Multi-Objective Personalization in Multi-Stakeholder Organizational Bulk E-mail: A Field Experiment

RUOYAN KONG, University of Minnesota - Twin Cities, USA
CHARLES CHUANKAI ZHANG, University of Minnesota - Twin Cities, USA
RUIXUAN SUN, University of Minnesota - Twin Cities, USA
VISHNU CHHABRA, University of Minnesota - Twin Cities, USA
TANUSHSRISAI NADIMPALLI, University of Minnesota - Twin Cities, USA
JOSEPH A. KONSTAN, University of Minnesota - Twin Cities, USA

Bulk email is often used in organizations to communicate "important-to-organization" messages such as policy changes, organizational plans, and administrative updates. However, normal employees may prefer messages more relevant to their jobs or interests. Organizations face the challenge of balancing prioritizing the messages they prefer employees to know (tactical goals) while maintaining employees' positive experiences with these bulk emails, then they continue to read these emails in the future (strategic goals).

Could personalization help organizations achieve these tactical and strategic goals? In an 8-week field experiment with a university newsletter, we implemented a 4x5x5 factorial design on personalizing subject lines, top news, and message order based on both the employees' and the organization's preferences. We measured these designs' influences on the open/interest/recognition/read-in-detail rate of the whole newsletter and the single messages within it.

We found that "important-to-organization" messages only got higher recognition rates when being put on subject lines / top news (tactical goal). Mixing them with employee-preferred messages in top news did not bring further improvement to their own recognition rates but could improve the whole newsletter's recognition rate. Only when the top news solely contained the employee-preferred messages were the employees slightly more interested in the newsletter (strategic goal). We further analyze on which topics the employees and the organization's preferences conflicted. Finally, we discuss the design suggestions for organizational bulk email.[1]

CCS Concepts: • **Human-centered computing** → **Empirical studies in collaborative and social computing**.

Additional Key Words and Phrases: organizational communication, email, personalization, field experiment

---

[1]This is a pre-print version of a paper accepted to CSCW 2022 — The 25th ACM Conference On Computer-Supported Cooperative Work And Social Computing.

Authors' addresses: Ruoyan Kong, kong0135@umn.edu, University of Minnesota - Twin Cities, USA; Charles Chuankai Zhang, University of Minnesota - Twin Cities, USA; Ruixuan Sun, University of Minnesota - Twin Cities, USA; Vishnu Chhabra, University of Minnesota - Twin Cities, USA; Tanushsrisai Nadimpalli, University of Minnesota - Twin Cities, USA; Joseph A. Konstan, konstan@umn.edu, University of Minnesota - Twin Cities, USA.

Proc. ACM Hum.-Comput. Interact., Vol. 6, No. CSCW2, Article 528. Publication date: November 2022.
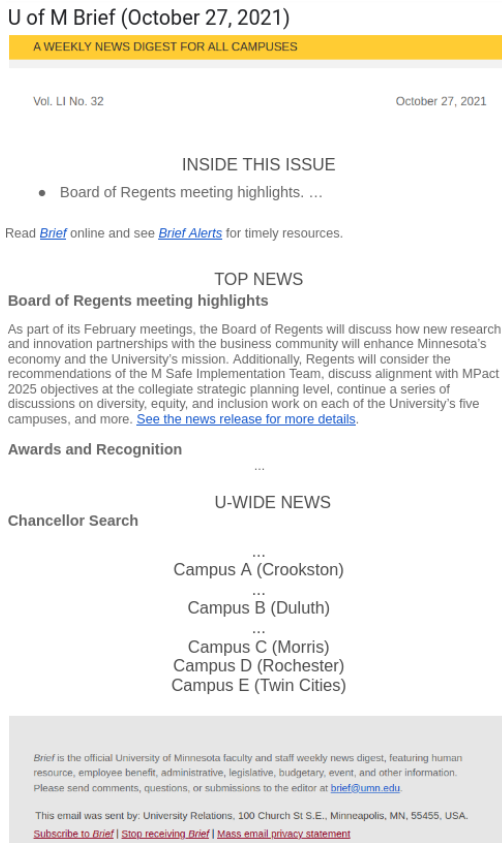
528

Fig. 1. A sample Brief on Oct 27, 2021. The messages within each section are hidden in this picture.
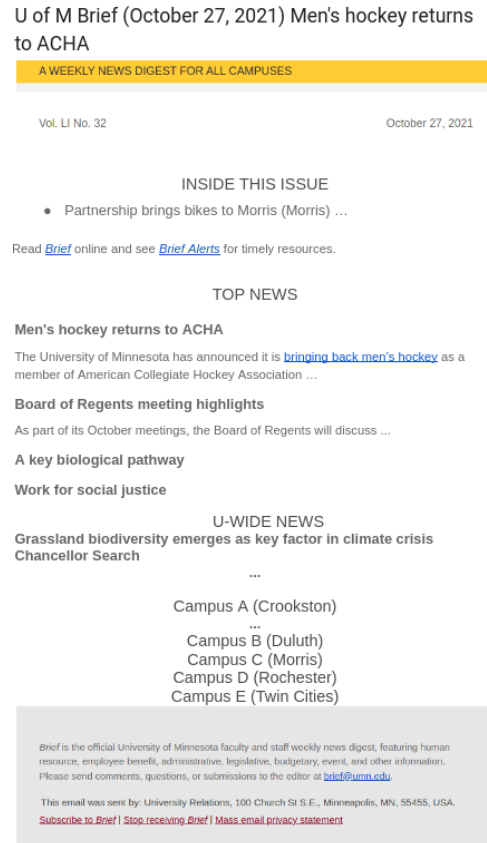
Fig. 2. A sample personalized Brief. The messages within each section are hidden in this picture.

1 INTRODUCTION

Organizations often send **bulk emails** (e.g., newsletters, all-company emails) to employees [13] to announce events, policy changes, and administrative updates. In the university we conducted this study with, an employee receives 30 bulk emails per week from the central offices on average. Many of them contain tens of **single messages**, which means make each employee receives over 250 single pieces of information from these bulk emails per week. Figure 1 is an example university bulk email of our study site University of Minnesota: U of M Brief. It is a weekly newsletter sent to all the employees across all 5 campuses with around 30 messages and 7 sections (top news, u-wide news, and each campus' news) in each newsletter.

These bulk emails (**organizational bulk email**) have **multiple stakeholders**, including the information sources (e.g., organizations' leaders), the **communicators** (the communication staff in charge of designing and sending bulk emails, e.g., newsletter editors), the recipients (the employees who receive these bulk emails), managers, etc. In this paper, we focus on 2 kinds of stakeholders: the organization leaders and communicators who represent the organization's perspective, and the normal employees who represent the recipients' perspective on messages' work-relevance/interest/importance level. The organization and employees may have different perspectives

[35, 36, 59]. For example, the university's board of regents might want all the employees to know about their meeting updates; then they put it as the first message of the Brief with the hope that all the employees who opened this Brief will notice this message (see Figure 1). However, the employees might perceive it as unactionable **high-level information** and view this Brief as irrelevant to their job. They might stop reading Brief after seeing this message.

Therefore, designing organizational bulk emails is a **multi-objective problem** for organizations. Their **tactical** objective is to use bulk emails to make employees aware of "important" messages, such as the board of regents meeting updates. At the same time, they have the **strategic** objective of maintaining the effectiveness of their communication channels by ensuring employees see messages they perceive as relevant and continue to read in the future. Organizations need to balance their short-term tactical goals and long-term strategic goals in designing these emails.

Here we see the opportunity of exploring personalization based on both **organization's preference** (organization's view on message's work-relevance and importance to employees as assessed by the newsletter editors) and **employee's preference** (employee's view on message's work-relevance and interest level to themselves) to help organizations reach these communication goals. For example, a design we tried with the newsletter above (see Figure 2) is to put **organization-preferred messages** (e.g., board of regents meetings) near to **employee-preferred messages** (e.g., Men's hockey, in a case where the specific employee like sports) to balance these two interests. Specifically, we conducted an 8-week field experiment with a university newsletter and 141 employees. We employed a 4x5x5 factorial design in personalizing subject line/top news/message order. We measured these designs' influences on the open/interest/recognition rate of the whole newsletter (strategic goals) and the recognition/read-in-detail rate of the messages within it (tactical goals).

Regarding the tactical goals, we found that "important-to-organization" messages could only get higher recognition rates but not higher read-in-detail rates when being put on subject lines / top news. Mixing them with employee-preferred messages in top news did not bring further improvement to their own recognition rates but brought better performance on the strategic goals — they improved the whole newsletter's recognition rate. However, only when the top news solely contained the employee-preferred messages would the employees be marginally more interested in the newsletter. We looked into the organization's preferences and the employees' preferences on message topics and found that the conflicts over what is important/relevant existed widely.

Our contributions include two parts. Regarding the fundamental theoretical advances, we studied the unique personalization problem with organizational bulk emails where the short-term tactical goals might not align with the longer-term strategic goals, while commercial bulk emails' tactical goals (e.g., recipients purchase the recommended products) and strategic goals (e.g., recipients perceive their emails as useful and keep reading them next time) are aligned [48, 53, 58]. Also, we conducted an 8-week field experiment to enable employees to get used to the experimental newsletters, while the research on organizational emails is mainly lab experiments [43], dataset analysis (Enron and Avocado [6, 31, 65]), and observational field studies [23, 34, 38, 42]. Regarding the practical advances, we provided the tradeoffs and suggestions for organizations in designing bulk emails. We designed a personalization framework for organization's bulk emails, including the process to collect stakeholders' preferences, the algorithms to generate personalized bulk emails, and the mechanisms to evaluate their performance. The rest of this paper includes related work (2), background (3), methods (4), results (5), and discussion (6). This study was approved by the IRB of the University of Minnesota.

2 RELATED WORK & GAPS

2.1 Organizational Communication and Bulk Email's Goals

Organizational communication has been defined by various disciplines [28, 40]. Wrench and Punyanunt-Carter [64] define it as: *"the process whereby an organizational stakeholder(s) attempts to stimulate meaning in the mind of another organizational stakeholder(s) through the intentional use of verbal, nonverbal, and mediated messages."* Organizational communication is a problem with multiple stakeholders, including the information providers, the information gatekeepers, the information recipients, and the organization itself. Instead of prioritizing one single stakeholder's preferences, the goal of organizational communication is to maintain the whole organization's productivity [60] by developing the awareness of organizational tasks, promoting a positive sense of belonging, and developing the understanding of organizational needs and goals.

Several studies have shown that internal organization emails, especially the organizational bulk email system, usually fail the communication goals above. First, employees are often too overloaded to have a good recognition level of or experience with these emails. Dabbish and Kraut [13] conducted a nationwide organization survey and found that email volume is positively correlated with the feeling of email overload at work in 2006. Grevet et al.[17] interviewed 19 Gmail users and found the problem of email overload became more serious in 2014 as the size of inbox and the number of unread messages increased hugely compared to Whittaker and Sidner's findings in 1996 [61]. Second, the stakeholders naturally have different preferences on the information, which stops them from paying attention to these emails. Dabbish et al. surveyed 128 employees of a university and found that employees prioritized the emails from the senders that they had a direct work relationship with [14]. Kong et al. [36] interviewed the recipients, managers, and communicators of a university. They found that the managers and communicators thought that employees should know about what was going on in the university. But the employees felt that these bulk emails were unactionable and irrelevant. Then they questioned the credibility of these bulk emails and gradually decided to stop reading them.

