Towards an Effective Organization-Wide Bulk Email System

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Ruoyan Kong

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Advised by Prof. Joseph A. Konstan

August, 2023

Acknowledgements

Dedication

To my parents Hongxia Zeng and Zhangming Kong. To my grandmother Hewen Hui and grandfather Tengfang Zeng. To my partner Zitao Shen.

Abstract

Bulk email (emails sent to a large list of recipients) is widely used in organizations to communicate messages to employees. It is an important tool in making employees aware of policies, events, leadership updates, etc. However, in large organizations, the problem of overwhelming communication is widespread. Ineffective organizational bulk emails waste employees' time and organizations' money, and cause a lack of awareness or compliance with organizations' missions and priorities.

Prior research mainly studied commercial bulk emails from a single stakeholder's perspective, such as helping senders improve open rates or helping recipients filter unsolicited bulk emails. However, within organizations, bulk email communication involves multiple stakeholders (employees, communicators, managers, leaders, the organization itself, etc.) with different priorities. The goal of organizational bulk email system is to both reach organization-wide communication effectiveness and provide positive experiences for all the stakeholders.

This thesis focuses on understanding and improving organizational bulk email systems by 1) conducting qualitative research to understand different stakeholders' perceptions of the system and its current effectiveness; 2) proposing economic models to describe stakeholders' actions/cost/value; 3) conducting field studies to evaluate personalization methods' effects on getting employees to read bulk messages; 4) designing tools to support communicators in evaluating, designing, and targeting bulk emails.

We performed these studies at the University of Minnesota, interviewing 25 employees (both senders and recipients), and including 317 participants in our studies in total. We found that the university's current organizational bulk email system is ineffective as only 22% of the information communicated was retained by employees. The failure of this system was systemic — it had many stakeholders, but none of them necessarily had a global view of the system or the impacts of their own actions. Then to encourage employees to read high-level information, we implemented a multi-stakeholder personalization framework that mixed important-to-organization messages with employee-preferred messages and improved the studied bulk email's recognition rate by 20%. On the sender side, we iteratively designed and deployed a prototype of an organizational bulk email

evaluation platform (CommTool). In field evaluation, we found several features (such as bulk emails' message-level performance) of CommTool helped communicators in designing bulk emails. At the same time, to enable these message-level metrics, we collected ground-truth eye-tracking data and developed a novel neural network technique to estimate how much time each message is being read using recipients' interactions with browsers only, which improved the estimation accuracy from 54% (heuristics) to 73%. In summary, this work sheds light on how to design organizational bulk email systems that communicate effectively and respect different stakeholders' value.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background and Problem Statement

Bulk email (or mass email) is email that is sent to a large group of recipients [**14**].[1]   Bulk email is an essential part of organizational communication. Organizations, such as universities or companies, frequently use internal bulk emails (e.g., newsletters, single bulk emails) in delivering information to their employees. Example messages include changes of leadership, summaries of organizational meetings, announcements of organizational events, updates of organization's policies, etc. We refer to bulk emails within organizations as **organizational bulk emails** and the systems that generate, design, distribute, process, and evaluate organizational bulk emails as **organizational bulk email systems**. Figure 1.1 is an example organizational bulk email of our studied site – the University of Minnesota. It is from the university senate office, which wants to inform employees of the policies and plans the university senates are discussing. This office's communicator organizes, designs, and distributes the information. In this thesis, I study such organizational bulk email systems. Specifically, I am motivated by the three insights below.

First, organizational bulk emails too often are ineffective. For example, the email above is long, unpersonalized, and untargeted — it has 19 pieces of information (**organizational bulk messages**) within the email; the content is the same for all recipients; it is sent to all faculty and staff of the university. In this university, I found

---

[1]   See the appendix C.1 for a full list of terms we define in this thesis.

that in a typical week, employees on average receives over 30 such organizational bulk emails with over 250 pieces of information in total per week. Not just in universities, but also in many large organizations, the problem of overwhelming and ineffective communication is widespread. The size of inboxes, the number of unread messages, and the employees' feeling of email overload keep increasing [60]. Employees are receiving emails irrelevant to them while missing important ones [46, 78].

Second, organizational bulk emails have high costs in time and money to their organizations. An estimate of the cost of the email above would be 24,000 employees * 2 min (reading and interruption time) * $0.75 salary+fringe+overhead/min = $36,000. However, this cost is not calculated, not visible to the sender, and not considered or managed anywhere in the current organizational bulk email system.

Figure 1.1: An example organizational bulk email of the studied site. The whole email includes 19 messages and 1809 words.

Third, organizational bulk email systems are interesting theoretically as a multi-stakeholder system [6], within which stakeholders have different priorities. The stakeholders in an organizational bulk email system include:

- Communicators (information gatekeeper): the staff in charge of designing and distributing organizational bulk emails.

- Communicators' clients (information producer): the original sources of organizational bulk emails, who request the communicators to distribute the information.

- Employees (information recipient): the organization's staff who receive organizational bulk emails from the communicators.

- Management: the direct managers of the employees.

There is also a key stakeholder — the organization itself — that has priorities not always recognized by the stakeholders above. Communicators and their clients naturally focus on their own needs — getting the word out, establishing evidence of notice or compliance, or preserving a record of communication. But the organization, on another side, pays for employees' time in handling those emails and wants to protect the reputation of their communication channels. Employees, faced with more communication than they can handle, start to ignore these communications. In turn, organizational goals around compliance, informed employees, and employee productivity may suffer.

For example, an announcement of a new child care center would reflect the complexity of this multi-stakeholder problem. The stakeholders might have different perceptions of and actions on this message. The communicators need to complete this task quickly and might achieve that by sending this message as a single email to all the employees. However, those employees who do not have children at that age would have a time cost of reading this message and might question the relevancy of that communication channel. Then the organization's communication channels' reputation might suffer. But on another side, the communicators' clients, such as organization leaders, might perceive this message's value of showing that the organization is family-friendly and are willing to incur that cost. In summary, this communication system needs to carefully balance different stakeholders' value and cost to maximize its productivity.

Previous studies on bulk emails are mainly about external bulk emails, which focus on either the senders or the recipients' perspectives on bulk emails' value. Work on the sender-side [148, 176, 180] focused on exploring personalization or targeting to improve open rates, click-through rates, conversion rates, etc. Work on the recipient-side focused on helping recipients filter bulk emails [36, 126, 52]. However, within organizations, employees may be responsible for reading and knowing about some bulk messages even if they do not like them — we should not simply filter out organizational bulk emails that employees view as irrelevant. On the other hand, optimizing open rate or other metrics for the organizational bulk email senders could bring time/money cost to the organization when some bulk emails are actually irrelevant to their recipients.

Given the motivations above, I designed and conducted a series of studies on improving organizational bulk email system with the consideration of multiple stakeholders' perspectives.

1.2 Summary of Chapters

In this thesis, I will first introduce the related work around organizational communication, bulk email personalization, bulk email evaluation (performance tracking) in Chapter 2. Then in Chapter 3, I introduce the first step of my dissertation work to understand this system and identify research opportunities. I investigated this system's effectiveness, failings, and stakeholders' experiences through surveys and interviews. I found that organizational bulk emails were ineffective and expensive because of two main reasons — the employees do not like to read some information that is viewed as important by the organization; the senders do not have enough information on employees' interests and cost. In Chapter 4, based on the findings of the empirical study, I further proposed an economic model to describe the value, cost, and actions of this system's stakeholders and how their diverse perspectives cause ineffectiveness. The proposed model enables me to clarify the relationships between this system's stakeholders and identify a bunch of potential interventions, enable to do the next two studies.

With these 2 observations, this thesis explores the opportunity of a personalization tool (chapter 5) and a communication tool (chapters 6 & 7) for communicators. I started with communicators because they are important gatekeepers between the original sender

and the vast majority of employees and the control point of routing information in this system.

Specifically, the personalization tool I built in chapter 5 employed the mechanism to encourage employees to read the important-to-organization messages. I also built a rule-based model to estimate employees' and communicators' preferences on bulk messages. The communication tool prototype in chapter 7 (chapter 6 is a pre-requisite for chapter 7 in enabling the tool to estimate employees' reading time) evaluated different features' usefulness in supporting communicators to estimate bulk emails/messages' performance, cost, and reputation. I designed that prototype to encourage communicators and original senders to make editing/targeting decisions based on the whole system's effectiveness.

In summary, my dissertation work would provide 1) a better understanding of organizational bulk email system's effectiveness and its stakeholders' perceptions through a survey study, a case study, and an economic model; 2) a personalization tool that encourages employees to read important bulk messages; 3) a communication prototype with a field test that aims to find features that can enable communicators to consider bulk emails' performance, cost, and impact on communication channels' reputations. There are other topics left to be explored on this system, e.g., allocating communication budget, studying who should decide whether a message should be sent out or not, etc. Many of these studies will involve not only HCI community but also the disciplines of management, communication, psychology, etc. I will discuss these opportunities in the discussion section.

In the following, I introduce the studies of my dissertation work, including their background, related work, study methods, and findings. I concluded this thesis by discussing this series of studies' implications and future opportunities.

Chapter 2

Related Work

In this chapter, I summarize related work around organizational communication, organizational email systems, and bulk emails that inspire our study.

2.1 Organizational Communication

Morgan in the book *Images of organization* defined an organization (e.g., a company, a university) as a system with "a loose network of people with divergent interests who gather together to pursue common goals" [129]. From the perspective of information theory, Path proposed that an organization could be viewed as a communication system that processes and distributes information to coordinate internal tasks and adapt to external changes [137]. With this definition, **communication is a key factor in an organization's coordination and production process**. In fact, organizational communication has been studied for more than a century by various disciplines including sociology, psychology, anthropology [37], etc. Communication has been called *"the life blood of an organization"* [58], and *"the glue that binds it all together"* [87].

There are various theories on the process of organizational communication, including how to model the process of generating, distributing, and interpreting messages. Many studies aim to motivate organizational communication's various stakeholders to reach organizational goals. For example, Lewis proposed an organizational communication model that connects the selection of communication strategies with stakeholders'

concerns, assessments, and interactions with the system [109]. Rajhans interviewed employees of Vanaz Engineers Ltd., and found that employees' commitment to and trust in the organization increased if they felt that the organization's communication is effective [138]. Baker identified diverse stakeholders of organizational communication, including the people who "generate" information (information producers), people who "distribute" information (information gatekeepers), and people who "interpret" information (information recipients) [10]. Welch and Jackson also summarized the stakeholder groups of organizational communication and modeled how the achievement of organizational communication's goals is determined by all those stakeholders' actions [181].

Given the various stakeholder groups, Jones et al. proposed that handling the relationships between various stakeholders within organization is a major task for the discipline of organizational communication in the 21st century [84]. Several studies have shown that organizational communication affects employees' job performance, job satisfaction, motivation, and the feeling of job security. For example, Giri and Kumar surveyed 380 employees working at various organizations in India and found that job satisfaction and performance are very much dependent on the communication behavior of the organization [55]. Jiang and Probst surveyed 639 employees in six different companies and found that employees who perceived higher levels of positive organizational communication practices reported fewer negative consequences of job insecurity compared with employees who reported lower levels of organizational communication [83]. Ruck and Welch proposed that the goals of organizational communication should include motivating employee commitment, promoting a positive sense of belonging, and developing the awareness of organization needs [147]. For information producers, Randall [151] did a case study in a large manufacturing firm, finding that if information producers failed to provide clear information, it was unlikely that the information recipients would continue to seek out information. For management, Stohl and Redding [164] studied a steel construction corporation, finding that the employees' cognitive failures were significantly related to their managers' communication styles.

Motivated by these studies, for this thesis, we sought to design organizational bulk email systems that respect different stakeholders' preferences.

2.2 Organizational Emails

Organizational internal emails (we refer to them as organization emails in this thesis) are emails whose senders and recipients are within the same organization. Organizational emails create what Sproull and Kiesler called a "networked organization" [162], in which people can be available even when they are physically absent.

There are many studies on organizational emails focusing on employees (recipients) and employee-perceived effectiveness, including:

- How organizational email burdened employees: Dabbish and Kraut proposed the concept **email overload** to describe employees' perceptions that their own use of email has gotten out of control because they receive more emails than they can handle effectively, in a study on white-collar workers [43]. Whittaker and Sidner interviewed 20 office workers in 1996 and found that the problem of email overload is overwhelming [182]. Fisher et al. revisited this study in 2006 and found that employees' email archives have grown tenfold compared to 1996 [50]. Merten and Gloor [124] found that employee job satisfaction went down as internal email volume increased in a case study in a 50 people company. Huang and Lin [73] surveyed three universities and found that knowledge workers were "ruled by email."

- Datasets on recipient's data: employees' log data [187], inbox data, and behavior data with organizational emails were collected, including Avocado dataset [186], Enron Corpora [13], and Outlook Dataset [5], etc.

- How the information producers and communicators should change their email actions according to recipients' needs to reduce email overload: Jackson et al. [78] studied 16 employees at the Danwood Group and proposed that email frequency should be controlled. Lu et al. [114] described EPIC, an email prioritization tool that combined global priority with individual priorities. Reeves et al. [144] experimented with attaching virtual currency to emails to signal importance in a large company.

These studies show that information overload is a widespread challenge for organizations in email communications. In this thesis, we aim to reduce the information

overload caused by a specific type of email — organizational bulk emails.

2.3   Bulk Email

The development of email is accompanied by the usage of bulk emails (or mass email) [14]. There are many studies about **external bulk emails** — the bulk emails sent to the recipients outside organizations. External bulk emails usually have goals like advertising products to customers, promoting services to clients, or creating the awareness of the brand to the general public [62]. According to the Data & Marketing Association's annual email marketing report, the average return on investment (ROI) of email marketing is 3,800% in 2020 ($38 for every $1 invested) [153].

2.3.1   Filtering External Bulk Email

Many external bulk emails are unsolicited bulk emails — their recipients do not opt-in to receive them, which we often called as spam [38]. According to Rao and Reiley's estimation, American companies and individuals lose $20 billion annually due to spam for their time filtering these messages [139]. Many studies focused on helping recipients filter unsolicited bulk emails by 1) building white lists and black lists [4, 36] based on email addresses, IP addresses, domain names, etc.; 2) content-based filtering based on extracted features and words of emails and classification models [126, 31, 96]; 3) agents that check whether the email is from a human, a robot, or malicious senders by peer-to-peer networks [125], CAPTCHA [67], etc. It is worth noting that many organizational bulk emails are also unsolicited — employees would often automatically be subscribed to some internal mailing lists and receive bulk emails from them.

2.3.2   Personalizing / Targeting External Bulk Email

The studies on the sender-side focused on developing personalization and targeting methods [153] to get higher open rates, click-through rates (CTR), conversion rates, return on investment (ROI), etc. The personalization content of bulk emails could be demographic information like names [148, 180], majors, departments [176]; or preference information like browsing history [180], deals or tools recommendation [159, 65]. Though Sahni et al.'s experiments found adding recipients' names to subject lines useful [148],

many other studies supported that the personalization based on preferences performed better than the personalization based on demographics. Wattal et al. [180] 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 [176]. 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 [65]. Wattal et al. [180] personalized email content based on customers' purchasing preferences and received positive responses. Many personalization designs have been tested on external bulk emails. We summarized them into the following five categories:

1. Subject line: An informative subject line was believed to be a key factor in successful email marketing [178]. Sahni et al. added recipients' names to subject lines [148], 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 [149, 25]. The "Long vs. Short Email Subject Line Test" of WhichTestWon.com in 2011 [178] 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 [184]. 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 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 [184, 80].

2. Top section: The traditional theory is that users would pay more attention to the top positions during browsing [156]. Wattal et al. [180], and Trespalacios and Perkins [176] 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. [180] put products that the customer might like most in the email. By analyzing 30 email-marketing

campaigns, Rettie [145] found that response rate is negatively correlated with email length. Carvalho [27] 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 [156], there were also many studies supporting different arrangements. Wojdynski and Nathanie [183] examined 12 web page designs and found that the advertisements in the middle or bottom positions got better recognition. Heinz and Mekler [70] found that banner placement did not influence recognition and recall.

5. Visual designs: Several studies focused on how to highlight the important content. Rettie [145] found that response rate is positively correlated with the number of images. Wojdynski and Nathanie [183] found that *"users tend to gaze at a target object that is surrounded by objects with weaker "demand for attention" values."*

Informed by these studies, we also explore personalization in the context of organizational bulk emails.

2.3.3   Bulk Email Platforms

As a part of CRM (customer relationship management), many large business platforms are supporting bulk email senders in designing and targeting bulk emails, such as Salesforce, Mailchimp, Revue, Constant Contact, etc [16, 127].

In terms of designing, targeting, testing, and evaluating bulk emails, the features supported by the two most popular CRM platforms, Salesforce and MailChimp, include:

- Targeting: 1) target by user fields that already existed in the database or predicted demographics based on their behavior logs on the websites,

  https://www.salesforce.com/products/marketing-cloud/best-practices/

  email-audience-segmentation/,

  https:// help.salesforce.com/s/articleView?

  id=sf.networks_audiences_rbc.htm&type=5;

2) target by behavioral segmentation, which usually needs to connect to the users'
behavior logs on websites,

https://mailchimp.com/features/predicted-demographics;

3) target by users' active status, https://www.salesforce.com/news/stories/

salesforcemakes-email-marketing-smarter-with-new-einstein-ai-innovations/.

- Personalization: 1) personalize strings based on user fields; 2) trig emails by users'
  actions on the websites,

  https://help.salesforce.com/s/articleView?language=en_US

  &type=5&id=sf.mc_es_available_perso nalization_strings.htm,

  https://mailchimp.com /features/personalization/.

- Costs: 1) unsubscribe rates; 2) the trend of open rates/click rates/number of
  subscriptions

  https://help.salesforce.com/s/articleView?id=sf.pardot_emails_open_ tracking

  .htm&type=5.

- A/B Tests: test the emails on several subscribers and compare their open rates
  /click rates/unsubscribe rates

  https://help.salesforce.com/s/articleView?language=en_US &type=5&

  id=sf.mc_es_ab_testing.htm.

There are still challenges for these bulk email platforms in data integration, person-
alization, and evaluation [62]. Several CRM platform designs have been proposed to
resolve these challenges:

- Data Integration, Targeting, and Personalization: Sukarsa et al. integrated the
  bulk email platform with a hotel's customer reservation system [167]. They also re-
  ported their whole Software Development Life Cycle for this integration. Kokkalis
  et al. developed MyriadHub, a crowd-sourced mail client where the crowd work-
  ers extracted conversational patterns and got matching templates and suggested
  responses from MyraidHub [93]. Du et al. proposed EventAction, a marketing

tool that could identify similar records based on the customers' previous product usage and recommend marketing actions like sending promotional emails [48].

- A/B Testing: Xu et al. introduced XLNT, a large-scale A/B test platform on Linkedin [185]. XLNT focused more on testing the experiments on Linkedin.com instead of emails.

- Customer Feedback: Garbett et al. [53] implemented a promotion platform for the campaigns of mobile applications. They sent promotional emails to "the most influential" community members and reached 11 successes in 27 campaigns.

These studies on external bulk emails show that personalization and evaluation tools helped bulk email senders complete their tasks and improved bulk email recipients' experience. In this thesis, we will also develop these tools within the context of organizational bulk emails to improve this system's effectiveness.

Chapter 3

Mixed-methods Study on Understanding An Organizational Bulk Email System

3.1 Introduction

First and foremost, we want to **understand an example organizational bulk email system**. In this chapter, we introduce a mixed-methods study (interviews and surveys) we conducted to understand our study site's (University of Minnesota) organizational bulk email system's effectiveness and its various stakeholders. [1]

As we discussed in the introduction, a feature of organizational bulk email system is its multiple stakeholders. Previous research on organizational emails mainly studied general emails from recipients' perspectives, such as which emails employees would delay reading [150], the probability of an email being retained/deleted [42], and recipients' productivity [123]. However, bulk email communication within organizations is not only a problem involving email recipients but is also an example of a multi-stakeholder problem [6]. The stakeholders in organizations include communicators, their clients, recipients

---

[1]   Ruoyan Kong, Haiyi Zhu, and Joseph A Konstan.   Learning to ignore: A case study of organization-wide bulk email effectiveness. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW1):1–23, 2021. [98]; Ruoyan Kong, Haiyi Zhu, and Joseph Konstan. Organizational bulk email systems: Their role and performance in remote work. New Future of Work, 2020 [95].

(employees), and their direct managers. Not only are there different goals for communicators, their clients, and recipients, but also there is a key fourth stakeholder — the organization itself — which has its priorities not always recognized by communicators, their clients, and recipients. Communicators and their clients naturally focus on their own needs — getting the word out, establishing evidence of notice or compliance, or preserving a record of communication. But recipients, faced with more communication than they can handle, have to scan, filter, or simply ignore organizational bulk emails. In turn, organizational goals around compliance, informed employees, and employee productivity may suffer.

Thus to **understand the multiple stakeholders' perspectives** within organizational bulk email systems, we conducted a mixed-methods case study. Case studies are the predominant method of conducting organizational studies [169, 22] and organizational communication studies [150, 42, 123, 151, 164]. A case study enables us 1) learn domain experts' (communicators) information management techniques and how the bulk email system is used within the organization; and 2) understand the links between different stakeholders within the same organization [107]. We selected a representative organization (University of Minnesota) with hierarchical structures, and centralized and decentralized communication offices. We aim to 1) inform the design of the organizational bulk email systems [188]; 2) provide interview protocols for future studies on such systems; and 3) be generalizable to similar organizations.

We conducted a survey study and an interview study at the University of Minnesota, to learn the effectiveness of its organizational bulk email system, the communicators' practice of designing and distributing organizational bulk emails, and the communicators/recipients/managers' experience and assessments. We analyzed self-reported data, logged inbox data, and inbox-review data collected from surveys and interviews (artifact walkthroughs [141]). This study inspired us to develop personalization tools and evaluation tools to support communicators in designing / targeting in Chapters 5 to 7.

3.2   Gaps and Research Questions

From the related work on organizational communication and bulk email (see Chapter 2), we identified 2 major gaps:

- The studies on organizational emails focused on recipients: though studies on organizational communication pointed out that when measuring organizational communication's effectiveness, we should consider not only recipients, but also information providers (communicators and their clients) and the organization itself [163, 84]. Previous studies on organizational emails have not investigated whether recipients and information providers within an organization have different preferences about emails, and how these mismatches affect an organization.

- The studies on organizational bulk emails have focused on external bulk emails [62]: few studies of bulk emails have examined instances where the bulk email sender and recipients are part of the same organization, where the ultimate goal is maximizing the whole organization's interests rather than the information providers' or the recipients' interests [43, 124]. Within organizational context, whether a recipient (employee) should receive an organizational bulk email may not be determined by whether the recipient likes the email. Sometimes employees may be responsible for knowing about this email though they might not be interested in it. We see a need to understand bulk emails from an organizational context.

Based on these gaps we identified a need to undertake a systemic investigation of bulk email communications from a whole organization's view. Notice that we referred to organizational bulk email as **bulk email** and organizational bulk email system as **bulk email system** in the rest of this chapter. We posed the following research questions:

Q1: How effective is the organizational bulk email system at communicating information to employees?

Q2: What are communicators' current practices for designing and distributing bulk emails?

Q3: What are the experiences with and assessments of bulk emails of different stakeholders?

We conducted a survey study and an in-depth case study of a representative organization. We interviewed both communicators and recipients within an organization using

an artifact walkthrough approach. We delved into email cases and got their assessments of email values. This chapter will discuss communicators, recipients, communicators' clients, and recipients' managers.

3.3 Study Site

We carried out our studies at the University of Minnesota, a large university with approximately 24,000 employees and five campuses. We met with a group of 9 communicators within this university at the beginning of this project. In the meeting, we found agreement that the current bulk email system was not what the university wanted it to be, since the communicators felt that bulk emails were ignored too often. They encouraged us to move forward.

Figure 3.1 is the structure of the organizational bulk email system. Like many large universities, it has both centralized and decentralized governing structures. Each major unit has a communication office or communicators (communication director/staff). There are central units led by university leaders responsible for the core business functions. The University Relations office (under the Vice President for University Relations) is in charge of internal communications for the university. Each Vice President's office has a communication unit or several communication staff, who draft and send emails for university leaders and conduct university-wide bulk email communications through newsletters. Many central units run their own newsletters. Employees are subscribed to many of these newsletters by default.

The non-central units like colleges' Deans' offices also perform bulk communications locally. Their communicators conduct local communications and also work with central communication offices, such as collecting and submitting content to the central newsletter editors, or redistributing the messages from the central units.

The IT infrastructure provides technical support for the bulk email system, including an email system (Gmail), an email marketing platform (Salesforce), and a human resource system (Peoplesoft). The email marketing platform controls the access to sending bulk emails to campus-wide groups via mailing lists or query-based databases. Only trained communicators have access to the corresponding authorized mailing lists. It is

worth noting that querying the exact group of employees can often be difficult and time-consuming. For example, there is no clear definition of "researchers" in the PeopleSoft database. To send a message to all the "researchers" of the university, the communicators need to read the definitions of hundreds of job codes to judge whether each job code should be included in the mailing list.
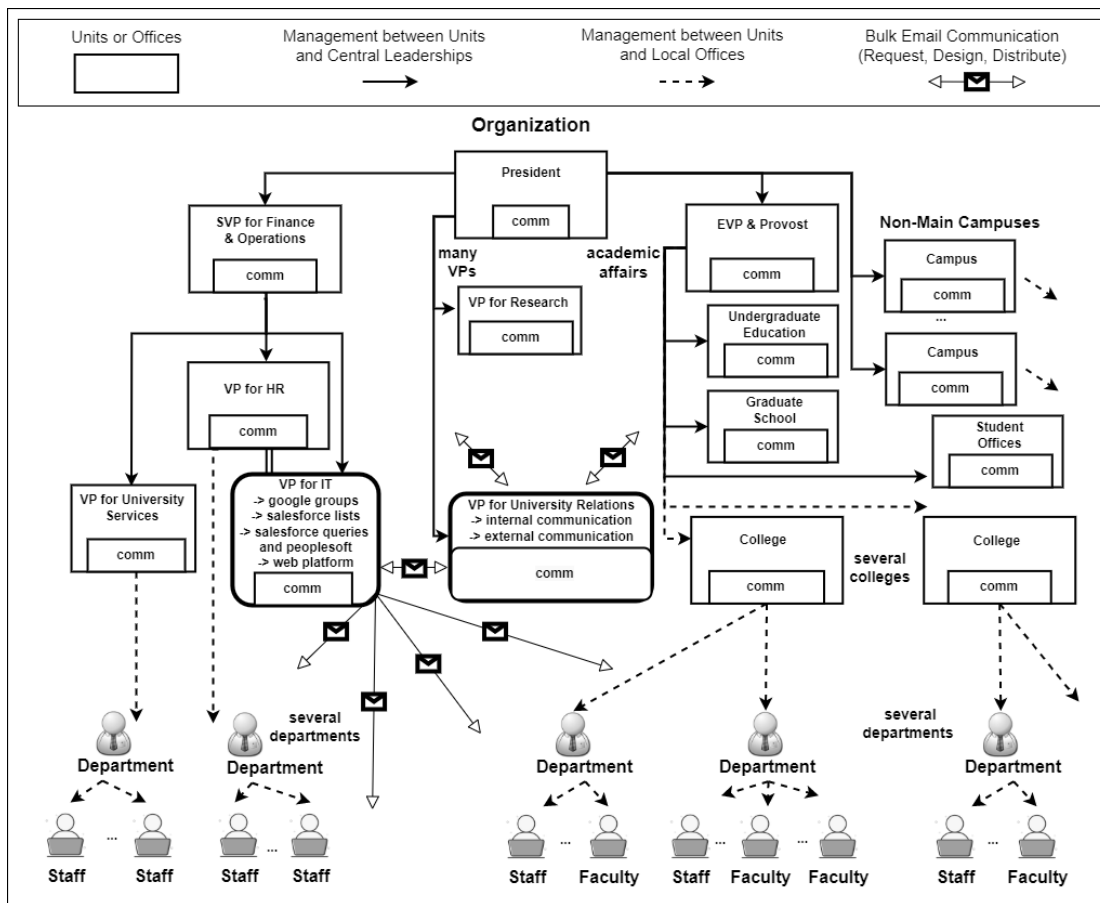
Figure 3.1: Organization structure of the University of Minnesota related to bulk communication; "SVP", "EVP", "VP": (Senior/Executive) Vice President's offices; "comm": communication units or communicators. Details of other campuses and many central units are not shown. Some units also conduct external bulk communication (e.g., emails to alumni and the general public), which is not shown here.

3.4   Survey Study Design

To test the effectiveness of channels without the influence of message content, we designed survey questions in pairs: a real message and a corresponding fake message. We built 11 such message pairs and asked the participants to indicate whether they could recognize each message. 162 responses were collected.[2]

The participant pool was defined as employees who had received the real messages we selected in the past 2 weeks before the survey was distributed, were not senior leaders of the university, and were part of the university's volunteer pool for conducting usability studies on university systems. We randomly selected 3000 potential participants from the participant pool.

We distributed surveys to all of the potential participants through invitation emails with survey links. Participation was voluntary and uncompensated. We collected 162 completed responses, and there were 30 additional responses that were incomplete. The data we show below are calculated based on the completed responses. It is worth noting that the participants who completed the survey were a population likely to be more engaged with organization email than average. The effectiveness calculated based on their data would likely be higher than the effectiveness based on the whole recipient population within this university.

To test the effectiveness of channels without the influence of message content, we designed survey questions in pairs: a real message and a corresponding fake message. We wanted to test whether the recipients could recognize the real messages from the fake messages. The real and fake messages had similar content features. We built 11 such message pairs with different channel and content features.

For the real message, the standards of selection were:

- It was received by all the potential participants (employees).

- It had general importance to the university and the participants.

For the fake message, the standards of selection were:

- It had a similar importance level as the corresponding real message.

---

[2]   Ruoyan Kong, Haiyi Zhu, and Joseph Konstan. Organizational bulk email systems: Their role and performance in remote work. New Future of Work, 2020. [95]

• It had the same actionable, relevancy features as the corresponding real message.

After we designed the message pairs, we worked with the communicator group at the university to verify the messages we selected matched the standards above.

In the survey, we asked the participants to indicate whether they could recall the message on a 5 point likert scale where 1 is "I Have Not Seen it" and 5 is "I Have Seen it". The purpose of this investigation was to test how well participants recalled the message they received, with the corresponding fake message with similar content features serving as a control. When analyzing survey responses, a score of 4 or 5 was considered as having seen the message, while 1 to 3 indicated having not seen it.

3.5   Survey Study Findings

Employees did not retain most of the bulk messages. On average, the real messages were claimed seen by 38% participants, but the fake messages were also claimed seen by 16% participants. That suggested a $22\% = 38\% - 16\%$ average effectiveness. 22% of the real bulk messages could be recalled by participants. 62% of real messages were not claimed as seen. 16% of real messages claimed as seen could not be discerned from the fake messages. See Table 3.1 for effectiveness and percentage of recognition of all messages.

Table 3.1: Effectiveness of each test message. For each message group, effectiveness of the real message in it was defined as %*real message claimed seen by participants* − %*fake message claimed seen by participants*.

message no.	3	7	1	2	4	average	11	10	5	9	8	6
effectiveness	0.94	0.39	0.34	0.3	0.25	0.22	0.22	0.13	0.12	-0.01	-0.02	-0.1
%real claimed seen	0.98	0.42	0.51	0.38	0.41	0.38	0.36	0.17	0.18	0.08	0.31	0.38
%fake claimed seen	0.04	0.03	0.16	0.08	0.16	0.16	0.14	0.03	0.06	0.09	0.34	0.49

The message that was most effective was the single message that was relevant and actionable to most of the recipients ("Your 2019 W-2 tax-reporting form is now available online. (message 3)") — 98% of the recipients claimed that they had seen it, and it was the only one that had effectiveness over 50%. The real message that the least recipients claimed seen is the message about another campus in the newsletter (message 9) —

only 8% of participants claimed that they had seen this message. The message that was most ineffective was from the president (message 6) — though 38% participants claimed they had seen the real message, 49% claimed they had seen the fake one, which might mean that the participants remember the types of bulk messages they received, but did not retain these messages' content.

3.6   Artifact Walkthrough Design

What are the practices, experiences, and assessments of different stakeholders in the bulk email system at this university? We conducted artifact walkthroughs with 17 stakeholders — 6 communicators, 9 recipients, and 2 managers within the university.[3]

An artifact walkthrough is used to walk the interviewer through the process of completing a task. It helps the interviewer infer possible design intent by asking probing questions regarding the decision between various options [17]. In our study, for example, recipients used certain queries in their inbox to retrieve specific types of emails; answered questions about their actions (trashed/opened/scanned/read in detail) and assessments (sometimes they were asked to reread the email) on each email case retrieved.

3.6.1   Recruitment and Participants

We interviewed 17 participants in the same university using the following process:

1. We worked with the same group of communicators in section 3.3 to generate a list of potential participants by a stratified sampling approach, which included 8 communicators and 20 recipients, based on length of work experience, nature of the office, and job responsibility.

2. We began inviting participants for a 30 to 60 minute interview in our lab or their own offices.

3. We stopped inviting after we interviewed 6 communicators and 9 recipients as we found the addition of new interviewees did not add any new observations (i.e., we had reached data saturation).

---

[3]   Ruoyan Kong, Haiyi Zhu, and Joseph A Konstan. Learning to ignore: A case study of organization-wide bulk email effectiveness. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW1):1–23, 2021. [98]

4. To avoid recipients' answers being influenced by their awareness of manager involvement, we waited until after the interview was complete to ask the recipient if they would allow us to invite their direct manager to discuss the non-personal emails we had collected from them. Two recipients agreed and we invited their managers for an extra 30-minute interview.

This study was reviewed and approved by the IRB of the University of Minnesota [4] . 17 individuals participated (Table 3.4). These include:

1. 6 communicators — 2 communication directors who had over 10 years of experience and 4 communication staff, from 5 different central offices and 1 college office; their responsibilities included editing newsletter, writing drafts for university leaders, leading technical support, etc.

2. 9 recipients — 5 staff (1 from a central office, 2 from college offices, 2 from departmental offices) and 4 faculty; 5 recipients had been at the university for over 10 years.

