Prioritizing Unread Emails: An Experiment Showing that People Send Urgent Responses Before Important or Short Ones

Running Head: Prioritizing Unread Emails

ABSTRACT

People are overwhelmed by the volume of email that they receive. To ensure their emails are read, senders sometimes use explicit inbox-level cues in an attempt to garner the receiver's attention. We report the results of a field experiment that investigates whether and how such cues influence recipients' email processing behavior. Forty-five participants were sent 360 emails each over a three-week period. Inbox-level cues were given to indicate: (1) the urgency of responding, (2) the time that would be required to work on a response, (3) the importance of responding, (4) and the salience of that importance. Results show that email prioritization is influenced by an interaction between these cues. When emails were not time-sensitive, participants sensibly prioritized responses to messages that were most important and required the least effort to respond to. This rational triaging strategy faltered when emails required a time-sensitive response; urgent messages were responded to quickly regardless of other cues. The results are discussed with reference to Kahneman's dual-process theory of judgment and decision making.

CONTENTS

1. INTRODUCTION

Although email interfaces have not changed much in the past decade, how we manage our inbox has changed a lot, as a result of the growing number of messages we receive (The Radicati Group, 2015) across a growing collection of accounts (Cecchinato et al., 2015). This growth has increased the workload associated with tasks necessary to filter and manage the inbox and makes for an overwhelming experience known as email overload (Dabbish & Kraut, 2006; Eppler & Mengis, 2004). Email triage –"*the process of going through unhandled email and deciding what to do with it*" (Neustaedter et al., 2005, p. 1997)– is particularly challenging because, while some emails are critical, the majority are either irrelevant or do not require immediate attention (Buthpitiya et al., 2009). Thus, researchers have been arguing for more work that helps users reduce the burden of decision making associated with managing emails (Grandhi & Lanagan-Leitzel, 2016).

Whittaker et al. (2005) argued that people need help in identifying important messages and called for "*systems that support users in detecting and processing messages associated with important tasks" (p. 6).* However despite current solutions that rely on machine learning algorithms to cluster messages into priority/non-priority groups (e.g. Microsoft Outlook's Clutter and Gmail's Priority Inbox), users may still be receiving large numbers of messages within each of those clusters and thus may still need to "[sift] *through multiple messages attempting to determine how each message might relate to their outstanding tasks*" (Whittaker et al., 2005, p. 6).

Email etiquette often requires using clear and directive subject lines. Wainer et al. (2011) demonstrated that people use the subject line as a filter mechanism to help them decide which messages to prioritize. However, there are several ways in which the subject line can influence the decision process. For example, although previous work (Porter & Whitcomb, 2005; Sappleton & Lourenço, 2016; Wainer et al., 2011) has found that leaving a blank subject line or omitting information can help feed the receiver's curiosity to react to the message, Sappleton and Laurenço (2016) found that leaving a subject line blank was not always enough though to warrant a response.

To date, very little work has looked at how various email factors, such as importance, urgency and cost of reply, aid users in sifting through their emails and making appropriate decisions. Moreover, when these factors have been investigated it has often been done in an isolated manner, in the lab using simulated inboxes. The reality of managing one's inbox is more complicated and messier than is simulated in such investigations. It is therefore timely to explore how these factors interact to influence how people prioritize their emails.

To address our lack of understanding about how different factors in a subject line interact to affect the decision-making process of email replies (especially in a situated context), in this paper we report on a study with 45 participants, who collectively over a three-week period replied from their own inboxes to 16,200 emails we sent them. For each email we manipulated four email subject line cues (urgency, cost, importance, salience of importance) and measured the number of responses and response times. Our findings show that people prioritized responses to important messages, except when emails required a time-sensitive response; urgent messages were responded to quickly regardless of other cues. We make three novel contributions. First, we provide empirical evidence of how responses are prioritized based on explicit subject line cues. Second, we demonstrate how our field experiment can be used as a valid method to investigate the daily triage of emails in a situated context. Finally, we elaborate on implications for design that can inform how future email management systems should be designed.