2.2 Bulk Email Personalization

2.2.1 Bulk Email Personalization's Content. Personalization has been pointed as a solution for email overload [11] and several studies used personalization to improve commercial bulk emails' performance. The personalization content could be demographic information like names [48, 58], majors, departments [56]; or preference information like browsing history [58], deals or tools recommendation [19, 53]. Though Sahni et al.'s experiments found adding recipients' names to subject lines useful [48], many other studies supported that the personalization based on preferences performed better than the personalization based on demographics. Wattal et al. [58] found that customers responded negatively to emails with identifiable information. Trespalacios and Perkins also found the effect of adding identifiable demographic information insignificant in an experiment with a university email [56]. Hawkins et al. pointed out that personalized messages need to provide the recipients with new information about themselves instead of simply adding names or addresses [19]. Wattal et al. [58] personalized email content based on customers' purchasing preferences and received positive responses. We personalize based on preferences instead of demographics in this study. Because we could not simply add the recipients' names to every internal newsletter of the organization — the employees could quickly learn that seeing their names in the organizational bulk emails means nothing special.

2.2.2 Bulk Email Personalization's Design. In this section, we looked at the categories of the personalization designs of commercial bulk emails. Besides that, we also searched the work around the advertising placements of web pages as we considered a similarity between attracting users'

attention to advertisements and attracting employees' attention to the "important-to-organization" bulk messages.

(1) Subject line: An informative subject line was believed to be a key factor in successful email marketing [57]. Sahni et al. added recipients' names to subject lines [48], and that method increased a marketing email's open rate by 20%. However, there are also studies showing that an uninformative subject line could create an information gap that attracts recipients to open emails [9, 49]. The "Long vs. Short Email Subject Line Test" of WhichTestWon.com in 2011 [57] found that a longer subject line led to a higher open rate. But Alchemy Worx's test on a discount promotion email found that a longer subject line led to lower open rates [63]. An explanation for the contradictory results is that the longer subject lines only influence by providing more information. Under this theory, the factor that actually matters here is whether the topic of the subject line matches with the recipients' preferences. The longer subject lines might get lower open rates but higher action rates, because only those who are interested in it will open it [24, 63].

(2) Top section: The traditional theory is that users would pay more attention to the top positions during browsing [51]. Wattal et al. [58], and Trespalacios and Perkins [56] tried adding recipients' names, majors, or departments to the greetings or the first paragraphs of emails in their studies and Wattal et al. found that greetings influenced their customers' response rates significantly.

(3) Selection of contents: Many studies personalized commercial bulk emails by selecting the most interesting content for recipients. Wattal et al. [58] put products that the customer might like most in the email. By analyzing 30 email-marketing campaigns, Rettie [46] found that response rate is negatively correlated with email length. Carvalho [10] proposed a personalization algorithm that put news liked by similar users in e-newsletters.

(4) Order of contents: Besides the theories supporting putting important information on top [51], there were also many studies supporting different arrangements. Wojdynski and Nathanie [62] examined 12 web page designs and found that the advertisements in the middle or bottom positions got better recognition. Heinz and Mekler [21] found that banner placement did not influence recognition and recall.

(5) Visual designs: Several studies focused on how to highlight the important content. Rettie [46] found that response rate is positively correlated with the number of images. Wojdynski and Nathanie [62] found that *"users tend to gaze at a target object that is surrounded by objects with weaker "demand for attention" values."*

We focused on leveraging personalization to improve bulk email's performance in this paper. But it is worth noting that there are other potential factors influencing recipients' engagement with bulk emails. For example, the email marketing platform Mailchimp found that sending frequency, writing (like the use of emojis), and sender's industry influence bulk email's open rate (ranging from 15% to 28% according to their report) [12]. Bulk email's from lines, signature lines [25], and sending time and day [1, 7] are also found to influence open rate by around 10% in the previous experiments from marketing, communication, and management science.

2.3 Gaps

Personalizing organizational bulk email is different from personal, commercial bulk email in several ways. Most important, employees may have an obligation to read, know, and act upon information from their employer — even information they may personally not find interesting — in a manner that does not apply to typical commercial bulk email. While commercial bulk email may use one-off

branding (to focus on one-time response rates) or recurring branding (to build a reputation and encourage repeat reading), organization bulk email nearly always uses recurring branding tied to the structure and leadership of the organization. Accordingly, organizations always have to balance the tactical goal of having employees read the messages they choose to send (and view as important) while also maintaining the effectiveness of the communication channel by having employees perceive the messages as relevant. There is no study on how different ways of personalization could help organizations achieve these two types of goals. In the following, we will introduce a field experiment we conducted to bridge this gap.

3 BACKGROUND

3.1 Study Site and Newsletter

The study site is a public university with over 25,000 employees and several campuses. There are communication offices in both the central units (e.g., presidential offices) and decentralized units (e.g., collegiate offices). There are newsletter editors in central units in charge of collecting news, designing, and distributing university-wide newsletters.

Before the study, we met with three communicators from the central offices and decided to experiment with the newsletter U of M Brief (see figure 1), which is sent to all the employees weekly (with subject line *"U of M Brief <Date>"*). Each brief contained around 30 messages and 7 sections. Brief encourages people to submit information about "need-to-know" administrative news, making the university more accessible and creating connections among faculty and staff, promoting healthy lives, or the university's mission of outreach, research, teaching, and education.

3.2 Communication Goals

Based on the studied site and newsletter above, we now frame the organization's communication goals with Brief. The tactical goals are about how well the specific group of messages are recognized and read. The strategic goals measure employees' overall experiences with Brief. Specifically, we measured these 6 metrics:
★ Recognition/read-in-detail rate: the percentage of the investigated messages being self-reported as seen/read-in-detail by the employees in the study's end survey (tactical goal). For example, the percentage of the messages in Top News the employees reported "seen" when Top News were all organization-preferred messages. [2]
★ Open rate: the percentage of the investigated Briefs being opened by the employees (strategic goal). For example, the percentage of a Brief being opened when we put the organization-preferred messages on subject lines.
★ Interest rate: the percentage of the investigated Briefs being rated as "interesting" by the employees (strategic goal).
★ Reading time: the average reading time of the investigated Briefs (strategic goal).
★ Overall recognition rate: the average of the recognition rates of all the investigated Briefs' messages (strategic goal).

3.3 Designs

We considered 5 kinds of personalization designs — original/random/employee-preferred/organization-preferred/mixed designs. We will discuss how we selected the employee/organization-preferred

---

[2]We could not measure the reading time of a single message in Brief because of the technical challenge of tracking specific regions' reading time naturally (our future work). First, many browsers (e.g., Chrome and Gmail) block access to the exact loading time of invisible pixels. Second, there is a lack of low-cost eye-tracking technology (e.g., eye-tracking based on a single computer camera) [15]; and employees might also pay more attention to bulk emails when being recorded by camera.

messages in 4.2. Let us now use Figure 1's Brief as an example. Suppose a faculty member is interested in biology stories and sports while the organization wants them to know about administrative and social justice updates.

A Subject line: which message to be added to the Brief's subject line.

    A1 Original subject line: *"U of M Brief (October 27, 2021)"*.

    A2 Random subject line: the original subject line with a random message, e.g., *"U of M Brief (October 27, 2021) - Fall 2021 Capstone presentations"*.

    A3 Organization-preferred subject line: the original subject line with the message that the organization mostly preferred the faculty member to read, e.g., *"U of M Brief (October 27, 2021) - Board of Regents meeting highlights"*.

    A4 Employee-preferred subject line: the original subject line with the message that the faculty member mostly preferred, e.g., *"U of M Brief (October 27, 2021) - Men's hockey return to ACHA"*.

B Top news: which 4 messages are to be selected as the Brief's top news.

    B1 Original top news: the same top news as Figure 1.

    B2 Random top news: use 4 random messages.

    B3 Organization-preferred top news: use the 4 messages the organization most preferred the faculty member to know. E.g., Board of Regents meeting highlights, Work for social justice, etc.

    B4 Employee-preferred top news: use the 4 messages the faculty member mostly preferred as top news. E.g., Men's hockey return to ACHA, A key biological pathway, etc.

    B5 Mixed top news: mix 2 employee-preferred messages and 2 organization-preferred messages in top news. E.g., Men's hockey return to ACHA, Board of Regents meeting highlights, A key biological pathway, Work for social justice.

C Message Order: how to sort the messages in the non-top sections of this Brief.

    C1 Original order: use the original Brief's messages' order.

    C2 Random order: sort the messages randomly.

    C3 Organization-preferred order: sort the messages by the organization's preference.

    C4 Employee-preferred order: sort the messages by the faculty member's preference.

    C5 Zipper order: repeat this process — select the message with the highest employee-preference score (see 4.2), the message with the highest organization-preference score, the message with the 2nd highest employee-preference, etc.

If a message is added to the subject line but not selected to top news, we add it to the end of top news to avoid the employees feeling deceived if they click into the Briefs because of the subject lines. If a message from the campus sections was selected to top news, the name of its campus would be added to its title.

Within each treatment, we had two control groups: a good original control group (A1, B1, C1) which used the original subject lines/top news/message order — as we discussed, these were carefully selected by an experienced editor (the communicator of Brief) according to their criterion on how to design Briefs; a bad random control group (A2, B2, C2) which used random subject lines/top news/message order generated by the system. Figure 2 is a sample personalized Brief for this faculty member if we assigned them to A4 x B5 x C5.

3.4 Research Questions and Hypotheses

We proposed hypotheses and questions on these designs' influences on the communication goals.

A Subject lines. When subject lines match the employees' preferences, Brief might achieve these strategic goals:

Adding employee-preferred messages on subject lines will increase the newsletter's interest rate (**H1.1**) / reading time (**H1.2**) / overall recognition rate (**H1.3**) / open rate (**H1.4**).

On tactical goals, the messages on subject lines should have a greater chance of being seen by the employees:

(**H1.5**) Putting messages on subject lines will increase these messages' recognition rates.

We expect only to see an improvement in the read-in-detail rate for those employee-preferred messages, as the content should be interesting to employees to make them click/read in detail [29, 30]:

(**H1.6**) Putting employee-preferred messages on subject lines will increase their read-in-detail rates.

B Top News. When top news matched the employees' preferences, Brief might achieve a better interest rate:

H2.1 Putting employee-preferred messages in top news will increase the newsletter's interest rate.

We do not have theories to predict the effect of placing employee-preferred messages in top news on the reading time and overall recognition rate. For example, when putting employee-preferred messages in top news, employees might be motivated to read the rest of the newsletter or only read top news and leave. We proposed these questions:

What is the effect of putting organization-preferred messages/employee-preferred messages/mixing employee-preferred messages and organization-preferred messages in top news on the newsletter's reading time (**Q2.2**) and overall recognition rate (**Q2.3**)?

We hope that the mixed top news let employees read the organization-preferred messages when reading interesting messages: Mixing organization-preferred messages with the employee-preferred messages in top news will increase the organization-preferred messages' recognition rates (**H2.4**).

Besides that, we also had the hypotheses similar to those for the subject lines on tactical goals:

(**H2.5**) Putting messages in top news will increase their recognition rates.