3. 2 managers — R5's manager (a college office's director) and R6's manager (a program director) were invited.

3.6.2   Interview Protocol

A. Communicators. The artifact-walkthroughs with communicators were composed of 3 parts:

- General practice questions on their duties, goals, and mechanisms of distributing bulk emails.

- Email case questions on how an email was designed, sent and measured. We asked communicators to select important/unimportant cases from their points of view.

- General assessment questions, see Table A.1 in the appendix.

B. Recipients. The artifact-walkthroughs with recipients were composed of 3 steps:

---

[4] IRB ID: STUDY00006417

Questions	Command
How many emails did the participant receive within 1 week?	newer_than:7d,in:anywhere AND NOT from:me
How many emails did the participant receive and not read within 1 week?	newer_than:7d,in:anywhere AND label:unread AND NOT from:me
How many emails did the participant receive from their organizations within 1 week?	newer_than:7d from:umn.edu AND NOT from:me ,in:anywhere
How many emails did the participant receive from their organizations and unread within 1 week?	newer_than:7d,in:anywhere AND NOT from:me AND label:unread AND from:umn.edu
How many mass emails did the participant receive from the organization within 1 week?	newer_than:7d,from:umn.edu AND NOT from:me AND (list:(local) OR list:(list) OR to:(lists) OR (category:(Forums|Promotions))) ,in:anywhere
How many mass emails did the participant receive from the organization and unread within 1 week?	newer_than:7d,from:umn.edu AND NOT from:me AND label:unread AND (list:(local) OR list:(list) OR to:(lists) OR (category:(Forums|Promotions))) ,in:anywhere
How many newsletters did the participant receive from the organization within 1 week?	newer_than:7d,from:umn.edu AND NOT from:me AND (list:(local) OR list:(list) OR to:(lists) OR (category:(Forums|Promotions))) AND (subject:(news|update|brief)) ,in:anywhere
How many newsletters did the participant receive from the organization and unread within 1 week?	newer_than:7d,from:umn.edu AND NOT from:me AND label:unread AND (list:(local) OR list:(list) OR to:(lists) OR (category:(Forums|Promotions))) AND (subject:(news|update|brief)) ,in:anywhere
How many personal emails did the participant receive from the organization within 1 week?	newer_than:7d,from:umn.edu AND NOT from:me AND to:umn.edu AND NOT (list:(local) OR list:(list) OR to:lists OR (category:(Forums|Promotions))) ,in:anywhere
How many personal emails did the participant receive from the organization and unread within 1 week?	newer_than:7d,from:umn.edu AND NOT from:me AND to:umn.edu AND NOT (list:(local) OR list:(list) OR to:lists OR (category:(Forums|Promotions))) ,in:anywhere ,label:unread

Table 3.2: Email searching commands for Gmail. The University of Minnesota used a Gmail system based on G Suite for Education.

- Collect inbox logged data: They were asked to copy and paste 10 email queries in their Gmail's search box (see Table 3.2) to retrieve subsets of organizational emails, non-mass emails, mass emails, and newsletters they had received in the previous week. We recorded the number of emails left unread/opened of each category (see the actions' definitions in Table 3.3 — logged left unread/opened).

  Notice that we retrieved emails in all categories, including those archived and in trash/spam, by command "in:anywhere", except those permanently deleted from the trash. By default, only emails that have been in the trash for more than 30 days would be automatically deleted by Gmail. We did not observe the case that the trash was manually cleaned by the recipients. We assumed that we could retrieve all emails received in the past 7 days.

  These queries were pretested in one of the authors' inbox to ensure they retrieve the right subset of emails. We recorded the size of each subset, and how many were opened.

- Collect self-report data: For the mass/newsletters we retrieved above, we investigated up to 12 emails for each category. For each email, we recorded the recipient's actions on: 1) whether the email was left unread/opened/scanned/read in detail/trashed (see the actions' definitions in Table 3.3 — self-report left unread/opened/ scanned/read in detail/trashed); 2) whether the email asked for actions/whether the recipient took actions; 3) importance/urgency/relevance on scale 1 to 5; 4) reasons for the answers/actions, and 5) whether they would change their mind about answers/actions if they reread it now (see appendix A). We investigated 163 cases in total.

- General questions: We asked recipients some general questions such as how frequently they checked their email accounts, how many emails accounts they had, did they feel that the number of emails they received was too many to read all of them, and how often they did not read an email.

C. Managers. The answers from the employee-manager pair were confidential from each other. We showed the managers the emails we collected from their employees, and asked them to give their preferred actions that their employees ought to have done with

those emails — trashed/left unread/opened/scanned/read in detail, and estimations on importance/urgency from 1 to 5. We investigated 31 emails in total with the two managers. Note that the term "manager" here means the direct supervisor of the recipients, and is not a reflection of a particular position in the organization. Furthermore, the recipients listed as "director" in Table 3.4 are heads of their offices or programs but are not listed as "manager" because we interviewed them about their own received messages, not about the messages of one of their employees.

To summarize, for participants invited as communicators/recipients/managers, we discussed the bulk emails they sent/received/their employees received correspondingly.

The Actions of Recipients	Definition
Logged Left Unread	The email was labeled unread by Gmail.
Self-Report Left Unread	The recipient claimed that they did not open the email.
Logged Opened	The email was labeled read by Gmail.
Self-Report Opened	The recipient claimed that they opened the email.
Self-Report Scanned	The recipient claimed that they read the email quickly to get its general idea only.
Self-Report Read in Detail	The recipient claimed that they read the whole email thoroughly and carefully.
Self-Report Trashed	The recipient claimed that they archived the email or moved it to the trash folder; untrashed email was left in the inbox or moved to a non-trash folder.
Self-Report Opened and Trashed	The recipient claimed that they opened the email first then trashed it.

Table 3.3: The definitions of recipients' actions with emails. Logged data was collected by using queries to retrieve subsets. Self-report data was collected by asking recipients their actions directly. Each participant provided both of the logged data and self-report data.

#	Stakeholder Type	Years at University	Position	Level of Office	#	Stakeholder Type	Years at University	Position	Level of Office
C1	Communicator	6 - 10	Staff	Central	R1	Recipient	1 - 5	Staff	Central
C2	Communicator	11- 20	Director	Central	R2	Recipient	1 - 5	Staff	Departmental
C3	Communicator	1 - 5	Staff	College	R3	Recipient	11 - 20	Director	Departmental
C4	Communicator	11 - 20	Director	Central	R4	Recipient	20	Staff	College
C5	Communicator	1 - 5	Staff	Central	R5	Recipient	1 - 5	Staff	College
C6	Communicator	6 - 10	Staff	Central	R6	Recipient	20	Professor	\
M1	R5's Manager	6 - 10	Director	College	R7	Recipient	11 - 20	Professor, Director	\
M2	R6's Manager	20	Professor	\	R8	Recipient	11 - 20	Professor	\
					R9	Recipient	1 - 5	Professor, Director	\

Table 3.4: Demographics of participants. C1 — C6, R1 — R5, M1 are staff and the rest are faculty. Central-level offices are in charge of university-wide affairs, like university services, information technology; college-level or departmental offices are located within a college or department, such as the dean's office.

3.6.3   Data Analysis

To compare the experience and assessments of the effectiveness of the bulk email system from different stakeholders, we took a grounded theory approach [29], specifically with the following iterative procedure:

1. Identifying the themes from the communicators' transcripts.

2. Searching for relevant text and themes in the recipients' transcripts.

3. Comparing the actions taken by the recipient and preferred by the manager (if applicable).

4. Inviting new participants and repeating the steps above until we reached data saturation. We stopped when we interviewed 17 stakeholders because we found strong repetition in the themes identified.

3.7   Artifact Walkthrough Findings

3.7.1   Number of Emails Received and Actions Taken

Table 3.5 reported the statistics of the 9 recipients for emails received, opened, read, and trashed by category. The recipients, on average, received 376 emails in the week prior: 153 of them were organizational emails and 30 of them were bulk emails (25 mass emails and 5 newsletters). The number of messages in one of the 55 newsletters investigated in the interview could be as many as 35, with an average of 8.5. Participants (recipients) reported that they, in general, received too many bulk emails. R6 said *"Sometimes I felt overwhelmed."*. R4 and R8 said that a large number of bulk emails became a burden to them.

Faculty received 175 organizational emails a week on average (148 non-mass emails, 23 mass emails, 4 newsletters), and for staff, this average number was 136 (103 non-mass emails, 27 mass emails, 6 newsletters). The average logged open rate of mass/newsletters of faculty recipients ($> 90\%$) was higher than their self-report open rate ($¡70\%$). Sometimes they clicked an email's title and removed its unread tag, but they did not think that they "opened" the email.

Though many bulk emails were opened, few of them were read in detail. In fact, 58% of mass emails were reported being opened by staff while only 28% of them were read in detail; 67% of mass emails were reported as being opened by faculty although only 13% of them were read in detail.

Many bulk emails were trashed, 52% of mass emails and 22% of newsletters were reported as being trashed by faculty, while for staff the percentages were 27% and 15%.

Faculty	Email Type	Overall	Organizational	Organizational Non-Bulk	Organizational Mass	Organizational Newsletter
no.= 4	# Received Per Week	560	175	148	23	4
	% Logged Open Rate	55%	96%	97%	92%	94%
	% Self-report Open Rate	\	\	\	67%	59%
	% Self-report Read in Detail Rate	\	\	\	13%	9%
	% Self-report Trash/Archive Rate	\	\	\	52%	22%
Staff	Email Type	Overall	Organizational	Organizational Non-Bulk	Organizational Mass	Organizational Newsletter
no. = 5	# Received Per Week	229	136	103	27	6
	% Logged Open Rate	54%	73%	78%	52%	50%
	% Self-report Open Rate	\	\	\	58%	58%
	% Self-report Read in Detail Rate	\	\	\	28%	30%
	% Self-report Trash/Archive Rate	\	\	\	27%	15%

Table 3.5: Average per-capita email volume by category; % logged open rate is calculated by the inbox logged data; % self-report open/read-in-detail/trash/archive rate is calculated by the self-report data, based on all the bulk emails investigated in the corresponding category, for example, self-report read in detail rate of mass email = # mass email read in detail/# mass email investigated.

3.7.2 Nature of Bulk Email System as an Organizational System

The bulk email system was part of the university's administrative system, with communicators trying to achieve the communication goals of this university. The communicators in this system had clients, performance metrics, and tools.

Communicators as a gateway for organizational leaders to reach employees.
Communicators had clients from their own offices or other offices who were not professionals in communication and who wanted to promote events, notify staff of changes of policies, etc. Many clients were the communicators' leaders:

*We work with our leadership whoever sends the message, so for example, we
send a few messages for the associate vice president for research, who's also
the head of responsive administration. (C1)*

Naturally, clients wanted to send more content and reach more people. A communicator talked about their clients:

*We're working with other departments coming up with certain final messages
that may be close to the amount of information that people can digest. But
some of them think that people have a more of an appetite for reading. (C1)*

Thus communicators usually needed to narrow mailing lists and shorten the length
of emails. One mechanism they used was letting clients fill in templates:

*The U Relations developed a partnerships with emergency management.
When there is a safety alert, there are certain blanks in the template they fill
in. That's why you can see the safety messages [emails] are often similar.
Because we give them the authority to send them, so we want to limit the
scope of the content they can choose to put in there. (C4)*

Sometimes communicators could not modify the emails. Some bulk emails were
required by law:

*We have things called safety advisories. Maybe somebody is grabbed from
Ford Hall last night and we just found out about it. That person is gone but
we are required by federal law to tell you about that. (C4)*

Communicators are anchored to the "open rate" metric

Optimizing a system depends on using appropriate metrics. The metrics used by
communicators to measure bulk email's effectiveness did not match recipients' assessments. We asked communicators about the metrics they used in the bulk email system, and measured recipients' assessments for each bulk email from various metrics like
whether trashed/opened/scanned/read in detail, and ratings on urgency/importance
/relevance. For communicators, the priority order of the metrics used to report a bulk

email's performance was: open rate > click rate > replies and others. They only needed to report a good open rate to their clients. Communicators indicated that 50% was a good open rate to them:

> *For all campus messages [emails], open rates are usually at least 50%, which is good for us, so pretty high. We look at click rate but not closely. Most times we look at open rates. (C5)*

However, "open rate" can be a flawed metric for measuring recipients' feedback, because:

1. "Being opened" did not always mean "being read". As we discussed above, the logged open rate, which was used by communicators, might be higher than the self-report open rate — as sometimes recipients clicked an email's title and removed its unread tag, without even reading any of the email. Some faculty recipients, like R7, opened 17 of the 22 bulk emails we investigated, but only read 2 of them in detail, see Table 3.6.

   Recipients were not opening bulk emails because they considered them relevant but because they were unsure of their relevance and wanted to verify:

   > *I read the first line then deleted it because I've seen similar information in other news. (R7)*

2. Most "opened" bulk emails were only scanned and got low ratings in importance/urgency/ relevance. As shown in Figure 3.2 (3), 43/54/36 bulk emails got 1 or 2 on importance/urgency/ relevance ratings out of 66 scanned emails.

3. Some bulk emails were opened, found irrelevant, and then trashed. As shown in Table 3.6, R1, R3, R5, R6, R7, R9 all trashed some bulk emails after they opened them.

   Among participants, trashing email as an inbox tool was a matter of personal preference. For example, R1 trashed almost every bulk email after they opened and read it (see Table 3.6), *"If it doesn't require actions, or not related to my work, I just trash it.".* In contrast, R2 did not trash any email.

Trashing a bulk email did not necessarily mean that it was useless. For example, R5 trashed a relevant bulk email after reading it: *"I skimmed it because I was going to see whether we have all clear on the water issue ... I delete it because the water is good, no action needed."*, and trashed a relevant bulk email without opening it: *"I didn't need to open it because all the useful information are in the title — come and get [person name] ... I trashed it."*.
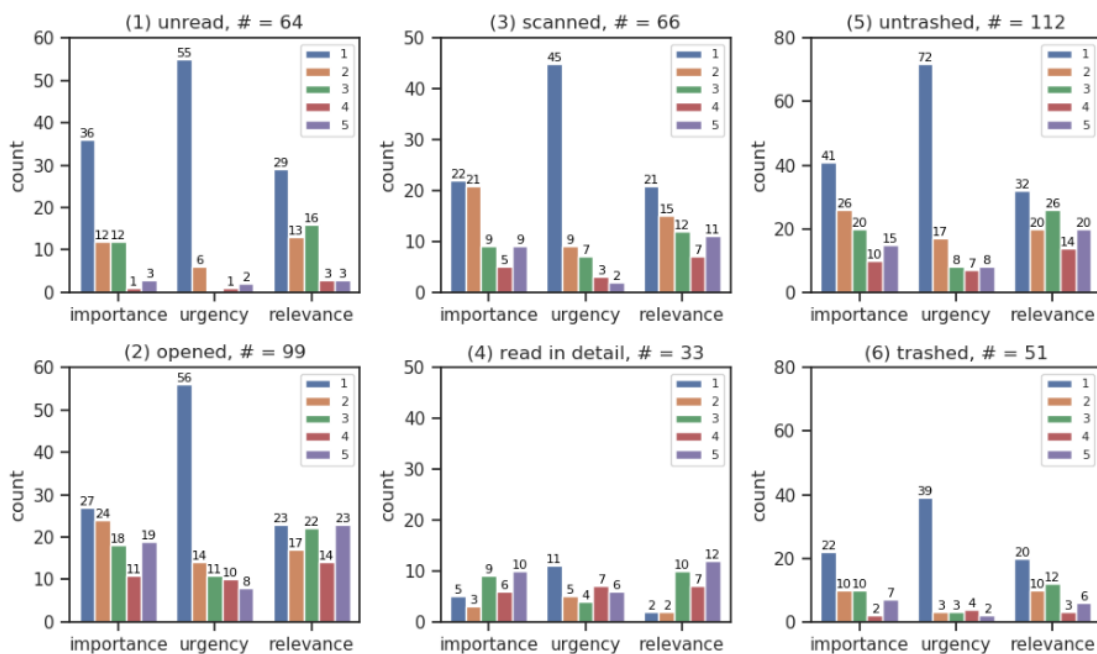
Figure 3.2: Counts of ratings on importance/urgency/relevance of the investigated bulk emails which were reported being left unread/opened/scanned/read in detail/left un-trashed/trashed by the recipients.

Though "open rate" was a flawed metric, communicators had few tools in the bulk email system to measure performance beyond an email's open/click rate: *"I don't know, how long people spend, whether they share with other people." (C1)*

Implication 1: Provide End-to-End Metrics. Focusing on "open rate" distracted communicators from real measurements of both channel and bulk email's effectiveness. Communicators usually were not involved in the transactional process of the bulk emails.

They did not know about the recipients' level of awareness of the content — the recipients read the emails and took corresponding actions to complete the organization's goals. Thus showing metrics on the transactional process of bulk emails to communicators could be helpful. For example, a measure showing how many people missed the due dates of events that were sent to them via bulk emails. When connecting heterogeneous systems is difficult, we could build systems similar to some human resources analytic systems [140], such as collecting assessments from clients/leaders for each bulk email.

#	type	# investigated	# opened	# read in detail	# trashed	# opened and trashed
R1	staff	20	5	3	14	5
R2	staff	23	19	12	0	0
R3	staff	12	4	1	1	1
R4	staff	8	5	2	1	0
R5	staff	20	15	6	3	2
R6	faculty	17	12	1	7	4
R7	faculty	22	17	2	8	5
R8	faculty	24	10	2	3	0
R9	faculty	17	12	4	14	9

Table 3.6: The number of bulk emails investigated/opened/read in detail/trashed/opened and trashed. The data was self-reported by participants on the specific bulk emails we investigated during the interviews.

Diverse types of communication with different content and delivery channels, and different matchings between content and channels.

A. Bulk email's contents. Communicators mentioned three types of bulk email's content:

- Transactional content: This was content aimed at facilitating the work of the university, and which asked recipients to take action. Communicator C4 talked about the goals of transactional bulk emails: *"Generally, we want you to change your behavior or be aware of something happening, or think differently about behavior or things like that."*

  Some transactional content was urgent and required immediate actions:

  > *Around something we want people to immediately know about, we usually send those transactional versus commercial, we have that option*

within the salesforce. [5]   *(C1)*

- Highlighted News: These were the most important updates that offices wanted employees to be aware of (e.g., announcements by the university president of new officials), though for most employees these would not be actionable.

- Good-to-know News: These were other updates that offices felt employees would benefit from knowing, or that the institution would benefit from having more people aware of. Communicators recognized that individual employees could miss this component without consequence. A communicator introduced a e-newsletter with good-to-know and highlighted news:

  > *The broadest thing we do is that we have a blog, that we produce content and stories every month, and every month we sent highlights of the blog, that we want people to know about, to be aware of. (C1)*

A bulk email's content might be good-to-know for some recipients, while being transactional for others. We asked a communicator whether all the recipients (all students, faculty, and staff) of a bulk email about campus-safety tips should read it: *"Students for sure. For faculty and staff, it's a reminder, but it's not something you have to know." (C5)*

B. Bulk email's distribution mechanisms.   There were 3 types of distribution mechanisms of bulk emails: single bulk emails (mass emails), newsletters (newsletter emails), and redistribution (sent to the communicator's contacts in decentralized units). Rather than being universal, the matching of type of content to type of distribution mechanism varied across communicators. Some communicators sent good-to-know news through newsletters and some sent it through single bulk emails. Some communicators sent highlighted news and transactional content through single bulk emails and some communicators sent it through newsletters:

> *We send single messages [individual bulk emails] when something is really timely we need to get it out, and when it's critical that people receive and*

---

[5]   *See Saleforce's transactional message API: `https://developer.salesforce.com/docs/ marketing/marketing-cloud/guide/transactional-messaging-api.html`.*

> *see it. If something is a story, less critical then that might find its way into our newsletter instead. (C1)*

> *If there is an actionable thing we want people to do, that typically will go into a newsletter bulletin. The single message [individual bulk email] would be like a topic that needs to have a lot more detail but not necessarily an action needed to come out of it. So it's an awareness thing. (C5)*

As communicators use bulk email distribution mechanisms differently, good-to-know news, highlighted news, or transactional content might be put into a newsletter at the same time. We asked C5 whether the recipients of an email should read all of the 10 messages in it: *" Not really. So we really want the message "have the ID card with you" get out, that was important, the most actionable thing, but the rest of it is to communicate that we care." (C5)*

Recipients needed to go through all the messages in each email to screen messages relevant to them. A recipient talked about a newsletter with over 30 messages:

> *What I do is I go through it, ok, [program name] … I'm not interested in it, upcoming programs … if I see something interesting then I'll read and click on that. (R6)*

This lack of consistency between bulk email's content and distribution mechanisms across different communicators caused confusion about what type of content newsletters should be used for and whether recipients should view newsletters as important.

Communicators' various roles — Communicator Directors vs. Communicator Staff.

Communicators' concerns regarding bulk emails varied based on their roles. When asked about improving the bulk email system, the communicator directors prioritized the engagement of the community, *"We struggle on the engagement, make people know the service we have." (C2)*, *"Our hardest things are how to let people feel that we all have a part of responsibility in safety, each individual have to make effort, it's much easier for us to keep the campus safe." (C4)*

Communicator staff, however, would like more tools within the bulk email system to support their operational work, *"It's great to have people to say 'I don't want this*

*type of thing' or 'Please send more content like this."' (C5), "From a user perspective,
I think that maybe adapting some common templates could be helpful." (C6)*

3.7.3 Different Perspectives on Bulk Email's Values

Communicators' assessments — high-level bulk email is important.

High-level bulk emails were source of general university information, such as changes
in benefits, leadership, public safety, administration, or working tools, and were usu-
ally from the university-wide offices (see Table 3.8 for the definitions of email levels).
They were classified as bulk emails with general importance by communicators. Com-
municators usually sent these types of emails to all faculty and staff: *"Public safety
is everybody who works at the U should know about, and any big change in benefits or
leadership will affect everybody." (C1)*

Communicators viewed bulk emails from university leaders as important, as we asked
C5 what kind of emails were considered more important: *"Message from the president,
from the vice president of university services about campus safety, those types of things."
(C5)*

Managers' assessments — have awareness about high-level information.

Managers sometimes agreed with their employees (recipients) that some bulk emails
were not important. Within the 31 emails we asked managers to assess, there were 5
emails that the recipient and the manager both agreed should be scanned; and there were
13 emails that both the recipient and their manager agreed should be left unread. There
were times when managers disagreed with their employees. There were 8 cases where
employees opened these emails while their managers thought they should skip them.
These 8 cases involved emails of interest to employees, but that were not particularly
job-related, such as alumni marketing, a company's on-campus recruitment feedback
questionnaire, etc.

Managers thought that employees should have some sense about what was going on
at the high level of this university, specifically leadership and administration changes, as
in the 5 cases shown in Table 3.7. That table represents the bulk emails that managers
1) thought their employees should read while the employees did not, or 2) thought the
employees should read in detail while the employee scanned. For example, M2 thought

R6 should read the bulk email with the title *"Your Help Matters Now"* and content *"Legislative session ends Monday. Please take 60 seconds to make sure that when the final negotiations are done."* in detail, but R6 left it unread.

Employees' assessments — high-level bulk email is irrelevant.

Employees often skipped bulk emails about legislation of the university and leadership changes (see Table 3.7) because 1) they had time pressure; 2) they felt these emails were too high level to be related to their role as non-leader staff: *"I skimmed and deleted it. It's a program that's important to my boss (the dean), but not directly related to my office." (R7)*

Surprisingly, the higher the bulk email's level, the lower the probability that the bulk email would be opened/read in detail; Figure 3.3 is a graph showing the levels of the investigated bulk emails (see Table 3.8) with their open/read in detail rates. A bulk email from a university-wide office had 48% chance of being opened and 14% chance of being read in detail, while those statistics for a bulk email from a department's office were 95% and 57%.

Ignoring high-level bulk emails made employees miss some useful emails. When asked to read previously unread bulk emails their managers had deemed important, employees found some of these emails helpful. R6 found that bulk email #1, on Table 3.7, should be read and was actually easily actionable: *"When I read it now, I may click on here and take 30 seconds to make sure the final negotiation. So I didn't realize that it was so easy to click this button. I was thinking it might involve more like writing a short letter."*

3.7.4   Costs of Communication

Recipients' feeling of burden and their reading strategies.

Controlling the burden on recipients was not one of the goals of the current bulk email system, as stated by a communicator when asked whether a recipient's time cost was considered: *"No, the email we sent are most not nice-to-know emails. They are required information for people." (C2)*

Receiving many emails did not necessarily make recipients feel burdened. R9 received over 774 emails per week but did not get a sense of email overload. R9 talked

#	Manager & Recipient	Manager Expectation	Employee Reaction	Title	Sender	Employee's Reason
1	M2 - R6	Read in Detail	Unread	Your Help Matters Now	University Government Relations	I just deleted it because I think it's important to advocate for the university but I don't do that for myself.
2	M2 - R6	Scan	Unread	[College Name]'s Update ● Summer 2019	Newsletter Center of [College Name]	I was too busy, this email is not like so important.
3	M1 - R5	Scan	Unread	Reminder - Incorporating GoldPASS into your Career Courses session tomorrow	University Employer Relations	I will not open it, because it's for undergrad department.
4	M1 - R5	Scan	Unread	PeopleSoft Navigation Update: MyU PeopleSoft Outage This Weekend	PeopleSoft Update Team	I already learn it from other sources.
5	M2 - R6	Read in Detail	Scan	[[College Name] - Staff] [College Name] Highlights, Leadership Announcements and Resources to Share, 05/20/19	[College Name] Leader's Office	I am busy and I will spend less time on this.

Table 3.7: Title/sender of the bulk emails which managers thought their employees should read while they did not, or thought the employees should read in detail while they just scanned, and the employees' reasons.

#	Level	Meaning
1	University	Sent from university wide offices.
2	College	Sent from college wide offices.
3	Department	Sent from department wide offices.

Table 3.8: The email levels' definitions. For example, "university-wide" offices indicates that the office was in charge of sending bulk emails to recipients across the university.
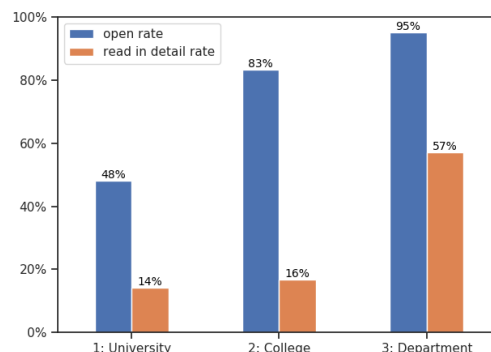
Figure 3.3: The self-report open/read in detail rate of bulk emails of different levels collected in the interviews with recipients.

about their strategy: *"I can handle it (bulk email) ... I scan it right away ... I delete the things I know I don't have to read ... I look at the subject ... Mentally you do it in hierarchy — you determine what's the next, what's the first."*

The reading strategy "checking the content whenever a new email came in" might bring on a sense of burden, like R6 said: *" I checked my email pretty quickly. Even when I am working on something, I may go back to my email to see is there anything important that's coming, and I respond pretty quickly ... To be honest, sometimes I felt overwhelmed.".* This result mirrored Mark's findings on work emails and productivity [123].

Retrieving the correct bulk email from inboxes can be a burden: *"I might miss something. Sometimes my supervisor ask me whether you received it — I trashed it then I try to find it. That's frustrating because in Gmail sometimes I can't find it. You try to think what's in the subject line, date ..." (R4).*

Implication 2: Make Recipients' Burden Visible to Communicators. The time cost caused by bulk email and the collective burden put on recipients was inapproachable for the communicators. They did not know whether they sent a bulk email to too many recipients and let them spend unnecessary time, or collectively let some recipients feel overloaded. For example, employees' time spent on bulk emails and the corresponding financial costs could be estimated [79].

Communicators' Costs — Loss of channel credibility (organizational communication capital).

Communicators have, as requested by their clients, had to send some bulk emails at the cost of the sender's credibility. We asked a communicator whether the credibility of bulk email channel was considered:

> *Yes, we definitely think about it (the credibility of the sender's name). But we work with different clients, they request the audience they want. (C1)*

Recipients learned from experience and stopped opening bulk emails from low-credibility senders. For example, R9 trashed an email from the university fitness center without opening it: *"I recognized the sender and trashed it because I am not interested in it."*

Implication 3: Make the Credibility of Bulk Email Channels Visible to Communicators. Without a tool to tell communicators how credible a channel is to its recipients, communicators discarded ineffective channels and replaced them with new channels, which was a time consuming strategy. Quantifying channels' credibility and setting standards of usage for channels could be a helpful tool. For example, calculating the credibility and success probability of communication [47], or avoiding sending optional bulk emails via high-credibility channels.

3.7.5   Understanding Current Practice and Its Failings

Lack of personalization tools

Communicators tried to help recipients find important messages in a newsletter by putting them in the top position or putting their keywords in the subject of the email: *"Sometimes we put important messages in the subject line. It's not the same importance (to all recipients), but they can recognize that it's important or not. We do try to keep it relevant and make it clear at the top." (C1)*

However, as C1 said, importance of the messages in a bulk email may vary for different recipients. Recipients tended to close a newsletter whose first message was not relevant; in the process they often missed the later messages which were useful to them.

R6 told us why a bulk email was unread and R6 found it useful when we asked them read the email:

> *They started with sports, then I thought it's not relevant to me. But when I go back with you, I found out that it has research that I am interested in. (R6)*

Given this situation, communicators thought that personalization of email designs would be helpful, but they did not have the technology to do so. Currently, all recipients receive the same subject lines/subtitles/content for each bulk email. We asked a communicator whether they had felt the need for personalization of emails):

> *We haven't, although it's something we would like to explore doing. I don't think our team has technical expertise. I don't know where and how to get that type of thing. (C6)*

Implication 4: Explore Personalization. Develop personalization tools for communicators to design bulk emails. For instance, have users estimate the relevance of messages and rank messages from high to low relevance [176].

Flawed Email Targeting Tools

A. Target through mailing lists — recipients cannot opt-out. There were some mailing lists built for certain groups, e.g. all faculty of the university, all staff of a department. Communicators assumed that recipients would unsubscribe from mailing lists they were not interested in. Communicators also designated certain lists without opt-out choice to avoid core recipient opt-out. For example, a communicator talked about a monthly newsletter:

> *For the people who work as administrators, they cannot opt-out of that. I check this. Everyone else is optional, they can unsubscribe. (C3)*

However, there were cases where a mailing list was used to send both important and unimportant bulk emails at different times. Recipients could not opt-out from this mailing list, even though unimportant bulk emails were sent through it, because they would not receive important emails sent through that mailing list in the future. For

example, a communicator talked about an optional bulk email sent to communication directors:

> It's optional for them to take action to do what I ask them to do. This is meant to be kind of helpful service I'm giving ... They could (opt-out of this mailing list) but this is a mailing list for the communication directors so they probably won't. (C5)

B. Target through querying position-database — limited to the human resource system. If there were no mailing lists existing for the target population, communicators would query the database of employees when sending emails. However, the current bulk email system only has access to the human resource system. The HR system solely supports querying based on positions, not on recipients' networks or interests. A communicator introduced this process:

> We will go to HR, and ask them to pull a list from PeopleSoft. Sometimes, for example, we were trying to reach everyone with family, so we go to HR, and they were not able to pull out. They can only do if the data is in PeopleSoft and we can get it by job code. (C5)

A communicator talked about a bulk email whose target population was all employees who had not downloaded documents from Webex. But the communicator did not have such information in the bulk email system, it was sent to all Webex users instead:

> This is the final reminder to Webex users: action required to download reports. Receivers are people who have been using Webex in the last year. That went to 9000 receivers.

> We cannot tell who has downloaded their file or not. The audience is who is using Webex and have files. Anybody who has already moved to Zoom, downloaded their files don't need to take action, but we don't know who are them. (C2)

C. Redistribution mechanisms — no tracking of bulk emails. Communicators without enough information to target recipients often assumed that lower-level units

would have more employee information asking unit-level contacts to forward emails to the related recipients:

> *We have in each unit one person in charge of the facility things. And they may have people in their department, in their buildings. Sometimes we will send information to the unit facility leads, which is a small set of people (author's note: this is a set of 400 to 500 people). That's when we want other people to distribute something more locally versus from the University Relations. We want people in [college] to get some messages [emails] from their unit. We may ask them to distribute that message [email]. (C4)*

However, communicators could not track whether and how the contacts sent these bulk emails, nor the performance of these bulk emails. When we asked a communicator if they could track the performance of the bulk emails sent to the contacts, the response was: *"Unfortunately we can not track that." (C6)*

D. When it is hard to target — overwhelm everyone. As both mailing lists and querying were limited to certain scenarios, precise targeting was difficult for communicators. C6 introduced a group difficult to narrow: *"Open enrollment is tough because we don't know how many people need to take action. That's kind of a dilemma. Because we try to make that system easier for people so you don't have to do anything. But that makes it harder for us to do communications."*

Motivated by "getting the proof of delivery", communicators thought that it was better to overwhelm everyone than to miss a single person. The bulk email mentioned above finally went to all employees: *"We moved everyone to do it to ensure that no one is left behind." (C6)*

"Sent to all" was one of the reasons that recipients received irrelevant bulk emails. R4 talked about a bulk email they classified as not useful: *"This is what I call blanket message [email] — it goes to everyone, it's common knowledge, it's not that related to me."*

Implication 5: Precise targeting is difficult. Previous research has shown that an email's label can change recipients' behaviors [144]. Given those findings, a mechanism could be designed to hint at an email's value to its recipients before they open it, for example, labeling whether a bulk email is optional or mandatory, action-needed or not

[177].

Communicator transferred the responsibility of being aware

Communicators transferred the responsibility of knowing bulk emails' content to their recipients. A communicator talked about a newsletter from the college's dean that they sent to all employees weekly, with 17 messages and 5 messages from the dean:

> *They are all important emails, sent once a week. So our expectation is —*
> *"This is the important email you get from your college, regarding your job.*
> *This email is the business of the college, you should read it." (C3)*

However, recipients found the bulk emails were too many or too long for them to filter out unimportant ones, find important ones, and be aware of all of the content. They felt that it was not the recipient's responsibility to be aware of all the content. A recipient talked about a weekly newsletter sent to all employees about government activities:

> *It took so much time to read, 10 paragraphs, nobody gonna read that, it's*
> *gambling. People use emails like a safety cover, but there is no understand-*
> *ing from the recipients' side. The responsibility shifts from the sender to the*
> *recipient, that's unfair. (R3)*

Vicious cycle — recipients did not read then communicators sent more.