2. RELATED WORK

2.1. Current Email Systems

Over the past few years, a variety of email support tools have been developed to help people quickly and easily identify important unread messages in their inbox e.g., Gmail's Priority Inbox, VIP Lists on iOS Mail, and Microsoft Outlook's Clutter. These approaches try to identify important emails based on user-defined information about the importance of the sender or system-based assumptions about a sender's importance based on the user's prior responsiveness to messages from that sender (indicated by whether or not a response was given and the speed of response). VIP Lists on iOS identifies emails coming from people that are important to the user (e.g., a boss, a close colleague, or

family member), so that these messages can be given attention first. Gmail's Priority Inbox extends this concept by using automated machine learning techniques to assign greater importance to emails that users have responded to quickly in the past. Microsoft Outlook's Clutter moves emails that are usually ignored or not responded to into a separate inbox. These new tools have moved beyond threading or clustering of related messages to making judgements about what is important to the receiver.

Whilst these tools are a step forward compared to those advocated by Whittaker et al. (2005), they are as yet unable to differentiate between messages from the same sender that relate to tasks of different importance. As a consequence, when people send an email, they often try to give explicit cues about the urgency, cost and importance of replying to their message in order to influence the responsiveness of the sender. For example, using a subject line, "URGENT: Can we meet today?" is giving a clear and explicit signal that the receiver can pick up on as they triage their unread messages, hopefully eliciting a timely response. Most email clients also allow senders to use and set explicit Priority Flags (e.g., "*!! High Priority*" vs. "*- Low Priority*"). A third but perhaps less frequently used cue is an indication of the type of response required: subject lines such as "*For Information Only*" and "*NNTR*" (No Need To Respond) indicate that the message can be read and then simply filed or deleted. Other cues such as "Quick short response required" are sometimes included in email subject lines with the aim of indicating that the response required will only take a very short time to compose. All of these cues are explicitly created by the sender of an email in an attempt to influence how the receiver prioritizes their responses to the many unread emails in their inbox.

Many factors can potentially determine if and when people decide whether to file, defer, delete, or respond to a message. In the email literature there are four particularly important factors that influence this triaging behavior (Wainer et al., 2011): the perceived *urgency* of giving a timely response; the *cost* of composing a response; the perceived *importance* of responding to an email; and the *salience* of that importance to a recipient. In the following subsections we discuss each of these in turn and argue that it is necessary to develop a detailed understanding of how these factors *interact* for the development of more efficient email management practices and support tools.

2.2 The Urgency of a Message

Prior research suggests that people prioritize emails that are time-sensitive and require an urgent response (Tyler & Tang, 2003). Receivers likely use information such as the identity of the sender in order to determine the urgency required (Siu et al., 2006). For example, an email from a senior colleague might elicit an urgent response. Indeed, this assumption underpins many email support tools, which mark messages from particular people as important. For instance, Google Priority Inbox ranks messages from people that are usually responded to as important, whereas VIP lists on iOS Mail allow the user to explicitly define 'important people'. However, a study by Karagiannis and Vojnovic (2009) found very little evidence of email response times being influenced by organizational seniority. This would suggest that when prioritizing responses to emails, people have a more nuanced strategy than to simply respond to their senior colleagues quickly. Instead, other features of the message must be used to determine the urgency of response required.

People sometimes give clear and explicit inbox-level cues about the urgency of response required to their message. For example, by using the subject line "*URGENT: Please review final draft of CHI paper (attached) prior to submission deadline*", the sender makes it clear to the receiver that they require a prompt and timely response. While urgency and importance often align, there are many cases when these two features are orthogonal. For instance, one might be cc'd on a message marked as "*URGENT*". While an urgent response is required by the primary recipient, the message does not require a timely response from those listed in the 'cc' line.