(**H2.6**) Putting employee-preferred messages in top news will increase their read-in-detail rates.

C Message Order. We were uncertain about the direction and scale of message order's effect. We proposed the following questions: what is the effect of sorting messages by employee's preference/organization's preference/zipper order on the overall recognition rate (**Q3.1**) / reading time (**Q3.2**) of the newsletter?

(**Q3.3**) What is the effect of interleaving messages (sorting messages by the zipper order of employee/organization's preference) on the recognition rates of the organization-preferred messages?

We hope that sorting by employee's preference would make them feel the Brief is more interesting:

(**H3.1**) Sorting messages by employee's preference will increase the interest rate of the newsletter.

We summarized our hypotheses and research questions in Table 1. There were blanks (marked as to be observed) in this table when we did not find any related work or reasons to make a guess on a significant effect. We just observed what happened in these blanks. Besides the questions above.

4 METHODS

We collaborated with the editor of Brief to conduct the experiment in the university with 141 employees. The experiment took 8 weeks. The steps were (see Figure 4):

Step 1 (week 1), recruiting and assigning participants: through Brief and a communication newsletter. Each selected participant would be assigned to a treatment combination $A_i \times B_j \times C_k$ through the whole study.

Step 2 (week 1), collecting and calculating the employees' preferences: the employees filled in preference surveys and we collected their work-relevance/interest scores for each topic.

Step 3 (week 2 to 7), collecting and calculating the organization's preferences: the editor sent the draft of the newsletter to the experiment system every week. The system extracted text and html, then sent the editor a survey to collect each message's topics (up to 4) and work-relevance/importance scores from the organization's perspective.

Table 1. Summary of the newsletter's communication goals, hypotheses, and research questions. org-pref messages: organization-preferred messages. emp-pref messages: employee-preferred messages. A1/2, B1/2, C1/2 are control groups.

Group	Treatment	Strategic Goal				Tactical Goal	
		Interest Rate of this Brief	Reading time of this Brief	Recognition Rate of this Brief	Open Rate of this Brief	Recognition Rate of this message	Read-in-detail Rate of this message
A: Subject lines	1: Original	To be observed	To be observed	To be observed	To be observed	To be observed	To be observed
	2: Add a random message	To be observed	To be observed	To be observed	To be observed	To be observed	To be observed
	3: Add an org-pref message	To be observed	To be observed	To be observed	To be observed	H1.5 Increase recognition rate.	To be observed
	4: Add an emp-pref message	H1.1 Increase interest rate.	H1.2 Increase reading time.	H1.3 Increase overall recognition rate.	H1.4 Increase open rate.	H1.5 Increase recognition rate.	H1.6 Increase read-in-detail rate.
B: Top news	1: Original top news	To be observed	To be observed	To be observed		To be observed	To be observed
	2: Put random messages	To be observed	To be observed	To be observed		To be observed	To be observed
	3: Put org-pref messages	To be observed	Q2.2 How does it affect reading time?	Q2.3 How does it affect overall recognition rate?		H2.5 Increase recognition rate.	To be observed
	4: Put emp-pref messages	H2.1 Increase interest rate.	Q2.2 How does it affect reading time?	Q2.3 How does it affect overall recognition rate?		H2.5 Increase recognition rate.	H2.6 Increase read-in-detail rate.
	5: Mix emp-pref /org-pref messages	To be observed	Q2.2 How does it affect reading time?	Q2.3 How does it affect overall recognition rate?		H2.4 Increase recognition rate of org-pref messages. H2.5 Increase recognition rate.	To be observed
C: Order	1: Original order	To be observed	To be observed	To be observed		To be observed	To be observed
	2: Random order	To be observed	To be observed	To be observed		To be observed	To be observed
	3: Sort by org -preference	To be observed	Q3.1 How does it affect reading time?	Q3.2 How does it affect overall recognition rate?			
	4: Sort by emp -preference	H3.1 Increase interest rate.	Q3.1 How does it affect reading time?	Q3.2 How does it affect overall recognition rate?			
	5: Sort by mix-order	To be observed	Q3.1 How does it affect reading time?	Q3.2 How does it affect overall recognition rate?		Q3.3 How does it affect recognition rate of org-pref messages?	

Step 4 (week 2 to 7), generating newsletters: the system sent the original non-personalized newsletter to collect base performance data in week 2 and generated personalized newsletters based on the employees' experimental groups, organization's preference, and employee's preference during weeks 3 to 7.

Step 5 (week 8), collecting performance metrics and feedback: the system sent out the end surveys to collect the recognition/read-in-detail/interest data, and the log files of a plugin which tracked open rates and reading time.

4.1 Recruitment (step 1, week 1)

The goal of this experiment is to study newsletter reading in a natural context. The scope of this experiment is organizations' newsletters sent to a large number of employees and their regular and occasional readers. We did not study employees who never read Briefs because reading the Brief is not their natural behavior — being in the experiment could force their open rate and recognition

Domain	Path	Start timestamp	Finish timestamp	URL	Title
https://mail.google.c	/mail/u/1/#search/brief/	1631920592000	1631920756000	https://mail.google.c	U of M Brief (July 28, 2021)
https://mail.google.c	/mail/u/1/#search/brief/	1631920765000	1631920820000	https://mail.google.c	U of M Brief (July 28, 2021)
https://mail.google.c	/mail/u/1/#search/brief/	1631921339000	1631921350000	https://mail.google.c	U of M Brief (July 28, 2021)

Fig. 3. A sample csv file recorded by the plugin. It records the domain, path (which helps us to discriminate whether it is the experimental Brief), start/end reading timestamp, url, and tab titles.

rate to increase from 0 to a substantial level, which could largely interfere with the main outcomes of our interventions. Accordingly, we posted the recruitment message as the first message of a Brief. In the hope of broadening our participants, we also reached out to the communicator network and posted the message at the top of its newsletter [3]. The newsletters only had their brands without specific messages on their subject lines that could be used to target specific audiences [4].

We planned to have at least 100 participants. This number was estimated through 1000 simulations on the generalized linear mixed model power analysis tool SMIR [16]. The analysis of open rate/interest rate, where we could only collect 5 data points from each employee, has the highest requirement. We target to observe a 15% change in these metrics with a 20% standard deviation and an 80% power. Considering the dropout rate, we targeted 140 to 150 participants.

In the signup form, we asked the potential participants whether they were employees of the university and mainly used Gmail and Chrome in reading Briefs and selected those confirmed. We also got their campuses and job categories. To make our surveys more concise, 20 job families of the university's human resource system were summarized into 7 job categories by two researchers and the Brief editor, according to whether these categories were considered different audience groups of Brief (see Appendix B).

We received 304 responses to our recruitment message and selected participants from this pool. We balanced the number of selected participants from different campuses, job categories, and recruiting sources and contacted 181 employees to set up the study (38 did not reply and were not enrolled). Each employee filled in a preference survey on message's topics and had a 20-minute 1-to-1 zoom meeting with a research team member. In this meeting, we helped the employee: 1) set up a filter rule in their university Gmail, which archived all the original Briefs they received during the experiment into a separate folder. They were told to avoid checking this folder during the study; 2) install a plugin on their Chrome browser. The plugin only recorded the time they spent when they were on a tab with the text "U of M Brief" (see figure 3). 141 participants completed the setup process (2 employees could not install the plugin and were not enrolled). The participants were compensated with a $20 Amazon gift card after setting up the study.

4.2 Personalization Procedure

To define and collect preferences data within limited survey questions, we used "topic" as a bridge to connect messages and preferences (we assume that each message could have at most 4 topics). 20 topics were summarized by a thematic analysis [8] of 140 messages from 5 previous Briefs (see

---

[3]We did two things to understand the generalizability in view of the Brief-centric recruitment — 1) compare the participants from different sources: the 10 other employees recruited from the communication newsletter had a base recognition rate of the week 2's original Brief (29%) similar to the other 107 employees recruited from Brief (30%); 2) checking the number of occasional Brief readers: 44 of our participants did not open all the experimental Briefs, and 22 participants opened less than or equal to 60% of the experimental Briefs (Brief's average open rate).
[4]The messages closest to our recruitment message in that recruitment Brief had topics on Administrative News/Student Stories/Awards and Recognition. According to Table 5, our participants were not more interested in these topics compared to other topics.

Table 2. The variables used in the personalization procedure and their definitions. Except that $org\_importance_k$ is the same for all the employees, the employee/organization's preferences were calculated by the inputs from the employees/editor on the left column.

Inputs from Employees	Definition	Employee's Preference	Definition
$campus_i$	$employee_i$'s campus	$emp\_interest_{i,k}$	$employee_i$'s interest score for $message_k$ (0 to 1)
job_i	$employee_i$'s job category (Appendix B)	$emp\_relevance_{i,k}$	$employee_i$'s work-relevance score for $message_k$ (0 to 1)
$interest_{i,j}$	whether $employee_i$ looks for interesting messages from $topic_j$ (0 or 1)	$emp\_pref_{i,k}$	$employee'_i$s preference on $message_k$ $(= (emp\_interest_{i,k} + emp\_relevance_{i,k})/2,\ 0\ \text{to}\ 1)$
$relevance_{i,j}$	whether $employee_i$ looks for work-relevant messages from $topic_j$ (0 or 1)		
Inputs from Editor	**Definition**	**Organization's Preference**	**Definition**
$important_k$	$message'_k$s general importance to all the employees from the organization's view (1 to 4)	$org\_importance_k$	$message_k$'s general importance score to all the employees from the organization's view $(= (important_k - 1)/3,\ 0\ \text{to}\ 1)$
$target\_campus_k$	$message_k$'s target campuses (list)	$org\_relevance_{i,k}$	organization's work-relevance score for $message_k$ given $employee_i$ (0 to 1)
$target\_job_k$	$message_k$'s target job categories (list)	$org\_pref_{i,k}$	organization's preference on $message_k$ given $employee_i$ $(= (org\_relevance_{i,k} + org\_importance_k)/2,\ 0\ \text{to}\ 1)$
$topic\_list_k$	$message_k$'s relevant topics (list, up to 4, Appendix A)		

Appendix A). 5 research team members grouped them, labeled the clusters, and identified the hierarchy. The Brief editor checked the list and suggested two special topics — "news from my campus" and "news from other campuses". We summarized the symbols we used in the personalization procedure and their definitions in Table 2.

4.2.1 Collecting and calculating employees' preferences (step 2, week 1). At the set up of this study, we asked the participants to fill out a preference survey (figure 4 - week 1). For each topic, we asked the participants to check whether the statements "I would look up this category for messages interesting to me" or "I would look up this category for messages work-relevant to me" applied to them separately. Let's call $employee_i$'s answers to $topic_j$ as $interest_{i,j}$ and $relevance_{i,j}$, which took the value 0 or 1. We also asked the campus ($campus_i$) and job category (job_i) in the survey.