Similar to Randall's findings in organizational communication [151], the ineffectiveness of the bulk email system fed into a vicious cycle — 1) recipients did not read bulk emails on time; 2) to get recipients to read bulk emails/take actions on time, communicators sent bulk emails multiple times/more widely; 3) recipients received more irrelevant bulk emails, lost trust in bulk email channels, and read fewer bulk emails in the future. A communicator talked about the collection process of the employee engagement survey: *"Those messages [emails] start from September, October, in terms of emails, go through January ... we do have to do emails more than once because our email open rate is about 61 or 62 percent." (C6)*

Communicators did not realize the existence of the cycle and thought repeated bulk emails would be appreciated: *"My assumption is if somebody is emailing me about a service that is going away, and I have taken actions, I will never need to open it, but I will be grateful if I get reminders." (C2)*

However, recipients found they usually opened a bulk email and found that they already read it/still not interested/already took action. For example, a recipient trashed a bulk email about reimbursement:

> *It is important and relevant but this is not the first time she sent. I already picked up what I need ... I only read the first email a couple of months ago ... She even attached the same document again ... so I just deleted it. (R9)*

Implication 6: Provide Feedback to Communicators. To interrupt the vicious cycle, recipients' feedback could be collected. This could include whether the recipient wanted more like this/less like this (similar to the applications in social medias [89]), how much time the recipient had spent on it, whether they had already seen/replied to this bulk email. Communicators could use that feedback to decide whether to keep sending these bulk emails to the recipient, or only distribute reminders to people that have not read or confirmed them.

Limitations: The observations might not be generalizable to all kinds of organizations with different cultures/structures.

3.8   Discussion

3.8.1   Align Different Stakeholders' Priorities

We found a large gap between different stakeholders' priorities. Communicators thought they were sending important high-level information about the organization, and communicators' clients wanted their messages to be sent broadly. Regardless of those intentions, recipients did not view high-level bulk emails as important. This gap in perception was caused by different stakeholder priorities.

On one side, the major task for communicators was to get the "proof of delivery" of bulk emails to their clients, whose first priority was to make the target population aware of certain information and take appropriate actions. On the other side, recipients, who

were collectively burdened by the communicators, wanted to receive bulk emails that were interesting and relevant to them, and fewer messages overall.

Outside of the organizational environment, recipients can unsubscribe or not open email — forcing senders to modify their practices to reach audiences. However, within the organization, where employees have responsibilities to complete the organization's tasks and minimize the organization's costs, this misalignment of priorities brings costs to the organization. The communication channels' credibility end up harmed, and the organization's tasks do not get done.

Thus it is important to align different stakeholders' priorities, while also including organizational priorities in the bulk email system. Recipients should be reminded that besides information they find interesting, they have the responsibility of being aware of what is going on in the organization. Communicators and their clients should be reminded that, beyond distributing their information, they have the responsibility of maintaining the effectiveness of organizational communication channels and should avoid wasting employees' time on unnecessary emails. At an extreme, a Key Performance Indicator System [104], which takes the organization's priorities into consideration, could be integrated into the bulk email system. For communicators, there could be indicators of the time cost, the credibility cost, as well as clients' feedback for each bulk email. For recipients, there could be indicators of the speed of taking actions required by the senders.

3.8.2   Limitations — Organizational Culture/Structure's Influence on Bulk Email System

This is a qualitative case study of one study site, as such the observations may not be generalizable across organizations. It is possible that stakeholders in other organizations have different practices/perspectives on bulk email systems with respect to their organizations' cultures/structures.

The organization we studied had a hierarchical management structure — from university leadership (Board of Regents, Office of the President, Office of the Provost), to the college offices (deans, academic affairs), down to the department offices (heads, program directors). Thus the information flow of the bulk email system was mostly one directional — low-level employees received information from the high-level leaders

via bulk emails, while high-level leaders rarely needed to get information from low-level employees via bulk emails.

As for stakeholders within an organization, bidirectional interaction proved to be an important factor in the engagement levels of organizational affairs [134]. Due to the hierarchical management structure within the university, the bulk email recipients may have less willingness to participate, causing the bulk email system to be ineffective. It is possible that recipients in organizations with a flat structure have a higher willingness to participate in bulk email communication.

Though hierarchical, the organization's communication system was also decentralized. Each communication office worked for its own units, had its own goals, and served its own clients. There were no central alignments of these communication offices and information would often be sent multiple times from different offices. These repeated bulk emails became a burden that multiple communication offices collectively placed on recipients. As the previous research found, *"decentralized organization is not a very efficient organization model for work that requires a lot of intensive interaction between different employees."* [122]. It is possible that a bulk email system with a central communication office to arrange different communication offices' tasks and collaboration would work more efficiently.

3.9   Conclusion and Future Work

Conclusion: This study indicated a systemic failure of the organizational bulk email system. The organizational bulk email system had many stakeholders, but none of them had a global view of the system or the impacts of their own actions. First, the communicators' high-level clients wanted their messages sent broadly. Second, the communicators felt they were sending important high-level information through bulk emails; they got high "open rates" and only reported that to their clients. Third, the recipients viewed most of the bulk emails not relevant; they often opened and then rapidly discarded bulk emails without reading the details. Last, though the communicators across the organization worked together in a network that attempted to improve the quality of communication practice, there was little consistency in how bulk email channels were used and there was a limited set of targeting and feedback tools to support their task.

In theory, the answer is simple; there is a long history of information filtering technology to help recipients avoid spam based on their preferences, and email marketing technology to help communicators advertise products to external recipients. But in an organizational setting, recipients' or communicators' preferences are not enough; the organization's priorities matter. Recipients may not have an interest in certain information, but their employer expects an awareness of that information regardless. At the same time, communicators and their clients may want to send bulk emails broadly, but they need to consider the cost to the organization.

Impact: We hope this case study provides detailed information on how an organizational bulk email system works and fails, with respect to this organization's structure and nature. To the best of our knowledge, this is the first work focusing on multi-stakeholders' perspectives on bulk emails within an organization. We hope our protocols on inbox artifact-walkthroughs are useful for future research.

Based on these findings, we plan to 1) frame an economic model for organizational bulk email systems, including the value, cost, and action of its stakeholders (chapter 4); 2) design an organizational bulk email system to support multi-stakeholder prioritization. This system would include personalization tied to both the interests of the recipient and the needs of the organization (chapter 5). Finally, we plan to 3) support communicators in evaluating their bulk communications' value and cost (chapters 6, 7).

Chapter 4

Economic Model

4.1 Introduction

In this chapter, we propose an economic model to describe the organizational bulk email system to identify this system's control points, potential interventions, and research opportunities. [1]

We consider this system's economics from this simple scenario. Imagine that the CIO of a university wants to inform everyone that voicemail will be down Sunday from 5-7 am, and does so by asking a communicator to send an email to 20,000 employees. The CIO may find this email free and convenient. But it is not free to the organization. An estimate of the cost of employee time would be 20,000 * 2 min (reading and interruption time) * $0.75 salary+fringe+overhead/min = $30,000. Plus, there is a future cost in the sender's reputation. If only 3% of the employees find this message useful, then the other 97% who read this email might view future emails from CIO as possibly irrelevant and fail to open them. In this case, the stakeholders (CIO, communicator, employees) interact with the system differently, and each perceives various costs to and value from the operation of the system. Moreover, the organization as a whole is a stakeholder, seeking to maximize both its employees' productivity and their organizational knowledge.

[1] Joseph A Konstan and Ruoyan Kong. The challenge of organizational bulk email systems: Models and empirical studies. In The Elgar Companion to Information Economics. Edward Elgar Publishing, 2023. [102]

In this chapter, we frame the ongoing economics (such as the value, cost, and actions of the stakeholders above) within organizational bulk email systems based on our empirical studies (Chapter 3).

We will first look at related work on organizational communication models. We then propose an economic model based on our empirical studies and explain the model with specific bulk email cases. Finally, we discuss design implications and research opportunities identified from the proposed model.

4.2   Related Work and Gaps

Several economic models around organizational communication have been proposed. Davenport and Beck proposed an economic model of employees' attention [45]. It described an "attention budget" for each person and applied the law of supply and demand to it — the larger the amount of information employees received, the greater the demand for their attention, the higher the likelihood more information would be ignored, and ultimately the greater the likelihood business goals will not be attained. Blazenaite treated organizational communication as a series of treatments of measurable communication variables and relationships. They framed the value organizations receive from communication at each stage and pointed out that communicators are the most powerful components of this system [19]. Ruck and Welch suggested that organizational communication should consider employees' information needs on designing the contents instead of focusing on the communication process only. They proposed a model to reflect employees' individual (job) and social (organizational identification) communication needs [147]. It is also worth noting that the stakeholders in these economic models might not be fully rational. For example, Glynn surveyed a symphony orchestra and found the stakeholders' perceived value ("capability") of communication is nonlinear, nonrational, and socially constructed [57].

Though there are several economic models for organizational communication, the economics within organizational bulk email systems have not been specifically explored. What makes organizational bulk email interesting is the variety of choices that exist for distribution and targeting (including newsletters, individual messages, and routing

messages through trusted individuals), and the degree to which some of these choices become known "channels" with **reputations.** The reputation of a channel is defined as the recipients' perception of the credibility, quality, and relevance of the emails sent through this channel. It is associated with the identifiable characteristics of the channel—the sender's name, the sender's address, or the newsletter brand in the subject line. The recipients can identify this channel from these characteristics before opening its messages. Then the recipients make decisions on which messages to open and read. Reputation can be analyzed either accumulatively (the overall reactions of all the audience of a channel) or individually (each employee's individual reaction to different channels). Bulk email also comes with a set of tools that are better capable of tracking the opening and reading of messages (compared with hardcopy distribution, for instance) allowing more detailed study of the practices of recipients and more direct feedback to senders.

To bridge this gap, in this chapter, we propose an economic model for an organizational bulk email system based on our study in chapter 3 which frames 1) the value, costs, and preferences of the stakeholders in the organizational bulk email system and how they lead to different actions; 2) how these actions affect the organization's cost, value, and communication channels' reputations. For simplicity, we refer to organizational bulk email as bulk email, and organizational bulk email system as bulk email system below.

4.3 Channels and Distribution Mechanisms

What makes organizational bulk email interesting is the variety of choices that exist for distribution and targeting. In this section, we introduce the available channels and distribution mechanisms for bulk emails. The **channel** of a bulk email refers to its sender and brand. They could be newsletters' names or university leaders' (offices') names. They are characteristics visible to recipients before opening.

Let's use the email "*U of M Brief (January 26, 2022)*" and "*Senior Leader Search Update*" in Figure 4.1 and 4.2 as examples. They are sent to all the employees across all the university campuses. "*Brief*" is a weekly newsletter from the University Relations Office. It shows its brand "*U of M Brief*" in its subject line and template. Employees
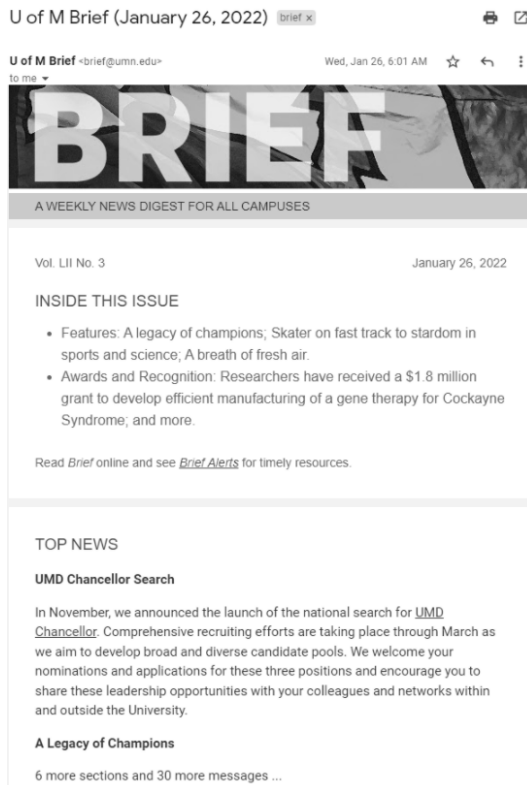
Figure 4.1: An example message "*UMD Chancellor Search*" in an example email "*U of M Brief*". Several paragraphs are omitted in this figure.
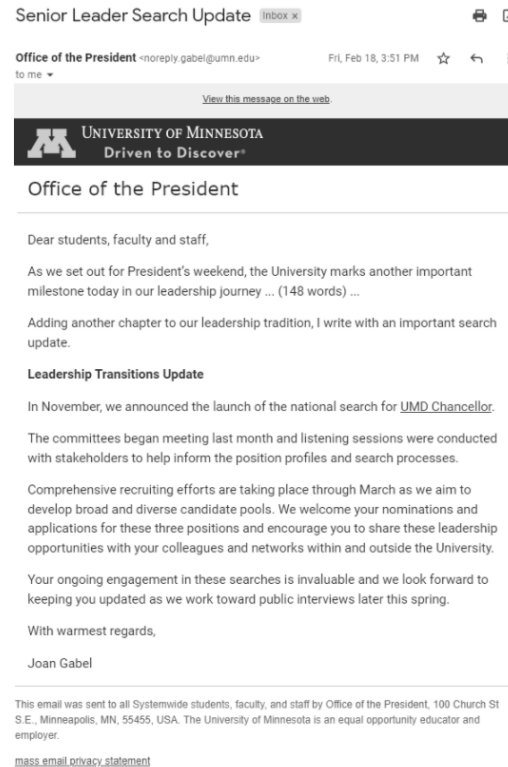
Figure 4.2: An example email *"Message from the President"*. It contains two messages.

will observe that this email is sent from "U of M Brief <brief@umn.edu>" in their inboxes. *"Senior Leader Search Update"* has its brand "*Office of the President*" in its template. This email's sender is shown as "Office of the President <noreply.gabel@umn.edu>". The senders, brands, and templates of these bulk emails mean they usually do not look like personal emails from the same "people" (probably intentionally). Different channels have different levels of "**grouping**" — the number of messages in that channel's email. "*Brief*" is high-grouping (around 30 messages embedded in each email). "*Office of the President*" is low/medium-grouping (1 to 5 messages in each email).

The **distribution mechanism** of a bulk email refers to how communicators send

it to recipients. Communicators use different distribution mechanisms for different channels. With different distribution mechanisms, channels have different levels of "**targeting**" — the granularity of matching between the messages in that channel and its audiences. There are three major distribution mechanisms: newsletter, mailing to a list, email tree.

4.3.1 Newsletter

Suppose that a communicator at the President's office wants to announce the search for the Chancellor of the UMD campus (University of Minnesota – Duluth [UMD]) and feels that this message has a broad audience (see Figure 4.1). They could submit the content to a central newsletter editor (Figure 4.3, step 1). The editor then puts the message into a newsletter, such as the U of M Brief above (step 2). The communicator distributes the Brief to its subscribers (— all the staff and faculty) through Salesforce (step 3). From a low level to a high level of targeting, the communicator has choices within all-employee newsletters (e.g., *Brief* or *Message from the President*), more narrowly targeted university-wide newsletters (E.g., Synthesist, a biweekly newsletter of the Graduate School sent to all graduate faculty & staff with around 5 messages embedded), and affinity group newsletters (E.g., the research newsletter (a google group) of the college of science and engineering, see Figure 4.6, 2c)

4.3.2 Mailing to a List

Suppose the communicator above wants to be more targeted by only sending the message to the employees of that specific campus UMD. The communicator first searches for the mailing list of UMD employees in Google Groups or Salesforce. If there is no such list, the communicator queries the human resource system to pull a list (Figure 4.4, step 1). The email is sent to the mailing list by Gmail or Salesforce (step 2). From a low level to a high level of targeting, the communicator has choices within all-employee lists, departments or units' employee lists, and filtered lists based on more fields (such as job codes). For example, the communicator could choose send the message only to directors, department heads, and deans on the UMD campus.

Step 1: The communicator sends the message "*UMD Chancellor Search*" to a university-wide newsletter editor.
Step 2: The editor puts the message into a Brief.
Step 3: Salesforce distributes the Brief to all the faculty and staff.

The employees (or the employees in the corresponding units) that received the message.
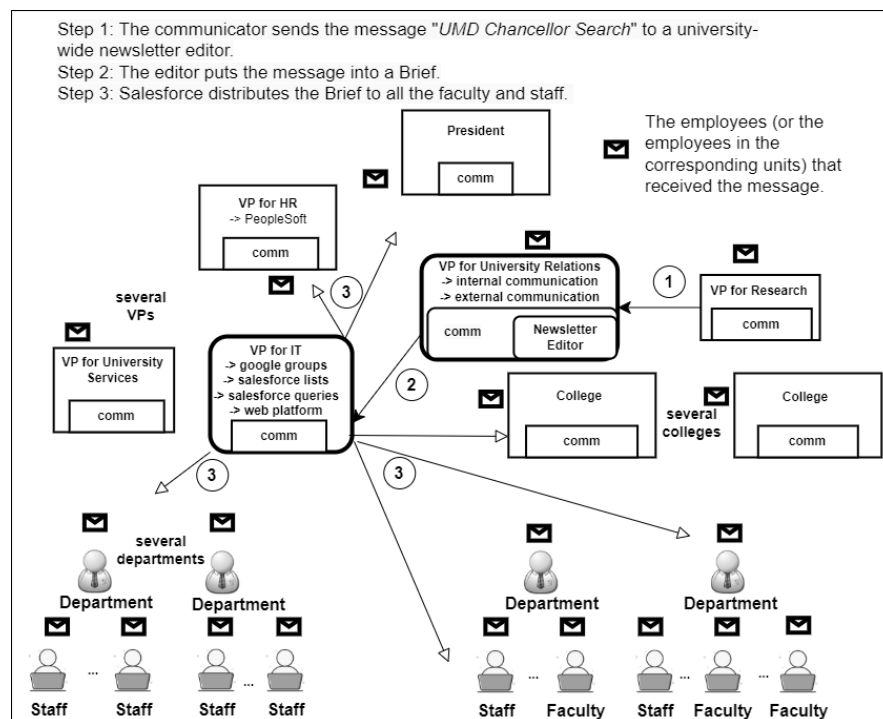
Figure 4.3: Distribute bulk messages through newsletters.
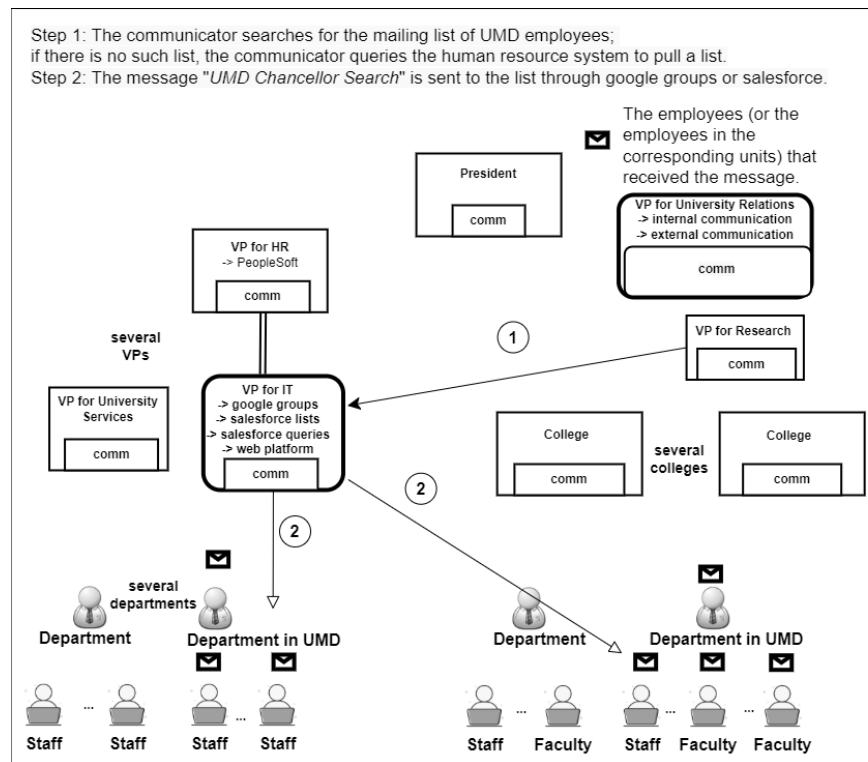
Figure 4.4: Distribute bulk messages through mailings lists or querying databases.

4.3.3 Email Tree (Send to Local Units to Redistribute)

We use the term **email tree** to refer to the process of distributing a message through the management structure from higher-level managers down to lower-level management and eventually to individual employees. The goals of using such a tree are to gain employees' attention (because they listen to their local managers) and sometimes also to filter the recipients to those who need to receive the message (since managers know which units or employees are likely to need it). For example, to reach all faculty in the university, a communicator could send the message to Deans and Chancellors, asking them to forward it to their faculty; in turn they would send it on to Department Heads who would send it to the faculty (who would in turn pay more attention because they know their department head), or send to Vice Presidents to forward within their staffs to keep parallel.

For filtering, suppose that a communicator from the Office of Vice President for Research has a message on a grant opportunity for the early-career faculty in the field of cognitive science (see Figure 4.5) and could not get this list from their databases. They could distribute the message through an email tree. The communicator sends the message to several relevant Deans' offices, such as the Associate Deans for research of the college of biological sciences, the college of science and engineering (Figure 4.6, step 1a), etc. If the college is large, the Associate Dean might ask several department heads to redistribute the message to the relevant faculty (step 2a). If the college is small enough that the Associate Dean knows each faculty member's interests, they might send individual messages to all the early-career cognition faculty they know (step 2b). The Associate Dean could also put the message in a college research newsletter (step 2c). The faculty and staff who subscribe to the newsletter's google group will receive the newsletter. The degree of grouping and targeting depends on the local units' actions.

To consider how an email tree can borrow channels of higher reputation and hence increase the likelihood that a message is read, let us examine the case of a communicator from the Office of Vice President for Human Resources who has a message about face mask requirements (see Figure 4.6), and they would like everyone to know about this message. The communicator could use an email tree to utilize local attention since employees are more likely to open emails from their own units, as discussed in the mixed-methods study section. The communicator sends the message to all the colleges

Faculty Nominations

I am writing to invite you to nominate early-career faculty to participate in a new, three-year Scialog (science dialog) series that we will be launching in October 2022, *Molecular Basis of Cognition* (https://rescorp.org/scialog/molecular-basis-of-cognition), co-sponsored by the Research Corporation for Science Advancement, the Frederick Gardner Cottrell Foundation, and the Canadian Institute for Advanced Research. Each annual meeting of this Scialog will bring together about 50 Fellows (faculty from their first year through soon after tenure) and 10 Facilitators (senior faculty who guide the discussions) over 3½ days, with the goal of launching innovative, collaborative, interdisciplinary projects that we and our foundation partners support with one-year grants of $50,000 per Fellow on each funded team. In 2021, Research Corporation and our partners provided over $7M in grants to Scialog Fellows.

We expect to include Fellows from a broad array of fields and who bring a variety of methodological approaches to the study of cognition. To nominate one or more faculty as a Scialog Fellow (no limit to the number of nominations from an institution), please use the form available at:

https://www.surveymonkey.com/r/2022_ScialogFellow_nomination_form

Figure 4.5: An example message "*Nominations for Cognition Faculty*".

and administrative units (Figure 4.7, 1b), and asks them to forward this message to all of their employees (step 2d).

Changes to our requirements
For our faculty, staff, and student workers in instructional, clinical, and transit settings, face coverings will continue to be required. Additional information on where and when masks may be required is available from Safe Campus.

Those who work in other settings—including residential housing, dining facilities, and offices—will not be required to wear a mask while at work. However, you may continue to wear a mask in these settings based on your personal preference and expect support from your coworkers and leaders in creating a positive workplace that is welcoming and respectful.

Figure 4.6: An example message "*Changes to Face Mask Requirements*".

We summarize various channels' distribution mechanisms and levels of grouping and targeting in Table 4.1. When a communicator receives a message request, they will have to decide which channels and distribution mechanisms (the **control point** that influences the corresponding stakeholders' cost and value retrieved from this message) to employ. We propose an economic model below to discuss how this control point influences the actions, costs, and values of various stakeholders and the reputations of communication channels.

Table 4.1: Summary of distribution mechanisms and their levels of targeting, and the example channels with the corresponding distribution mechanisms and their levels of grouping.

Distribution Mechanism	Low Targeting à High Targeting		
Newsletter	All-employee newsletters. E.g., *Brief* (high grouping), *Message from the President* (low/medium grouping)	University-wide newsletters to all employees. E.g., *Synthesist* (medium grouping)	Affinity-group newsletters. E.g., the research newsletter (a google group) of the college of science and engineering (Figure 4.6, 2c) (medium grouping)
Mailing to a List	All-employee lists E.g., the voice-mail down message from the CIO to all the employees (low grouping)	College / Department / unit lists, E.g., sending the message "*Nominations for Cognition Faculty*" to the list of faculty in the college of science and engineering (low grouping)	Filtered/selected listsE.g., sending the message "*Nominations for Cognition Faculty*" to the early-career faculty with research interests in cognitive science (low grouping)
Email Tree	Use local status / attention to reach all employees, E.g., *Changes to Face Mask Requirements* (Figure 4.6, 2d – low grouping)	Use hierarchy to identify all relevant employees, E.g., *Nominations for Cognition Faculty* (Figure 4.6, 2a /2b – low grouping)	

Step 1a: The communicator sends the message "*Nominations for Cognition Faculty*" to all or several Deans for Research, such as the Dean of the College of Science and Engineering, the College of Biological Sciences, or certain related offices, etc.

Step 2a: If the college is large, the Dean might ask several department heads to redistribute the message to the relevant faculty.

Step 2b: If the college is small, the Dean might send individual messages to all the relevant faculty (early-career cognition faculty).

Step 2c: The Dean puts the message in a college research newsletter. The faculty and staff who subscribe to the newsletter's google group will receive the newsletter.

Step 1b: The communicator sends the message "*Changes to Face Mask Requirements*" to all colleges and asks them to forward it to all their employees.

Step 2d: The college sends the message to all its employees.
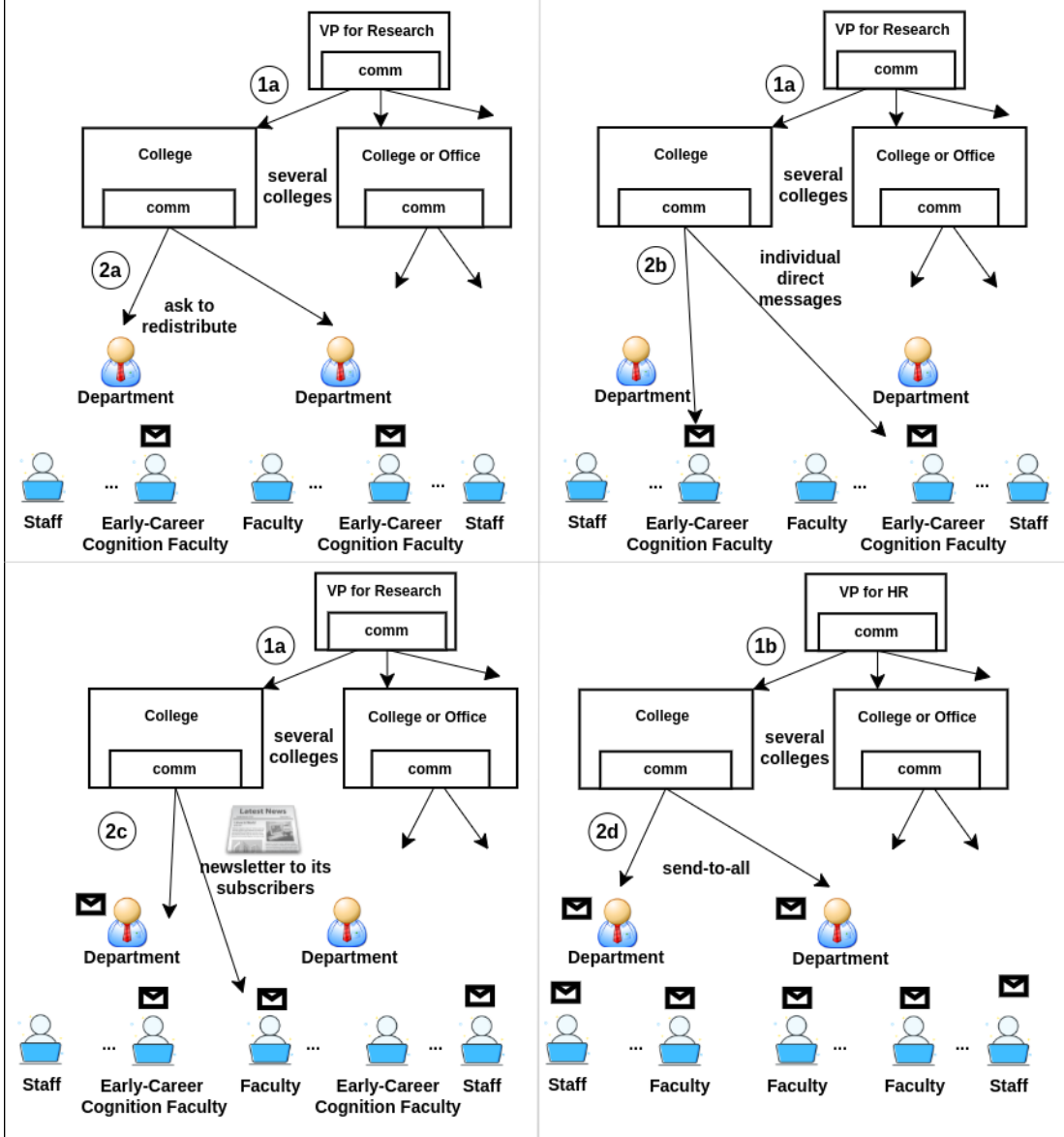
Figure 4.7: Distribute bulk messages through an email tree.

4.4   Economic Model

In this section, we propose an economic model for organizational bulk email systems based on our study site's organization structure and bulk email structure above. Figure 4.8 shows the correlation between the stakeholders' costs and value in the organizational bulk email system. One information producer sends a request to the communicator to distribute a message (step 1). The communicator then makes decisions on their actions, including whether, how, and whom to send it (step 2). The employees who receive the messages then decide whether to read or not (step 3) and then receive related costs and value. These cumulative value and costs contribute to the organization's total value and cost. The time cost of the communicator also contributes to the organization's total cost.

Table 4.2 summarizes the stakeholders' costs and value that they could receive from their actions with bulk emails. These value and cost are inspired by our interview study above — communicators want to get their tasks done within reasonable efforts; employees want to get relevant information; information producers want their information to reach all the relevant employees). $t_{e,m}$ refers to whether employee e is targeted as message m's recipient ($t_{e,m} = 1$ if employee e receives message m otherwise 0). $p_{e,m}$ refers to employee e's probability of reading message m (we'll discuss its relationship with channel reputation below). In general, the information producers value that employees receive their messages; the communicators want to meet their clients' requirements with minimal time and effort; the employees would like to get useful information efficiently. The stakeholders try to maximize their returns (g= value – cost). The model in Table 4.2 has the following assumptions:

1. Different stakeholders have different understandings of the value of a message being work-relevant / important / interesting to an employee. For example, we use $v_{e,m}^{\mathrm{wr,p}}$ to represent information producer p's assessment on how work-relevant message m is for employee e to know about it. This value could be different from $v_{e,m}^{\mathrm{wr,b}}$ — the manager b's assessment on message m's work-relevance to employee e. And these two values might all be different from $v_{e,m}^{\mathrm{wr}}$ — the assessment on message m's work-relevance from employee e themselves.

2. The time cost c^{tim} contains multiple elements. For the information producers, it
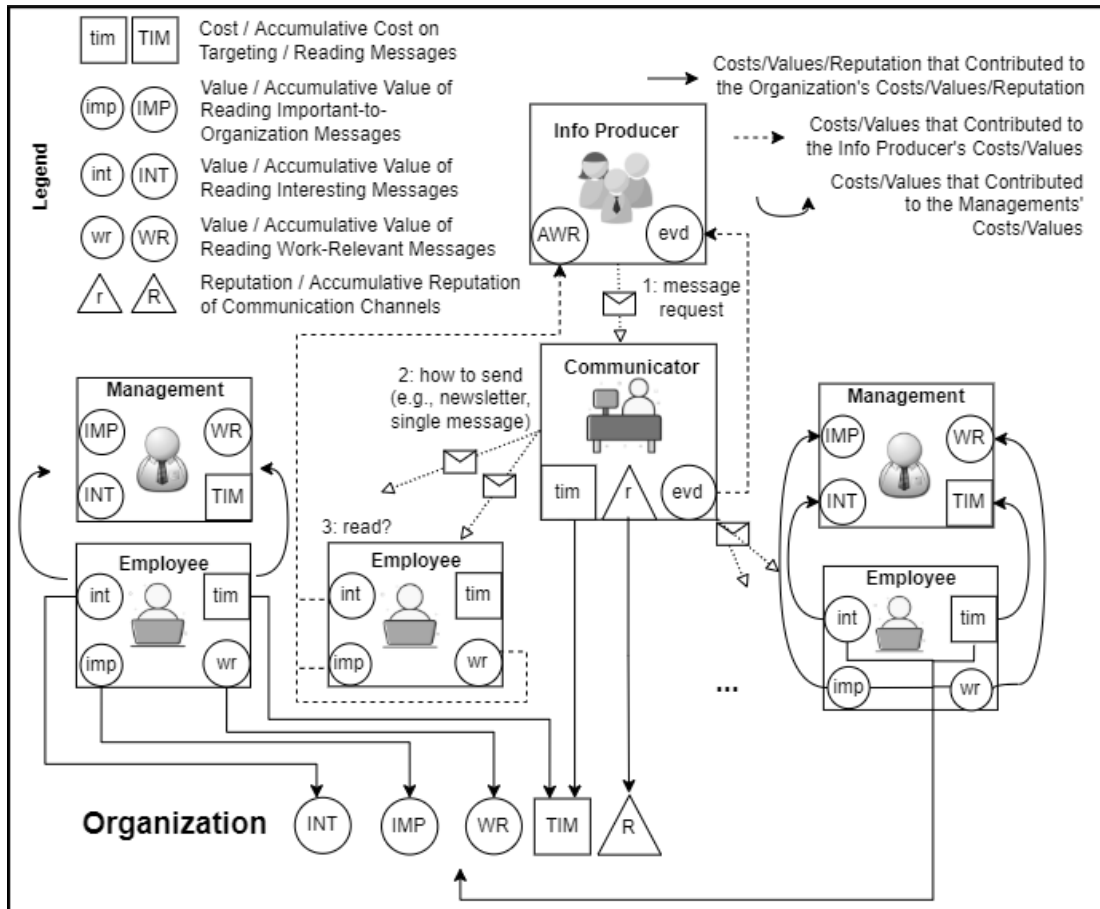
Figure 4.8: The correlation between the stakeholders' costs and value in the organizational bulk email system.

contains the time needed for composing and communicating with the communicators. For the communicators, it contains the time cost for designing and targeting the messages.