The question of how urgency cues affect email response behavior has not been given sufficient attention in the HCI literature. Is it that people assess *both* the urgency and the importance of a message when deciding whether to respond, or do urgent but less important messages garner undue attention? Research on human decision-making would suggest that decisions made in haste are instinctive and automatic, whereas decisions that are made slowly are rational and deliberative (Kahneman, 2011). Similarly, research on how people multitask has found that 'urgent' events are often prioritized over less urgent but more important ones (Bogunovich & Salvucci, 2011; Kerstholt, 1994). This suggests

that urgency trumps importance. However, it is not clear whether or how this cognitive bias extends to how people triage email; do people prioritize responses to messages that a sender has indicated require an urgent response? The results of the study presented here provide a detailed, empirically grounded understanding of how people trade-off explicit inbox-level cues about the expressed urgency and importance of responding to an email.

2.2. The Cost of Making a Response

Not all emails demand a response (Di Castro et al., 2016), but for those that do, responses can vary considerably. Some require only a one-word response, while others require a lengthy and engaged answer. This difference can then be compounded by the fact that email is also triaged on a variety of devices (Cecchinato et al., 2016; Collins et al., 2015). Writing a one-word response on a tiny smartphone keyboard is easy, but writing a long and complex response on the same device is difficult and time consuming. Matthews et al. (2009) show that the difficulty of composing messages on small mobile keyboards means that users often defer writing longer replies. Cecchinato et al. (2015) found that people tended to use their smartphones to reply to messages that were urgent and required only a short, quick response. Similar behaviors have been observed by (Kooti et al., 2015), who found that email responses from smartphones were the quickest and shortest, followed by tablets and then laptops/PCs.

In addition to the physical costs of typing replies, people consider the amount of work required in order to be able to respond to a given message. Karagiannis and Vojnovic (2009) found that people responded particularly slowly to emails that had an attachment, suggesting that people are sensitive to variations in the time costs of reading emails. Sinking a lot of time into responding to one complex email necessarily takes away from time that could be given over to responding to many messages that require only a brief and simple response. Human decision-making research shows that people can make these kinds of trade-offs about how to allocate effort over time in isolation (Jarvstad et al., 2012). We investigate here how people trade-off the cost of responding against multiple other critical factors (i.e., response urgency and importance).

2.3. The Importance of a Message

People seem to prioritize emails that have been designated as 'important' by the sender (Wainer et al., 2011). A sender can use inbox-level cues to try and signal the relative importance of a message, for instance, by communicating this in the subject line of the email or by using importance indicators found in many email clients (e.g., Priority Flags). In an experiment by Wainer et al. (2011), participants had to process emails to organize a fictional event. Messages that were marked as important by the sender always contained important information that was relevant to the organization of the event. Their study showed that people were more likely to respond to messages if they included this *reliable* indicator of importance.

However, importance flags can be misused. What is important to the sender of an email may not always be important to the recipient of that message. In another lab study, Kraut et al. (2002) found that sender-determined indicators of importance *tended not to be* used by the receiver to prioritize attention to messages. A possible explanation for this finding is that there is often considerable variability in whether and how people use importance indicators when sending email. As a result, it can be difficult to know whether an email that is flagged as important really is of importance to the recipient of that message, or whether it is an indicator that the sender thinks the email is important, or whether, in fact, the sender is trying to indicate that an urgent response is required. This prior research would therefore suggest that people can and do use importance indicators when prioritizing the unread emails that they receive, but only when these flags are used in a consistent and meaningful way by senders.

2.4. The Salience of a Message's Importance

The importance of a message is sometimes obvious to a recipient from the subject line. Sometimes, though, more engagement (such as reading the body of the message) is required before the importance of a message becomes clear. The relationship between the salience of a message's importance and people's processing is not always as intuitive as one might expect. When sending email, many strive to give accurate and explicit information in the subject line about the importance and urgency of their message. Paradoxically, Porter and Whitcomb (2005) found that blank subject lines actually

yielded the highest response rates. This finding was corroborated by Wainer et al. (2011), whose participants received messages with ambiguous subject lines. The 'information gap' created by the ambiguous subject lines in their study influenced how emails were prioritized: people attended to messages with ambiguous subject lines more quickly. Receiving a handful of emails with ambiguous or empty subject lines might effectively pique one's curiosity. But would this still be the case if many emails were received that had ambiguous or blank subject lines? We are interested in further investigating how people respond to how these different kinds of inbox level cues are used to indicate the importance of an email.