Then for $employee_i$ and $message_k$, we calculated the employee's preference on this message ($emp\_pref_{i,k}$) with 3 steps. First, $employee_i$'s interest score for $message_k$ is defined as (see 4.2.2 for $topic\_list_k$ and $target\_campus_k$):

$$emp\_interest_{i,k} = \begin{cases} 0, \text{if } employee_i \text{ is not interested in other campuses' messages and } campus_i \\ \quad \notin target\_campus_k \\ avg(\{interest_{i,q}|topic_q \in topic\_list_k\}), \text{otherwise} \end{cases} \quad (1)$$

Second, we calculated $employee_i$'s work-relevance score $emp\_relevance_{i,k}$ for $message_k$ as:

$$emp\_relevance_{i,k} = \begin{cases} 0, \text{if } employee_i \text{ does not look for other campuses' relevant messages or} \\ \quad campus_i \notin target\_campus_k \\ avg(\{relevance_{i,q}|topic_q \in topic\_list_k\}), \text{otherwise} \end{cases} \quad (2)$$

Third, $employee_i$'s preference on $message_k$ is defined as

$$emp\_pref_{i,k} = (emp\_interest_{i,k} + emp\_relevance_{i,k})/2 \quad (3)$$

4.2.2 Collecting and calculating the organization's preferences (step 3, week 2 to 7). During weeks 2 to 7, the Brief editor provided the organization's preference with each message weekly (Figure 4, week 2 to 7 (1) - (5)). The Brief editor sent the draft Brief to the system every week. The system

then retrieved the messages (subject lines, titles, content, html, etc.) from the draft Brief, generated the editor survey, and sent it to the editor. The system listened to the editor's responses and loaded the responses to the database. After the responses were successfully loaded, a verification message would be sent to the editor to ensure that the original Briefs could only be sent after the system got the data to generate personalized Briefs. The editor survey contained the following questions for each $message_k$:

(1) How relevant this message is in building community, pride, common understandings of excellence and mission of the university (from 1: not relevant to 4: very relevant)? ($importance_k$).
(2) Select the employee categories that might find the message above work-relevant ($target\_job_k$).
(3) Specify the message's relevant topics (select no more than 4 topics) ($topic\_list_k$).
(4) There is an implicit question on target campus ($target\_campus_k$), as the editor suggested that the original Briefs' campus sections had already represented it.

Then we calculate the organization's preference ($org\_pref_{i,k}$) on $message_k$ given $employee_i$ by 3 steps. First, the organization's work-relevance score for $message_k$ given $employee_i$ with job category (job_i) and campus ($campus_i$) is:

$$org\_relevance_{i,k} = 1, if\ campus_i \in target\_campus_k\ and\ job_i \in target\_job_k;\ 0, otherwise \qquad (4)$$

Second, $message_k$'s general importance score to all the employees is just the standardization of $important_k$. This value is the same for all the employees, as the editor suggested that the important messages should apply to all:

$$org\_importance_k = (important_k - 1)/3 \qquad (5)$$

Third, the organization's preference on $message_k$ given $employee_i$ is defined as:

$$org\_pref_{i,k} = (org\_relevance_{i,k} + org\_importance_k)/2 \qquad (6)$$

4.2.3 Generating newsletters (step 4, week 2 to 7). After we calculated these preference data, the system generated personalized Briefs from weeks 3 to 7 (original Brief for week 2) and sent these Briefs to the employees. We give the employees the choice to select to receive the email at 6 AM or 9 AM, given which time is better for them to receive Briefs and read it on their laptop or desktop's Chrome with the plugin we installed (eventually, we did not observe significant differences on the performance metrics between the 6 AM and 9 AM group).

With a 4 x 5 x 5 factorial design on the treatments A (subject lines), B (top news), C (message order) below, each participant would be assigned to a treatment combination $A_i \times B_j \times C_k$ through the study randomly. Their Briefs' would be generated according to the criterion in 4.2, based on the employee preference emp_pref and organization preference org_pref we calculated for each message (Figure 4, week 2 to 7 (6)).

4.3 Collecting Performance Metrics and Feedback (step 5, week 8)

At week 8, we sent out the end survey (Figure 4, week 8). Participants were compensated with another $20 Amazon gift card after submitting the end surveys. We collected the recognition data for the message below: 1) week 2 and week 7's messages in the top news, up to the top 10 messages in the u-wide news, up to the top 2 messages in the participant's campus news; 2) weeks 5 to 7's messages in the top news; 3) weeks 3 to 7's messages on the subject lines.

The recognition data was collected by the question "Have you seen it in recent Briefs? No/Not Sure/Skimmed/Read fully". We defined $employee_i$'s *recognition* as 1 if the answer is skimmed or read fully, and *read_in_detail* as 1 if the answer is read fully. After that, the survey asked the participants to indicate how interesting each Brief is to them in general from "1 Not interesting" to "4 Very interesting". The survey also collected plugin data. We would know the reading time of each Brief, and whether the participants opened a Brief or not. We also collected the interest

Preference Survey

	$interest_{i,j}$	$relevant_{i,j}$
$topic_1$	1	0
$topic_2$	1	1
...		
$topic_j$	0	0
$campus_i, job_i$		

$employee_i$

Week 2 to 7: (6) Generate personalized Briefs

Week 1: (1) Send preference surveys; install filter rules and plugins

Week 1: (2) Employee fill in surveys; Load responses to database

Week 8: Send end surveys; collect plugin data

	$employee_i$ preference		communicator preference	
	$emp\_interest_{i,k}$	$emp\_relevance_{i,k}$	$org\_relevance_{i,k}$	$org\_importance_k$
$message_1$	1	0.5	0	1
$message_2$	1	0	1	0.33
...				
$message_k$	...	...		...

Experiment System

Week 2 to 7: (5) Generate employee/organization preference for each employee-message pair

Week 2 to 7: (4) Send verification message

Week 2 to 7: (1) Send Brief's drafts

Week 2 to 7: (2) Extract Messages from original Briefs; Generate Editor survey and send it

Week 2 to 7: (3) Brief editor fill in surveys; Load responses to database

Actions of Server

Actions of Employees

Actions of Brief Editor

Brief editor

Editor Survey

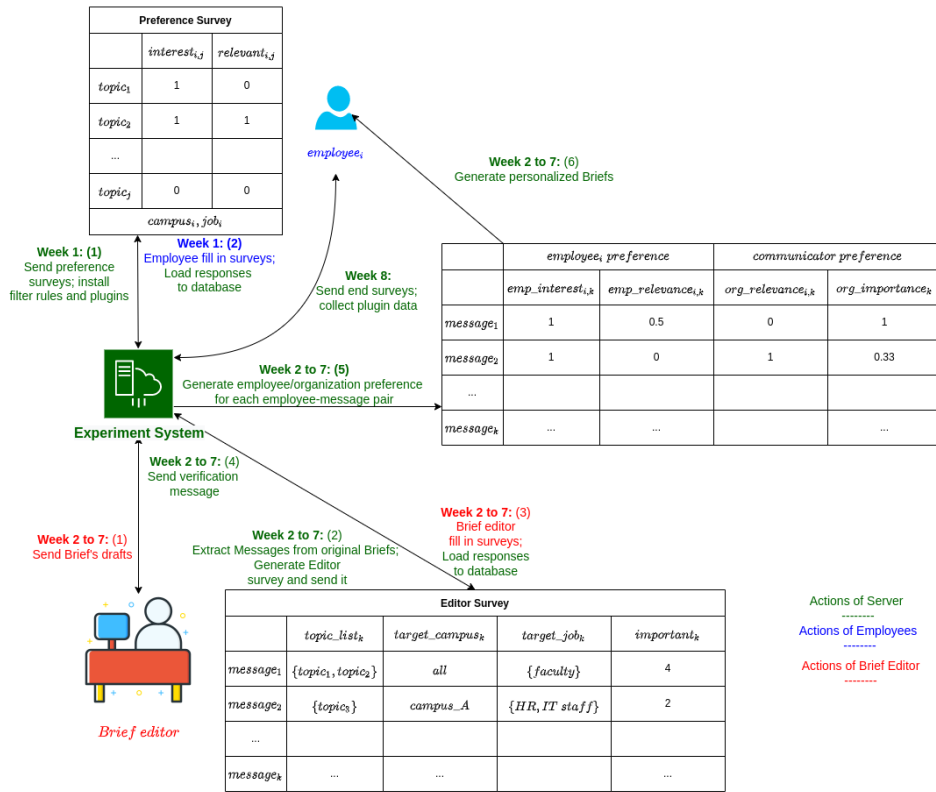|	$topic\_list_k$	$target\_campus_k$	$target\_job_k$	$important_k$
$message_1$	$\{topic_1, topic_2\}$	all	$\{faculty\}$	4
$message_2$	$\{topic_3\}$	$campus\_A$	$\{HR, IT\ staff\}$	2
...				
$message_k$	...	...		...

Fig. 4. The outline of the experiment procedure. Weeks 2 to 7's steps were repeated weekly. We sent the original Brief in week 2.

scores (scale 1 to 4) and work-relevant scores (scale 1 to 4) for week 2's messages to study the size of conflicts. The order of these questions is randomized. The participants were asked not to search their inbox while answering these questions. For each experimental group (e.g., the Briefs with random subject lines), we reported and tested:

★ *recognition_rate* and *read_in_detail_rate*: the percentage of the considered messages (e.g, the messages in the top news, the message on subject lines) in the experimental group that got *recognition* = 1 or *read_in_detail* = 1.

★ *overall_recognition_rate*: the percentage of the messages in this experimental group's Briefs with *recognition* = 1.

★ *interest_rate*: the percentage of Briefs that got an interesting level ≥ 3 in that experimental group.

★ *open_rate*: the percentage of Briefs that got *open* = 1 in that experimental group.

★ *reading_time*: the average of the Briefs' reading time in that experimental group.

We received 132 responses for the end survey, and 117 of them were complete. There were 15 incomplete responses either because the participants did not complete the surveys, the plugin was deleted or blocked by a Chrome update, or the participants lost access to their devices. This dataset contained the recognition data and read-in-detail data of 4242 messages, and the recognition data,

open data, and reading time data of 702 Briefs in total. We did received 2 reports of participants forgetting to read in the browser, and their data was excluded.

To avoid spillover effects [52], our participants were scheduled for separate 1-on-1 meetings in the setup, and they were not aware of each other's participation nor their experimental groups. We observed no communication or sharing that would have led to spillover effects. To avoid Hawthorne effects [26], we sent out the original Briefs in week 2, and measured the base performance data, which would be later included in our models as a factor. And our participants were in the experiment for 6 weeks, to avoid incentivizing participants to pay more attention to read/remember these Briefs, we only analyze the performance data collected in the last week of this experiment.

5   RESULTS

5.1   Analysis and Overview

We summarized the performance of each experimental group in Table 3 and the results on the hypotheses and questions in Table 4. We built mixed logistic models to evaluate categorical performance metrics (interest rate, recognition rate, open rate, and read-in-detail rate) and mixed linear models to test numerical performance metrics (reading time) by the afex package, which provided ANOVA table with likelihood-ratio tests for both linear and logistic models [54]. We had a random effect based on subjects (from which employee we collected this data point) [5], and we selected likelihood-ratio tests because we had many levels on the random effect (number of participants) [4]. The independent variables include the corresponding experimental groups and the base performance metrics (the average of that performance metric given the corresponding employee's reactions with week 2's original Brief). For the base open rate, we used the number of Briefs they opened in 2021/the number of Briefs they received in 2021 before the experiment. We asked the employees to input queries in their Gmail in the preference survey to retrieve this number. If they have deleted Briefs, they reported their approximate numbers. We excluded the employees who gave the same interest scores to all the experimental Briefs from the analysis of interest scores and excluded the employees who opened all or did not open any of the experimental Briefs from the analysis of open rates.