For employees, it contains the time cost for interruption (same for the messages in each email), deciding (same for the messages in each email), and reading (proportional to the corresponding message's length and position in the email). An employee's time cost with a message will be zero if they do not receive it. If they receive it, they might be interrupted and take several seconds to decide to read it or not based on a message's sender and subject line, even if they decide not to read it later. Time costs for employees also depend on their reading strategy. For example, the employees who process their email in batches might spend less total time due to less interruption.

3. Communication channel ch's reputation r_{ch} is the weighted sum of its reputation to each employee e. r_{ch} also contributes to the reputation R of the organization's communication channels as a whole. r_{ch} is not only influenced by the messages sent through channel ch but by also other communication channels.

4. Each information producer, communicator, manager is also an employee. For the simplicity of this model, we do not consider the interactions between employees and do not consider the nonrationality of stakeholders' actions. It would be helpful to validate these simplifications or explore these factors to build a more complete model in future work.

We will discuss how we come up with the return formulas in this table below in this section.

4.4.1 Employee's Value and Cost

Employee e will only receive value from a message m if 1) they receive that message ($t_{e,m}$ = 1), and 2) they read that message, or they can get value from reading the title of that message ($p_{e,m}$ = 1), otherwise they receive zero value. Whether they receive message m or not depends on communicator c's actions. Their probability of reading a message m they receive ($p_{e,m}$) is positively correlated with the message's communication channel's

Table 4.2: Summary of the costs (c), value (v), return (g), and reputations (r) of the organizational bulk email system's stakeholders/communication channels.

Information Producer p's Value and Cost on Message m	Communicator c's Value and Cost on Message m
$v_{p,m}^{awr}$: the relevant employees know what they are trying to disseminate ($=\sum_e t_{e,m}p_{e,m}(v_{e,m}^{wr,p}+v_{e,m}^{imp,p}+v_{e,m}^{int,p})$) $v_{p,m}^{evd}$: provide the evidence that the required information was disseminated to the relevant employees ($=\sum_{e\in \text{relevant employees}} t_{e,m}$) $c_{p,m}^{tim}$: time needed for composing and communicating with the communicator $v_{p,m}$: information producer p's total value from message m ($=v_{p,m}^{awr}+v_{p,m}^{evd}$) $g_{p,m}=v_{p,m}-c_{p,m}^{tim}$	$v_{c,m}^{evd}$: meet their clients' requirements (send the message to all the relevant employees, $=\sum_{e\in \text{relevant employees}} t_{e,m}$) $c_{c,m}^{tim}$: the time cost for sending Δr_{ch}: communication channel ch's reputation's change $g_{c,m}=v_{c,m}^{evd}-c_{c,m}^{tim}+\Delta r_{ch}$
Employee e's Value and Cost on Message m	**Manager b's Value and Cost**
$v_{e,m}^{wr}$: get work-relevant information (if they read it) $v_{e,m}^{int}$: get enjoyment/interesting information (if they read it) $v_{e,m}^{imp}$: get important-to-organization information (if they read it) $c_{e,m}^{tim}$: the time cost of interruption, deciding, and reading (would be 0 if they do not receive this message) $v_{e,m}$: employee e's total value from message m ($=t_{e,m}p_{e,m}(v_{e,m}^{wr}+v_{e,m}^{imp}+v_{e,m}^{int})$) $g_{e,m}=v_{e,m}-c_{e,m}^{tim}$	$v_{b,m}^{wr}$: the employees maintain awareness of the work-relevant messages sent through communication channels ($=\sum_{e\in E} t_{e,m}p_{e,m}v_{e,m}^{wr,b}$), E= the employees who report to b $v_{b,m}^{imp}$: the employees have good feelings about the organization through getting important-to-organization information ($=\sum_{e\in E} t_{e,m}p_{e,m}v_{e,m}^{imp,b}$) $v_{b,m}^{int}$: the employees have good feelings about the organization through getting interesting information ($=\sum_{e\in E} t_{e,m}p_{e,m}v_{e,m}^{int,b}$) $c_{b,m}^{tim}$: the employees' time cost ($=\sum_{e\in E} c_{e,m}^{tim}$) $v_{b,m}$: manager b's value from message m ($=v_{b,m}^{wr}+v_{b,m}^{imp}+v_{b,m}^{int}$) $g_{b,m}=v_{b,m}-c_{b,m}^{tim}$
Organization o's Value and Cost on All Messages	
$v_{o,m}$: the weighted sum of all stakeholders' value ($=\sum_{p,m} w_p v_{p,m}+\sum_{e,m} w_e v_{e,m}+\sum_{c,m} w_c v_{c,m}+\sum_{b,m} w_b v_{b,m}$) $c_{o,m}^{tim}$: the accumulative cost of all stakeholders ($=\sum_{e,m} c_{e,m}^{tim}+\sum_{c,m} c_{c,m}^{tim}+\sum_{p,m} c_{p,m}^{tim}$) ΔR: the accumulative change of communication capital ($=\sum_{ch} \Delta r_{ch}$) $g_o=\sum_{p,m} w_p v_{p,m}+\sum_{e,m} w_e v_{e,m}+\sum_{c,m} w_c v_{c,m}+\sum_{b,m} w_b v_{b,m}-(\sum_{e,m} c_{e,m}^{tim}+\sum_{c,m} c_{c,m}^{tim}+\sum_{p,m} c_{p,m}^{tim})+\sum_{ch} \Delta r_{ch}$	

reputation r_{ch} and the reputation R of the organization's communication channels as a whole. For example, an employee might trust and tend to read an email sent within their own unit compared to an email sent from central units; or an employee might trust *President's Message* more compared to *Brief*.

Message m's chance of being read might also depend on its position in the email. For example, a message in the Top News (the top section of *Brief*) might have a higher probability of being read compared to the messages in the later sections of *Brief*.

Employee e's time cost with message m would be 0 if they do not receive it (if $t_{e,m} = 0, c_{e,m}^{tim} = 0$). If they receive it, they will pay the corresponding time cost $c_{e,m}^{tim}$ given their actions with this message, including the time of interruption, deciding, and reading. The factors that influence the size of $c_{e,m}^{tim}$ are still left to be studied. For example, will employees spend more time on the channels they trust? Will they spend more time on the messages they feel work-relevant or interesting to them, or important to the organization? Will they spend more time reading a single-message email compared to a message in a newsletter? Will they spend more time reading the Top News of a newsletter compared to the later sections? These questions need to be answered for quantifying employee's costs.

For the value employee e receives from reading message m, there are three types of potential value. The first type is the value of message m being work-relevant to employee e ($v_{e,m}^{wr}$). The second type is the value of message m being interesting to employee e ($v_{e,m}^{int}$). The third type is the value of the message m being important-to-organization $v_{e,m}^{imp}$. This value generally seems to be smaller than the value from the organization's perspective ($v_{e,m}^{imp} \leq v_{e,m}^{imp,o}$). As according to our prior interview study in chapter 3, some employees feel that this type of information, such as the messages from the University Senate above, is too high-level to be actionable or relevant to them.

In summary, the final return $g_{e,m}$ the employee receives from that message would be

$$g_{e,m} = t_{e,m}p_{e,m}(v_{e,m}^{wr} + v_{e,m}^{imp} + v_{e,m}^{int}) - c_{e,m}^{tim} \tag{4.1}$$

4.4.2   Information Producer's Value

The information producer would like the relevant employees to know about their messages ($v_{p,m}^{awr} = \sum_e t_{e,m}p_{e,m}(v_{e,m}^{wr,p} + v_{e,m}^{imp,p} + v_{e,m}^{int,p})$). For example, if they are announcing

a museum exhibit, they would like all the employees who might be interested in it to know about the message.

Driven by compliance, the information producer sometimes needs to get the evidence that the required information was disseminated to all the potentially relevant employees ($v_{p,m}^{evd} = \sum_{e \in \text{relevant employees}} t_{e,m}$).

The information producer spends time composing the message and communicating with communicators on their message request. In summary, the information producers' return of a message would be

$$g_{p,m} = v_{p,m}^{evd} + \sum_e t_{e,m} p_{e,m} (v_{e,m}^{wr,p} + v_{e,m}^{imp,p} + v_{e,m}^{int,p}) - c_{p,m}^{tim} \tag{4.2}$$

4.4.3 Communicator's Value and Cost

When communicator c receives a message request, they first decide between sending it versus rejecting it/asking their clients to narrow the recipients or shorten the contents (control point 1). However, they are very likely to receive pushback from their clients if they reject the requests.

If they select to send this message, they have several channels across the dimensions of targeting and grouping (control point 2, see Table 4.1). Generally speaking, with a higher level of targeting, the communicator will spend more time sending message m ($c_{c,m}^{tim}\uparrow$), and it reduces the problem of **oversend** (employees receive less irrelevant messages) and saves employee's time ($c_{e,m}^{tim}\downarrow$). The communicator's time cost may be higher if they want to put the message in a channel under other communicators' controls. For example, the communicator from the President's office needs to spend time communicating with the editor of *Brief* to put the message "*UMD Chancellor Search*" in it.

The communicator is an important control point in the organizational bulk email system. Through communicator c cares about the influences of their actions on channel ch's reputation r_{ch}, they do not have access to this influence, and they need to meet the information producer's requirements ($v_{c,m}^{evd}$). Especially when they have limited information in the databases and targeting tools, the high-targeting choices might bring a problem of **undersend** (miss relevant employees), causing them to fail to get $v_{c,m}^{evd}$. Therefore, from the communicator's return formula with message m (equation 3), they

might tend to get $v_{c,m}^{evd}$ with a minimal time cost by some low-targeting actions.

$$g_{c,m} = v_{c,m}^{evd} - c_{c,m}^{tim} + \Delta r_{ch} \qquad (4.3)$$

4.4.4 Management's Value and Cost

Management is the direct manager/supervisor of the information recipients. They basically share the same value/costs with their employees. They care about whether their employees read and get the information they need to complete organization tasks ($v_{e,m}^{wr,b}$) while controlling the time cost of processing bulk messages ($c_{e,m}^{tim}$). For example, for an employee who often uses voicemails during the weekend, their manager would like them to read the voicemail down notification email to prepare earlier. But managers and employees might disagree on "what is relevant or important to know", which means that $v_{e,m}^{wr,b}$, $v_{e,m}^{int,b}$, $v_{e,m}^{imp,b}$ might not equal to the employee's own assessments $v_{e,m}^{wr}$, $v_{e,m}^{int}$, $v_{e,m}^{imp}$. In our interviews with managers, they indicated that high-level information, like legislation news, what is going on generally in the organization, and leadership updates, could be helpful, and the employees have responsibilities in knowing about these messages. Managers might also disagree with the organization / information producers on these values since they know better about their employees' job responsibilities. All these factors make manager b's return with message m as

$$g_{b,m} = \sum_{e \in E} t_{e,m} p_{e,m} (v_{e,m}^{wr,b} + v_{e,m}^{imp,b} + v_{e,m}^{int,b}) - \sum_{e \in E} c_{e,m}^{tim} \qquad (4.4)$$

4.4.5 Organization's Value and Cost

The organization's value is a weighted sum of all stakeholders' value. We represent the organization's perspective by the weights w_p, w_e, w_c, w_m. The organization may not weigh each w_p, w_e, w_c, w_m the same. For example, they may decide that the President's weight is higher than Vice Presidents, or it cares more about the value of higher-paid employees. An important question left to be studied here is who and how to decide these weights.

The organization's cost is an accumulative sum of all stakeholders' costs. At the same time, the organization would also like to maintain or improve the reputation of its communication channels ($\sum_{ch} r_{ch}$). With these factors, the organization's total return

with all messages is defined as

$$g_o = (\sum_{p,m} w_p v_{p,m} + \sum_{e,m} w_e v_{e,m} + \sum_{c,m} w_c v_{c,m} + \sum_{b,m} w_b v_{b,m}) - (\sum_{e,m} c_{e,m}^{tim} + \sum_{c,m} c_{c,m}^{tim} + \sum_{p,m} c_{p,m}^{tim}) + \sum_{ch} \Delta r_{ch}$$

(4.5)

Equation (4.5) indicates that, though designing and targeting bulk emails is time-consuming for communicators, it might make sense for the organization to encourage a communicator to spend 1000 minutes targeting and designing a message, compared to spending 1 minute direct sending that message to a list of 20, 000 employees, each of whom might spend over 1 minute reading it on average. However, the organization's costs and value are not measured, reflected, or incorporated anywhere in current organizational bulk email systems.

4.5   Case Study

Let's use the example message "*UMD Chancellor Search*" (Figure 4.1) and "*Nominations for Cognition Faculty*" (Figure 4.5) to discuss the scales of the stakeholders' value and costs given different channels.

4.5.1   "*UMD Chancellor Search*"

This message is sent from the President's office to advocate the search for the Chancellor of the Duluth campus. The employees who are interested in the nomination or application (mainly from the Duluth campus) would view this message as important or interesting. For most employees from other campuses, it might be considered irrelevant high-level information. For the information producer (the President), they want 1) all the employees who might be interested in it to receive and read this message (get $v_{p,m}^{awr}$); 2) to show that they value this search. Therefore, they tend to send this message from the central offices. It is worth noting that **"show the importance of a message" might be different from "a message is really important"**. Because if a message is really important, communicators might use an email tree to utilize the local attention to reach more employees. And for the communicator c who receives this request, they will need to select a channel for this message which could show that they send the message to all the potentially relevant employees (get $v_{c,m}^{evd}$).

"Message from the President": If the communicator sends it as a single-message email (*"Message from the President"*), they will definitely get the proof-of-delivery $v_{c,m}^{evd}$ because all the employees across all the campuses will receive this message ($t_{e,m} = 1$).

For the information producer, a single-message email from this channel could also show the importance of this message. They will get a higher value of awareness because whoever opens this single-message email will be exposed to this message ($p_{e,m}$ and $c_{e,m}^{tim}$ might be large). From equations (4.2) and (4.3), the information producer and communicator would get a positive return.

However, the majority of the recipients will not perceive the value of this message ($v_{e,m}^{wr} + v_{e,m}^{imp} + v_{e,m}^{int}$ is small). From equation (4.1), most employees would get a negative return from this message, with a high time cost and low value in general. And the small part of employees interested in this message will get a positive return, as they have a large chance to open and read this email.

For the organization as a whole, from equation (4.5), the return might be negative as most employees receive a negative return from this message. This choice also hurts the reputation of its communication channel — employees might think that *"Message from the President"* sends irrelevant information and stop reading them in the future. The reputation of the organization's communication channels as a whole might also be hurt if the employees then also start to untrust the messages from other university leaders (to be studied).

It is worth noting that the damage to the channel's reputation might affect every stakeholder in the end. Because it would gradually make employees stop reading messages, then the information producer's message will be neglected in the future ($v_{p,m}^{awr}$ will be small), and the employees might also miss the messages that could be interesting or work-relevant to them.

"U of M Brief": The communicator could also put the message in *Brief*, they will also get the proof-of-delivery $v_{c,m}^{evd}$ as *Brief* will be sent to all the employees ($t_{e,m} = 1$). With a minimal cost on designing and targeting, the communicator would get a positive return from equation (4.3).

To show that this message is viewed as important by the organization, the communicator might put the message in Top News. Then the effects of this message would be similar to the effects of putting it in a single-message email.

If the communicator puts the message in the later sections of *Brief*, the employees who are interested in it might miss this message if they skip the later sections. Then the information producer might get less value on employee's awareness ($v_{p,m}^{awr}$ is small). From equation (4.1), with low value and low time cost, most employees who are not interested in this message would get a zero or small negative return. For the organization as a whole, its return would be a small (negative) value — less time cost and less attention. ***Precisely targeted mailing list***: To save employee's time, the communicator could query the database to pull a list of employees from the Duluth campus and directly send a single-message email to them. In this case, the employees outside that campus would not spend time on the message. Then the total time cost of all the employees would be smaller.

However, there is also a risk for the communicator and the information producer. There might be employees from other campuses who are also interested in nominating the UMD's chancellor. They might fail to get the proof-of-delivery if they miss these relevant recipients. Moreover, even with a powerful database, this approach is time-consuming, and communicators may not want to spend days generating a mailing list.

The best approach here might be combining Brief and targeted mailing lists. We could put the message in a Brief to collect the proof-of-delivery, while sending it as a single-email message to the UMD employees to catch their attention. By this approach, the organization might receive a large value with a small cost because the employees who spend much time processing and reading this message would mainly be the potentially relevant employees.

4.5.2 *"Nominations for Cognition Faculty"*

This message is sent from the Office of the Vice President for Research to notify a fellowship for the early-career faculty in the field of cognitive science — only those faculty would receive a value on work-relevance if they read this message. For the information producer, they want all these faculty to be aware of this message. But they won't view this message as of general importance to the organization. Therefore this message should not be sent as a single-message email to all the employees from the central units. For the communicator who receives this message request, their task is to select a channel to guarantee that they deliver the message to all the relevant employees.

Precisely targeted mailing list: The communicator might not be able to get a precisely targeted list for this message if the database does not have a complete record of faculty's research interests. To guarantee that the recipient lists will cover all the relevant employees (get $v_{c,m}^{evd}$), the communicator's best option is to send the message to the list of faculty in several colleges, such as all the faculty in the college of science and engineering, or all the faculty in the college of biological sciences (or the college of liberal arts, medical school, etc.), which would still be a large list. Sending the message as a single-message email might increase its chance of being read by the several early-career cognition faculty. However, this approach would bring time costs to a large list of irrelevant employees, and the organization as a whole. Besides, it will hurt the reputation of this channel. The employees who receive this message might feel the messages from the Office of the Vice President for Research as possibly irrelevant and stop reading them. In summary, the organization's return would be negative given the limited capability of targeting by querying the database.

"UMN Research": The communicator could contact their office's newsletter editor of UMN research, which is sent to all faculty and research-related staff monthly. For the communicator, they could get the proof of delivery within a minimal time cost. However, the information producer and relevant faculty might not get the value of awareness if the relevant faculty are too busy to read this newsletter. In short, this method would get a small value (less attention) with a small cost for the organization. The relevant faculty are affected most if they miss this information.

Email tree: Another way is to distribute the message to several relevant College Deans, such as the Dean of the College of Biological Sciences, the College of Science and Engineering, etc. Then the employees' value and costs depend on how their college distributes this message. If the Dean's office just send-to-all, the employees' time cost would still be high. Suppose the college is small and the Dean knows about the faculty's research interests and years of experience. In that case, they could directly send individual messages to the relevant employees, which are then very likely to be opened and read. Then the information producer and the relevant employees will get the value of awareness, and the organization suffers a minimal cost.

However, the risk for the communicator and information producer is that they can't guarantee that all Deans will notice this request and forward this message — Deans

are busy. There might be relevant faculty being missed. This is why there are relevant people in each college — for example, the communicator could send the message to all the Research Associate Deans to show openness / inclusion.

In summary, we could not find an optimal method to deal with this type of message in the current organizational bulk email system. Our databases do not have enough information to support precisely targeting the recipients. All the current channels might cause a negative return or risk to some stakeholders.

4.6   Opportunity and Future Research Directions

From an economic perspective, we should only send a bulk email when the organization's return with that email is positive from equation (4.5). We should improve the system's mechanism towards maximizing this return. Proposing a model is not the same as instantiating/calibrating it. Therefore we propose several research directions and potential interventions below, which might make organizational bulk email system's economic mechanism more effective.

4.6.1   Preliminary step: operationalize and measure value and cost

Many of the value and costs proposed in the model are not operationalized and measured in the current CRM platforms. The CRM platforms mainly track open rates by inserting an invisible pixel in the email and track click rates by rewriting the links uniquely.[2]   However, awareness, engagement, time, money, and reputation are not recorded and measured anywhere. Operationalizing these effects is necessary for building mathematical models assessing bulk email system performance.

Measuring these effects is also challenging. For example, due to privacy concerns, many browsers do not allow access to the precise time of loading an image and the access to the time gap between loading two images, which makes it difficult to measure the reading time of emails. Testing the bulk messages' awareness with minimal effort from the recipient's side also needs to be solved. After that, calculating money cost would need a model to link reading time and salary.

---

[2]   `https://help.salesforce.com/s/articleView?id=sf.pardot_emails_open_tracking.htm`

4.6.2   Interventions on the sender-side

The motivation of getting the proof-of-delivery ($v_{c,m}^{evd}$) within a minimal time cost ($c_{c,m}^{tim}$) is an important part of the return formulas of communicators and clients. This is why they often select to send-to-all and then employees receive many irrelevant organizational bulk emails. To incentivize communicators and clients to spend more time targeting and designing bulk emails, there are three kinds of interventions from the equations (4.2) and (4.3): reducing the targeting cost (decrease $c_{c,m}^{tim}$), making the change of reputation observable to communicators (measure Δr), and incorporating employees' costs ($\sum_{m,e} I_{e,m} p_{e,m} c_{e,m}^{tim}$) into the return formula.

Reducing the targeting/filtering cost. The databases of the current bulk email systems only provide minimal data like units and job titles. However, this information is not enough to represent employees' preferences. If we can measure open rates, click rates, and reading time, we could predict employees' preferences through collaborative filtering techniques [32], deep learning models [111, 41], weakly supervised learning [112, 40], etc. More importantly, we need to understand the organization's preferences. A possible direction is to learn from the employees' job descriptions. When the organization recruits an employee for a position, it posts a job description to introduce the responsibility of this job. We could build a knowledge management system [174] along with natural language processing models like Bert, XLM [68, 69, 173, 172] to learn which messages would be work-relevant to the corresponding employees with these job descriptions. After we build such models, we could either choose to help communicators better target messages and narrow the range of recipients or implement better filtering on the recipient-side to help them select those relevant messages. For example, we could send a message in several rounds. And in the later rounds, we only send the message to the employees who are similar to those who have clicked the message in the previous rounds.

Making the change of reputation observable to communicators. As we discussed in the model, the outcomes of organizational bulk emails influence themselves, their channels' future emails, other bulk emails, and other channels. Each time we send a message, we could measure its influence on the channels' reputations (e.g., by the difference between the open rates/reading time of this email and the next email from this channel). And when we detect a decrease in reputation, we remind the communicators

and clients that they should put more effort in targeting and designing the email.

Incorporating employees' reading cost into communicators and clients' return. As we discussed above, for the communicators, it is convenient to select "send-to-all" and expensive to select "filter and target". Besides the technical reasons in targeting, selecting "send-to-all" is also caused by the invisibility of the organization's cost. For example, though the communicators need to spend 300 minutes to filter the recipients while "send-to-all" only costs them 3 minutes, "send-to-all" would cost over 10,000 minutes for the whole organization. We need to measure and present the overall cost to the communicators given their actions on "send via newsletter" versus "send via single-message email", "send-to-all" versus "precisely targeting". The cost could be utilized in 3 ways: first, showing the cost of different approaches in the training sessions with communicators on incentivizing them to pay more effort in targeting; second, showing the estimated communication cost to support communicators in selecting communication channels for each message; third, building a budget system to allocate time budgets to departments and their communicators centrally. For example, an employee could at most spend 20 minutes in bulk emails per week, or a department could at most use 5 minutes of each employee per week.

Employing a human-centered approach. The discussion in this chapter indicates the necessity of employing a human-centered approach [61] to the problem of organizational bulk email communication. For example, in the case study of "*UMD Chancellor Search*", we pointed out that for the communicators and information producers, "showing a message is important" is different from "a message is really important". This inconsistency indicates senders' disagreements and misperceptions on the optimal communication channels for the corresponding messages — there is no common criterion for the communicators to understand the value of a message, and which channel they should use. Therefore, besides technical solutions, human-centered approaches are necessary for such social-technical systems, like educating the senders on the bulk email practices they should apply.

4.6.3 Intervention on the recipient-side

From equation (4.1), we could see that employees care more about the messages that are interesting or work-relevant to them but less about the messages that are viewed as

important-to-organization. Then many high-level messages, such as the messages from the University Senate (Figure 1), would not be read. There is a need to coordinate different stakeholders' preferences. At first, we need to study the size of the preference conflicts between the organization and the employees, i.e., the value of μ and λ in equations (4.1) and (4.5) — to how much extent the employees perceive the value of a message that is important to the organization and to how much extent the organization perceives the value of a message that is interesting to an employee. After that, there are three possible mechanisms for incentivizing employees to read the "important" messages. First, we could incentivize employees to read those messages by putting them next to the "interesting messages". Second, we could explain why a bulk message is important to employees by explainable recommendation [190, 34, 33]. Third, we could design performance evaluation systems [119] to encourage employees to read those important messages [3], such as giving badges or rewards to employees if they paid enough attention and sending reminders if they forgot to read.

4.7 Conclusion

We proposed an economic model to describe different stakeholders' actions, value, and costs in organizational bulk email systems. The model shows that, with different goals, none of the stakeholders have the ability to see the global effects of their actions on the system's effectiveness and the organization's costs and value. Based on the model, we identified two main opportunities: 1) design mechanisms for encouraging employees to read the "important" messages (Chapter 5); 2) make the change of reputation and employees' cost visible to senders (Chapters 6 & 7).

Chapter 5

Multi-Objective Personalization of Organizational Bulk Emails

5.1 Introduction

As we found in the interviews (Chapter 3), many bulk messages are viewed as "important-to-organization" information by organization leaders and communicators. They build the awareness of the community and promote the organization's missions. However, normal employees may prefer messages more relevant to their jobs or interests [98]. 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). In this study, we evaluate the use of personalization to help organizations reach these goals. [1]

Figure 5.1, the newsletter U of M Brief of our study site, is an example of the preference mismatch above. 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. For example, the university wants employees to know what its governing board is doing, so it puts Boards of Regents updates as the

---

first message of the Brief with the hope that all the employees who opened this Brief will notice this message (see Figure 5.1). 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 (the value of protecting channel reputation in chapter 4). 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 5.2) is to put **organization-preferred messages** (e.g., board of regents meetings) adjacent 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).

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 the previous studies are most about commercial bulk emails, whose tactical goals and strategic goals are often aligned — these studies usually prioritize the messages that recipients would like, such as the products the recipients might be interested in [148, 180]. Also, we conducted an 8-week field experiment to enable employees to get
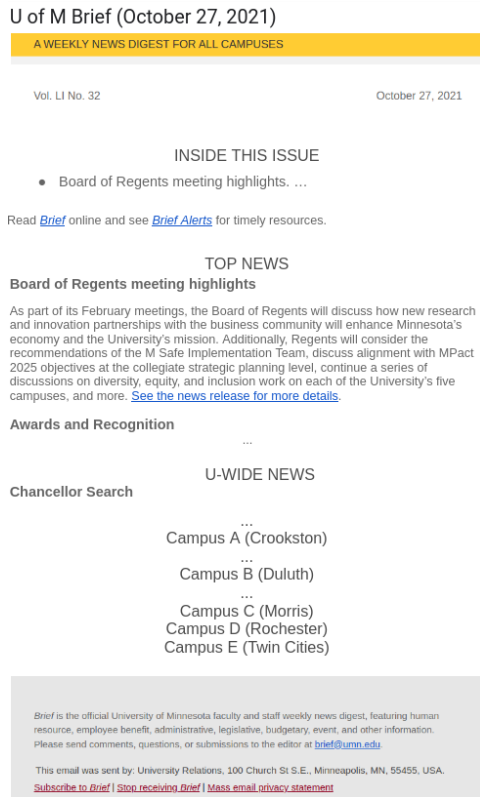
Figure 5.1: A sample Brief on Oct 27, 2021. The messages within each section are hidden in this picture. For a full email of a personalized Brief, please see Appendix B.
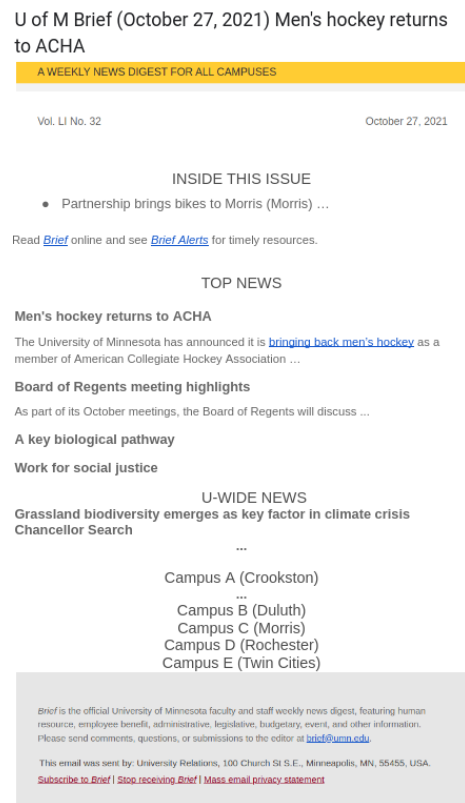
Figure 5.2: A sample personalized Brief. The messages within each section are hidden in this picture.

used to the experimental newsletters, while the research on organizational emails is mainly lab experiments (for example, Park et al. invited participants to their lab to try an email visualization tool [136]), dataset analysis (the datasets often used include Enron [92] and Avocado dataset [186]), and observational field studies (for example, Mark et al. observed how email use influences employee's stress [123]). Regarding the practical advances, we provided 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 chapter 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 (STUDY00012816).

5.2   Related Work & Gaps

5.2.1   Bulk Email Personalization's Data Sources

Personalization has been pointed out as a solution for email overload [28] and several studies used personalization to improve commercial bulk emails' performance. The personalization data sources could be demographic information like names [148, 180], majors, departments [176]; or preference information like browsing history [180]. Though Sahni et al.'s experiments found adding recipients' names to subject lines useful for improving open rate [148], Wattal et al. [180] 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 [176]. 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 [65]. Wattal et al. [180] personalized email content based on customers' purchasing preferences and received positive responses. We personalize based on preferences instead of demographics in this study. We do not want to 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.

5.2.2 Bulk Email Personalization's Design Choices

We reviewed different ways to personalize commercial bulk emails in Chapter 2, including subject line, top section, selection of contents, order of contents, and visual designs. In this chapter, 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) [35]. Bulk email's from lines, signature lines [81], and sending time and day [1, 18] are also found to influence open rate by around 10% in the previous experiments from marketing, communication, and management science.

5.2.3 Gaps in Personalizing Organizational Bulk Emails

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. We introduce a field experiment to bridge this gap.

5.3 Study Overview and Background

This study is aimed to use personalization to help organizations achieve their tactical goals (getting employees to know important messages) and strategic goals (keeping employees interested in the newsletters) of bulk communication. Next, we first introduce

the newsletter we personalize. After that, we introduce the different ways of personalizing the newsletter we tried, and propose our hypothesis on these designs' influence on the communication goals. Then in the method section, we introduce the 8-week field experiment we carried out to evaluate these hypotheses, including the participant recruiting process, how we calculate the employees' and organizations' preferences with the newsletter, and how we implement the different personalization mechanisms and measure employees' feedback. Figure 5.3 is the outline of this experiment.

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

Experiment System

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

	$employee_i$ 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
...				
$message_k$	...	...		...

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

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
...				
$message_k$	...	...		...

$Brief\ editor$

Actions of Server

Actions of Employees

Actions of Brief Editor

Figure 5.3: The outline of the experiment procedure. Weeks 2 to 7's steps were repeated weekly. We sent the original Brief in week 2.

5.3.1   Studied Newsletter

Before the study, we met with three communicators from the central offices of the study site and decided to experiment with the newsletter U of M Brief (see figure 5.1), which is sent to all the employees weekly (with subject line *"U of M Brief [Date]"*). Each brief contains around 30 messages and 7 sections. Regarding the organization's preference for messages, Brief encourages people to submit information about:

- "need-to-know" administrative news

- messages that aimed to make the university more accessible and that create connections among faculty and staff

- promoting healthy lives

- the university's mission of outreach, research, teaching, and education

5.3.2   Designs

We considered several personalization designs, including original/random/employee-preferred/organization-preferred/mixed designs. It is worth noting that we focus on personalizing the order of messages in a newsletter in this chapter, because we are asked to send out all the messages. There are other personalization designs we have not tried, such as filtering out irrelevant messages, highlighting important messages, personalizing the time of sending messages, etc.

Let us use Figure 5.1'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 from the newsletter, 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 (see 5.4.2 for the definition of organization/employee's preference), 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 5.1.

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. B3's messages might be different from B1's messages as B3 is personalized — the organization might set different priorities for different employees.

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 (see 5.4.2 for the definition of organization / employee'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 6.4.2), 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 5.2 is a sample personalized Brief for this faculty member if we assigned them to A4 x B5 x C5.

5.3.3 Research Questions and Hypotheses

We propose 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 compared to the messages not on subject lines:

(**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 [91, 88]:

(**H1.6**) Putting employee-preferred messages on subject lines will increase their read-in-detail rates compared to the messages not on subject lines.

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 propose these questions:

What is the effect on the newsletter's reading time (**Q2.2**) and overall recognition rate (**Q2.3** when we

- put organization-preferred messages in top news

- put employee-preferred messages in top news

- mix employee-preferred messages and organization-preferred messages in top news

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 propose 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 hypothesize that sorting by employee's preference would make them feel the Brief is more interesting as they will see interesting messages in advance:

(**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 **??**. 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.

