screen transition rates of the applications defined as an engagement in the previous section. To calculate the transition rate, we obtained the number of subjects who performed the screen transition after they opened the notification

from the log and divided it by the number of subjects who actually opened the notification. The results of the screen transition rate are shown in Table 10. The numbers in parentheses in Table 10 indicate the number of people who made screen transitions. Table 10 shows that the screen transition rate of the EN Text was lower than that of the normal notification text in Period 1. In Period 2, there were no subjects who made screen transitions after receiving either the normal notification text or the EN Text.

Table. 9 CTR for notification text and exaggerated notification text by period.

	Normal notification texts	EN Texts
Period1(n=220)	2.27% (5 subjects)	5.45% (12 subjects)
Period2(n=220)	2.27% (5 subjects)	1.36% (3 subjects)
Entire period(n=440)	2.27% (10 subjects)	3.41% (15 subjects)

Table. 10 Screen transition rates for notification text and exaggerated notification text by period.

	Normal notification texts	EN Texts
Period1	60.00% (3 subjects)	33.33% (4 subjects)
Period2	0.00% (0 subjects)	0.00% (0 subjects)
Entire period	30.00% (3 subjects)	26.67% (4 subjects)

5.4 Test of the difference in CTR

In the results of this experiment presented in the previous section, we obtained the result that the CTR of the normal notification texts is higher than the EN Texts. In this section, we describe the results of statistical tests.

First, Table 11 shows the frequency table of the number of subjects who opened the notification in period 1. Based on the values in Table 11, we conducted a one-tailed Fisher's exact test [20], which is often used for small samples. The significance level was set at 5%. The x^2 values, degrees of freedom, *p-values*, and next *N* are shown in Table 12. Table 12 shows that the difference between the number of subjects who opened the normal notification text and the EN Text in period 1 was at $.05<p<.10$. It is not significant but has a significant trend. Here, the next *N* is the number of samples required to confirm a significant difference, and it would be ideal to meet the number of samples in the next experiments. Therefore, in this test, a significant difference can be confirmed if the number of subjects is increased, and the sample size satisfies *N* in the next experiments.

Next, Table 13 shows the frequency table of the number of subjects who opened the notification in period 2. In Table 13, Fisher's exact test was conducted as in Period 1. The significance level was set to 5% as well. The results are shown in Table 14. Table 14 shows that the difference between the number of subjects who opened the normal notification text and the EN Text in period 2 was $p<.10$, which was not significant, nor did it show a significant trend as in period 1. In addition, the next *N* is "unknown", and even if we increase the number of the samples, no significant difference is confirmed. Based on these results, we found that the difference in CTR between normal notification text and EN Text tended to be significant in period 1, but it became less significant in period 2.

Table. 11 Frequency tabulation table of the number of people who opened the notification (period 1).

	Opened	Number of notifications	CTR
Normal notification texts	5	220	2.22%
EN Texts	12	220	5.17%

Table. 12 Results of the test for the difference in independent ratios (period 1).

x^2 values	Degree of freedom	*p-value*	Next *N*
2.013	1	0.078	1782

Table. 13 Frequency tabulation table of the number of people who opened the notification (period 2).

	Opened	Number of notifications	CTR
Normal notification texts	5	220	2.22%
EN Texts	3	220	1.35%

Table. 14 Results of the test for the difference in independent ratios (period 2).

x^2 values	Degree of freedom	*p-value*	Next *N*
0.118	1	0.365	unknown

5.5 Discussion

In this section, we will discuss the results. In this verification, we conducted an experiment on EN Texts, for which showed significant differences in more than two indicators out of four. In the experiment, we compared the CTR and screen transition rate of normal notification texts and EN Texts to confirm the impact of EN Texts.

First, the CTR of EN Texts was higher than that of normal notification texts in Period 1. On the other hand, the CTR of EN Texts was lower than that of normal notification texts in Period 2. In addition, the CTR of normal notification texts did not change in periods 1 and 2, but the CTR of EN Texts decreased in period 2. Based on these results, we tested the difference in CTRs statistically by using Fisher's exact test, and it showed that the difference in CTRs between normal notification texts and EN Texts tended to be significant in Period 1. In period 1, the difference in CTR between normal notification texts and EN Texts tended to be significant, while in period 2, there was no significant difference or trend in CTR. Therefore, the exaggeration may only have affected the CTR in Period 1. There is a possibility that similar results were not obtained in Period 2, because the experiments in Period 2 were conducted successively after Period 1. The notification was new to the subject in Period 1, but in Period 2 they were already familiar with it, and this also affected the CTR over time. In particular, the CTR of the EN Texts decreased significantly between Period 1 and Period 2. This is assumed to be due to both the loss of positive impression and strengthened negative impression given by the exaggeration instead. In other words, although EN Texts increase CTR in the initial period, they may cause a sharp decrease in CTR in the long run. Therefore, exaggeration of notification text should not be used in situations where notifications are made regularly.

Next, a statistical test for significant differences in screen transition rates could not be conducted due to the insufficient amount of data. Although we were not able to show a

significant difference, we obtained a result that showed the rate of normal notification texts was higher than that of EN Texts in Period 1. As mentioned earlier, there was a significant tendency for the CTR of EN Texts to be higher in Period 1, and since more subjects opened EN Texts than normal notification texts, it is natural that more subjects would make a screen transition. However, the results showed that there was almost no difference in the number of subjects who made screen transitions between normal notification texts and EN Texts, and the screen transition rate of EN Texts was lower than that of normal notification texts. It is likely that when the subjects opened the EN Texts and started the application, they judged that the application did not meet their heightened expectations and quit the application without making any screen transitions. Therefore, exaggeration to improve CTR when creating notification messages may result in lower engagement and user disengagement. In this study, we were not able to prove this claim due to the small amount of data, so it is necessary to conduct experiments with more data in the future.