For the mixed logistic models in this paper, we checked [27] 1) whether the numeric independent variables were linearly associated with the dependent variable in logit scale by visually plotting the line of predictor's value - logit of predicted probabilities; 2) multicollinearity (whether GVIF < 2); 3) whether outlier exists (by package dharma [18]). For the mixed linear model on reading time, we transformed $reading\_time$ and $base\_reading\_time$ by $log10(1+x)$. 3 outliers (the 3 emails were read for more than 30 minutes) were removed. The transformation and outlier removing were needed to satisfy the normality requirement of the model's residuals. We checked the homogeneity of variance by Levene's Test and checked the normality of residuals by QQPlot [47].

For each model and effect, first, we calculated the average of that performance metric of each experimental group. Second, we checked whether the effect was (marginally) significant by its ANOVA table to see whether there existed significant differences among different experimental groups (treatments) of this effect. If it was significant and we did observe large differences between some experimental groups with the control groups, we conducted pairwise tests between these experimental groups with its control groups (see Table 4 for the effects, treatments, and control groups we examined). The P-values of the pairwise tests were adjusted by the holm-bonferroni method [22]. We got the following marginal / significant results on the hypotheses and questions (see Table 4 for the numbers):

Interest rate: H2.1 Putting employee-preferred messages in top news marginally increased Brief's interest rate versus putting random messages.

Table 3. **The performance metrics of each experimental group**, and the standard deviations of that group's participants' performance metrics. The blank cells are not applicable or not of interest. The employees who gave the same interest scores to all the experimental Briefs were excluded from the analysis of interest rates; those who opened all or did not open any of the experimental Briefs were excluded from the analysis of open rates. non-s: messages not on subject lines. non-t: messages not in top news.

Group	Treatment	Strategic Goal				Tactical Goal	
		Interest Rate of Brief (%)	Reading Time of Brief (seconds)	Recognition Rate of Brief (%)	Open Rate of Brief (%)	Recognition Rate of Message (%)	Read-in-detail Rate of Message (%)
Non-s messages						**41%**, 20%	**15%**, 13%
A: Subject lines	1: Original subject line	**68%**, 22%	**143s**, 92s	**44%**, 25%	**65%**, 21%		
	2: Add a random message	**69%**, 27%	**150s**, 105s	**39%**, 20%	**63%**, 14%	**40%**, 27%	**15%**, 17%
	3: Add an org-pref message	**67%**, 29%	**149s**, 96s	**40%**, 23%	**60%**, 24%	**60%**, 24%	**18%**, 19%
	4: Add an emp-pref message	**68%**, 31%	**141s**, 76s	**36%**, 20%	**76%**, 8%	**56%**, 27%	**24%**, 19%
Non-t messages						**37%**, 25%	**13%**, 14%
B: Top news	1: Original top news	**68%**, 38%	**146s**, 98s	**38%**, 25%		**44%**, 27%	**17%**, 15%
	2: Put random messages	**58%**, 29%	**123s**, 68s	**30%**, 17%		**32%**, 16%	**9%**, 12%
	3: Put org-pref messages	**63%**, 25%	**158s**, 81s	**46%**, 23%		**49%**, 16%	**18%**, 12%
	4: Put emp-pref messages	**76%**, 21%	**142s**, 82s	**37%**, 21%		**49%**, 21%	**22%**, 18%
	5: Mix emp-pref /org-pref messages	**72%**, 23%	**162s**, 123s	**49%**, 19%		**53%**, 17%	**18%**, 14%
C: Message order	1: Original message order	**66%**, 24%	**136s**, 66s	**40%**, 23%			
	2: Sort messages randomly	**69%**, 28%	**127s**, 77s	**34%**, 20%			
	3: Sort messages by org-pref	**59%**, 30%	**166s**, 124s	**46%**, 25%			
	4: Sort messages by emp-pref	**74%**, 31%	**146s**, 84s	**41%**, 20%			
	5: Sort messages by mix-order	**74%**, 19%	**157s**, 104s	**39%**, 22%			

Table 4. Results of hypotheses and research questions. The bold texts show (marginally) significance.
Format: experimental group mean = control group and its mean + difference between experimental and control groups (p.val). Blanks: not applicable or not of interest. Signif. codes: '*' 0.05, '+' 0.1, 'NS' no significant effect was found. Control groups:
rnd: random control group; org: original control group;
rnd-s: random subject lines' messages; non-s: messages not on subject lines;
rnd-t: random top news' messages; org-t: original top news' messages; non-t: messages not in top news.
P-values were adjusted by Holm-Bonferroni correction for controlling Type I error [22].

Group	Treatment	Strategic Goal				Tactical Goal	
		Interest Rate of Brief (%)	Reading Time of Brief (seconds)	Recognition Rate of Brief (%)	Open Rate of Brief (%)	Recognition Rate of Message (%)	Read-in-detail Rate of Message (%)
A: Subject lines	Anova P.val	(0.867)	(0.553)	(0.628)	(0.349)	**(0.001*)**	**(0.009*)**
	3: Add an org-pref message	NS	NS	NS	NS	H1.5 Increase recognition rate. **60%=non-s(41)+19 (0.001*)** **60%=rnd-s(40)+20 (0.007*)**	NS
	4: Add an emp-pref message	H1.1 Increase interest rate. NS	H1.2 Increase reading time. NS	H1.3 Increase overall recognition rate. NS	H1.4 Increase open rate. NS	H1.5 Increase recognition rate. **56%=non-s(41)+15 (0.001*)** **56%=rnd-s(40)+16 (0.026*)**	H1.6 Increase read-in-detail rate. **24%=non-s(15)+9 (0.008*)** rnd-s: NS
B: Top news	Anova P.val	(0.182)	(0.809)	**(0.008*)**		**(0.001*)**	**(0.001*)**
	3: Put org-pref messages	NS	Q2.2 How does it affect reading time? NS	Q2.3 How does it affect overall recognition rate? org: NS **46%=rnd(30)+16 (0.000*)**		H2.5 Increase recognition rate. **49%=non-t(37)+12 (0.002*)** rnd-t: NS org-t: NS	NS
	4: Put emp-pref messages	H2.1 Increase interest rate. org:NS **76%=rnd(58)+18 (0.075+)**	Q2.2 How does it affect reading time? NS	Q2.3 How does it affect overall recognition rate? NS		H2.5 Increase recognition rate. **49%=non-t(37)+12 (0.001*)** **49%=rnd-t(32)+17 (0.008*)** org-t: NS	H2.6 Increase read-in-detail rate. **22%=non-t(13)+9 (0.001*)** **22%=rnd-t(9)+13 (0.005*)** org-t: NS
	5: Mix emp-pref /org-pref messages	NS	Q2.2 How does it affect reading time? NS	Q2.3 How does it affect overall recognition rate? **49%=org(38)+11 (0.069+)** **49%=rnd(30)+19 (0.002*)**		H2.4 Increase recognition rate of org-pref messages. orn_B3: NS H2.5 Increase recognition rate. **53%=non-t(37)+16 (0.001*)** **53%=rnd-t(32)+21 (0.050+)** org-t: NS	NS
C: Message order	Anova P.val	**(0.088+)**	(0.674)	(0.446)			
	3: Sort messages by org-preference	NS	Q3.1 How does it affect reading time? NS	Q3.2 Increase/decrease overall recognition rate? NS			
	4: Sort messages by emp-preference	H3.1 Increase interest rate. NS	Q3.1 How does it affect reading time? NS	Q3.2 How does it affect overall recognition rate? NS			
	5: Sort messages by mix-order	NS	Q3.1 How does it affect reading time? NS	Q3.2 How does it affect overall recognition rate? NS		Q3.3 How does it affect recognition rate of org-pref messages? NS	

Overall recognition rate: Q2.3 Mixing organization/employee-preferred messages in top news increased Brief's overall recognition rate significantly versus putting random messages, marginally versus putting original messages.

Putting organization-preferred messages in top news significantly increased Brief's overall recognition rate versus putting random messages.

Recognition rate: H1.5 Putting messages on subject lines significantly increased their recognition rates versus the messages not on subject lines or putting random messages.

H2.5 Putting messages in top news significantly increased their recognition rates versus the messages not in top news.

Read-in-detail rate: H1.6 Putting employee-preferred messages on subject lines significantly increased their read-in-detail rates versus the messages not on subject lines.

H2.6 Putting employee-preferred messages in top news significantly increased their read-in-detail rates versus the messages not in top news or putting random messages.

5.2 Strategic Goals

Interest Rate. Strategically, we could marginally make employees perceive Brief as more interesting by personalizing top news with their preferred messages. We plot the average of the employees' *interest_rate* on personalized Briefs of each experimental group on top news in Figure 5. It shows that the average *interest_rate* of the employee-preferred top news group (B4) was higher than the random control group of top news (B2) by 18%. The pairwise tests showed that **H2.1 was marginally supported versus the random control group**. In the end survey, a participant from B4 commented *"I like the headings or topics at the top"*. A participant from B4 recognized that the top messages were related to their answers in the preference survey and hoped that recipients could update their preferences in the system in the future: *"I liked having focused information. However, I think if you move forward with customized Briefs (which I support), people should get some what regular reminders with the ability to change what items they select to follow."* A participant from the random group B2 seems to be disappointed: *"I still think there's too much boosterism and fluff and I wish it was more work-related."* And a participant from group B3 (which prioritized organization-preferred messages) found the content boring: *"Most of it was skimmed. Most of the topics don't apply to me and/or my work. The content overall is generally uninteresting."* The subject lines and message order's effects on the interest rate are not significant.

Overall Recognition Rate. However, putting employee-preferred messages in top news seemed to be a bad choice for the overall recognition rate (see Figure 7, B4). The employees might close the Briefs early if they learned that most of the interesting messages would be at the top positions. As a participant from B4 said *"It was fun to see the things I was interested in at the top. It also let me pay more attention to the beginning of the briefs and then skim the rest. "*

To improve Brief's overall recognition rate, we could mix employee-preferred messages with organization-preferred messages in top news. With pairwise tests, we found that the overall recognition rates of group B5 (mixing employee/organization preferences) were significantly higher than group B2 (random top news) by 19% and marginally higher than group B1 (original top news) by 11% (**Q2.3**). Seeing interesting content both at the top and in other sections might keep employees reading, though they did feel some of the messages "irrelevant to them". A participant from B5 said *"I liked them. Overall I find things interesting; however, they are not really pertinent to my work always."* It is worth noting that the overall recognition rate of the organization-preferred message group was also significantly higher than the random control group. The employees seemed to keep searching for items of interest if they did not find them in top news. But this searching process might cause disappointment. A participant from B3 commented *"I was a little disappointed because I was expected slightly more tailored content."*
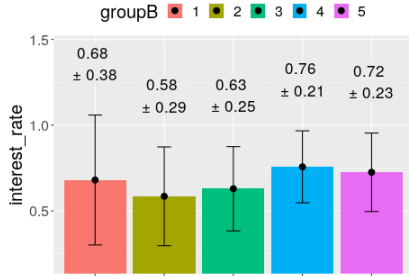
Fig. 5. The interest rates on personalized Briefs of each experimental group on top news and the standard deviations of the participants' interest rates in that group.
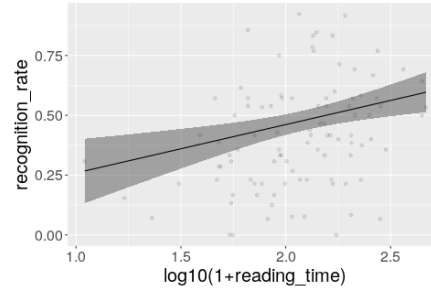

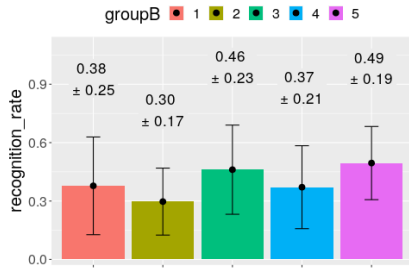Fig. 6. Brief's reading time versus its recognition rate.