Table 5.1: 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. There were blanks (to be observed) in this table when we did not find any related work or reasons to make a guess on a significant effect. NA: not applicable.

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	NA	To be observed	To be observed
	2: Put random messages	To be observed	To be observed	To be observed	NA	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?	NA	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?	NA	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?	NA	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	NA	To be observed	To be observed
	2: Random order	To be observed	To be observed	To be observed	NA	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?	NA	NA	NA
	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?	NA	NA	NA
	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?	NA	Q3.3 How does it affect recognition rate of org-pref messages?	NA

5.4   Study Procedure

5.4.1   Overview

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 5.3):

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.

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)

Figure 5.4: 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.

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.

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.

5.4.2 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 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.[2]   The newsletters only had their brands without specific messages on their subject lines which could bias the recruitment.[3]

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 [59]. 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 likely 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 collected their campuses and job categories. To make our surveys more concise, 20 job families from 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

---

[2] 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).

[3] The messages closest to our recruitment message in that recruitment Brief had topics on Administrative News/Student Stories/Awards and Recognition. According to Table 5.5, our participants were not more interested in these topics compared to other topics.

(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 5.4). 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.

5.4.3   Methods

To define and collect preference data within limited survey questions, we used "topic" as a bridge to connect messages and preferences (we limit each message to 4 topics). 20 topics were summarized by a thematic analysis [21] of 140 messages from 5 previous Briefs (see Appendix B). 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 5.2.

Collecting and calculating employees' preferences (step 2, week 1)

At the setup of this study, we asked the participants to fill out a preference survey (figure 5.3 - 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

Table 5.2: 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 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 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 to 1)
$topic\_list_k$	$message_k$'s relevant topics (list, up to 4, Appendix A)		

defined as:

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

$$(5.1)$$

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

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

$$(5.2)$$

Third, $employee_i$'s preference on $message_k$ is defined as the average of interest and relevance score, as belo. It is worth noting that max might also work here, but the author also tried using the maximum of interest and relevance score but found the selected messages less interesting and therefore decided to use average here.

$$emp\_pref_{i,k} = (emp\_interest_{i,k} + emp\_relevance_{i,k})/2 \tag{5.3}$$

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 5.3, 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 — the messages in the top news and U-Wide news are targeted at all campuses.

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:

$$org\_relevance_{i,k} = 1, if\ campus_i \in target\_campus_k\ and\ job_i \in target\_job_k;\ 0, otherwise \quad (5.4)$$

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:

$$org\_importance_k = (important_k - 1)/3 \quad (5.5)$$

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

$$org\_pref_{i,k} = (org\_relevance_{i,k} + org\_importance_k)/2 \quad (5.6)$$

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 5.3, week 2 to 7 (6)).

5.4.4   Collecting Metrics (step 5, week 8)

At week 8, we sent out the end survey (Figure 5.3, 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.

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. 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) [49]; and employees might also pay more attention to bulk emails when being recorded by camera.

⋆ 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).

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 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 [158], 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 [85], 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 encouraging participants to pay more attention to read/remember these Briefs, we only analyze the performance data collected in the last week of this experiment.

5.5 Results

5.5.1 Analysis and Overview

We summarize the performance of each experimental group in Table 5.3 and the results on the hypotheses and questions in Table 5.4. 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 [160]. We had a random effect based on subjects (from which employee we collected this data point) [11], and we selected likelihood-ratio tests because we had many levels on the random effect (number of participants) [12]. 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 to 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 [86] 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 [63]). 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 [146].

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)

Table 5.3: **The performance metrics of each experimental group**, and the standard deviations of that group's participants' metrics. The blank cells are not applicable or not of interest. non-s: messages not on subject lines. non-t: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 5.4: Results of hypotheses and research questions. Format: experimental group mean = control group and its mean + difference between experimental and control groups (p.val). 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 [71, 171, 170].

		Strategic Goal				Tactical Goal	
Group	Treatment	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	

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 5.4 for the effects, treatments, and control groups we examined). The P-values of the pairwise tests were adjusted by the holm-bonferroni method [71]. We got the following marginal / significant results on the hypotheses and questions (see Table 5.4 for the numbers):

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

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.

5.5.2 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 5.5. 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 5.7, 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."*

Figure 5.5: 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.

Figure 5.6: Brief's reading time versus its recognition rate.

Figure 5.7: 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.
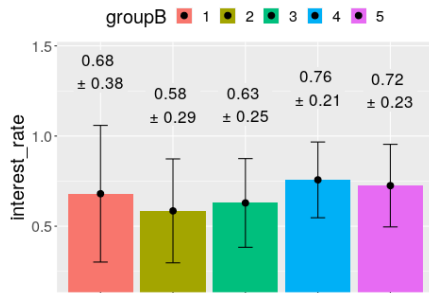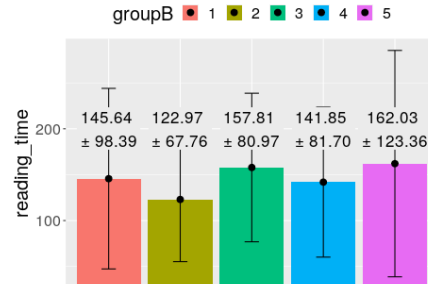

Figure 5.8: The reading time (seconds) on personalized Briefs of each experimental group on top news and the standard deviations of the participants' reading time in that group.
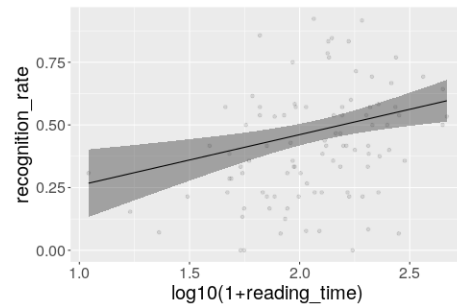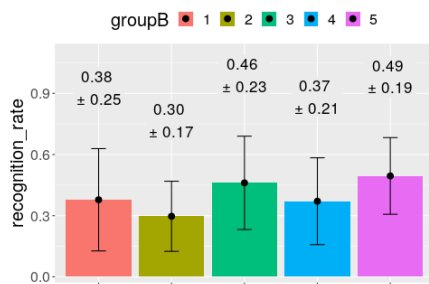
Reading Time. There were no significant effects of subject lines/top news/message order designs on reading time. On average, the Briefs were read for around 120 to 170 seconds, and the variation of reading time is large. However, we found that the patterns of reading time matched with the patterns of overall recognition rates (see 5.7 and 5.8). We plot reading time versus overall recognition rate in Figure 5.6. 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".

5.5.3 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 5.9). 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 5.11). 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.

Figure 5.9: 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.

Figure 5.10: 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 5.10). 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 5.10, 5.12). 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.

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.
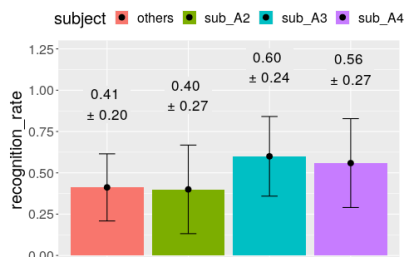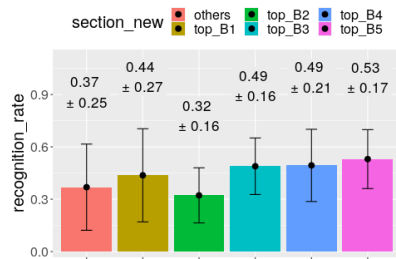
Figure 5.11: The recognition rates of the messages in top news with respect to group B and non-top messages and the standard deviations of the participants' recognition rates in that group.
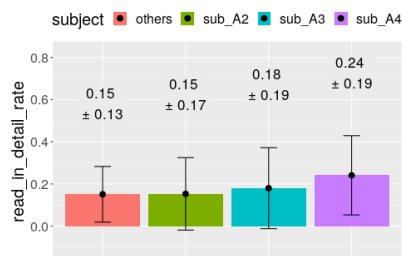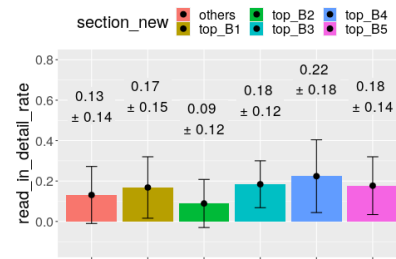
Figure 5.12: The read-in-detail rates of the messages in top news with respect to group B and non-top messages and the standard deviations of the participants' read-in-detail rates in that group.

5.5.4 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 5.5 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 5.5. 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 5.5. 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.

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

Table 5.5: 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%

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).

5.6 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 [132, 128, 136] 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 judgments 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.

5.6.1 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 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 [132].

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 [103, 144] (however Kraut et al. found that recipients still did not interpret the prices as emails' importance), indicating expected actions [5, 143], etc.

5.6.2   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 [123]. In that sense, organizations should also try to remove unnecessary messages and use personalization to put important/relevant messages upfront.

5.6.3   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. However, newsletters can also be personalized by filtering a subset of relevant messages and this method might work in organizations that allow taking this mechanism.

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. [4]

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,

---

[4]   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), we achieve a precision of 66% and a recall of 75%.

to how much extent the campus structure influenced the personalization's performance is still left to be studied.

5.6.4   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. It would be useful to run a study with eyetracking devices to collect such reading data, and to develop estimation algorithms 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.

3) Studying the effectiveness of fine-grained personalization models, enabling employees to update their preferences, exploring tools could better target the recipients (allowing excluding some messages), etc.

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

5) Auto-profiling employees: learning employees' job responsibilities and interests through their job descriptions, therefore we don't need to use surveys in the future.

5.7   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.

Besides encouraging employees to read important messages, we would also like to help senders consider each message's performance and cost. In the following chapter (6 & 7), we introduce a system design study on the evaluation platforms of organizational bulk emails.

Chapter 6

Message-Level Reading Estimation

6.1 Introduction

From Chapter 3's interviews, we found that to prevent communicators from overwhelming employees, we need to encourage communicators to consider employees' interests and the cost of bulk messages. To enable such consideration of interests and costs, in chapter 7, we would like to measure employees' time spent on bulk messages and show them to the commnicators. Chapter 6 serves as a prior study for Chapter 7. In this chapter, we learn how to estimate whether **each user** (e.g., an employee who receives a bulk email) read in detail, skimmed, or skipped each message (**read level**) by collecting necessary ground truth data and building estimation models. [1]

Making the estimations above is difficult because digital newsletters (such as the one shown in Fig 6.1) often include a variety of different **messages** within them, and how to estimate employees' time spent on each message is still left to be learned. Eye-tracking [54, 51, 135] would provide a good solution, but users lack the hardware (and willingness) to provide that data.

Instead, we want to rely on browser instrumentation to track time, scrolling, mouse

---

clicks, etc. However, there is no existing dataset on users web interactions and reading behaviors that we can use to build reading estimation models. In this chapter, we show how collecting a modest set of eye-tracking data and using it to train machine learning models for estimating reading behavior from browser logs can significantly increase the accuracy of classifying **message-level** reading behaviors.

Our core research question is **how to estimate message-level reading time and read level based on each user's logged interactions with browsers**. We looked progressively at heuristics, logistic regression, neural network and its variants to identify the best estimation model. For per message in a newsletter, we measured the models' error of estimating reading time and accuracy of classifying read-level category (skip/skim/read-in-detail), see Table 6.1.

Table 6.1: The Metrics' Definitions. Precision/recall is based on the specific skim/detail/read category. We evaluated various models' performance on the users they had not seen in training by cross validation.

Metric	Definition
percentage error of reading time (per_error)	$per\_error = |true\ reading\ time - estimated\ reading\ time|/true\ reading\ time$ when $true\ reading\ time \geq 10s$
absolute error of reading time (abs_error)	$abs\_error = |true\ reading\ time - estimated\ reading\ time|$ when $true\ reading\ time < 10s$
accuracy of read-level (accuracy)	$accuracy = \sum_{m \in \{messages\}} I(estimated\ category = true\ category)/|\{messages\}|$ $\{messages\}$ are all the messages from all the sessions.
precision of read-level (skim_precision, detail_precision, read_precision)	e.g., $skim\_precision = \frac{\sum_{m \in \{messages\}} I(true\ category=skim)I(estimated\ category=skim)}{\sum_{m \in \{messages\}} I(estimated\ category=skim)}$
recall of read-level (skim_recall, detail_recall, read_recall)	e.g., $skim\_recall = \frac{\sum_{m \in \{messages\}} I(true\ category=skim)I(estimated\ category=skim)}{\sum_{m \in \{messages\}} I(true\ category=skim)}$

The dataset (200k datapoints) is available at `https://github.com/ruoyankong/reading_region_prediction_dataset`. The proposed approach enables us to make relatively accurate (27% error in reading time) message-level reading estimations of bulk emails solely by user web interactions in Chapter 7.
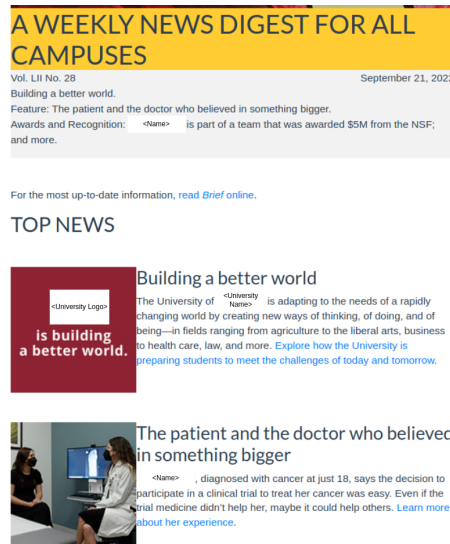
Figure 6.1: An example e-newsletter page of the study site with 30 messages. By message-level reading estimation, we estimate each message's (like "Building a better world") reading time and read level for each user.

6.2 Related Work

6.2.1 Reading Estimation Based on Eye-tracking Technology

Eye-tracking technology refers to the applications, devices, and algorithms that use user's eye movement to catch their attention trace [106]. It has been widely used in information processing tasks, including reading estimation of advertisement [142], literature [9], social media [131], recommendations [192], etc.

Remote eye-trackers and mobile eye-trackers are often used in reading behavior estimation. The remote eye-trackers use cameras, pupil center, and cornea reflection to track user's gaze position [20]. Example eye-trackers in this category include Tobii-Pro-TX300, which contains separate eye-tracking units, [2] or GazeRecorder, which simply uses Webcam.[3] The mobile eye-trackers are often head-mounted devices that users need to wear and use a camera on the visual path to record the view, such as Tobii Pro

---

[2] https://www.tobii.com/products/discontinued/tobii-pro-tx300
[3] https://gazerecorder.com/

Glasses.[4]

Though eyetrackers provide relatively accurate data on reading estimation, they either ask users to wear/use specific devices or allow access to their webcams. This accesse is hard to get on a large scale for platforms, most of which only have access to browser interaction data. **Therefore there is a need to build message-level reading estimation models based on available browser data**.

6.2.2   Eyetracking Data in Machine Learning

Beyond reading estimation, eye-tracking data has been used widely in training machine learning models for different goals. Zhao et al. predicted users' gaze on movie recommendations by eye tracking data through hidden markov models and reached an AUC of 0.823 [192]. Glaholt and Reingold studied participants' decision-making processes with eye-tracking data, and found that participants' preference on art images is positively correlated with their time spent gazing at corresponding types of images [56]. Buscher et al. used eyetracking data to generate a model for predicting the visual attention that individual page elements may receive [23]. All these studies show the potential of using eyetracking data to generate scalable machine-learning models for predicting user behaviors / preferences.

6.2.3   Reading Estimation Based on User Interactions

Several studies have been done in estimating reading behaviors (e.g., time, focus area) based on user interactions. Seifert et al. studied which regions of the web pages users focus on when they read Wikipedia and found using single features (mouse position, paragraph position, mouse activity) cannot reach enough accuracy in estimating reading regions [152]. Huang et al. and Chen et al. built linear models based on cursor position to estimate user's gaze coordinates during the web search / browsing [74, 30]. Huang et al. and Liu et al. used cursor position in ranking search results [75, 113] and found that 60% of the tested queries were benefited by incorporating cursor data or significantly improve clicking prediction accuracy by 5%. Arapakis et al. recorded how users respond to online news and found the distance between mouse cursor and the corresponding news

---

[4] https://www.tobii.com/products/eye-trackers/wearables/tobii-pro-glasses-3

is significantly correlated with user's interestingness [7].

However, most of these models did not consider user patterns. For example, the mouse position feature was treated the same for a user who moves mouse frequently versus a user who moves mouse infrequently [105]. Hauger et al.'s work proposes to adjust reading estimations with user behavioral patterns — but only a linear weight model was evaluated on the ground truth [64]. In fact, most of these studies focused on learning statistical correlations or only evaluated the proposed model's performance on groundtruth / heuristics. **We therefore found a need to evaluate various features and models' performance on reading estimation systematically**.

6.3   Methods

We defined **"reading estimation"** as 1) estimating reading time: how much time a user spends on reading each message per session (e.g., a user reads a message for 5 seconds); 2) classifying read level: whether a user skipped or skimmed a message, or read it in detail. Below, we introduced how we built features and models.

6.3.1   Features

To consider which interaction features a model solely based on user interactions can use, we identified the data a browser can collect in a reading session (from open to close a page). They can be classified into a 2x4 category: temporary/sessional features x pattern / user / message / baseline features (see Table 6.2).

Temporary features are collected per timestamp (each second for our case, e.g., the mouse position at time t). **Sessional features** are summaries of a reading session (e.g., the total number of seconds a message is visible on window in a reading session). **User features** represent user's status (e.g., mouse position). **Message features** represent message's status (e.g., message's position). **Pattern features** represent user's real-time behavioral patterns (e.g., mouse moving frequency until the latest timestamp). **Baseline features** represent the estimation given by the baselines — the heuristics found to be correlated in the previous literature [24, 152, 74] (e.g., the share of a message on window). See the supplementary material for feature definitions.

Table 6.2: The Definitions of Features in Temporary / Sessional Message / User / Pattern / Baseline Category

Feature Category	Temporary (collected per second)	Sessional features (collected per reading session)
Message (represent message's status)	the message's position on the window, the share of the window at that time--stamp; time gap to the latest time that the message is clicked by the user.	the message's average share of the window during the session, average position on window during the session; whether the user clicked the message during the session; the number of seconds the message is visible during the session.
User (represent user's status)	the user's mouse position; the time gap to the user's latest click.	not applicable
Pattern (represent user's behavioral patterns)	the user's mouse moving frequency in the past 2/5/10/infinite (since the beginning of the test) seconds in the horizontal/vertical direction, mouse scrolling frequency in the past 2/5/10 /infinite seconds; the percentage of messages in that newsletter the user clicked until that timestamp.	the user's average mouse moving frequency during the reading session in the horizontal/vertical direction, the average mouse scrolling frequency during the reading session; the percentage of messages in that newsletter the user clicked during the reading session.
Baseline (the heuristics found correlated in previous literature)	the probability of the message being read at that timestamp according to baselines (see 3.3.1).	baselines' estimations on a message's reading time of the session.

6.3.2 Models

We considered a list of reading estimation models that use the interaction features above (Figure 6.2). We explored models from simple heuristics baselines, and logistic regression, to neural networks. We also explored per-timestamp models (make estimations per second and sum the estimations at the end of the sessions) and per-session models (make estimations per session).

Per-Timestamp Models

Per-timestamp models estimate the probability $p_{m,t}$ of a message m being read at each timestamp t. We summarize per-timestamp models' estimations for each message at the end of each reading session to estimate the reading time and read level. Previous work found that message size, message position, and mouse position are usually found to be correlated with reading behavior [24, 152, 74]. Thus we started from 3 baselines:
Baseline 1: $p_{m,t}$ = message m's current share of window area.

Baseline 2: $p_{m,t}$ is weighted by the reverse of m's distance to the window center.

Baseline 3: $p_{m,t} = 1$ if m is closest to user mouse.

We started with simple logistic regression:
Logistic Regression Model: it takes temporary message/user features as inputs, $p_{m,t}$ as outputs.

After that, we considered neural networks to learn more complex decision boundaries.

Baseline-based NN: A fully connected feed-forward neural network that takes temporary baseline features as inputs and $p_{m,t}$ as outputs.

Then we tried using pattern features to learn which neurons' outputs are important.

Pattern+ Baseline-based NN: A two-tower feed-forward NN that takes baseline features in one tower and pattern features in another, then merges these two towers (multiply), and finally outputs $p_{m,t}$.

We also tried inputting user temporary interaction features directly as below.

Neural Network (NN): A fully connected feed-forward NN which takes temporary message / user features.

Pattern+ NN: A two-tower feed-forward NN that takes temporary message/user features in one tower and pattern features in another, then merges these two towers (multiply), and finally outputs $p_{m,t}$.
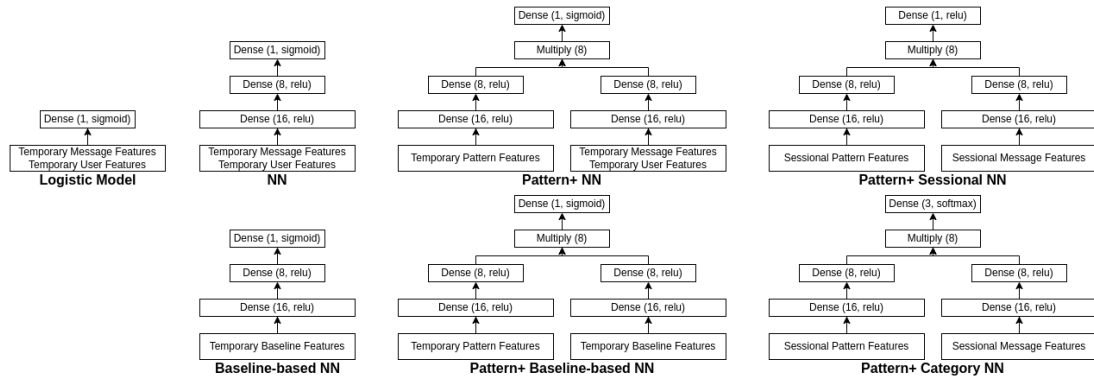
Figure 6.2: The estimation models and their configurations. Logistic / NN / Pattern+ NN / Baselin-based NN / Pattern+ Baseline-based NN output $p_{m,t}$ per second to optimize binary loss. Pattern+ Sessional NN outputs $time_m$ per session to optimize absolute error loss. Pattern+ Category NN outputs $read\_level_m$ per session to optimize crossentropy loss.

Read Level: The reading time $time_m$ of a message m in a reading session t_1 to t_2 is defined as $time_m = \sum_{t_1 \leq t \leq t_2} p_{m,t}$. Given message m's number of words ($word_m$), its $read\_level_m$ **(skip/skim/read-in-detail)** for a user is separated by the reading speed 400 words / minute and 200 words / minute. These limits are selected based on the 70% and 80% comprehension levels in [133]. And **read** means a message is either skimmed or read-in-detail by a user.

Per-Session Models

Instead of making one estimation at each timestamp (e.g., predicting whether the user is reading a message per second and summing the predictions of a reading session as the predicted reading time of that message), we can also make one estimation per reading session (e.g., directly estimating a user's reading time on that message).

Pattern+ Sessional NN: A two-tower NN that takes sessional message features in one

tower and sessional pattern features in another, then merges these two towers (multiply), and finally outputs reading time $time_m$ in a fully-connected ReLu layer.

Pattern+ Category NN: A two-tower NN that takes sessional message features in one tower and sessional pattern features in another, then merges these two towers (multiply), and finally outputs $read\_level_m$ in a fully-connected softmax layer.

We explore with the following questions:

Q1: Does logistic regression perform better than the heuristic baselines?

Q2: Do neural network models based on user interactions perform better than the logistic regression model?

Q3: Does incorporating user patterns improve neural network models' performance on?

Q4: Are there performance differences between the neural network models that take the user / message features as input with those that take baseline predictions as input?

Q5: Are there performance differences between the models which make one estimation per timestamp with the models which make one estimation per reading session? [5]

6.3.3 Eye-tracking Tests

We conducted eye-tracking tests to collect interaction data and ground-truth labels. Participants were recruited through 5 mailing lists of the university employees / in-person contacts of the research team [6]. We excluded the university's communication professionals, as they might have processed the test messages. The selected participants were invited to the university's usability lab.

Each participant then received 8 e-newsletters sampled from a diverse set of senders. The e-newsletters in the pool were selected randomly from the university-wide e-newsletters sent in 2022, containing 3 to 30 messages. During the test, the participants' gaze positions were tracked by Tobii-TX300. They were asked to read as naturally as possible and read at least 1 message of each e-newsletter in detail. They can move mouse, scroll pages, and click on any hyperlinks. All these interactions were caught by browser Javascript and sent to the backend database (Google Firebase) in real-time. They can spend at most 30 minutes reading these e-newsletters and can leave anytime if they

---

[5] We defined a reading session as the duration between a user opening a newsletter and closing it or going to other irrelevant tabs (which are not hyperlinks in the newsletter).

[6] This study was determined to be non-human subject research by the university's IRB

finish reading. The research team labeled the message located at the participants' gaze positions per second and calculated the reading time of each message as the ground truth.

After each eye-tracking test, we retrained our models to check whether we got improvements in model performance. We stopped recruiting when we did not observe further improvements in the models' performance (see section 4). We conducted 9 eye-tracking tests finally, which resulted in a 200k dataset (one data point for each message's reading status per second).

6.4 Results

After collecting and labeling the data, we trained the models and compared their performance. We used the Adam optimizer, a batch size of 64, 50 epochs with early stop. We sampled the split of train/validation/test set for 72 rounds — for each round, one participant's data were used as test set, and the rest participant's data were split into train set and validation set with a ratio of 7:1 based on reading sessions. The weight of positive samples is set as 20, given that each newsletter contains approximately 20 messages on average, and the user is reading one message at any time.

The average performance of all the proposed models are shown in Table 6.3, and the pair-wise t-tests' results between these models are shown in Table 6.4.

6.4.1 What can we do without eyetracking data

Without models trained by eyetracking data, the heuristic baselines have around 43-46% percentage error and 2.1 to 2.5 seconds absolute error in reading time estimation. At the same time, these baselines reach 66% to 69% precision and 48% to 59% recall on read level classification.

The baseline1 (estimated by the share of screen size) has the lowest error among baselines (43% percentage error and 2.1 seconds absolute error). It can be seen that heuristic baselines are not enough for us to provide an accurate metric on reading time (for Chapter 7). Below we show that with eyetracking data, we can improve the performance of estimations to a reasonable level.

6.4.2 What can we do with eyetracking data

In summary, **we can reach approximately 27% percentage error in reading time estimation with models based on user interactions and eyetracking data**, compared to the ground truth of using an eyetracker, while the heuristic baselines reach around 43-46% percentage error. Specifically, with Pattern+ NN, which uses user interaction features as input and adjusts their weights with user behavioral patterns, for reading time estimation, we reached 27% in percentage error and 1.7 seconds in absolute error; for read level classification, we reached 79% in read category precision and 68% in read category recall.

Also, we found that per-timestamp neural network models (Baseline-based NN, Pattern+ Baseline-based NN, NN, Pattern+ NN) have relatively higher overall performance compared to other models. Pattern+ Category NN's accuracy is only 47%, which might indicate that making one estimation on the read level category per session does not utilize information efficiently.

We then compared the relative pairs of models' performance according to questions Q1 to Q5 (see section 1). The comparison results are shown in Table 6.4. We summarized our answers below.

Q1: Do machine learning models perform better than the heuristic baselines?

Yes, the logistic model and neural network models perform better than the heuristic baselines on most of the numerical and classification metrics. For example, compared to baseline 1 (heuristic estimation based on window share), NN significantly improved per_error from 43% to 27%, read_precision from 69% to 81%, read_recall from 50% to 68%, etc. Though the logistic model does not perform better than baseline1 on abs_error, the NN model performs better than the baselines on all the metrics.

Q2: Do neural network models perform better than the logistic model?

Yes, the neural network model performs better than the logistic model on almost all the numerical and classification metrics (similar performance on detail_precision). For example, compared to the logistic model, NN significantly improves per_error from 38% to 27%, abs_error from 2.5s to 1.7s, accuracy from 82% to 87%, etc.

Q3: Do neural network models adjusted by user patterns perform better than the single-tower neural network models?

Table 6.3: The average performance of all the proposed models on the testsets (the users in testsets are not in the corresponding train / validation sets). Read: the message is either skimmed or read-in-detail by the user.

model	per_ error (%)	abs_ error (s)	accu- racy (%)	skim_ precision (%)	skim_ recall (%)	detail_ precision (%)	detail_ recall (%)	read_ precision (%)	read_ recall (%)
Baseline1	43%	2.1s	83%	47%	32%	39%	27%	69%	50%
Baseline2	45%	2.5s	81%	36%	24%	42%	31%	67%	48%
Baseline3	46%	2.3s	83%	49%	43%	33%	40%	66%	59%
Logistic Model	38%	2.5s	82%	45%	42%	64%	51%	64%	58%
Baseline-based NN	30%	1.8s	86%	59%	51%	67%	65%	77%	68%
Pattern+ Baseline-based NN	28%	1.7s	87%	63%	55%	69%	64%	79%	70%
NN	27%	1.7s	87%	65%	54%	71%	63%	81%	68%
Pattern+ NN	27%	1.7s	87%	64%	55%	71%	62%	79%	68%
Pattern+ Category NN	\	\	47%	22%	69%	34%	3%	32%	72%
Pattern+ Sessional NN	61%	3.1s	78%	36%	26%	30%	10%	57%	34%

Table 6.4: Model performance comparisons. Each cell is model1's metric value -> model2's metric value (pvalue). '*': pvalue ≤ 0.05; '.': pvalue ≤ 0.10. The pvalues were adjusted by the holm-sidak method [26].

que-stion	model1	model2	per_ error (%)	abs_ error (s)	accu-racy (%)	skim_ precision (%)	skim_ recall (%)	detail_ precision (%)	detail_ recall (%)	read_ precision (%)	read_ recall (%)
Q1	Baseline1	Logistic Model	43->38 (0.00*)	2.1->2.5 (0.00*)	83->82 (0.50)	47->45 (0.53)	32->42 (0.00*)	39->64 (0.00*)	27->51 (0.00*)	69->64 (0.32)	50->58 (0.00*)
	Baseline2	Logistic Model	45->38 (0.00*)	2.5->2.5 (0.62)	81->82 (0.28)	36->45 (0.06.)	24->42 (0.00*)	42->64 (0.00*)	31->51 (0.00*)	67->64 (0.62)	48->58 (0.00*)
	Baseline3	Logistic Model	46->38 (0.00*)	2.3->2.5 (0.23)	83->82 (0.60)	49->45 (0.53)	43->42 (0.95)	33->64 (0.00*)	40->51 (0.40)	66->64 (0.95)	59->58 (0.95)
	Baseline1	NN	43->27 (0.00*)	2.1->1.7 (0.00*)	83->87 (0.00*)	47->65 (0.00*)	32->54 (0.00*)	39->71 (0.00*)	27->63 (0.00*)	69->81 (0.00*)	50->68 (0.00*)
Q2	Logistic Model	NN	38->27 (0.00*)	2.5->1.7 (0.00*)	82->87 (0.00*)	45->65 (0.00*)	42->54 (0.00*)	64->71 (0.01*)	51->63 (0.00*)	64->81 (0.00*)	58->68 (0.00*)
Q3	NN	Pattern+ NN	27->27 (0.99)	1.7->1.7 (0.76)	87->87 (0.97)	65->64 (0.98)	54->55 (0.97)	71->71 (0.99)	63->62 (0.99)	81->79 (0.48)	68->68 (0.99)
	Baseline-based NN	Pattern+ Baseline-based NN	30->28 (0.00*)	1.8->1.7 (0.01*)	86->87 (0.17)	59->63 (0.12)	51->55 (0.10.)	67->69 (0.24)	65->64 (0.68)	77->79 (0.24)	68->70 (0.24)
Q4	Pattern+ NN	Pattern+ Baseline-based NN	27->28 (0.86)	1.7->1.7 (0.92)	87->87 (0.73)	64->63 (0.86)	55->55 (0.92)	71->69 (0.86)	62->64 (0.92)	79->79 (0.92)	68->70 (0.66)
Q5	Pattern+ Sessional NN	Pattern+ NN	61->27 (0.00*)	3.1->1.7 (0.00*)	78->87 (0.00*)	36->65 (0.00*)	26->55 (0.00*)	30->71 (0.00*)	10->62 (0.00*)	57->79 (0.00*)	34->68 (0.00*)
	Pattern+ Category NN	Pattern+ NN	\	\	47->87 (0.00*)	22->65 (0.00*)	69->55 (0.01*)	34->71 (0.01*)	3->62 (0.00*)	32->79 (0.00*)	72->68 (0.45)

Yes for the NN that takes baselines as input but not for the NN that takes message/user features as input. The baseline neural network adjusted by user pattern features further improves the baseline neural network on the absolute error and skim prediction (marginally) significantly: per_error (30% to 28%), skim_recall (51 to 55%). There is no significant performance difference between NN and Pattern+ NN.

Q4: Are there performance differences between the neural network models that take the user / message features as input with those that take baseline features as input?

No, the pattern+ neural network that takes baseline features as input performs similarly to the pattern+ NN that takes user / message features as input.

Q5: Are there performance differences between the models which make one estimation per timestamp with the models which make one estimation per reading session?

Yes, the models that make one estimation for each reading session perform significantly worse than the models that make one estimation each second (see Table 6.4). For example, Pattern+ Sessional NN's per_error is 61% compared to Pattern+ NN's 27%, accuracy is 78% versus Pattern+ NN's 87%.

Model Robustness: Here we report the trends of our models' performance during the data collection process. We recruited participants in 4 rounds. After each round, we retrained the models and evaluated their performance with the training/testing process above. Table 6.5 is the performance trend of the Pattern+ Baseline-based NN model. In these 4 rounds, we see improvements mainly in model per_error (36% to 28%), abs_error (2.1s to 1.7s). We stopped when we did not observe further improvements on per_error and abs_error.