6 Summary and future work

In this study, we proposed four indices of exaggerated notifications: "Unexpectedness," "Attractiveness," "Credibility," and "Usefulness," and conducted experiments to evaluate the validity of these indices. As a result, we found that 11 out of 14 notifications that significantly differed in any of the four indices met the definition of exaggeration, indicating that it was reasonable to evaluate them using the four indices.

In addition, we sent exaggerated notifications defined by these indices to subjects and investigated the effects of exaggeration on CTR and engagement. CTR showed a significant trend in period 1, but not in period 2 as a result. This suggests that exaggeration increases the CTR in the period when the notifications are first sent, but the effect is lost when users become accustomed to the notifications. Although the results of the engagement study did not show any significant differences, the subjects who opened the exaggerated notifications showed a tendency to exit the application without performing any screen operations, indicating that exaggeration may reduce engagement.

In conclusion, we found that (1) exaggerated notifications may temporarily increase the CTR but not continuously, and in some cases may have a negative impact on engagement, and (2) it is possible to determine and detect exaggerated notifications in advance by increasing or decreasing one or more of the four indicators.

In this study, we only confirmed the tendency of exaggerated notifications with a small data set of 22 subjects. Therefore, based on the results of this study, more experiments are needed to investigate the relationship between exaggerated notifications, CTR, and engagement in more detail. Specifically, the following two issues should be discussed. (1) Examination the impact of EN texts in depth. (2) Creation of a fully exaggerated EN text. The results of this study may have been insufficient in exaggerating the EN texts, so it is necessary to conduct the experiment (2) on a larger scale. We are currently planning to conduct experiments on a larger scale using the released version of the "Toyama Takeout Map".

References

1. Fuller, Inc., Mobile Market White Paper 2020, https://ja.appa.pe/reports/whitepaper-mobilemarket-2020
2. Kushal S. D and Vasudeva V.: Learning the click-through rate for rare/new ads from similar ads. In: Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval. 2010. p. 897-898.

3. Okoshi, T. Tsubouchi, K. Taji, M. Ichikawa, T and Tokuda, H.: Attention and engagement-awareness in the wild: A large-scale study with adaptive notifications, 2017 ieee international conference on pervasive computing and communications (percom). IEEE, pp. 100-110 (2017).
4. Fraser, K. and Conlan, O.: Enticing notification text & the impact on engagement, Proc. Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers. pp. 444-449, ACM (2020).
5. Khalid EI-Arini and Joyce Tang: Click-baiting, FACEBOOK (online), available from < https://about.fb.com/news/2014 /08/news-feed-fyi-click-baiting/> (accessed 2021-05-14)
6. Travis Yeh: Addressing Sensational Health Claims, FACEBOOK (online), available from <https://about.fb.com/news/2019/07/addressing-sensational-health-claims/> (accessed 2021-0515).
7. Google: Google Ads policies, Advertising Policies Help(online), available from <https://support.google.com/adspolicy/answer/6008942?hl=en&visit_id=637579653661401626-740926991&rd=1 >
8. Emily Bonnle: 250+ Push Notification Power Words to Skyrocket Mobile App Engagement, CleverTap(online), available from <https://clevertap.com/ blog/push-notification-power-words/ > (accessed 2021-05-16).
9. Burgers, C., Brugman, B. C., Renardel de Lavalette, K. Y., & Steen, G. J.: HIP: A method for linguistic hyperbole identification in discourse. Metaphor and Symbol, 31(3), 163-178. (2016)
10. Licata, J. W., Biswas, A., & Krishnan, B. C: Ambiguity and exaggeration in price promotion: perceptions of the elder and nonelder consumer. Journal of Consumer Affairs, 32(1), 56-81. (1998)
11. Ramiller, N.: Exaggeration in Information Systems: Charting an Inquiry into Its Functions, Processes and Paradoxes. AMCIS 2001 Proceedings, 390 (2001)
12. Olson, C.J. Toy, R.D. and Dover, A.P.: Do Cognitive Responses Mediate the Effects of Advertising Content on Cognitive Structure?, Journal of Consumer Research, Vol.9, No.3, pp. 245–262, 982 (1982).
13. Cowley, E.: Processing exaggerated advertising claims. Journal of Business Research, Vol.59, No.6, pp. 728-734 (2006).
14. Kunieda Y.: Impact of Exaggerated Expressions in Advertisements on Consumer Purchase Intentions (written Japanese), < http://c-faculty.chuo-u.ac.jp/~tomokazu/zemi_materials/4_kunieda.pdf> (2021-05-20).
15. Yamane, H., Hagiwara, M.: Tag Line Generating System Using Information on the Web. J. Adv. Comput. Intell. Intell. Informatics, 17(2), pp. 185-193. (2013)
16. Takeuchi R.: Consumers' Regulatory Focus and Advertising Avoidance. Japan Marketing Journal. 2018, Vol.38, No.2, p.39.
17. Toyama Takeout Map, available from < https://www.tulip-tv.co.jp /takeoutmap/> (accessed 2021-05-15).
18. Cialdini, R.B. Carl, A.K. and Raymond, R.R.: A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. Advances in experimental social psychology, Vol. 24, pp.201-234, (1991).
19. Gastwirth, J. L.: On the sign test for symmetry. Journal of the American Statistical Association, 66(336), pp. 821-823. (1971)
20. Fisher, R.A.: On the interpretation of χ 2 from contingency tables, and the calculation of P, Journal of the Royal Statistical Society, Vol.85, No.1, pp.87-94 (1922).