Fig. 7. The overall recognition rates on personalized Briefs of each experimental group on top news and the standard deviations of the participants' overall recognition rates in that group.
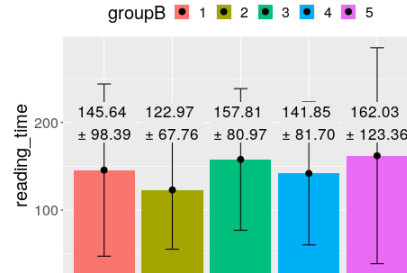

Fig. 8. The reading time (seconds) on personalized Briefs of each experimental group on top news and the standard deviations of the participants' reading time in that group.

Reading Time. There were no significant effects of subject lines/top news/message order designs on reading time. On average, the Briefs were read for around 120 to 170 seconds, and the variation of reading time is large. However, we found that the patterns of reading time matched with the patterns of overall recognition rates (see 7 and 8). We plot reading time versus overall recognition rate in Figure 6. The correlation between reading time (transformed by log10(1+10)) and the overall recognition rate was significant (Chisq=10.46, p.value = 0.001, coef = 0.20±0.06). This result shows that the gain in awareness is usually accompanied by time costs.

Open Rate. We did not observe significant differences among subject line groups' open rates. The average *open_rate* of group A4 (employee-preferred subject line group) was higher than other subject line groups but the pairwise tests were insignificant. The reason might be that our participants usually would take a quick check of these experimental Briefs during the experiment (though we asked them to treat these Briefs as naturally as possible). This is a limitation of this study — we only collected 6 weeks' datapoints, because we wanted to collect all the recognition data together by a survey with a reasonable amount of questions at the end. A longer study might find different results on the open rates. However, some participants did indicate that they decided whether to open a Brief or not based on its subject lines — a participant from A2 said that they left two Briefs unread because their subject lines were "not at all interesting".

5.3 Tactical Goals

Recognition Rate. Tactically, organizations could make those messages they view as important/relevant be recognized by more employees by putting them on subject lines or top news. The Anova tests showed that whether a message was on subject lines influences its recognition rate significantly. The pairwise tests showed that either putting organization-preferred messages or employee-preferred messages on subject lines would increase their recognition rates by over 15% compared to the recognition rates of the messages not on subject lines (see Figure 9). Similarly, either putting organization-preferred messages or employee-preferred messages in top news would increase their recognition rates by over 12% compared to the recognition rates of the messages not in top news (see Figure 11). It is worth noting that these recognition rates were not significantly higher than the recognition rates of the original top news groups (B1), which indicated an opportunity to learn from the human editor on their design and selection strategies.
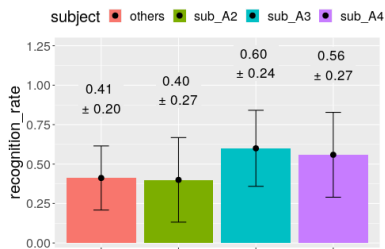
Fig. 9. The recognition rates of the messages on subject lines with respect to group A and non-subject line messages and the standard deviations of the participants' recognition rates in that group.
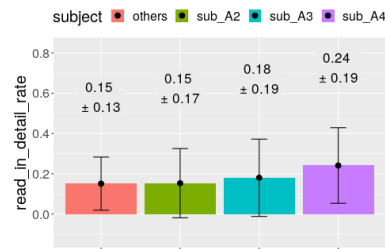
Fig. 10. The read-in-detail rates of the messages on subject lines with respect to group A and non-subject messages and the standard deviations of the participants' read-in-detail rates in that group.

Read-in-Detail Rate. However, organizations could not make employees read the organization-preferred messages in detail. The read-in-detail rates of those organization-preferred messages on subject lines were not significantly improved (see Figure 10). Actually, only the read-in-detail rates of those employee-preferred messages were significantly increased by 9% when being putting on subject lines or top news (see Figure 10, 12). The reasons might be that the employees tended to only click the messages they had some interest in. We might need stronger incentives if we would like the employees to read those important-to-organization messages thoroughly.
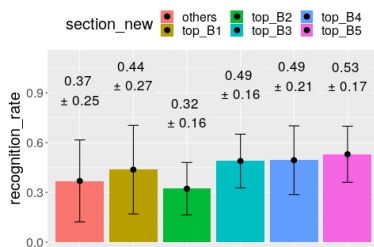
Fig. 11. The recognition rates of the messages in top news with respect to group B and non-top messages and the standard deviations of the participants' recognition rates in that group.
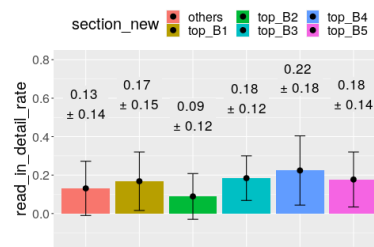
Fig. 12. The read-in-detail rates of the messages in top news with respect to group B and non-top messages and the standard deviations of the participants' read-in-detail rates in that group.

Mixing organization-preferred messages with employee-preferred messages in top news/other sections did not bring further improvements to their recognition rates. For the messages in top news, though the mixed group B5's organization-preferred messages' average recognition rate was 4% higher than B3's messages in the corresponding positions (45% versus 49%), the difference was not significant. And the difference (3%) between the recognition rates of group C3 and C5's top 2 organization-preferred messages in the u-wide news sections was also not significant.

5.4 Organization and Employee's Bulk Message Preferences

In this section, we discussed where are the preference conflicts on bulk messages between the organization and the employees. Table 5 shows a set of messages' topics. For each topic, it shows how the Brief editor labeled the messages in that topic (whether it was important to the organization) in the editor surveys, and the employees' assessments of the work-relevance and interest of a sample message representing that topic (we collected these data in the preference survey at the experiment setup). For example, for the topic fundraising & development, there were 9 messages during the study period. 5 of them were marked by the editor as important. 45.1% of the employees felt the corresponding message was relevant to their jobs (18.0% of the employees felt interesting and 38.5% of them felt work-relevant).

We arranged a meeting with the Brief editor to discuss Table 5. In that meeting, the editor told us that the frequency of topics (#messages) is basically a true reflection of the number of topics submitted to them. The editor rejected a small number of the submissions that were too narrowly focused: *"I really only reject maybe 10% of submissions. We have communicators (in each campus). You know, it's really up to those folks to determine what they feel is important."*.

We noticed a number of interesting things from Table 5. First, a large number of messages fell into the topic categories that the editor did not feel were usually important, and the employees generally would find unimportant and/or uninteresting, including award/recognition, student/alumni stories, faculty/staff stories. These were all work-relevant to fewer than 10% of the employees and interesting to fewer than 20% of them.

Second, there are topics that the editor viewed as very important while the employees felt not that interesting or relevant, including university history & celebrations, policy/admin news/governance, sports & spirit. Over 60% of the corresponding messages were viewed as important by the editor, while fewer than 40% of the employees viewed these categories as work-relevant, and fewer than 30% felt interesting.

Third, there are topics that the editor viewed as unimportant while the employees felt interesting. Some of them frequently appeared, including climate/eco, program awards/applications, health/covid. The editor told us that some contents were put because the employees might find them interesting: *"I probably select half of those (the messages about events) myself just based on what I think folks will find interesting. I'm thinking both in terms of readership like we'd like them to find something of interest, so they come back and read."* However, some of them appear only 3 or 4 times, including art & museums and engineering science research stories.

We further evaluated the preferences on the original Brief's messages. In week 2's Brief (the original Brief), 58% of the surveyed messages were tagged as neither interesting nor work-relevant by the employees. The editor identified one message with the title "U of M Public Engagement Footprint" very relevant in building community and common understanding ($org\_importance = 4$). This message is from the Provost's office to advocate employees to submit plans for the university's service, outreach, and community engagement. However, 58% of the employees found this message neither interesting nor work-relevant. The message "University and Faculty Senate Meetings" was tagged as work-relevant to all employees, while 39% of the employees found this message neither interesting nor work-relevant.

As we look at the results in total, it is clear that employee interest and editor judgment of importance are not perfectly-aligned. This finding reiterates the importance of considering the composition of the newsletter as a whole — how to have enough relevant and interesting content to encourage reading the important content as well.

Table 5. Summary of the organization and employees' preferences on bulk messages' topics. The topics are ordered by # messages (the number of messages that included a corresponding topic in the experiment). The editor selected ≤ 4 topics for each message.
times_imp: the number of times that the message got an importance score ≥ 3.
org_imp%: # times_imp / # messages.
emp_rel/int/pref%: the percentage of employees who tagged the message with the corresponding topic as work-relevant/interesting/either work-relevant or interesting in the preference survey.

topic	#messages	#times_imp	org_imp%	emp_pref%	emp_rel%	emp_int%
Talk/ Symposium/ Lectures Announcements	29	1	3.4%	39.3%	4.9%	38.5%
Student/ Alumni Stories	27	10	37.0%	18.0%	4.9%	14.8%
Community Service/ Social Justice/ Underserved Population	21	11	52.4%	78.7%	35.2%	73.0%
Faculty Staff Stories	20	4	20.0%	23.8%	9.0%	17.2%
Health/ Biology Research Stories	15	8	53.3%	64.8%	9.0%	60.7%
Climate/ Eco/ Agriculture	15	6	40.0%	71.3%	18.0%	69.7%
Health Wellness Resources/ COVID	12	2	16.7%	91.0%	67.2%	73.8%
Award/ Recognition to University, Faculty, Staff, Students	11	5	45.5%	23.0%	6.6%	19.7%
Program Award Applications/Announcements	10	2	20.0%	85.2%	60.7%	54.9%
Fundraising Development	9	5	55.6%	45.1%	18.0%	38.5%
History/ Social Science Research Stories	9	2	22.2%	45.9%	15.6%	38.5%
Policies/ Admin News/ Governance	8	5	62.5%	46.7%	39.3%	14.8%
Tech Tool Updates/ Workshops	8	0	0.0%	35.2%	26.2%	13.1%
Sports Spirit	6	5	83.3%	27.9%	5.7%	23.8%
University History/ Celebrations	6	4	66.7%	43.4%	29.5%	22.1%
Art Museums	4	0	0.0%	65.6%	6.6%	63.9%
University Program Success Stories	4	2	50.0%	39.3%	17.2%	27.9%
Operations Awareness/ Facility Closures	3	1	33.3%	89.3%	82.0%	49.2%
Engineering Science Research Stories	3	0	0.0%	54.1%	3.3%	52.5%
Youth, Children	0	0	0.0%	36.1%	8.2%	31.1%

For the engagement with topics versus campuses, only 25%/32% of the employees looked for work-relevant/interesting messages from other campuses. When being put in top news, the messages selected from other campuses got a recognition rate (21%, p-value=0.0001) significantly lower than the messages selected from the employees' own campuses (43%), and in this case, the employees' preference score *emp_pref* does not significantly influence these messages' recognition rate (this calculation has excluded the employees who indicated that they would not look at other campuses'

messages). For the messages selected from the employees' own campuses, their recognition rate (43%) is not significantly different from the messages originally selected from Top News (50%), and in this case, the employees' preference score is positively correlated with the messages' recognition rate (p-value=0.024, cohen-size=2.259).