Table 6.5: The performance trend of the Pattern+ Baseline-based NN model (its average performance in the train/test splits after each round of the data collection process).

round	#parti-cipants	per_error (%)	abs_error (s)	accu-racy (%)	skim_precision (%)	skim_recall (%)	detail_precision (%)	detail_recall (%)	read_precision (%)	read_recall (%)
1	2	36%	2.1s	87%	55%	54%	65%	65%	70%	67%
2	5	30%	2.0s	84%	61%	52%	39%	38%	74%	63%
3	7	28%	1.7s	86%	67%	52%	65%	67%	80%	66%
4	9	28%	1.7s	87%	63%	55%	69%	64%	79%	70%

6.5   Discussion

6.5.1   Per-Timestamp and Per-Session Models.

We found that the models which make one estimation per second (and summarize the estimations at the end of each session) perform significantly better than the models which make one estimation per session. The reason that per-timestamp models outperform per-session models might be that the per-session models fail to utilize a large part of the information that can be collected per timestamp. In the previous work on learning user's reading interests (either through duration data collected by eye-trackers or interaction data), the estimations are often made per session instead of per timestamp [7, 9, 113, 110]. **Therefore we suggested future work on reading estimation tasks to evaluate both the per-timestamp and per-session models.** Meanwhile, as per-timestamp models need more computing resources compared to per-session models, future studies also need to consider the tradeoffs between model performance and complexity [155, 118], online learning [116, 117], on-device learning [108, 76], etc.

6.5.2   Improve accuracy by more features and more data.

We also found that users' behavioral pattern features helped the Baseline-based neural network to further decrease its error on reading time estimation. This result shows that user contexts could be used in improving reading estimation. Besides user behavioral patterns, other potential contexts to be considered included user's intent (e.g., search queries [75]), user's preference (e.g., user's genre preference in a movie recommendation system [192, 191], user's friends [179]), item's features (e.g., product ratings [154]).

6.5.3   Limitations and Future Work

We only collected eye-tracking data on 9 users therefore the dataset we collected might not catch enough variance on user patterns. Future work should look at the data requirement and where the value of additional users starts to decay significantly [120, 115].

6.6   Conclusion

We studied how to estimate message-level reading time and read level of digital newsletters using user interactions extracted by browser JavaScript (e.g., mouse scroll, click, and hover actions). We applied the accurate but hard-to-scale eye-tracking data to learn the association between user interactions and reading time. We conducted 9 eye-tracking tests and collected 200k per-second participants' reading (gaze position) and interaction datapoints. The dataset will be public.

First, we found that **having a small set of eye-tracking ground truth data enabled us to build a model with relatively high accuracy on predicting message-level reading time based on user web interactions only.** Specifically, we reached a 73% accuracy in reading time estimation with a two-tower neural network based on user interactions, while the heuristic baselines have around 54% accuracy. Second, we found that **adding users' behavioral patterns into the neural network models further improved their performance on reading estimation**. This study 1) provides examples of generalizing eye-tracking data to build reading estimation approaches that use browser-extracted features only and thus can be applied on a large scale; 2) give insights for future reading estimation studies on designing features / models. We incorporate the reading estimation models in this study into a prototype platform that enables communicators to see the reading time of each bulk message they send in Chapter 7.

Chapter 7

Exploring Bulk Email Effectiveness Tools for Better Transparency

7.1 Introduction

This chapter is a design and evaluation study on how to better support communicators in evaluating organizational bulk emails with a prototype tool called CommTool. For example, whether knowing about the reading time of each message would help communicators decide which content to put, whether knowing about the interest rate of employees from different units would help communicators personalize bulk emails, etc. [1]

We conducted interviews, iterative designs [193], and field tests to study whether specific features are useful to communicators. We started with expert interviews (5 communicators of the studied university) to learn about their experiences with evaluating organizational bulk emails and proposed potential features. We found that besides open rate and click rate, communicators would benefit from various metrics to make decisions, such as **message-level** reading time, read level, interest rate, etc. [2] The

---

[1] Ruoyan Kong, Irene Ye Yuan, Ruixuan Sun, Charles Chuankai Zhang, and Joseph A Konstan. Commtool: Supporting organizations to evaluate bulk emails (under review). 2023 [101]

[2] "Message" here refers to single stories / pieces of information in bulk emails.

potential useful features of bulk email evaluation platforms were then summarized from the qualitative analysis of these interviews' transcripts. A prototype of a bulk email evaluation platform (CommTool) was designed iteratively through 7 usability tests. To evaluate the usefulness of the proposed features, we conducted a 2-month field test in the studied university with 5 communicators and 149 recipients (employees of the university).

Organizations now often send bulk emails via customer relationship management (CRM) platforms like Salesforce and Mailchimp, which provide metrics such as the open rate of emails and click rate of URLs [16, 127]. However, open rate and click rate may not be sufficient tools to support **communicators** (organizations' employees in charge of designing and sending organizational bulk emails) in making editing / targeting decisions on bulk emails. Open rate can be an overestimation of employees' awareness as employees often open and close bulk emails in a few seconds and skip their contents, or be underestimated if the email has a useful subject or preview. Many organizational bulk emails focus on sharing important information rather than clicking links [98]. Within organizations, there is a need to evaluate the time cost associated with bulk emails, as employees spend much time filtering and reading these messages. Organizations also strive to maintain communication channels' reputations.[3] Therefore we learn how communicators currently evaluate organizational bulk emails, identify design opportunities, and build an evaluation platform prototype (CommTool) that aims to support communicators in understanding their audience and designing bulk emails.

CommTool's workflow is shown in Figure 7.1 and 7.2) (similar to the workflow of the university's current bulk email evaluation platform Salesforce). In step 1, communicators upload emails to CommTool, preview them, and send them to recipients (participants). The recipients receive emails with the same title, sender name, and personalized links in their inboxes (step 2). Then CommTool records recipients' data when they click the links and summarizes various metrics and feedback (step 3). Communicators would be able to check metric reports in real-time (step 4) and also receive a reminder email after 24 hours (step 5).

---

[3] In this chapter, we use "channel" to refer to a series of bulk communications / newsletters with the same sender identity (the FROM field) and brand (in the title or format, e.g., xxx Brief Dec 1, 2022).
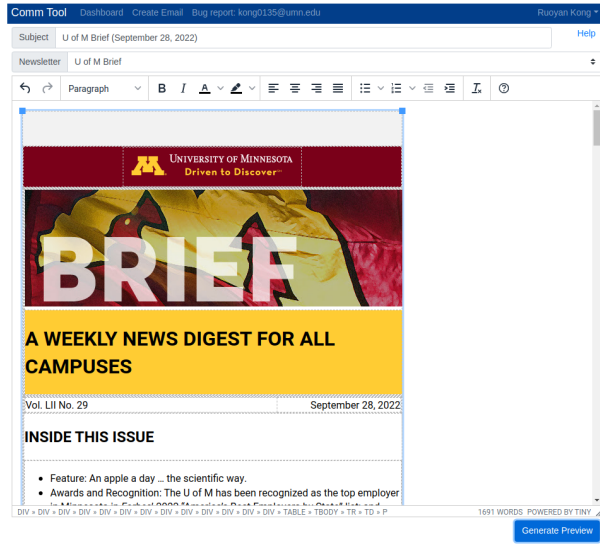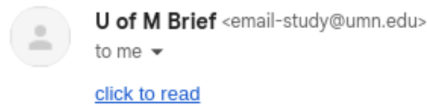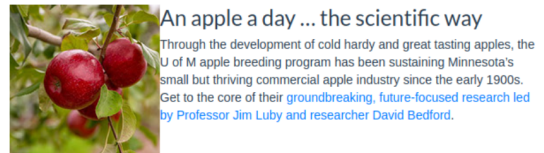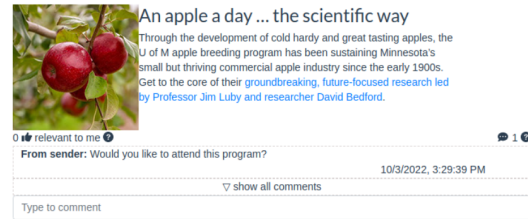
Figure 7.1: Communicator's workflow of evaluating organizational bulk emails through CommTool. Step 1 to 3.

Step 4: The communicator would be able to see per-email / per-message metrics in real-time.

Dashboard -- All Emails' Performance

Create Email

Newsletter	Subject	# Recipients	Open Rate	Click Rate	Read Rate	Detail Rate	Relevance Rate	# Comments	Reading Time	Est Cost	Reputation Change
Avg Performance of All Communicators			63%	31%	57%	21%	21%	31	58s	$12889	
+ U of M Brief	**NEW!** U of M Brief (September 28, 2022)	300	68%	23%	55%	19%	33%	67	53s	$11778	4%
	U of M Brief (September 28, 2022)	300	70%	27%	55%	20%	28%	50	59s	$13111	5%
	U of M Brief (September 21, 2022)	300	68%	29%	50%	22%	32%	45	49s	$10889	4%

Dashboard -- U of M Brief's Messages' Performance

Create an Email

Subject	Message	Click Rate	Read Rate	Detail Rate	Relevance Rate	# comments	Reading Time	Est Cost	Date
Avg Performance of All Communicators		11%	27%	8%	39%	15	26s	$5778	2022-10-0
U of M Brief (September 28, 2022) **NEW!**	An apple a day ... the scientific way	9%	16%	7%	10%	16	20s	$4444	2022-10-0
	Awards and Recognition	12%	26%	10%	16%	27	34s	$7556	
	Apply now: 2023-24 IAS Research and Creative Collaborative	11%	29%	8%	11%	20	24s	$5333	

An apple a day ... the scientific way

Through the development of cold hardy and great tasting apples, the U of M apple breeding program has been sustaining Minnesota's small but thriving commercial apple industry since the early 1900s. Get to the core of their groundbreaking, future-focused research led by Professor Jim Luby and researcher David Bedford.

38/150 click rate   29/150 relevant rate   18s $3999   3

From sender: Would you like to attend this program?

10/3/2022, 3:29:39 PM

Yes if it's at Twin Cities Campus.

10/3/2022, 3:53:50 PM

From sender: It will be held at Twin Cities and Morris.

10/3/2022, 3:55:08 PM

[Type to reply]

Step 5:
A reminder email will be sent to the communicator to check the metric reports and fill in research surveys after 24 hours.

Please take this survey each time:

Q1. Which of these metrics we show on this page help you make future editing / sending decisions? *

☐ Average performance of all communicators   ☐ Open Rate   ☐ Click Rate   ☐ Read Rate   ☐ Detail Rate   ☐ Relevance Rate   ☐ # Comments   ☐ Reading Time   ☐ Est Cost   ☐ Reputation Change   ☐ None of those are helpful.

Please select at least one option.

Q2. Which editing/sending decisions you have made / will make based on this report? *

☐ Send to more people   ☐ Send to fewer people   ☐ Share this report with my clients   ☐ Add more contents   ☐ Remove or shorten some contents   ☐ Move some contents to other newsletter   ☐ None of those

Please select at least one option.

Q3. Why certain metrics are helpful / unhelpful to you? Could you give an example of the editing / sending decisions you made based on this report?

Figure 7.2: Communicator's workflow of evaluating organizational bulk emails through CommTool. Step 4 to 5.

We posed the following research question:

RQ1: What are communicators' current practices of evaluating bulk emails and potential features that would help them?

Then we present CommTool, a prototype that implements the feature we proposed to evaluate bulk emails. We carried out a field test to evaluate these features' usefulness and proposed the following questions:

RQ2: (Feature Evaluation) How do recipients interact with CommTool in the field evaluation?

RQ3: (Feature Evaluation) What are communicators' experiences with CommTool in the field evaluation?

7.2 Related Work

As we discussed in Chapter 2, one challenge of organizational communication is employees' information overload — employees' perceptions that they receive more information than they can handle effectively [43]. Davenport and Beck proposed that employees have an "attention budget" — therefore, the more information they receive from communicators, the more information they will inevitably ignore [45]. Šliburytė proposed that in evaluating the effectiveness of organizational communication, the stakeholders' **cost** (e.g, time cost) needs to be considered [161]. For example, Jackson et al. monitored 15 employees' laptops and found they need 6 seconds to decide whether to open an email and 64 seconds to recover from an email interruption on average [77]. Dabbish and Kraut conducted a nationwide organization survey and found that email volume is positively correlated with the feeling of email overload at work [43]. Given these challenges regarding the effectiveness of organizational bulk emails, many organizations take action (see below) to evaluate their bulk emails' performance.

As reviewed in Chapter 2, many large business platforms, such as Salesforce, Mailchimp, Revue, Constant Contact, etc., support bulk email senders by evaluating their emails' performance, [16, 127]. The evaluation features supported by these platforms include the open rates, click rates, number of subscriptions and their trends, A/B Tests, etc.

Besides the business CRM platforms, several marketing studies also shed light on how to evaluate bulk emails (though most of them focused on bulk emails to external

customers). **"Awareness and impact"** of bulk emails are the two concepts most often mentioned. Cruz and Fill interviewed CRM users and learned that reaching and awareness (e.g., knowing the brand) is more important than behavioral metrics (e.g., hits and clicks) in evaluating marketing emails [39]. Morgan et al. reviewed the metrics of the current CRM platforms and pointed out that they focused on what can be measured quickly (e.g., open rate of marketing emails) instead of what should be measured (e.g., changes in the perceptions of the brand) [130]. Todor strengthened the importance of understanding bulk emails' consequence [175], like the return of interest, the number of businesses generated, etc. Siano et al. proposed a framework for senders to measure their reputation — recipient's perception of a specific sender's communication quality [157]. In the engineering field, there is work focusing on the API design and database integration of CRM platforms [165, 185].

We noticed a lack of user-centered work on designing platforms for evaluating organizational bulk emails' performance and cost. Their design requirements might be different from the commercial CRM platforms (which focus on external marketing emails). For example, organizations needs to consider their employees' time cost in reading bulk emails, and organizational bulk emails might focus less on conversion rate / return of interest but more on awareness of organizational goals. In this study, we understand organization communicators' experience with evaluating bulk emails, specifying potential useful features, and designing prototypes to better understand which features could support communicators in designing bulk emails.

In the following section, we introduce a formative study for the iterative design process of CommTool (RQ1), then we present the design of CommTool, and at last, we introduce a field test for evaluating CommTool (RQ2, 3).

7.3   Formative Study (RQ1)

We first conducted formative studies to 1) understand communicators' current practice of evaluating bulk emails; 2) identify potential useful features; 3) deploy CommTool iteratively. Our study site is the University of Minnesota, a public university with over 25,000 employees and several campuses. We collaborated with the communicators of its central units (e.g., presidential offices), who are in charge of organizing, designing, and

distributing university-wide bulk emails to all staff / faculty. This study is approved by the IRB of the University of Minnesota (STUDY00017138).

7.3.1 Expert Interviews

We conducted expert interviews with 5 university communicators for the purpose of RQ1 — learn potential features in evaluating bulk emails. We applied an artifact-walkthrough approach [17] by which we asked about communicators' practices with specific email cases. During each interview, the communicator selected 5 emails they sent out recently, then answered the following questions:

- Communication Goal: Do you come up with the content by yourself, or do your clients provide it? Who are you hoping to reach? How do you hope they react to these messages? What open rate/action rate would you be happy with? How much employee time would be taken according to your estimation?

- Communication Practice: How do you decide the ways of writing the content? How do you send it (e.g., in a newsletter, separate email, or distribution channels)? How do you evaluate the effectiveness?

- Expectations of Evaluation Platforms: Which information do you hope to have when you evaluate this email? Any challenges you have met with the current evaluation platform (Salesforce)?

In the end, we introduced several design ideas for the evaluation platform Comm-Tool, and asked the communicators their preferences on these ideas and the reasons.

We interviewed via zoom, and each interview took around 60 minutes. We recruited through the university's communicator forum. The forum director identified a list of potential participants working in the central communication offices. We contacted 6 communicators, and 5 (C1 to C5) agreed to participate in this expert interview.

For data analysis, I conducted the initial open coding of the first interview. The research team (see [101]) met weekly to discuss whether enough repeated themes emerged and whether more interviews were needed. Potential features were summarized using an affinity diagramming approach [72]. We summarized our findings around RQ1 below.

Observation 1: Communicators design / send / evaluate bulk emails for their clients.

Communicators sent those bulk emails for their clients within the university, such as their managers, university leaders, and central offices. They need to communicate with their clients first if they want to make some changes to bulk emails based on the evaluation results. For example, C2 described this editing process:

> *"So how it usually works is they (the clients) will have an email, and I'll review that and take that information to draft an email. And then they'll approve it, and then I'll send ... Especially when it comes to newsletters where there are multiple parties, it would be helpful to say: you know, based on the research that we've done on this newsletter you know, this kind of content doesn't isn't is appealing to this audience, so we're not going to include it." (C2)*

A bulk email evaluation platform can change the status quo of designing bulk emails only if the platform supports communicators to reach an agreement with their clients on the design. Based on this observation, we proposed the following feature:

Proposed feature 1 – Share results: The evaluation platform might enable communicators to share the results and communicate with their clients.

Observation 2: The metrics provided by the current platforms are not sufficiently measuring awareness and relevance.

Communicators noticed that their open rate was objectively good — usually above 60%. This open rate can be deceptive, as employees might not be actively reading those emails. For example, C1 talked about an email sent to over 8100 faculty:

> *"There's an open rate of 62%. It might count as an open rate if they just tap it in their inbox before archiving it. So I'm not sure if that means people actually put their eyes on it, or just like cleared it out." (C1)*

C3 talked about an email asking all employees to check their Human Resources dashboard, but employees seem to pay less attention to it:

> *"So it (the email about checking HR dashboard) has a high open rate, 80%, but only 10% of the audience that even got to that dashboard to see that*

data. It (the email) probably doesn't give them enough information to know if they've needed to pay attention, and we don't have enough information to know if those people are paying attention." (C3)

In summary, current performance metrics offer little to no insight into whether employees are aware of those bulk emails' contents, nor the perceived relevance of those emails. Based on this observation, we proposed the following feature:

Proposed feature 2 – Awareness and relecance: The evaluation platform might provide metrics around awareness and relevance besides open rate and click rate.

As each bulk email often contains multiple messages, the realization of these features implies that the granularity of those metrics should not only be limited to per email but also per message:

Proposed feature 3 – Message-level metrics: The evaluation platform might provide both email-level and message-level metrics.

Observation 3: Communicators sometimes need custom feedback.

Besides metrics in fixed formats, communicators also mentioned various formats of feedback they want to have. The examples mentioned by communicators include whether the recipients already knew the contents, whether the recipients were busy when receiving those emails, the job codes of the recipients who found those emails relevant, the recipients' devices, etc.

> *"We've (the communication unit) had discussions about like what time of day people want to receive emails, is it first thing in the morning, or are they overwhelmed by first thing in the morning." (C4)*

> *"The challenge we have for employee benefits, for example, and some other communications is really about which job codes are relevant, for example, some of the vaccine requirement information for specific employee groups." (C3)*

Based on this observation, we proposed the following feature:

Proposed feature 4 – Custom feedback: The evaluation platform might enable communicators to collect custom feedback.

Observation 4: Communicators need employee's interest to better target messages.

Communicators mentioned that, especially for communications sent to a large group, their audience is so mixed that communicators do not know which employees might be interested in which messages. With access to such interest information for employee groups, communicators may be able to better target their messages. For example, C2 talked about this challenge when they are sending communications widely:

> *"It's hard to know when you're kind of trying to send to a large group of people, a bunch of different stuff — you don't always know what it is they want to see, or what they're really interested in. If there is some way to fine-tune it (the newsletter) to opportunities, information, news, and stories that different people are really interested in, that would be helpful."*(C2)

Therefore we proposed the following feature:

Proposed feature 5 – Employee group's interests: The evaluation platform might enable communicators to understand different employee groups' (job categories, units, etc.) interests.

Observation 5: Communicators realize that bulk email has impacts but cannot measure them.

Communicators understood that bulk emails may have costs, both monetarily - in paying for employees' time - or impacts to the channel's reputation. Communicators also thought that sharing cost information may help them persuade clients to remove some content, while also expressing concern about the extent to which cost data would influence clients. Currently, communicators do not have access to such cost measures in the current bulk email evaluation platform. C2 talked about how they perceived the potential use of cost data:

> *"We don't know, how much time people spend on email and how much money the university loses — if some people (clients) are very driven by data, and that can help. But when you're working with somebody who just really wants, who really thinks whatever they have to say are important to their to the audience, I don't know if that is going to help."* (C2)

C5 mentioned that their audience stopped reading emails from some channels, and they used other channels instead.

> *"We found that if it's like a director of undergraduate studies that's sending out a survey, having it in that professor's name will increase opens tremendously rather than just from the department which people stop reading." (C5)*

With these observations in mind, we proposed the following feature:

Proposed feature 6 – Communication's impact: The evaluation platform might provide communicators with the impact of bulk communication (e.g., employee's time cost, reputation cost).

As the communicators mentioned, we are uncertain about the extent to which clients would give this cost data, and we plan on observeing that during the field test.

Observation 6: The evaluation platform should be light-weight.

Communicators mentioned that a challenge of getting deep evaluation results is that the process is complicated, and they are already overwhelmed by daily work.

> *"The current way of doing (deep evaluation) is focus group (usability lab). It is expensive, ineffective, and it's a challenge for us that we have no such internal service on feedback sessions." (C3)*

Therefore, a bulk email evaluation platform should match the following feature:

Proposed feature 7 – Lightweight use experience: The platform is lightweight — it should take no significant effort for communicators to evaluate bulk emails deeply beyond getting their main job done.

7.3.2 Design of CommTool.

Based on the features above, we designed a prototype platform to explore these features' usefulness in supporting communicators in evaluating bulk emails. An initial version of CommTool was iteratively deployed and tested by the research team.

1) Workflow: CommTool's workflow is shown in Figure 7.1, 7.2) (similar to the workflow of the university's current bulk email evaluation platform Salesforce). In step 1, communicators upload emails to CommTool, preview them, and send them to recipients (our study participants). As a prototype aimed at exploring diverse metrics' usefulness,

CommTool does not support present emails. All the emails will be sent out immediately when the communicators hit the send button. The recipients receive emails with the same title, sender name, and personalized links in their inboxes (step 2). The personalized links enable recipients to open these links on any device without login into CommTool, so we can record their data (step 3). Then CommTool collects data from recipients and summarizes metric reports. Communicators would be able to check metric reports in real-time (step 4). And also, after 24 hours (to give recipients enough time to react), communicator gets a reminder email to check reports and fill in research surveys (step 5).

2) Regular and Evaluation Recipient Group: An important step in calculating these metrics is that CommTool splits the recipients into a regular recipient group and an evaluation recipient group. The recipients stayed in one group throughout the whole study. The regular recipients were asked to react to the emails they received as normally as they could — CommTool uses their log data to calculate the open rate, reading time, read rate, read-in-detail rate, click rate, and estimated cost. The evaluation recipients were asked to read those emails and indicate whether each message was relevant to them and leave comments under the messages if applicable. We split the recipients into these two groups because "indicating relevance" and "commenting" are time-consuming and require more attention from the recipients. We do not want these actions to interfere with the calculation of awareness metrics. The definitions of the metrics we used are summarized in table 7.1.

In the following, we introduce the design details of CommTool. The match between those designs with the proposed features is summarized in Table 7.2.

3) Split Email into Messages: To provide message-level metrics, in CommTool, when communicators upload an html email, the email would be automatically split into sections in preview (by recognizing html tags like h2, h3, etc.), as displayed in Figure 7.3. Given the html length of each message, CommTool recognizes whether each section is just a title (like the "TOP NEWS" in Figure 7.3) or is a message with contents (like the "An apple a day ... " in Figure 7.3). CommTool adds surveys under each of those messages — "survey" here means we would collect metrics and feedback for these messages. CommTool also enables communicators to remove / add / merge sections, and add / remove surveys in case the split is incorrect.

Table 7.1: Metric Definitions. Email-level metrics are shown in the email dashboard. Message-level metrics are shown in the message dashboards and report dashboards.

Email-Level Metrics		Message-Level Metrics.	
Click Rate	The percentage of the regular recipients who at least clicked one link in this email.	Click Rate	The percentage of the regular recipients who at least clicked one link in this message.
Read Rate	The percentage of the regular recipients who at least skimmed one message of this email or read it in detail.	Read Rate	The percentage of regular recipients who skimmed this message or read it in detail.
Detail Rate	The percentage of the regular recipients who at least read one message of this email in detail.	Detail Rate	The percentage of the regular recipients who read this message in detail.
Reading Time	The average number of seconds the regular recipients who opened this email spent reading this email.	Reading Time	The average number of seconds the regular recipients who opened this email spent reading this message.
Estimated Cost	The estimated money cost of this email: the average reading time * open rate * $40 hour rate the actual number of recipients of this newsletter. 6 seconds are added for each recipient's time on making read / unread decisions.	Estimated Cost	The estimated money cost of this message: the average reading time * open rate * $40 hour rate * the actual number of recipients.
Relevance Rate	The percentage of the evaluation recipients who at least indicated one message of this email that is relevant to them.	Relevance Rate	The percentage of the evaluation recipients who indicated that this message is relevant.
# Comments	The number of comments the evaluation recipients left on this email.	# Comments	The average number of seconds the regular recipients who opened this email spent reading this message.
Open Rate	The percentage of the regular recipients who opened this email.	Who are Interested	The percentage of different categories (departments, job categories) of recipients who are interested (clicked/read/indicated relevant) in this message.
Reputation Change	This email's influence on this newsletter's reputation: this newsletter's predicted future open rate - this email's open rate.		

Table 7.2: Match between designs and proposed features

| Design	Proposed Feature
Split Email into Messages	3) Message-level metrics
Measure Message's Performance with Various Metrics	2) Awareness and relevance; 6) Communication's impact
Measure Message's Performance on Different Employee Groups	5) Employee group's interests
Collect Comments	4) Custom feedback
Share Reports	1) Share results
Automation of Evaluation Process	7) Lightweight use experience
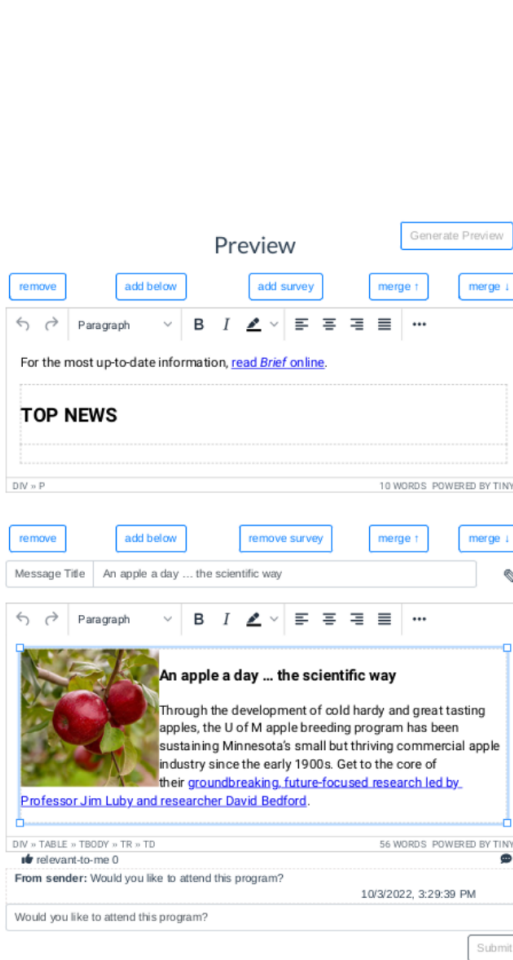
Figure 7.3: CommTool can automatically split an html email into single messages during editing.
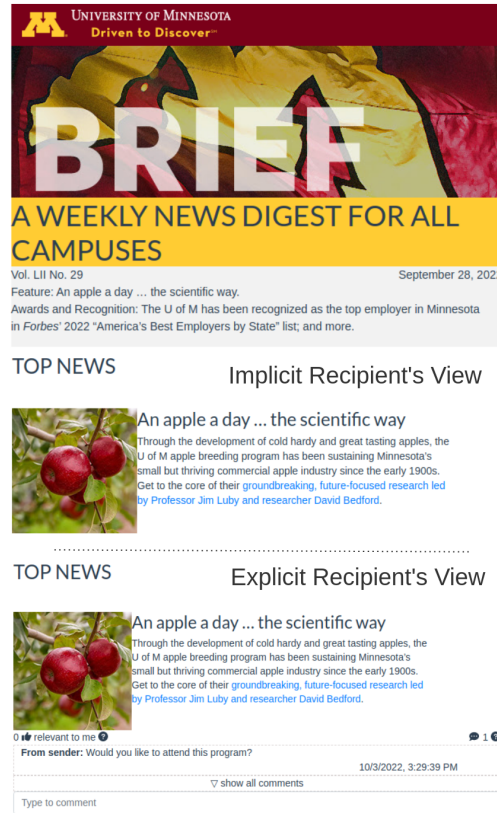

Figure 7.4: The email received by regular recipients is just like the original email. The email received by evaluation recipients would include "relevant-to-me" buttons and comment areas.

4) Measure Message's Performance with Various Metrics: CommTool supports communicators in collecting the awareness / impact / relevance metrics on both message level and email level. For awareness, we calculate the recipients' open rate for each email, average reading time, read (skim or read-in-detail) rate, and read-in-detail rate of each email / message. For impact, we calculate the estimated cost of each email/message and each email's estimated impact on its channel's reputation. For relevance, we calculate the percentage of recipients who viewed an email/message as relevant (either work-relevant or personally interesting) to them, or clicked them. The definitions of metrics are summarized in Table 7.1. In the dashboard of email's performance and dashboard of message's performance (see Figure 7.5), communicators would be able to see the metrics above in real-time (and they will receive a reminder email to check these metrics after 24 hours of sending the email).

Specifically, for email-level reading time, CommTool uses the average time the regular recipients are active on the email page.[4] 6 seconds are added for each recipient's time on making read / unread decisions, based on [77]'s finding that 70% of emails were opened within 6 seconds in the workplace. For message-level reading time, CommTool estimated a recipient's reading time for each message based on their interactions with the emails' pages (see Chapter 6 [100]).

5) Measuring Message's Performance on Different Employee Groups: On the dashboard page which shows the sample email (Figure 7.6), when communicators click "Who are interested?", a detailed view of different unit/job's interest rate will be shown. "Interested" here is defined as either a regular recipient clicked / read a message or an evaluation recipient indicating that the message is relevant.

6) Collect Comments: CommTool also enables communicators to collect custom feedback. Communicators can add their questions in the comment area (e.g., "Would you like to attend this program" in Figure 7.3). Communicators' questions would be pinned in the comment area and highlighted by "from sender" (anonymously). evaluation recipients would see these pinned questions and answer them anonymously if they want to (see Figure 7.7).

7) Share Reports: CommTool enables communicators to share the results with their

---

[4] "Active" here means that the window is visible, and if there are no interactions (scroll / click / mouse hover) in the past minute a window will pop up to check whether recipients would like to stay on the page.

Dashboard -- U of M Brief's Messages' Performance

Create an Email

Filter Newsletter / Subject | Type Any Key Words

Clear Sort X

Clear

→ Share this Page

Subject	Message	Click Rate	Read Rate	Detail Rate	Relevance Rate	# comments	Reading Time	Est Cost	Date
Avg Performance of All Communicators		11%	27%	8%	39%	15	26s	$5778	2022-10-03
+ U of M Brief (September 28, 2022) NEW!	An apple a day ... the scientific way	9%	16%	7%	10%	16	20s	$4444	2022-10-03
	Awards and Recognition	12%	26%	10%	16%	27	34s	$7556	
	Apply now: 2023-24 IAS Research and Creative Collaboratives	11%	29%	8%	11%	20	24s	$5333	
	Seeking proposals: Imagine Fund Special Events Grants	13%	24%	10%	12%	18	29s	$6444	

Dashboard -- All Emails' Performance

Create Email

Filter Subject / Message | Type Any Key Words

Clear Sort X

Clear

→ Share this Page

Newsletter	Subject	# Recipients	Open Rate	Click Rate	Read Rate	Detail Rate	Relevance Rate	# Comments	Reading Time	Est Cost	Reputation Change	Date
Avg Performance of All Communicators			63%	31%	57%	21%	21%	31	58s	$12889		2022-10-03
+ U of M Brief	NEW! U of M Brief (September 28, 2022)	300	68%	23%	55%	19%	33%	67	53s	$11778	4%	2022-10-03
	U of M Brief (September 28, 2022)	300	24, 70%	27%	55%	20%	28%	50	59s	$13111	5%	2022-10-02
	U of M Brief (September 21, 2022)	300	68%	29%	50%	22%	32%	45	49s	$10889	4%	2022-10-02

Figure 7.5: The email/message dashboards of CommTool. The message dashboard (top) shows message-level metrics. The email dashboard (bottom) shows email-level metrics.

The patient and the doctor who believed in something bigger

Sarah Ewald, diagnosed with cancer at just 18, says the decision to participate in a clinical trial to treat her cancer was easy. Even if the trial medicine didn't help her, maybe it could help others. Learn more about her experience.

| 30/150 click ❓ | 25/150 relevant ❓ | 23s ❓ | $5111 ❓ | ▽who are interested ❓ |
| Filter Type / Category | Type Any Key Words | | | |

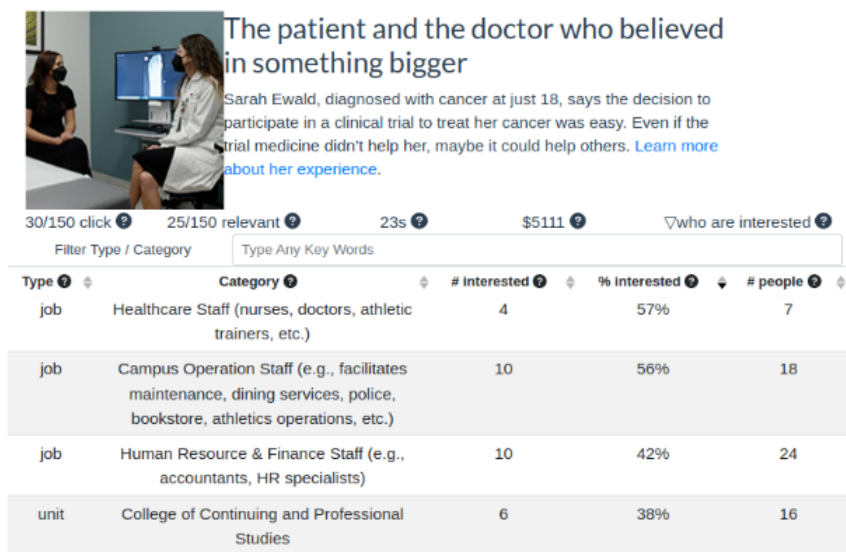| Type ❓ ⬍	Category ❓ ⬍	# interested ❓ ⬍	% interested ❓ ⬍	# people ❓ ⬍
job	Healthcare Staff (nurses, doctors, athletic trainers, etc.)	4	57%	7
job	Campus Operation Staff (e.g., facilitates maintenance, dining services, police, bookstore, athletics operations, etc.)	10	56%	18
job	Human Resource & Finance Staff (e.g., accountants, HR specialists)	10	42%	24
unit	College of Continuing and Professional Studies	6	38%	16

Figure 7.6: The report dashboard of a sample email. Communicators would be able to see the number of regular recipients who clicked the message's hyperlinks / the number of regular recipients, the number of evaluation recipients who viewed it as relevant / the number of evaluation recipients, average reading time, estimated cost, and the interest rate split by unit / job categories.

clients. On each dashboard (message performance, email performance, recipient view), there is a "share" button; communicators can click this button, add high-level summaries to that dashboard, and get the shareable links (see Figure 7.7). Their clients do not need to register in CommTool to view these reports but can also post comments as the senders of that email with the shareable links.