6 DISCUSSION

We explored 3 kinds of personalization (subject lines, top news, message order) based on 2 stakeholders' preferences (organization, employee), and investigated 2 types of goals (strategic / tactical goals). Overall, the tactical goals are easier to achieve than the strategic goals — the organization could put whatever they want to promote in the top position and would get a reasonable recognition rate. But for the strategic goals, the organization needs to also consider the employees' information needs, and some strategic goals (reading time, open rate) can't achieve with blanket newsletters.

Our work is different from many work in personalizing working emails [39, 41, 43] in that we try to address the challenge of bulk emails in this multi-stakeholder case — organizations have messages that they want their employees to be aware of while employees make individual judgements on which messages are relevant. Instead of prioritizing the recipients' preferences, we found that the best strategy for organizations is to mix messages they prefer with the messages their employees prefer. To the best of our knowledge, this is the first work focusing on this multi-objective personalization problem in the multi-stakeholder organizational bulk communication environment.

6.1 Organizations need to decide which messages to be sent and better communicate why.

Organizations and employees perceive different messages as important/relevant — this difference might have 2 outcomes. First, organizations might need to know more about their employees. For example, announcing the new Dean for the College of Biological Sciences through Brief is a convenient communication approach for the university leaders, because they do not need to spend time figuring out who would be interested in it and how to personalize the contents. However, many of its recipients would perceive this message as irrelevant, which might make them stop reading Brief in the future. In this case, organizations should collect more information to enable better targeting, such as collecting preferences based on message topics (this study), or allowing employees to select interesting message tags [41].

Second, if organizations decide that some messages are worthwhile for their employees to know about, they need to better communicate to their employees why they need to read those messages. For example, for the messages like "Board of Regents Meeting Highlights", the employees often skip reading them (62.5% of the administrative news was viewed as important by the organization, while only 14.8% of the investigated employees found this topic interesting). However, the employees might decide to read it if they are aware of, for example, that the board was discussing their salary plans. Potential approaches on this aspect include pricing emails [37, 45] (however Kraut et al. found that recipients still did not interpret the prices as emails' importance), indicating expected actions [2, 44], etc.

6.2 Organizations could use employee-preferred messages to attract their attention.

Though we found the only way to increase the recognition rates of those organization-preferred messages is to prioritize them, this does not mean that we can prioritize only these messages. Only prioritizing the organization-preferred messages would lose the gain in interest rate compared to putting employee-preferred messages on top, which might help protect this channel's long-term credibility — 13% more of the employees rated the Briefs with employee-preferred top news as interesting compared to the organization-preferred message group. Also, this choice might affect

this Brief's open rate — though insignificant, the open rate of the Briefs with organization-preferred subject lines is 16% lower than the open rate of the employee-preferred subject line group.

Only putting employee-preferred messages in top news might be a bad choice neither, though putting the contents interesting to its recipients is a common practice in commercial bulk emails. In our study, these Briefs' overall recognition rates were relatively lower (37%) compared to putting organization-preferred messages in top news (46%). The employees might be satisfied that they saw interesting messages early. However, this is undesirable for the organization because these employees might skip the rest of the newsletter, including the messages the organization wants them to know. In fact, when we mixed them in top news, the overall recognition rate would be the highest — 11% significantly higher than the original top news and 19% higher than the random top news. This is, to some extent, similar to Zhao et al.'s finding that to keep users browsing a recommender website, we can not put all the interesting items at the top [68]. Instead, we need to distribute it across the page and mix them with other items we would like to expose to users.

6.3 There are always tradeoffs — suggestions.

Within the current framework, we did not find any single optimal solution. Even with the mixed strategy, its interest rate was not as high as when we only put the messages the employees preferred on top news. The most interesting/efficient newsletter for employees would not be the newsletter that could best help organizations convey their messages. However, there are some decisions organizations can make when they know the priority of their communication goals:

★ Subject lines: for subject lines, organizations could put the messages they perceive as most important/relevant. This approach would bring these messages higher recognition rates. At the same time, at least in our (occasional) reader group, subject lines did not significantly affect the employees' open rate or interest rate.

★ Top news: to improve the overall recognition rate, a good approach for organizations would be mixing employee-preferred messages and organization-preferred messages in top news.

Besides, there is also a trade-off between bulk communication's cost and performance. Though we did not find any significant effect on reading time, reading time is significantly correlated with the overall recognition rate, which means that organization needs to pay for more employee's time if they want higher performance. Also, longer email reading time is correlated with lower working productivity [38]. In that sense, organizations should also try to remove unnecessary messages and use personalization [67] to put important/relevant messages upfront.

6.4 Limitations and Generalizability

The limitations of this study included: 1) Reordering only: because of the requirement of our collaborator, we did not exclude any message from the studied newsletter.

2) Measurement of recognition: we trusted our participants that they would select "Skimmed" or "Read fully" if they have seen a message and select "No" or "Not Sure" if they did not recognize this message or were uncertain.

3) Selection of participants: our participants were relatively active readers of Briefs. The employees who had stopped reading Briefs might have lower recognition rates, open rates, etc.

4) Technical issues: some plugins were deleted by a Chrome update during the experiment, the participants did not read Briefs with that browser, etc. Our personalization model is based on coarse-grained topics. [5]

---

[5]Among the 1404 messages we sent through the original Briefs in week 2 (our predicted employee preference based on topics versus these messages' employee preference calculated from the interest scores and work-relevance scores collected directly from the participants in the end survey, see 4.3), we achieve a precision of 66% and a recall of 75%.

In short, our study could only be generalizable when: 1) the organization newsletter is sent to a large list of employees;

2) studying newsletters' occasional/regular readers. We keep the newsletter's structure because the editor suggested that their audiences liked its campus structure; however, to how much extent the campus structure influenced the personalization's performance is still left to be studied.

6.5 Future Work

After conducting this study, we see the following future work in improving organizational bulk communication.

1) Measuring each message's reading time to better understand employees' preferences [33]. It would be useful to run a study with eyetracking devices to collect such reading data, and to develop estimation algorithms [32] based on recipients' interactions with bulk emails' webpages.

2) Exploring different designing strategies that could help employees understand why they need to read some messages: for example, encourage senders to tag the reasons for sending some messages [20].

3) Studying the effectiveness of fine-grained personalization models [3, 55], enabling employees to update their preferences, exploring tools could better target the recipients (allowing excluding some messages) and learn their preferences gradually [50, 66], etc.

4) Studying how to bring back nonreaders: restore nonreaders' trust on the bulk communication channels.

7 CONCLUSION

This work studied how to use personalization to help the studied organization lead their employees' attention to the bulk messages they perceive as important or relevant for the employees to know (tactical goals) while maintaining the employees' overall positive experiences with these emails (strategic goals). We conducted an 8-week 4x5x5 controlled field experiment with 141 employees of a university and a weekly university-wide newsletter.

We found that tactically, putting organization-preferred messages on subject lines or top news significantly increased their recognition rates, but did not increase their read-in-detail rates significantly. Only the employee-preferred messages' read-in-detail rates were improved. Strategically, mixing the employee-preferred/organization-preferred messages in top news significantly increased the overall recognition rate. Putting employee-preferred messages in top news increased their interest rates marginally. We further looked into where the preferences on bulk messages' topics conflicted between the employees and the organization. We discussed the limitations and generalizability of this study.

Besides the findings above, this work also provided a basic backend framework for coordinating multiple stakeholders' preferences on organizational bulk emails — employees could input their preferences through an onboarding survey; communicators and the organization leaders could input their preferences through weekly surveys; the system handles the transformation between text and html, listens to the survey responses, and generates personalized newsletters.

To the best of our knowledge, this is the first work focusing on this multi-objective personalization problem in the multi-stakeholder organizational bulk communication environment. We hope our study provides facts and possible directions for designing tools supporting organizational bulk communications.

ACKNOWLEDGMENTS

REFERENCES

[1] Alan S Abrahams, Tarun Chaudhary, and Jason K Deane. 2010. A multi-industry, longitudinal analysis of the email marketing habits of the largest United States franchise chains. *Journal of Direct, Data and Digital Marketing Practice* 11, 3 (2010), 187–197.

[2] Tarfah Alrashed, Chia-Jung Lee, Peter Bailey, Christopher Lin, Milad Shokouhi, and Susan Dumais. 2019. Evaluating User Actions as a Proxy for Email Significance. In *The World Wide Web Conference*. 26–36.

[3] Guy Aridor, Duarte Gonçalves, Daniel Kluver, Ruoyan Kong, and Joseph Konstan. 2022. The economics of recommender systems: Evidence from a field experiment on movielens. *arXiv preprint arXiv:2211.14219* (2022).

[4] DJ Barr, R Levy, C Scheepers, and HJ Tily. 2013. Random effects. In *Annual Conference of the Cognitive Science Society*. 197–202.

[5] Dale J Barr. 2021. Learning statistical models through simulation in R: An interactive textbook. *Retrieved from https://psyteachr.github.io/stat-models-v1* Version 1.0.0 (2021). https://psyteachr.github.io/ug3-stats/linear-mixed-effects-models-with-one-random-factor.html

[6] Pablo Bermejo, Jose A Gámez, and Jose M Puerta. 2011. Improving the performance of Naive Bayes multinomial in e-mail foldering by introducing distribution-based balance of datasets. *Expert Systems with Applications* 38, 3 (2011), 2072–2080.

[7] Antun Biloš, Davorin Turkalj, and Ivan Kelić. 2016. Open-rate controlled experiment in e-mail marketing campaigns. *Market-Tržište* 28, 1 (2016), 93–109.

[8] Virginia Braun and Victoria Clarke. 2012. Thematic analysis. (2012).

[9] Mario Callegaro, Yelena Kruse, Melanie Thomas, and Poom Nukulkij. 2009. The effect of email invitation customization on survey completion rates in an internet panel: A meta-analysis of 10 public affairs surveys. In *Proceeding of the AAPOR-JSM Conferences*, Vol. 5. 764–78.

[10] Carla Carvalho, Alipio M Jorge, and Carlos Soares. 2006. Personalization of e-newsletters based on web log analysis and clustering. In *2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2006 Main Conference Proceedings)(WI'06)*. IEEE, 724–727.

[11] Marta E Cecchinato, Jon Bird, and Anna L Cox. 2014. Personalised email tools: a solution to email overload?. In *CHI'14 Workshop: Personalised Behaviour Change Technologies*. ACM Conference on Human Factors in Computing Systems (CHI).