Figure 7.7: Share Report. Communicators can summarize the report in notes and share that link with their clients. Their clients can comment as senders with shareable links.

8) Automation of Evaluation Process: The whole workflow of evaluation (see Figure 7.1 and 7.2) only needs communicators to copy and paste their email to CommTool, click send, then they will receive results — in the usability tests, this process took less than few minutes.

After the research team reached an agreement on the usefulness of the platform, we

invited communicators to conduct a 30-min usability test. The usability test aims to understand CommTool's workflow's intuitiveness and its metrics' clarity. During the usability tests, the communicators were asked to explore the platform and complete the following tasks: 1) create a channel and send an email from that channel; 2) find the performance reports of those emails and tell us how they interpret those metrics; 3) share the reports with their clients. The value of the metrics was simulated during the usability tests. We observed whether they met any difficulties in completing the tasks, and whether they understood the metric reports. We then revised the platform according to their feedback. We stopped inviting communicators when the newly invited communicator used the platform smoothly without guidance. We conducted 7 usability tests in the end. We made the following changes to CommTool during the usability tests:

Visibility of Comment Area. Communicators wanted the platform to be more communicator-focused. One area of concern was that the recipients who commented might naturally be more disgruntled, or have more negative things to say. Based on that, the comment area was changed to only be fully accessible to the corresponding communicator, and recipients were not able to see each others' comments (except comments from senders).

Definition of Reading Time / Read Rate / Detail Rate. Communicators indicated that they would like the reading time calculation to exclude recipients who did not open emails, because that information is already included in the open rate. For the calculation of the read rate and the read-in-detail rate, however, they prefer to consider all the recipients; then they can intuitively compare the open rate, read rate, and read-in-detail rate and use those rates together to measure recipients' awareness.

Provide Definition Tips of Metrics. As CommTool provides various metrics, communicators could not recall definitions of each metric, and asked for explanations during the usability tests. Therefore, we provided tip buttons next to each metric in the reports to give detailed definitions.

7.4 Field Test (RQ 2,3)

To evaluate the proposed features' usefulness, we conducted a 2-month field test on CommTool with 5 communicators and 149 recipients (employees of the university).

7.4.1 Methods

The participating communicators were recruited from the communicators we contacted in the expert interviews and usability tests from the central communication offices. Before the field test, we asked each communicator to select a university-wide bulk email channel (sent at least monthly) to participate in the test. Before the field test, we asked the communicators in the past month, on average, how many times they communicated with their clients or shortened content during editing that channel.

The recipients were recruited through 1) the communicators putting the recruitment message on the selected channels; 2) other mailing lists of the university; 3) the research team distributing the study posters to 38 buildings' poster boards and offices; 4) the university's slack channels. We recruited 149 recipients in total. We split the recipients into evaluation/regular groups randomly (the recipients would stay in the assigned group throughout the whole study) and gave corresponding instructions (see 3.3). Then we gave the communicators the list of recipients to be excluded from the channels during the field test (to avoid these recipients receiving duplicate emails).

During the field test, each time when the communicators sent an email under the selected channel, they would go to CommTool to send the same one to the list of recipients. We logged the recipients' interactions with each email. At the first 9 a.m. after 24 hours after they sent out each email, CommTool will send the communicators a reminder email, asking them to check the reports of that email and answer the research surveys.[5]

The research surveys contained 3 questions: 1) Which of these metrics we show on this page help you make future editing / sending decisions? 2) Which editing/sending decisions have you made / will make based on this report? 3) Why are certain metrics helpful/unhelpful to you? Could you give an example of the editing / sending decisions you made based on this report? We also logged whether the communicators read each report, how much time they spent reading, and whether they shared the reports with

---

[5] The communicators asked for a regular time of receiving reports.

their clients.

After the field test, we conducted 30-min interviews with the communicators, with the following questions: 1) What's your general experience with CommTool? What's its pros and cons compared to your current evaluation platforms? 2) Which features of CommTool do you find helpful / useless? Why? 3) Which decisions do you usually make based on the reports (could you give a specific example)? How do you perceive the cost and performance tradeoff of these emails? 4) Do you have any other features you would like to see in such a bulk email evaluation tool? Would you like to keep using CommTool in the future? Why? The interviews were analyzed following the same practice of expert interviews.

7.4.2   Recipients' Interaction with CommTool (RQ2)

Figure 7.8 shows the statistics of recipients' interactions with per email. The level of open rate of each channel is close to their original level (around 70% for Brief, 60 % for Communication Blog, 50 % for Controller's Office, and Research is a brand new channel that is never sent before).

Most emails were skimmed. Recipients who opened these emails spent around 60 seconds reading long channels like (Brief, Research, and Controller's Office) and around 10 seconds reading the short channel (Communication Blog). The email read rates (of at least reading one message in that email) were below 70%, and the detail rates were below 60%. Besides Brief, the relevance rates were below 40%, and the comment rates were mostly below 10%. The read rate of these channels varied from 20% to 70%, while the detail rate of these channels varied from 10% to 60%. The click rates are below 35%. The relevance rates varied between 10% to 60%. The comment rates are generally below 20%. These results show that the channels' reputation needs to be protected.

Recipients' comments could be categorized into 3 topics: the reasons that they feel the corresponding messages are interesting / uninteresting personally; how the messages are work-relevant to them; feedback on the content / designs.

Why a message is interesting / uninteresting to me personally: employees told the senders about why they feel a message is interesting / uninteresting to themselves. The first reason is that the employees have some / no similar experience with the message topic. For example, an employee found a message about Alzheimer research interesting
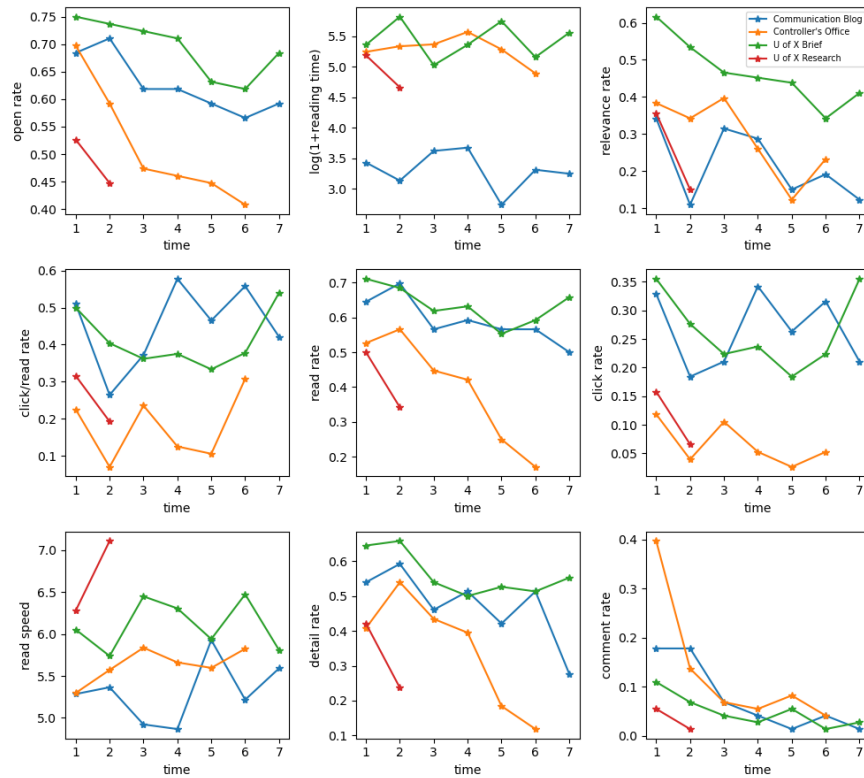
Figure 7.8: Summary statistics of recipients' reactions with each email in different channel. Time refers to the n-th email sent in that channel. Read speed is defined as n_word / log(1+reading time).

because *"My father-in-law has Alzheimer's so this is very relevant to my personal life."* An employee found a staff story from another campus uninteresting because *"Don't have a connection to that campus."*

The second reason for a message being interesting is that the employees like to see that the university is paying attention to the corresponding topics. For example, an employee is glad to see a message about first-generation student celebration because *"It's helpful to have visibility into central university focuses on first generation students."*

Why a message is work-relevant to me: employees tell the sender when they feel a message work-relevant to them. The work-relevance when mentioned in 2 cases: 1) the message is directly related to the employee's job: *"I find virtual events here that I attend for education hours as required by my Dept's Flexible Work Agreement"*; 2) the employee know someone in their units would be relevant: *"This is professionally relevant to me because the unit I work in has polar explorers."* (the message discusses climate-related studies)

Feedback on the content / design: Recipients give suggestions on the specific content they want to know / how they feel the content could be designed better. For example, an employee suggested *"It would be good to provide more information regarding HOW to set suppliers up with electronic payments - even if it is just "have the supplier contact us directly via phone (#) or email (address)."* under a message about special check handling. Another employee commented on a long message *"A lot of text with less bold/bullet/other ways to highlight. I would not have made it through the whole thing."*

7.4.3 Communicators' Experience with CommTool (RQ3)

In this section, we report communicators' experience with CommTool. In the below, Communication Blog's communicator is P1; Brief's communicator is P2; Controller's Office's communicator is P3; Research's 2 communicators are P4 and P5.

Overall Experience

Number of times of reading reports: Communicators used CommTool frequently. They sent 2 to 7 emails throughout the study and, on average, visited CommTool 5

times for each email. The message dashboards were visited most frequently (53 times) and shared with clients 7 times. The email dashboards were visited 40 times and shared with clients 4 times. The detailed report dashboards were visited 29 times and shared 2 times (see Table 7.3).

Table 7.3: Summary Metrics of Communicators' Feedback. The percentages show the rates that the metrics were selected as useful by the communicators in the surveys (averaged by all the channels' statistics).

Email Dashboard		Message Dashboard		Report Dashboard	
Total time of email dashboards being read	40	Total time of message dashboards being read	53	Total time of report dashboards being read	29
Total time of email dashboards being shared	4	Total time of message dashboards being shared	7	Total time of report dashboards being shared	2
Email Dashboard Metrics	**% selected as useful**	**Message Dashboard Metrics**	**% selected as useful**	**Report Dashboard Metrics**	**% selected as useful**
Click Rate	48%	Click Rate	57%	Comments	35%
Relevance Rate	44%	Read Rate	51%	Relevance Rate	24%
Open Rate	33%	Average performance of all communicators	46%	Who are interested?	15%
Average performance of all communicators	30%	Relevance Rate	38%	Reading Time	8%
Read Rate	23%	Reading Time	35%	Click Rate	4%
Detail Rate	18%	Detail Rate	15%	Est Cost	0%
Reading Time	15%	# Comments	12%	None of those are helpful.	0%
# Comments	12%	Est Cost	11%		
Reputation Change	11%	None of those are helpful.	0%		
Est Cost	4%				
None of those are helpful.	0%				
Email Dashboard Decisions	**% taken this action**	**Message Dashboard Decisions**	**% taken this action**	**Report Dashboard Actions**	**% taken this action**
Add more contents	27%	Add more contents	36%	Add more contents	24%
None of those	21%	Share this report with my clients	29%	Remove or shorten some contents	17%
Remove or shorten some contents	17%	Remove or shorten some contents	25%	Send to more people	7%
Share this report with my clients	15%	None of those	25%	Share this report with my clients	7%
Send to more people	12%	Send to more people	16%	None of those	7%
Send to fewer people	12%	Move some contents to other newsletter	13%	Send to fewer people	4%
Move some contents to other newsletter	8%	Send to fewer people	8%	Move some contents to other newsletter	0%

Ease of use: In interviews, communicators reflect that the platform is lightweight as

they only spent 5 to 10 minutes initiating an evaluation on CommTool:

"Everything about the tool, the basics, way things flow, the way you do. Things all seem logical and very useful. (P3)"

Communicators are generally satisfied with the readability of dashboards: *"This is a nice overview for quick scanning. (P5)"*

P1 felt that the designs of the 3 dashboards were too similar, but the reminder email sent to them after 24 hours provided a quick channel to check and differentiate the dashboards:

"I'm glad for the email the next morning that links to all 3 of those dashboards, because I don't know if I'd be able to figure out, because the first in the second dashboard looks so similar." (P1)

Comparison with existing platform Salesforce: Communicators feel that Salesforce provides more help in editing email templates, but CommTool provides more information in evaluating emails:

"Salesforce has templates which everybody loves. It has a glamorous interface. CommTool is simpler. It had metrics that people should care about and should be able to track, and at some point, some senior leader is going to ask them for them. I think it is helpful, especially for people who didn't have a lot of experience with their audience or who were working with completely unknown audiences. (P3)"

P1 mentioned that CommTool provided a channel to retrieve various data legally from employees:

"Apple messed things up last year. They pull out the tracking pixel before it gets to your inbox — so the open rates were wrong. In CommTool, people are those who agreed to share this information." (P1)

Sharing reports: Communicators shared the message dashboards 7 times and shared the email dashboards 4 times. These dashboards provided communicators with the reason for removing/extending specific messages. The per-message reading time specifically gets clients' attention:

"Because they are interested in how long people took to read it. They've always had the number of eyes on the open rates. They didn't know that that existed before I started working with them, so they were like, we can tell how many people read the message. Yes? you can. so now they're all comparing numbers with each other, our message did

this. They want to go further with metrics like that, how long people are engaged with the content, what their feedback is, and so forth." (P3)

The detailed report page was shared less because clients care more about numbers: *"When I'm talking with the people, they've all seen the message. They've helped write the message in most instances. So they're passed that part, and they want to see the numbers. Plus everyone I work with is a person they're all accountants. So when they see words and pictures, it freaks out a little bit, but when they see numbers they're like. Oh, I love it." (P3)*

Experience with Specific Features

Message - level versus email - level metrics (feature 3). Message metric dashboards were visited more times (53 visits) compared to email metric dashboards (40 visits), see Table 7.3. P2 mentioned that the message dashboard *"is useful if you wanted some more detailed analysis of your readership."* The report dashboards were visited less (29 visits) — P2 reflected that communicators who run large newsletters are too busy to look at the details on the report dashboard *"I would not have time to go through all those comments, especially on a list that's 20,000 strong. "*

Awareness vs relevance vs cost vs reputation metrics (feature 2, 6). The relevance metric was selected as useful most times, and the awareness metric followed. The cost and reputation metrics were less likely to be used.

Specifically, the email click rate is most usually to be selected as useful (48% of times) and followed by the relevance rate (44%). P2 told us that after seeing the email relevance rate, they took more time to *"consider relevance for audience members when developing content.".*

Besides, the message-level reading time and read rate are also often found to be helpful. In interviews, P3 said that:

"I pay attention to how much time they spent on it when that's available. That is the part of this tool that most interested me. So I could see, Did they engage with the content? And did they spend some time there?" (P3)

Communicators care less about detail rate because most messages only ask people

to have some sense of the content:

"In-depth reading doesn't really matter much just because it's (the email content) a summary. They (the recipients) can take action based on the email, click in or read the full article." (P1)

The finance newsletter cares more about read-in-detail rate because they are sending important information that every employee should know:

"I'm more interested in the detail than the I mean. All the tools will give you how many eyes looked at it? Right? I want to know how long they spent on it. You know, the University's budget is between 3 and 4 billion dollars a year. It's a lot of money, and most of it is spent on human costs, you know salaries and etc. Every employee should want to care, and the fact that person felt that it wasn't relevant to them meant that they didn't connect with the institution as an entity that has to watch its money, and watch how we behave with it." (P3)

Communicators do trust CommTool's estimation on per-message reading time:

"From what I could tell, and I had my own team members look at the email and sort of track. How long it took them to read the sections carefully, and I compared the 2 and I. It seemed very accurate, based on their experience." (P2)

"The top article in the top section has about a minute the one spent on it, and then the number gets kind of small as you go down, which makes sense to me because some of these things are meant to scan and other things. It's like you only have this much time to get those attention. So if they spend the most of their attention on the top thing, and then scroll the rest. That seems reasonable." (P4)

Communicators did review estimated cost but found it less helpful because of the organization's structure:

"(We) discussed estimated cost, and while interesting, it didn't seem to influence leaders' decision to send." (P5)

Also as we just discussed above, communicators felt some cost does make sense given the value of the information:

"It (the cost) was frightening. So now, every time I send out an email, I'm like, how much are we actually costing by making people read this? Then I remember we don't send out fluff, or like party invitations, or come to our sale we're sending out. You have work to do. This is related to your work. So we would incur that cost." (P3)

Communicators did not look at the reputation metrics directly but checked out the trend of metrics themselves and had their own judgments:

"I think it would be most effective in looking at trends over time." (P1)

"I'm going to have to see the metrics over time before I make any decisions about changing up the newsletter." (P2)

"I did look at the like week to week or message to message on my own like improvement, and if that went up a little bit, I was happy. If it went down, figure out what was in my content that made it go down. I like the predicted reputation change, although I'd want to know how it's actually calculated. But in really simple terms." (P3)

Communicators looked at the average performance of other communicators but did not use it to make decisions — they preferred to make decisions based on their own data:

"I knew there was a total average. I knew my messages were probably different than other messages, so I didn't pay attention to comparing them very much." (P1)

Custom feedback (feature 4). Comments gave communicators direct help with writing and formatting (see 4.2 for employees' comments). P3 indicates that based on the comments they received, they want to make the content clear: *"We've removed a layer of formality from our messaging. We try to be brief, be bright. Taxes shouldn't be scary." (P3)*

We also notice that bi-directional communication gets more feedback than single-direction communication. In our study, an email that the communicator has back-and-forth conversations with recipients gets 107 comments while the others are below 40. Some communicators did not use this feature because they did not notice this feature and said that they would reply:

"If they're (recipients) asking a direct comment like a direct question. (P1)"

P3, who got much feedback, gave us suggestions on how to get more replies:

"If you ask them unexpected questions, or you ask them in a less formal way, they're more likely to respond. So I think that's why I got a lot of comments." (P3)

Group-level interest (feature 5). P4 and P5 from the research newsletter particularly look at "who are interested" to "Consider relevance/interest for audience members and which audiences respond to the types of messages when developing content". But they felt that the "who are interested" feature could be more useful if they could setup groups

themselves:

"Whenever you have that level of detail, it's interesting. I'm seeing all these different colleges. And that's great. But it's not maybe necessarily useful for us to know that. For us, it would be interesting to see the breakdown between staff and students and faculty." *(P4)*.

These findings suggest that 1) current bulk email evaluation platforms, besides open / click rate, should take more active actions on collecting the relevance data (like by adding a small relevance to me button under each email), skim-level awareness, and trend of metrics; 2) the evaluation platform should allow message-level custom feedback, and especially encourage the bi-direction communication between senders and recipients; 3) if we want to control bulk emails' time cost, we need to build the mechanism that can influence organization leaders.

7.4.4 Email's Performance versus Reputation

In this section, we reported the correlation between email's performance and its channel's reputation. We defined reputation's change as $reputation_t = open\_rate_{t+1} - open\_rate_t$ (the difference between an email sent at time t's open rate to its next open rate). We run ordinary least square models (with constant) between $reputation_t$ and each $metric_t$. The coefficients' values, p-values (for the hypothesis $coefficient \neq 0$), and confidence intervals are reported in Table 7.4.

We found click rate positively correlated with reputation change (coef=0.221, p-val=0.049). Though none of the reading time, read rate, or detail rate correlated with reputation change, the ratio between click rate to read rate is positively correlated with reputation change (coef=0.169, p-val=0.026). This result matches the intuition — the emails that interest the recipients who read them can build up their reputation. Interestingly, the comment rate was negatively correlated with reputation change (coef=-0.260, p-val=0.041*). According to the contents of the comments we received, a potential explanation is that most recipients only commented when they found an email irrelevant or confusing (about contents, designs, intentions, etc.); therefore, these recipients are less likely to open the next email.

Table 7.4: Coefficients between email's performance with its reputation's change. CI: the [0.025,0.975] confidence interval of the corresponding coefficients.

metric	coef	p-val	metric	coef	p-val	metric	coef	p-val
open rate	-0.038	0.768	log(1+reading time)	-0.007	0.542	relevance rate	0.026	0.775
click/read rate	0.169	0.026* CI[0.023, 0.315]	read rate	0.027	0.811	click rate	0.221	0.049* CI[0.001, 0.441]
read speed	0.007	0.786	detail rate	0.055	0.629	comment rate	-0.260	0.041* CI[-0.508, -0.012]

7.5   Discussion

In this paper, we studied how to support communicators in evaluating organizational bulk emails. We conducted expert interviews to understand potential useful features. Then with an iterative design approach, we deployed the system CommTool. At last, we evaluated these proposed features in a 2-month field test with 149 employees and 5 communicators.

We found that communicators wanted more details like message-level metrics to help them better evaluate and target bulk messages (RQ1). Employees were not paying enough attention to the tested channels (RQ2). Communicators used CommTool to better understand their audience and also suggested improvements on features like cost estimation, group-level interest, etc (RQ3) In the below, we discussed the observations we found interesting.

7.5.1   Influence of Organization Structure on Evaluating Bulk Emails

We observed that the organization's structure influences communicators' use of bulk email evaluation platforms. In the field test, though communicators had access to the evaluation reports and felt that certain email designs could be changed, they did not always have the final say on whether a message should be sent out and they needed to get back to their clients, the organization leaders. The separation of information access and decision-making brought extra costs for CommTool's users in communicating back and forth with their clients if they want to make changes. CommTool tried to reduce this cost by enabling communicators to share reports with their clients. However, this approach does not significantly influence leaders' decisions. Communicators suggested

that leaders would need the reports to be more appealing:

"If that (the report) was going to be used to show your leadership, it might provide something easier to prepare, more appealing, like a presentation" (P5).

Another high-level solution to this challenge is to transfer part of the decision power to communicators. As Jensen and Meckling proposed in their paper about organization's knowledge control and structure [82], *"If the knowledge valuable to a particular decision is to be used in making that decision, there must be a system for partitioning out decision rights to individuals who already have the relevant knowledge and abilities or who can acquire or produce them at the lowest cost"*. For example, the organization leaders could allow communicators to remove some messages themselves if the history reports show that those types of messages are not interesting to the audience.

7.5.2 Making positive suggestions

One interesting observation is that communicators found the message-level reading time useful but cost less helpful, while cost is approximately a rephrase of reading time in CommTool. Communicators reflected that the cost displayed is too "frightening" to lead them to make any decisions based on that. Intuitively, reading time and cost can be viewed as the positive side and negative side of the same concept. The fact that users are more likely to take actions based on the positive side match with the previous findings on psychology — for example, Hayashi and Kim et al. both found that students/participants were more motivated to learn when they received positive feedback from conversational agents [66]. This observation suggests that bulk email evaluation platforms could benefit from making the reports' presentation more positive to encourage communicators to take action in designing / targeting.

7.6   Conclusion

In this chapter, we studied how to support communicators evaluating organizational bulk emails to improve organizational communication's effectiveness. We started with expert interviews with 5 communicators. We found that communicators wanted more detailed information like message-level reading time, relevance rates, comments, etc to help them understand their audience. Then with an iterative design approach, we

deployed an organizational bulk email evaluation platform CommTool. CommTool gives communicators convenient real-time access to not only email-level but also message-level reading time, read rate, read-in-detail rate, relevance rate, comments, group-level interests, etc. We then conducted a 2-month field test with 5 communicators and 149 employees to evaluate the proposed features. We collected employees' interactions through the field test and communicators' feedback through surveys during the field tests, as well as interviews after the test. We found that communicators liked the message-level reading time, read rate, and relevance rate provided by CommTool, which helped them to better understand their audience's interests in each message. Comments helped communicators in designing messages though communicators might be too busy to check all of the comments. Communicators trust the reputation and cost algorithm, but felt that they preferred to observe the trends themselves and use that as reputation instead of a number, and their clients felt that some messages should be sent out at such costs.

The limitations of this study include that 1) though the tested channels are selected from the newsletters that target the whole organization, their mailing list is not the same as our participant list and the measured metrics would not be the same as these channels' performance on their real mailing list; 2) we conducted our study in a large organization with a top-down communication system; organizations with different structures (such as a flattened management structure) might have different findings (for example, their communicators might have more say in deciding which messages to be sent out).

With these limitations, this study contributes to improving organizational communication by understanding communicators' expectations, and deploying and testing a bulk email evaluation platform with diverse message-level and email-level metrics on awareness, relevance, cost, and reputation. Based on the test, this study suggests bulk email platforms add certain message-level metrics such as reading time and read rate, and gives specific recommendations on how to present these metrics to make them helpful for communicators.

Chapter 8

Discussion and Conclusion

In this chapter, I summarize the contributions of the dissertation, the implications of our findings, and identify future work. By conducting surveys, interviews, controlled field experiments, and iterative design studies, this work sheds light on how organizational bulk email system works, and at which control points we can insert interventions to make this system more effective.

8.1 Contribution Summary and Future Work

We started this thesis from our personal experience with the study site — why, as an employee, we receive hundreds of pieces of non-targeted and non-personalized information from the university through bulk emails every week. In a follow-up meeting with 9 communicators of the study site, we learned that the study site's organizational bulk email system might be ineffective and not what the university wanted it to be. Then we looked back to previous work, and found much knowledge on how organizational communication's effectiveness is impacted by various stakeholders' actions, but less knowledge on organizational bulk emails, especially from a multi-stakeholder perspective. Therefore we identified a need to understand this system's effectiveness and its stakeholders, and to dig into the technology opportunities for making this system better.

In Chapter 3, we first conducted a mixed-methods study to understand organizational bulk email system and its stakeholders. Through a survey of 162 employees of our

study site, we learned that the current organizational bulk email system is ineffective as only 22% of the information communicated was retained by those employees. Then through the artifact walkthroughs with 6 communicators, 9 recipients, and 2 managers, we learned that the failure of the study site's organizational bulk email system was systemic — the organizational bulk email system had many stakeholders, but none of them necessarily had a global view of the system or the impacts of their own actions. The senders would like their messages to be distributed broadly. The communicators lacked tools for personalizing, targeting, and evaluating bulk emails, and as a result, often distributed messages too widely. The employees, overwhelmed by large quantities of information, only want to read messages they perceive as relevant to themselves. And nobody was watching the millions of dollars of people's time consuming in this process. The limitation of this work is that this is a qualitative case study of one study site. The observations may not be generalizable across organizations. It is possible that stakeholders in other organizations have different practices/perspectives on bulk email systems with respect to their organizations' cultures/structures. Therefore a question to be studied in the future is to collaborate with more organizations to study whether our findings in this case study are common.

In Chapter 4, based on the findings of the empirical study, we further proposed an economic model to describe the value, cost, and actions of this system's stakeholders and how their diverse perspectives cause ineffectiveness. We summarized two promising interventions. On the recipient side, we need mechanisms for encouraging employees to read high-level information. On the sender side, we need to let senders consider the value and cost of their bulk messages, and design tools for communicators to support them in making targeting and designing decisions. The limitation is that we only propose a model without justification. However, the proposed model enables us to clarify the relationships between this system's stakeholders and identify a bunch of potential interventions, enable to do the next two studies. There is future work to be explored in the calibration and implementation of this model, including 1) calibrating and measuring the value and costs proposed in the model; 2) designing tools to make the change of reputation observable to communicators; 3) predicting employees' preferences on each bulk message, such as to learn from employees' job descriptions; 4) learning how to weigh different stakeholders' opinions.

In Chapter 5, we studied the recipient-side interventions — the personalization designs of organizational bulk emails that can lead employees' attention to the messages important to the organization, while maintaining employees' positive experiences with these bulk emails, then they continue to read these emails in the future. 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 found that mixing important-to-organization messages with employee-preferred messages in top news could improve the whole newsletter's recognition rate. This work also provided a basic backend framework for communicators in personalizing organizational bulk emails. The limitation of this work is that the personalization is reordering only — because of the requirement of our collaborator, we did not exclude any message from the studied newsletter. The future work includes 1) personalizing newsletters by filtering a subset of relevant messages in organizations that allow taking this mechanism; 2) exploring different designing strategies that could help employees understand why they need to read some messages: for example, encouraging senders to tag the reasons for sending some messages; 3) enabling employees to update their preferences; 4) studying how to restore nonreaders' trust on the bulk communication channels

In Chapters 6 & 7, we studied the sender-side interventions — a communication prototype tool that supports communicators in making editing and targeting decisions on organizational bulk emails. To enable such evaluation, we first developed a novel neural network technique to estimate how much time each message is being read using recipients' interactions with browsers only, which improved the estimation accuracy from 54% (heuristics) to 73%, based on 200k ground truth data points we collected through eye-tracking tests. For limitation, we only collected eye-tracking data on 9 users therefore the dataset we collected might not catch enough variance on user patterns. Future work should look at where the value of additional users starts to decay significantly.

Then we iteratively designed and deployed a prototype of an organizational bulk email evaluation platform (CommTool), which enables communicators to learn the performance and cost of each bulk message. We evaluated the usefulness of different features in CommTool through a 2-month field test with 5 communicators and 149 organization employees. We found that 1) the message-level metrics such as reading time and read

rate helped communicators in designing; 2) email's performance is correlated with its channels' future reputation; 3) CommTool does not influence the organization leaders' decisions. We summarized with suggestions on designing organizational bulk email evaluation platforms that consider leadership and provide positive feedback. The limitations of this study include that 1) though the tested channels are selected from the newsletters that target the whole organization, their mailing list is not the same as our participant list; 2) organizations with different structures (such as a flattened management structure) might have different findings. As we found that showing the cost to bulk email senders does not influence their decisions, there is future work left on reducing overwhelming bulk communication in organizations, such as allocating communication budget according to communication channels' performance, enabling better filtering mechanisms on the recipient side, etc.

In short, organizational bulk email system is a complex multistakeholder system where different participants have limited capabilities to see the impact of their actions on the organization as a whole. We present a set of studies aimed at improving this system toward an integrated system that considers diverse stakeholders' actions and opinions. This thesis provided: 1) in-depth knowledge of how an organizational bulk email system works and its stakeholders' perspectives, including its stakeholders' roles, practices, and perspectives; 2) an empirical evaluation of several mechanisms for encouraging organizational bulk email recipients to pay more attention to messages deemed as important by the organization; 3) the design and empirical evaluation of a tool that incorporates various feedback features on bulk email's relevance to employees and the attention they receive.

8.2   Discussion

Based on our studies, we highlighted that organizational bulk email systems should be designed differently from the existing commercial bulk email systems based on their nature as a system in the workplace and for organizations.

8.2.1   Nuances, Fluidity, and Exceptions as a Workplace System

As a social technology system in workplaces, organizational bulk email system is full of nuances, fluidity, and exceptions [2].

"Nuances" refers to the depth of details needed for understanding the contexts to make predictions and decisions [2]. All the stakeholders in organizational bulk email system have very nuanced perceptions of the context. For employees, whether they perceive a message as work-relevant depends on their job responsibilities, which can be complicated and obscure. For example, in Chapter 3, we discussed a bulk email reminding Webex users to transfer their documents. Strictly, this information is only relevant to those employees who have not transferred their files and would need to use those files in the future. This granularity of information is hard to collect (especially implicitly without asking employees), and is case-by-case — emails with different contents would definitely need different context. It is obvious that an effective organizational bulk email system, ultimately, would build on certain knowledge management systems that can identify the required contexts smartly and collect them automatically [121]. For example, a knowledge management system that stores employees' job responsibilities and preferences to decide which bulk messages are relevant to them.

"Fluidity" refers to the changes in stakeholders' perspectives and practices [2]. For example, employees' actions toward bulk emails might depend on their schedules. In Table 3.7, employee R6 skipped a bulk email because they were too busy that month and found the email relevant when we asked them to read it. Employees' job range is also fluid. Using the "voicemail down" message in Chapter 5 as an example, this kind of message might be relevant to an employee if they are planning for a voicemail message on Sunday but might be irrelevant if they have no such plans next time. Also, for the senders, there are less common criteria on how to send a bulk message (see section 3). One communication channel can be used to send important and good-to-know messages at the same time, which would cause employees' confusion about the importance of the messages from that communication channel. Given the fluidity of organizational bulk email system, it is important for the knowledge management system to collect contexts in real time. For example, updating employees' time schedules and controlling the quantities of information they receive according to that.

"Exceptions" — the revision of general guidelines in practice — are normal in workplace [166]. Indeed, the stakeholders in organizational bulk email systems do not completely follow rational standards. Whether and how a bulk message would be sent not only depends on its objective value, but also the perceptions and level of its senders. For example, in Chapter 5, the message *"UMD Chancellor Search"*, though mainly relevant to a single campus, was sent to the employees across all the campuses as a single email, because the sender perceived the value of their message being distributed widely. Therefore organizational bulk email systems should have back-channels that allow stakeholders to negotiate when an exception can be made and evaluate the impacts of exceptions (e.g., the impact on channel reputation of sending a good-to-know message to all the employees).

8.2.2 Key Aspects for Designing Organizational Bulk Email Systems [94]

Given the nuances, fluidity, and exceptions in organizational bulk email system, the agents designed to support this system should be socially-embedded, which means that it will 1) analyze the social contexts of this system; 2) give the right incentive to each stakeholder that can steer their actions toward the whole organization's benefit. It should consider the employees' time cost of reading bulk emails, the communicators' time cost of designing and targeting bulk emails, and the leaders' value of getting important information out. Specifically, the agent should consider these aspects: organizational structure, economics, data sources, and decision-making process.

For organizational structure, the agent need be aware of different units and their responsibilities (e.g., the University Relation department is in charge of our study site's internal / external communications), the management style (top-down, bottom-up, distributed, etc.), the key stakeholder groups, and the IT infrastructure and its limitations (see Chapter 3).

Then the agent needs to understand the general economics of the bulk email system — all the stakeholders' value and costs with the corresponding bulk messages, such as employees' value of receiving job-relevant or interesting messages, employees' time costs, communicators' time costs, information producers' value of letting employees know about important messages, etc (see Chapter 4).

Given all the data collected above, a socially-embedded agent would look at each of these costs and value in providing feedback and suggestions on how to achieve the disparate goals of the organization — maximizing the desired value while minimizing both the cost and the loss of reputation (and long-term value) if employees become less likely to read future messages. For example, a bulk email agent can select communication channels for a message according to the channels' effectiveness and the messages' potential impact on the channels' reputation. The agent can also personalize bulk emails based on all the involved stakeholders' value (Chapter 5). We mainly explored personalizing message order in this thesis. Besides that, the agent can also personalize distribution mechanisms by sending message to different defined communication groups via different channels. For example, Beringer designed a system with a user interface that enabled the organization senders to send a message of a certain message type within one of the collaborative conversation channels [15]. The agent can also personalize the sending time by predicting how busy or interrupted the corresponding employee is when they receive the message Danninger et al..

Besides personalization, the agent can influence the university leaders' and central offices' communication behaviors by tracking performance and controlling the budget (Chapter 7). We explored visualizing costs in this thesis. The other way is to formalize this through budget allocations – units could have communications budgets that force them to make decisions about what's worth communicating broadly (and what's not).

In short, with this thesis, we now understand organizational bulk email systems' stakeholders' value, and how personalization and visibility tools could be used to influence employees' and senders' behavior toward the whole organization's benefits. And give the nuances, fluidity, exceptions, and conflicts in organizational bulk email system, there is still work to be done to manage this system's knowledge, quantify its metrics, and make decisions on designing / targeting organizational bulk emails.

References

[1] Alan S Abrahams, Tarun Chaudhary, and Jason K Deane. 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): 187–197, 2010.

[2] Mark S Ackerman. The intellectual challenge of cscw: the gap between social requirements and technical feasibility. *Human–Computer Interaction*, 15(2-3): 179–203, 2000.

[3] Imtiaz Ahmed, Ineen Sultana, Sanjoy Kumar Paul, and Abdullahil Azeem. Employee performance evaluation: A fuzzy approach. *International Journal of Productivity and Performance Management*, 2013.

[4] Keno Albrecht, Nicolas Burri, and Roger Wattenhofer. Spamato-an extendable spam filter system. In *CEAS*. Citeseer, 2005.

[5] Tarfah Alrashed, Chia-Jung Lee, Peter Bailey, Christopher Lin, Milad Shokouhi, and Susan Dumais. Evaluating user actions as a proxy for email significance. In *The World Wide Web Conference*, pages 26–36, 2019.

[6] Jorg Andriof, Sandra Sutherland Rahman, Sandra Waddock, and Bryan Husted. Introduction: Jcc theme issue: Stakeholder responsibility. *The Journal of Corporate Citizenship*, pages 16–19, 2002.

[7] Ioannis Arapakis, Mounia Lalmas, and George Valkanas. Understanding within-content engagement through pattern analysis of mouse gestures. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, CIKM '14, page 1439–1448, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450325981. doi: 10.1145/2661829.2661909. URL `https://doi.org/10.1145/2661829.2661909`.

[8] Guy Aridor, Duarte Gonçalves, Daniel Kluver, Ruoyan Kong, and Joseph Konstan. The economics of recommender systems: Evidence from a field experiment on movielens. In *ACM Transactions on Economics and Computation (EC2023).*, 2023.

[9] Olivier Augereau, Kai Kunze, Hiroki Fujiyoshi, and Koichi Kise. Estimation of english skill with a mobile eye tracker. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, pages 1777–1781, 2016.

[10] Kathryn A Baker. Organizational communication. *Management Benchmark Study*, pages 1–3, 2007.

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

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

[13] Ron Bekkerman. Automatic categorization of email into folders: Benchmark experiments on enron and sri corpora. 2004.

[14] Hal Berghel. Email—the good, the bad, and the ugly. *Communications of the ACM*, 40(4):11–15, 1997.

[15] Joerg Beringer. Collaborative conversation channels, February 21 2012. US Patent 8,122,084.

[16] Gregg Bernstein and Laurissa Wolfram-Hvass. Research methodologies, data collection, and analysis at mailchimp: A case study. In *Contemporary Research Methods and Data Analytics in the News Industry*, pages 170–189. IGI Global, 2015.

[17] Hugh Beyer and Karen Holtzblatt. Contextual design. *interactions*, 6(1):32–42, 1999.

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

[19] Aukse Blazenaite. Effective organizational communication: In search of a system. *Social Sciences*, 74(4):84–101, 2012.

[20] Martin Bohme, André Meyer, Thomas Martinetz, and Erhardt Barth. Remote eye tracking: State of the art and directions for future development. In *Proc. of the 2006 Conference on Communication by Gaze Interaction (COGAIN)*, pages 12–17, 2006.

[21] Virginia Braun and Victoria Clarke. Thematic analysis. 2012.

[22] David A Buchanan. Case studies in organizational research. *Qualitative organizational research*, pages 351–370, 2012.

[23] Georg Buscher, Edward Cutrell, and Meredith Ringel Morris. What do you see when you're surfing? using eye tracking to predict salient regions of web pages. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 21–30, 2009.

[24] Georg Buscher, Ralf Biedert, Daniel Heinesch, and Andreas Dengel. Eye tracking analysis of preferred reading regions on the screen. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '10, page 3307–3312, New York, NY, USA, 2010. Association for Computing Machinery. ISBN 9781605589305. doi: 10.1145/1753846.1753976. URL https://doi.org/10.1145/1753846.1753976.

[25] Mario Callegaro, Yelena Kruse, Melanie Thomas, and Poom Nukulkij. 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, American Statistical Association*, volume 5, pages 764–78, 2009.

[26] G Cardillo. Holm-sidak t-test: a routine for multiple t-test comparisons. *Disponible en Matlab Central, en la página de la Red Mundial: http://www. mathworks. de/matlabcentral/fileexchange/12786 (Consultado el 22 de Noviembre 2009)*, 2006.

[27] Carla Carvalho, Alipio M Jorge, and Carlos Soares. 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)*, pages 724–727. IEEE, 2006.

[28] Marta E Cecchinato, Jon Bird, and Anna L Cox. 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), 2014.

[29] Kathy Charmaz. *Constructing grounded theory.* Sage, 2014.

[30] Mon Chu Chen, John R Anderson, and Myeong Ho Sohn. What can a mouse cursor tell us more? correlation of eye/mouse movements on web browsing. In *CHI'01 extended abstracts on Human factors in computing systems*, pages 281–282, 2001.

[31] Nina Chen, Cassandra Kane, and Elisa Zhao Hang. *Evaluating Automated System Interventions Against Email Harassment.* Association for Computing Machinery, New York, NY, USA, 2021. ISBN 9781450380959. URL https://doi.org/10.1145/3411763.3451524.

[32] Ziheng Chen and Hongshik Ahn. Item response theory based ensemble in machine learning. *International Journal of Automation and Computing*, 17:621–636, 2020.

[33] Ziheng Chen, Fabrizio Silvestri, Gabriele Tolomei, Jia Wang, He Zhu, and Hongshik Ahn. Explain the explainer: Interpreting model-agnostic counterfactual explanations of a deep reinforcement learning agent. *IEEE Transactions on Artificial Intelligence*, 2022.

[34] Ziheng Chen, Fabrizio Silvestri, Jia Wang, He Zhu, Hongshik Ahn, and Gabriele Tolomei. Relax: Reinforcement learning agent explainer for arbitrary predictive models. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pages 252–261, 2022.

[35] Mail Chimp. Average email campaign stats of mailchimp customers by industry, 2018.

[36] V Christina, S Karpagavalli, and G Suganya. A study on email spam filtering techniques. *International Journal of Computer Applications*, 12(1):0975–8887, 2010.

[37] Charles Conrad and Michael Sollitto. History of organizational communication. *The international encyclopedia of organizational communication*, pages 1–32, 2017.

[38] Lorrie Faith Cranor and Brian A LaMacchia. Spam! *Communications of the ACM*, 41(8):74–83, 1998.

[39] Danilo Cruz and Chris Fill. Evaluating viral marketing: isolating the key criteria. *Marketing Intelligence & Planning*, 2008.

[40] Yiming Cui. Faq: Feature aggregated queries for transformer-based video object detectors. *arXiv preprint arXiv:2303.08319*, 2023.

[41] Yiming Cui, Liqi Yan, Zhiwen Cao, and Dongfang Liu. Tf-blender: Temporal feature blender for video object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8138–8147, 2021.

[42] Laura Dabbish, Gina Venolia, and JJ Cadiz. Marked for deletion: An analysis of email data. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '03, page 924–925, New York, NY, USA, 2003. Association for

Computing Machinery. ISBN 1581136374. doi: 10.1145/765891.766073. URL https://doi.org/10.1145/765891.766073.

[43] Laura A. Dabbish and Robert E. Kraut. 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, page 431–440, New York, NY, USA, 2006. Association for Computing Machinery. ISBN 1595932496. doi: 10.1145/1180875.1180941. URL https://doi.org/10.1145/1180875.1180941.

[44] Maria Danninger, Leila Takayama, Qianying Wang, Courtney Schultz, Jörg Beringer, Paul Hofmann, Frankie James, and Clifford Nass. Can you talk or only touch-talk: A voip-based phone feature for quick, quiet, and private communication. In *Proceedings of the 9th international conference on Multimodal interfaces*, pages 154–161, 2007.

[45] Thomas H Davenport and John C Beck. The attention economy. *Ubiquity*, 2001 (May):1–es, 2001.

[46] Roger Dawkins. Mass email at university: current literature and tactics for future use. *Open Learning: The Journal of Open, Distance and e-Learning*, 0(0):1–17, 2018. doi: 10.1080/02680513.2018.1556090. URL https://doi.org/10.1080/02680513.2018.1556090.

[47] Mathias Dewatripont and Jean Tirole. Modes of communication. *Journal of political economy*, 113(6):1217–1238, 2005.

[48] Fan Du, Sana Malik, Eunyee Koh, and Georgios Theocharous. Interactive campaign planning for marketing analysts. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI EA '18, page 1–6, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450356213. doi: 10.1145/3170427.3188531. URL https://doi.org/10.1145/3170427.3188531.

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

[50] Danyel Fisher, AJ Brush, Eric Gleave, and Marc A Smith. Revisiting whittaker & sidner's" email overload" ten years later. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, pages 309–312, 2006.

[51] Gregory Funke, Eric Greenlee, Martha Carter, Allen Dukes, Rebecca Brown, and Lauren Menke. Which eye tracker is right for your research? performance evaluation of several cost variant eye trackers. In *Proceedings of the Human Factors and Ergonomics Society annual meeting*, volume 60, pages 1240–1244. SAGE Publications Sage CA: Los Angeles, CA, 2016.

[52] Tushaar Gangavarapu, CD Jaidhar, and Bhabesh Chanduka. Applicability of machine learning in spam and phishing email filtering: review and approaches. *Artificial Intelligence Review*, 53(7), 2020.

[53] Andrew Garbett, Rob Comber, Edward Jenkins, and Patrick Olivier. App movement: A platform for community commissioning of mobile applications. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 26–37, 2016.

[54] Agostino Gibaldi, Mauricio Vanegas, Peter J Bex, and Guido Maiello. Evaluation of the tobii eyex eye tracking controller and matlab toolkit for research. *Behavior research methods*, 49(3):923–946, 2017.

[55] Vijai N Giri and B Pavan Kumar. Assessing the impact of organizational communication on job satisfaction and job performance. *Psychological Studies*, 55(2): 137–143, 2010.

[56] Mackenzie G Glaholt and Eyal M Reingold. Eye movement monitoring as a process tracing methodology in decision making research. *Journal of Neuroscience, Psychology, and Economics*, 4(2):125, 2011.

[57] Mary Ann Glynn. When cymbals become symbols: Conflict over organizational identity within a symphony orchestra. *Organization science*, 11(3):285–298, 2000.

[58] Gerald M Goldhaber. Organizational communication. Technical report, WCB,, 1990.

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

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

[61] Xiaolei Guo, Weihuang Xu, Yiming Cui, Jason Bonnette, Roser Matamala, Felix B Fritschi, Thomas E Juenger, and Alina Zare. Human-centered machine learning for plant root analysis and characterization. In *ASA, CSSA, SSSA International Annual Meeting*. ASA-CSSA-SSSA, 2022.

[62] Mari Hartemo. Email marketing in the era of the empowered consumer. *Journal of Research in Interactive Marketing*, 2016.

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

[64] David Hauger, Alexandros Paramythis, and Stephan Weibelzahl. Using browser interaction data to determine page reading behavior. In *International conference on user modeling, adaptation, and personalization*, pages 147–158. Springer, 2011.

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

[66] Yugo Hayashi. On pedagogical effects of learner-support agents in collaborative interaction. In *Intelligent Tutoring Systems: 11th International Conference, ITS 2012, Chania, Crete, Greece, June 14-18, 2012. Proceedings 11*, pages 22–32. Springer, 2012.

[67] Peizhou He, Yong Sun, Wei Zheng, and Xiangming Wen. Filtering short message spam of group sending using captcha. In *First International Workshop on Knowledge Discovery and Data Mining (WKDD 2008)*, pages 558–561. IEEE, 2008.

[68] Yunzhong He, Yuxin Tian, Mengjiao Wang, Feier Chen, Licheng Yu, Maolong Tang, Congcong Chen, Ning Zhang, Bin Kuang, and Arul Prakash. Que2engage: Embedding-based retrieval for relevant and engaging products at facebook marketplace. In *Companion Proceedings of the ACM Web Conference 2023*, WWW '23 Companion, page 386–390, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9781450394192. doi: 10.1145/3543873.3584633. URL `https://doi.org/10.1145/3543873.3584633`.

[69] Yunzhong He, Cong Zhang, Ruoyan Kong, Chaitanya Kulkarni, Qing Liu, Ashish Gandhe, Amit Nithianandan, and Arul Prakash. Hiercat: Hierarchical query categorization from weakly supervised data at facebook marketplace. In *Companion Proceedings of the ACM Web Conference 2023*, WWW '23 Companion, page 331–335, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9781450394192. doi: 10.1145/3543873.3584622. URL `https://doi.org/10.1145/3543873.3584622`.

[70] Silvia Heinz and Elisa D Mekler. 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*, pages 803–804, 2012.

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

[72] Karen Holtzblatt and Hugh Beyer. *Contextual design: defining customer-centered systems*. Elsevier, 1997.

[73] Chien-Chih Huang, Tung-Ching Lin, and Kuei-Ju Lin. Factors affecting pass-along email intentions (paeis): Integrating the social capital and social cognition theories. *Electronic Commerce Research and Applications*, 8(3):160–169, 2009.

[74] Jeff Huang, Ryen White, and Georg Buscher. User see, user point: gaze and cursor alignment in web search. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1341–1350, 2012.

[75] Jeff Huang, Ryen W White, Georg Buscher, and Kuansan Wang. Improving searcher models using mouse cursor activity. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 195–204, 2012.

[76] Bashima Islam, Yubo Luo, and Shahriar Nirjon. Amalgamated intermittent computing systems. *arXiv preprint arXiv:2303.13000*, 2023.

[77] Thomas Jackson, Ray Dawson, and Darren Wilson. The cost of email interruption. *Journal of systems and information technology*, 2001.

[78] Thomas W. Jackson, Ray Dawson, and Darren Wilson. Understanding email interaction increases organizational productivity. *Commun. ACM*, 46(8):80–84, August 2003. ISSN 0001-0782. doi: 10.1145/859670.859673. URL `http://doi.acm.org/10.1145/859670.859673`.

[79] Thomas W Jackson, Anthony Burgess, and Janet Edwards. A simple approach to improving email communication. *Communications of the ACM*, 49(6):107–109, 2006.

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

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

[82] Michael C Jensen and William H Meckling. Specific and general knowledge, and organizational structure. In *Knowledge management and organizational design*, pages 17–38. Routledge, 2009.

[83] Lixin Jiang and Tahira M Probst. Organizational communication: A buffer in times of job insecurity? *Economic and Industrial Democracy*, 35(3):557–579, 2014.

[84] Elizabeth Jones, Bernadette Watson, John Gardner, and Cindy Gallois. Organizational communication: Challenges for the new century. *Journal of Communication*, 54(4):722–750, 2004.

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

[86] A Kassambara. Logistic regression assumptions and diagnostics in r. *Statistical tools for highthroughput data analysis*, 2018.

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

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

[89] M Laeeq Khan. Social media engagement: What motivates user participation and consumption on youtube? *Computers in human behavior*, 66:236–247, 2017.

[90] Yanghee Kim, Amy L Baylor, and Entong Shen. Pedagogical agents as learning companions: the impact of agent emotion and gender. *Journal of Computer Assisted Learning*, 23(3):220–234, 2007.

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

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

[93] Nicolas Kokkalis, Chengdiao Fan, Johannes Roith, Michael S. Bernstein, and Scott Klemmer. *MyriadHub: Efficiently Scaling Personalized Email Conversations with Valet Crowdsourcing*, page 73–84. Association for Computing Machinery, New York, NY, USA, 2017. ISBN 9781450346559. URL `https://doi.org/10.1145/3025453.3025954`.

[94] Ruoyan Kong and Joseph A Konstan. Socially-embedded agents in organizational contexts – bulk email as an example. In *Intelligent Systems in the Workplace: Design, Applications, and User Experience*. Springer Publishing, 2023.

[95] Ruoyan Kong, Haiyi Zhu, and Joseph Konstan. Organizational bulk email systems: Their role and performance in remote work. *New Future of Work*, 2020.

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

[97] Ruoyan Kong, Ruobing Wang, and Zitao Shen. Virtual reality system for invasive therapy. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 689–690. IEEE, 2021.

[98] Ruoyan Kong, Haiyi Zhu, and Joseph A Konstan. Learning to ignore: A case study of organization-wide bulk email effectiveness. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1):1–23, 2021.

[99] Ruoyan Kong, Charles Chuankai Zhang, Ruixuan Sun, Vishnu Chhabra, Tanushsrisai Nadimpalli, and Joseph A. Konstan. Multi-objective personalization in multi-stakeholder organizational bulk e-mail: A field experiment. *Proc. ACM Hum.-Comput. Interact.*, 6(CSCW2), nov 2022. doi: 10.1145/3555641. URL https://doi.org/10.1145/3555641.

[100] Ruoyan Kong, Ruixuan Sun, Charles Chuankai Zhang, Chen Chen, Sneha Patri, Gayathri Gajjela, and Joseph A. Konstan. 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*, ETRA '23, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9798400701504. doi: 10.1145/3588015.3588404. URL https://doi.org/10.1145/3588015.3588404.

[101] Ruoyan Kong, Irene Ye Yuan, Ruixuan Sun, Charles Chuankai Zhang, and Joseph A Konstan. Commtool: Supporting organizations evaluate bulk emails (under review). 2023.

[102] Joseph A Konstan and Ruoyan Kong. The challenge of organizational bulk email systems: Models and empirical studies. In *The Elgar Companion to Information Economics*. Edward Elgar Publishing, 2023.

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

[104] Peter Kueng. Process performance measurement system: a tool to support process-based organizations. *Total quality management*, 11(1):67–85, 2000.

[105] Dmitry Lagun and Eugene Agichtein. Inferring searcher attention by jointly modeling user interactions and content salience. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 483–492, 2015.

[106] Meng-Lung Lai, Meng-Jung Tsai, Fang-Ying Yang, Chung-Yuan Hsu, Tzu-Chien Liu, Silvia Wen-Yu Lee, Min-Hsien Lee, Guo-Li Chiou, Jyh-Chong Liang, and Chin-Chung Tsai. A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational research review*, 10:90–115, 2013.

[107] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. *Research methods in human-computer interaction*. Morgan Kaufmann, 2017.

[108] Seulki Lee, Bashima Islam, Yubo Luo, and Shahriar Nirjon. Intermittent learning: On-device machine learning on intermittently powered system. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(4):1–30, 2019.

[109] Laurie K Lewis. An organizational stakeholder model of change implementation communication. *Communication Theory*, 17(2):176–204, 2007.

[110] Yixuan Li, Pingmei Xu, Dmitry Lagun, and Vidhya Navalpakkam. Towards measuring and inferring user interest from gaze. In *Proceedings of the 26th International Conference on World Wide Web Companion*, pages 525–533, 2017.

[111] Dongfang Liu, Yiming Cui, Wenbo Tan, and Yingjie Chen. Sg-net: Spatial granularity network for one-stage video instance segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9816–9825, 2021.

[112] Dongfang Liu, Yiming Cui, Liqi Yan, Christos Mousas, Baijian Yang, and Yingjie Chen. Densernet: Weakly supervised visual localization using multi-scale feature aggregation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 6101–6109, 2021.

[113] Yiqun Liu, Ye Chen, Jinhui Tang, Jiashen Sun, Min Zhang, Shaoping Ma, and Xuan Zhu. Different users, different opinions: Predicting search satisfaction with mouse movement information. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '15, page 493–502, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450336215. doi: 10.1145/2766462.2767721. URL https://doi.org/10.1145/2766462.2767721.

[114] Jie Lu, Zhen Wen, Shimei Pan, and Jennifer Lai. Epic: a multi-tiered approach to enterprise email prioritization. In *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces*, pages 199–202, 2012.

[115] Xiaoling Luo, Xiaobo Ma, Matthew Munden, Yao-Jan Wu, and Yangsheng Jiang. A multisource data approach for estimating vehicle queue length at metered on-ramps. *Journal of Transportation Engineering, Part A: Systems*, 148(2):04021117, 2022.

[116] Yubo Luo and Shahriar Nirjon. Spoton: Just-in-time active event detection on energy autonomous sensing systems. *Brief Presentations Proceedings (RTAS 2019)*, 9, 2019.

[117] Yubo Luo and Shahriar Nirjon. Smarton: Just-in-time active event detection on energy harvesting systems. In *2021 17th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, pages 35–44. IEEE, 2021.

[118] Yubo Luo, Le Zhang, Zhenyu Wang, and Shahriar Nirjon. Efficient multitask learning on resource-constrained systems. *arXiv preprint arXiv:2302.13155*, 2023.

[119] Xiaobo Ma. *Traffic Performance Evaluation Using Statistical and Machine Learning Methods*. PhD thesis, The University of Arizona, 2022.

[120] Xiaobo Ma, Abolfazl Karimpour, and Yao-Jan Wu. Statistical evaluation of data requirement for ramp metering performance assessment. *Transportation Research Part A: Policy and Practice*, 141:248–261, 2020.

[121] Ronald Maier and Thomas Hadrich. Knowledge management systems. In *Encyclopedia of Knowledge Management, Second Edition*, pages 779–790. IGI Global, 2011.

[122] Eerikki Mäki, Eila Järvenpää, Kirsi Ziegler, and Kirsi Ziegler. Communication and knowledge sharing in a decentralized organization. 2004.

[123] Gloria Mark, Shamsi T Iqbal, Mary Czerwinski, Paul Johns, Akane Sano, and Yuliya Lutchyn. 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*, pages 1717–1728, 2016.

[124] Frank Merten and Peter Gloor. Too much e-mail decreases job satisfaction. *Procedia-Social and Behavioral Sciences*, 2(4):6457–6465, 2010.

[125] Jörg Metzger, Michael Schillo, and Klaus Fischer. A multiagent-based peer-to-peer network in java for distributed spam filtering. In *International Central and Eastern European Conference on Multi-Agent Systems*, pages 616–625. Springer, 2003.

[126] Sabah Mohammed, Osama Mohammed, Jinan Fiaidhi, Simon Fong, and Tai Hoon Kim. Classifying unsolicited bulk email (ube) using python machine learning techniques. *International Journal of Hybrid Information Technology*, 6(1):43–56, 2013.

[127] Campaign Monitor. Campaign monitor, 2012.

[128] Paul B Moody. Reinventing email. *IBM Research, Cambridge. CSCW*, 2002.

[129] Gareth Morgan. Images of organization. thousands oaks. *Cal.: Sage Publications*, 1997.

[130] Nigel Morgan, Ella Hastings, and Annette Pritchard. Developing a new dmo marketing evaluation framework: The case of visit wales. *Journal of vacation marketing*, 18(1):73–89, 2012.

[131] Moses Namara and Curtis John Laurence. What do you see? an eyetracking study of a tailored facebook interface for improved privacy support. In *ACM Symposium on Eye Tracking Research & Applications (ETRA)(2019)*, volume 7, 2019.

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

[133] Francesco Di Nocera, Orlando Ricciardi, and James F Juola. Rapid serial visual presentation: Degradation of inferential reading comprehension as a function of speed. *International Journal of Human Factors and Ergonomics*, 5(4):293–303, 2018.

[134] Jill M Norris, Deborah E White, Lorelli Nowell, Kelly Mrklas, and Henry T Stelfox. How do stakeholders from multiple hierarchical levels of a large provincial health system define engagement? a qualitative study. *Implementation Science*, 12(1):98, 2017.

[135] Youssef Othman, Mahmoud Khalaf, Ahmed Ragab, Ahmed Salaheldin, Reham Ayman, and Nada Sharaf. Eye-to-eye: Towards visualizing eye gaze data. In *2020 24th International Conference Information Visualisation (IV)*, pages 729–733. IEEE, 2020.

[136] Soya Park, Amy X. Zhang, Luke S. Murray, and David R. Karger. Opportunities for automating email processing: A need-finding study. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–12, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450359702. doi: 10.1145/3290605.3300604. URL `https://doi.org/10.1145/3290605.3300604`.

[137] Shortest Path. Communication systems. 1968.

[138] Kirti Rajhans. Effective organizational communication: A key to employee motivation and performance. *Interscience Management Review*, 2(2):81–85, 2012.

[139] Justin M Rao and David H Reiley. The economics of spam. *Journal of Economic Perspectives*, 26(3):87–110, 2012.

[140] Thomas Rasmussen and Dave Ulrich. Learning from practice: how hr analytics avoids being a management fad. *Organizational Dynamics*, 44(3):236–242, 2015.

[141] Mary Elizabeth Raven and Alicia Flanders. Using contextual inquiry to learn about your audiences. *SIGDOC Asterisk J. Comput. Doc.*, 20(1):1–13, February 1996. ISSN 0731-1001. doi: 10.1145/227614.227615. URL `https://doi.org/10.1145/227614.227615`.

[142] Keith Rayner and Monica S Castelhano. Eye movements during reading, scene perception, visual search, and while looking at print advertisements. 2008.

[143] Kyle Rector and Joshua Hailpern. Minemail: Sms alert system for managing critical emails. CHI '14, page 783–792, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450324731. URL `https://doi.org/10.1145/2556288.2557182`.

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

[145] Ruth Rettie. Email marketing: success factors. 2002.

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

[147] Kevin Ruck and Mary Welch. Valuing internal communication; management and employee perspectives. *Public relations review*, 38(2):294–302, 2012.

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

[149] Natalie Sappleton and Fernando Lourenço. 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):611–622, 2016. doi: 10.1080/13645579.2015.1078596. URL https://doi.org/10.1080/13645579.2015.1078596.

[150] Bahareh Sarrafzadeh, Ahmed Hassan Awadallah, Christopher H. Lin, Chia-Jung Lee, Milad Shokouhi, and Susan T. Dumais. Characterizing and predicting email deferral behavior. In *WSDM '19*, 2019.

[151] Randall S. Schuler. A role perception transactional process model for organizational communication-outcome relationships. *Organizational Behavior and Human Performance*, 23(2):268 – 291, 1979. ISSN 0030-5073. doi: https://doi.org/10.1016/0030-5073(79)90058-8. URL http://www.sciencedirect.com/science/article/pii/0030507379900588.

[152] Christin Seifert, Annett Mitschick, Jörg Schlötterer, and Raimund Dachselt. Focus paragraph detection for online zero-effort queries: Lessons learned from eye-tracking data. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, pages 301–304, 2017.

[153] Stephan Serrano. 2020 email marketing roi statistics: Open rate to revenue. 2020. URL https://www.barilliance.com/email-marketing-statistics/.

[154] ABM Fahim Shahriar, Mahedee Zaman Moon, Hasan Mahmud, and Kamrul Hasan. Online product recommendation system by using eye gaze data. In *Proceedings of the International Conference on Computing Advancements*, pages 1–7, 2020.

[155] Zitao Shen, Dalton Schutte, Yoonkwon Yi, Anusha Bompelli, Fang Yu, Yanshan Wang, and Rui Zhang. 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):1–11, 2022.

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

[157] Alfonso Siano, Agostino Vollero, and Maria Palazzo. Exploring the role of online consumer empowerment in reputation building: research questions and hypotheses. *Journal of brand management*, 19(1):57–71, 2011.

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

[159] Lavneet Singh and Girija Chetty. 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*, pages 302–309. Springer, 2015.

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

[161] Laimona Šliburytė. Internal communication in organizations undergoing change. *Management of Organizations: Systematic Research*, (29), 2004.

[162] Lee Sproull and Sara Kiesler. Computers, networks and work. *Scientific American*, 265(3):116–127, 1991.

[163] Cynthia Stohl. *Organizational communication*. Number 5. Sage, 1995.

[164] Cynthia Stohl and W Charles Redding. Messages and message exchange processes. 1987.

[165] Igor M Stubarev, Anton I Belov, and Olga K Alsova. Development of the analytical platform for crm-system. In *2018 XIV International Scientific-Technical Conference on Actual Problems of Electronics Instrument Engineering (APEIE)*, pages 546–551. IEEE, 2018.

[166] Lucy Suchman and Eleanor Wynn. Procedures and problems in the office. *Office Technology and People*, 1984.

[167] I Made Sukarsa, Putu Wira Buana, and I Putu Krisna Arynasta. Software as a service: Design and build lower usage cost email marketing for hospitality industry. *Scientific Journal of Informatics*, 7(2):189–202, 2020.

[168] Ruixuan Sun, Ruoyan Kong, Qiao Jin, and Joseph A. Konstan. Less can be more: Exploring population rating dispositions with partitioned models in recommender systems. In *Adjunct Proceedings of the 31th ACM Conference on User Modeling, Adaptation and Personalization*, UMAP '23 Adjunct, New York, NY, USA, 2023. Association for Computing Machinery.

[169] Peter Swanborn. *Case study research: What, why and how?* Sage, 2010.

[170] Jinjin Tian and Aaditya Ramdas. Addis: an adaptive discarding algorithm for online fdr control with conservative nulls. *Advances in neural information processing systems*, 32, 2019.

[171] Jinjin Tian and Aaditya Ramdas. Online control of the familywise error rate. *Statistical methods in medical research*, 30(4):976–993, 2021.

[172] Yuxin Tian, Xueqing Deng, Yi Zhu, and Shawn Newsam. Cross-time and orientation-invariant overhead image geolocalization using deep local features. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2512–2520, 2020.

[173] Yuxin Tian, Shawn Newsam, and Kofi Boakye. Fashion image retrieval with text feedback by additive attention compositional learning. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1011–1021, 2023.

[174] Amrit Tiwana. *The knowledge management toolkit: practical techniques for building a knowledge management system*. Prentice hall PTR, 2000.

[175] Raluca Dania Todor. Marketing automation. *Bulletin of the Transilvania University of Brasov. Economic Sciences. Series V*, 9(2):87, 2016.

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

[177] Jaclyn Wainer, Laura Dabbish, and Robert Kraut. Should i open this email? inbox-level cues, curiosity and attention to email. In *Proceedings of the SIGCHI*

Conference on Human Factors in Computing Systems, CHI '11, page 3439–3448, New York, NY, USA, 2011. Association for Computing Machinery. ISBN 9781450302289. doi: 10.1145/1978942.1979456. URL `https://doi.org/10.1145/1978942.1979456`.

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

[179] Yu Wang, Yuying Zhao, Yi Zhang, and Tyler Derr. Collaboration-aware graph convolutional network for recommender systems. In *Proceedings of the ACM Web Conference 2023*, WWW '23, page 91–101, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9781450394161. doi: 10.1145/3543507.3583229. URL `https://doi.org/10.1145/3543507.3583229`.

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

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

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

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

[184] Alchemy Worx. Subject line gold–originality is the key to sustainable success. *Alchemy Worx*, 2016.

[185] Ya Xu, Nanyu Chen, Addrian Fernandez, Omar Sinno, and Anmol Bhasin. From infrastructure to culture: A/b testing challenges in large scale social networks. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 2227–2236, 2015.

[186] Liu Yang, Susan T Dumais, Paul N Bennett, and Ahmed Hassan Awadallah. Characterizing and predicting enterprise email reply behavior. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 235–244, 2017.

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

[188] Robert K Yin. Case study research: Design and methods (vol. 5). 2003.

[189] Charles Chuankai Zhang, Mo Houtti, C. Estelle Smith, Ruoyan Kong, and Loren Terveen. Working for the invisible machines or pumping information into an empty void? an exploration of wikidata contributors' motivations. *Proc. ACM Hum.-Comput. Interact.*, 6(CSCW1), apr 2022. doi: 10.1145/3512982. URL `https://doi.org/10.1145/3512982`.

[190] Yongfeng Zhang and Xu Chen. Explainable recommendation: A survey and new perspectives. *arXiv preprint arXiv:1804.11192*, 2018.

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

[192] Qian Zhao, Shuo Chang, F Maxwell Harper, and Joseph A Konstan. Gaze prediction for recommender systems. In *Proceedings of the 10th ACM Conference on Recommender Systems*, pages 131–138, 2016.

[193] John Zimmerman and Jodi Forlizzi. Research through design in hci. In *Ways of Knowing in HCI*, pages 167–189. Springer, 2014.