[12] Mail Chimp. 2018. Average Email Campaign Stats of Mailchimp Customers by Industry.

[13] Laura A. Dabbish and Robert E. Kraut. 2006. Email Overload at Work: An Analysis of Factors Associated with Email Strain. In *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work (CSCW '06)*. 431–440.

[14] Laura A Dabbish, Robert E Kraut, Susan Fussell, and Sara Kiesler. 2005. Understanding email use: predicting action on a message. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 691–700.

[15] Onur Ferhat and Fernando Vilariño. 2016. Low cost eye tracking: The current panorama. *Computational intelligence and neuroscience* 2016 (2016).

[16] Peter Green and Catriona J MacLeod. 2016. SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution* 7, 4 (2016), 493–498.

[17] Catherine Grevet, David Choi, Debra Kumar, and Eric Gilbert. 2014. Overload is overloaded: email in the age of Gmail. In *Proceedings of the sigchi conference on human factors in computing systems*. 793–802.

[18] Florian Hartig. 2019. DHARMa: residual diagnostics for hierarchical (multi-level/mixed) regression models. *R package version 0.2* 4 (2019).

[19] Robert P Hawkins, Matthew Kreuter, Kenneth Resnicow, Martin Fishbein, and Arie Dijkstra. 2008. Understanding tailoring in communicating about health. *Health education research* 23, 3 (2008), 454–466.

[20] Yunzhong He, Cong Zhang, Ruoyan Kong, Chaitanya Kulkarni, Qing Liu, Ashish Gandhe, Amit Nithianandan, and Arul Prakash. 2023. HierCat: Hierarchical Query Categorization from Weakly Supervised Data at Facebook Marketplace. In *Companion Proceedings of the ACM Web Conference 2023*. 331–335.

[21] Silvia Heinz and Elisa D Mekler. 2012. The influence of banner placement and navigation style on the recognition of advertisement banners. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*. 803–804.

[22] Sture Holm. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics* (1979), 65–70.

[23] Thomas W. Jackson, Ray Dawson, and Darren Wilson. 2003. Understanding Email Interaction Increases Organizational Productivity. *Commun. ACM* 46, 8 (aug 2003), 80–84. https://doi.org/10.1145/859670.859673

[24] Kokil Jaidka, Tanya Goyal, and Niyati Chhaya. 2018. Predicting email and article clickthroughs with domain-adaptive language models. In *Proceedings of the 10th ACM Conference on Web Science*. 177–184.

[25] Simms Jenkins. 2008. *The truth about email marketing*. FT Press.

[26] Stephen RG Jones. 1992. Was there a Hawthorne effect? *American Journal of sociology* 98, 3 (1992), 451–468.

[27] A Kassambara. 2018. Logistic Regression Assumptions and Diagnostics in R. *Statistical tools for highthroughput data analysis* (2018).

[28] Daniel Katz and Robert L Kahn. 2008. Communication: the flow of information. *Communication theory* (2008), 382–389.

[29] Sabrina Heike Kessler and Ines Engelmann. 2019. Why do we click? Investigating reasons for user selection on a news aggregator website. *Communications* 44, 2 (2019), 225–247.

[30] Yoojung Kim, Mihyun Kang, Sejung Marina Choi, and Yongjun Sung. 2016. To click or not to click? Investigating antecedents of advertisement clicking on Facebook. *Social Behavior and Personality: an international journal* 44, 4 (2016), 657–667.

[31] Bryan Klimt and Yiming Yang. 2004. The enron corpus: A new dataset for email classification research. In *European Conference on Machine Learning*. Springer, 217–226.

[32] Ruoyan Kong, Zhanlong Qiu, Yang Liu, and Qi Zhao. 2021. NimbleLearn: A scalable and fast batch-mode active learning approach. In *2021 International Conference on Data Mining Workshops (ICDMW)*. IEEE, 350–359.

[33] Ruoyan Kong, Ruixuan Sun, Charles Chuankai Zhang, Chen Chen, Sneha Patri, Gayathri Gajjela, and Joseph A. Konstan. 2023. Getting the Most from Eye-Tracking: User-Interaction Based Reading Region Estimation Dataset and Models. In *Proceedings of the 2023 Symposium on Eye Tracking Research and Applications* (Tubingen, Germany) *(ETRA '23)*. Association for Computing Machinery, New York, NY, USA, Article 10, 7 pages. https://doi.org/10.1145/3588015.3588404

[34] Ruoyan Kong, Ruobing Wang, and Zitao Shen. 2021. Virtual Reality System for Invasive Therapy. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 689–690.

[35] Ruoyan Kong, Haiyi Zhu, and Joseph Konstan. 2020. Organizational Bulk Email Systems: Their Role and Performance in Remote Work. (2020).

[36] Ruoyan Kong, Haiyi Zhu, and Joseph A Konstan. 2021. Learning to Ignore: A Case Study of Organization-Wide Bulk Email Effectiveness. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–23.

[37] Robert E Kraut, Shyam Sunder, Rahul Telang, and James Morris. 2005. Pricing electronic mail to solve the problem of spam. *Human–Computer Interaction* 20, 1-2 (2005), 195–223.

[38] Gloria Mark, Shamsi T. Iqbal, Mary Czerwinski, Paul Johns, Akane Sano, and Yuliya Lutchyn. 2016. Email Duration, Batching and Self-Interruption: Patterns of Email Use on Productivity and Stress. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*.

[39] Paul B Moody. 2002. Reinventing email. *IBM Research, Cambridge. CSCW* (2002).

[40] Michele Tolela Myers and Gail E Myers. 1982. *Managing by communication: An organizational approach*. McGraw-Hill College.

[41] Les Nelson, Rowan Nairn, and EH Chi. 2010. Mail2tag: Efficient targeting of news in an organization. In *CSCW Workshop "Collective Intelligence in Organizations*. Citeseer.

[42] William F Paczkowski and Jason Kuruzovich. 2016. Checking email in the bathroom: monitoring email responsiveness behavior in the workplace. *American Journal of Management* 16, 2 (2016), 23.

[43] Soya Park, Amy X. Zhang, Luke S. Murray, and David R. Karger. 2019. *Opportunities for Automating Email Processing: A Need-Finding Study*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300604

[44] Kyle Rector and Joshua Hailpern. 2014. MinEMail: SMS Alert System for Managing Critical Emails *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 783–792. https://doi.org/10.1145/2556288.2557182

[45] Byron Reeves, Simon Roy, Brian Gorman, and Teresa Morley. 2008. A marketplace for attention: Responses to a synthetic currency used to signal information importance in e-mail. *First Monday* 13, 5 (2008).

[46] Ruth Rettie. 2002. Email marketing: success factors. (2002).

[47] Richard J Roiger. 2020. *Just Enough R!: An Interactive Approach to Machine Learning and Analytics*. CRC Press. https://benwhalley.github.io/just-enough-r/checking-assumptions.html

[48] Navdeep S Sahni, S Christian Wheeler, and Pradeep Chintagunta. 2018. Personalization in email marketing: The role of noninformative advertising content. *Marketing Science* 37, 2 (2018), 236–258.

[49] Natalie Sappleton and Fernando Lourenço. 2016. Email subject lines and response rates to invitations to participate in a web survey and a face-to-face interview: the sound of silence. *International Journal of Social Research Methodology* 19, 5 (2016), 611–622.

[50] Zitao Shen, Dalton Schutte, Yoonkwon Yi, Anusha Bompelli, Fang Yu, Yanshan Wang, and Rui Zhang. 2022. Classifying the lifestyle status for Alzheimer's disease from clinical notes using deep learning with weak supervision. *BMC medical informatics and decision making* 22, 1 (2022), 1–11.

[51] Sav Shrestha and Kelsi Lenz. 2007. Eye gaze patterns while searching vs. browsing a Website. *Usability News* 9, 1 (2007), 1–9.

[52] Betsy Sinclair, Margaret McConnell, and Donald P Green. 2012. Detecting spillover effects: Design and analysis of multilevel experiments. *American Journal of Political Science* 56, 4 (2012), 1055–1069.

[53] Lavneet Singh and Girija Chetty. 2015. Email Personalization and User Profiling Using RANSAC Multi Model Response Regression Based Optimized Pruning Extreme Learning Machines and Gradient Boosting Trees. In *International Conference on Neural Information Processing*. Springer, 302–309.

[54] Henrik Singmann, Ben Bolker, Jake Westfall, and Frederik Aust. 2015. Package 'afex'.

[55] Ruixuan Sun, Ruoyan Kong, Qiao Jin, and Joseph Konstan. 2023. Less Can Be More: Exploring Population Rating Dispositions with Partitioned Models in Recommender Systems. In *Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization*. 291–295.

[56] Jesús H Trespalacios and Ross A Perkins. 2016. Effects of personalization and invitation email length on web-based survey response rates. *TechTrends* 60, 4 (2016), 330–335.

[57] DJ Waldow and Jason Falls. 2012. *The Rebel's Guide to Email Marketing: Grow Your List, Break the Rules, and Win*. Que Publishing.

[58] Sunil Wattal, Rahul Telang, Tridas Mukhopadhyay, and Peter Boatwright. 2012. What's in a "name"? Impact of use of customer information in e-mail advertisements. *Information Systems Research* 23, 3-part-1 (2012), 679–697.

[59] Mary Welch. 2012. Appropriateness and acceptability: Employee perspectives of internal communication. *Public Relations Review* 38, 2 (2012), 246–254.

[60] Mary Welch and Paul R Jackson. 2007. Rethinking internal communication: a stakeholder approach. *Corporate communications: An international journal* (2007).

[61] Steve Whittaker and Candace Sidner. 1996. Email overload: exploring personal information management of email. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 276–283.

[62] Bartosz W Wojdynski and Nathaniel J Evans. 2016. Going native: Effects of disclosure position and language on the recognition and evaluation of online native advertising. *Journal of Advertising* 45, 2 (2016), 157–168.

[63] Alchemy Worx. 2016. Subject Line Gold–originality is the key to sustainable success. *Alchemy Worx* (2016).

[64] Jason Wrench and Narissa Punyanunt-Carter. 2012. An Introduction to organizational communication. *Creative Commons: Mountain View, CA, USA* (2012).

[65] Liu Yang, Susan T. Dumais, Paul N. Bennett, and Ahmed Hassan Awadallah. 2017. Characterizing and Predicting Enterprise Email Reply Behavior. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17)*.

[66] Yoonkwon Yi, Zitao Shen, Anusha Bompelli, Fang Yu, Yanshan Wang, and Rui Zhang. 2020. Natural language processing methods to extract lifestyle exposures for Alzheimer's disease from clinical notes. In *2020 IEEE International Conference on Healthcare Informatics (ICHI)*. IEEE, 1–2.

[67] Hongke Zhao, Qi Liu, Yong Ge, Ruoyan Kong, and Enhong Chen. 2016. Group preference aggregation: A nash equilibrium approach. In *2016 IEEE 16th International Conference on Data Mining (ICDM)*. IEEE, 679–688.

[68] Qian Zhao, Gediminas Adomavicius, F Maxwell Harper, Martijn Willemsen, and Joseph A Konstan. 2017. Toward better interactions in recommender systems: cycling and serpenting approaches for top-n item lists. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*. 1444–1453.