Whilst factors such as importance, urgency and length of response have been previously studied in independently, in real life settings these cues are often used in combination. However, no work to date has looked at how these factors interact with each other to influence email response behavior. Our research question is therefore: To what extent do people respond to competing subject line cues about importance, urgency, and cost of replying when managing their situated daily email?

3. AN INVESTIGATION OF THE INFLUENCE OF MULTIPLE INBOX LEVEL CUES ON EMAIL RESPONSE BEHAVIOR

In our review of the literature we identified four different types of inbox-level cue that people use when deciding how to prioritize responding to unread email: (1) the urgency of responding, (2) the cost of responding, (3) the importance of the message, and (4) the salience of that importance. These factors have often been studied separately and often in a lab setting, with participants doing an artificial email task that involves processing messages that have no personal meaning. However, we want to know how these factors interact with one another and their impact on *situated* email behaviour. This is important because the senders of emails often use multiple cues in combination to try and garner a fast response to the emails that they send. Taking inspiration from Wainer et al.'s (2011) lab-based experiment on email behavior, we developed and deployed a field experiment in which we sent people emails to their primary existing email addresses to find out which combination of inbox-level cues resulted in the most responsive behaviour, measured by whether or not they responded and how fast they responded. To do this, we

sent 45 participants 360 emails each over a three-week period and recorded their email response patterns. These experimenter-generated emails were sent to participants' main existing email accounts so that these emails would sit amongst the usual variety and quantity of email that our participants received. The external pressures on participants' lives limited their ability to respond to our emails and we therefore expected their responses to be selective, focused on those emails that captured their attention.

Our primary concern in conducting this field experiment was to answer our research question and learn to what extent people respond to competing subject line cues about urgency, cost of replying and importance when managing their situated daily email. Given the variety of factors that might influence people's decision making around inbox-management, and the need to understand these factors individually and collectively, we first make the following hypotheses about how these individual factors are likely to impact response rate (number of responses) and response time based on previous literature:

Urgency:

H1a– participants will respond to more emails where a less urgent response is required than those that require fast responses. This is because participants are more likely to miss the opportunity to respond, on time, to incoming emails that require fast responses.

H1b – participants will respond more quickly to incoming emails that require fast responses than those that have a longer response time.

Cost:

H2a– participants are more likely to respond to incoming emails that require a low cost response (i.e. require less effort to respond to) than those where the cost of making a response is higher.

H2b – participants will respond more quickly to incoming emails that require a lower cost response than those that require a high cost response.

Importance:

H3a – participants are more likely to respond to incoming emails that are marked as important than those marked as low importance.

H3b – participants will respond more quickly to incoming emails with a higher marked importance than those with a lower marked importance.

Salience:

H4a – participants are more likely to respond to incoming emails without an indicator of importance in the subject line than those with an indicator of importance due to the curiosity to discover how many points the message is worth.

H4b - similarly, participants will respond more quickly to incoming emails without an indicator of importance.

As we have already noted, these individual factors do not operate in isolation. Understanding the way that these factors interact with one another is therefore critical if we are to understand both the relative importance of each factor and the way that factors interact to amplify or diminish one another. We therefore pay close attention to the interactions in our analyses.

Our field experiment was designed to assess the relative influence of inbox-level cues that can be used by a sender in the subject line to communicate the importance, urgency, and cost of responding to their email. To do this, we needed an objective and simple way to operationalize each of these variables for the purposes of conducting an experiment. Sappleton and Laurenço (2016) suggested using incentives alongside blank subject lines to investigate response rates. We arrived at an approach in which participants received *points* for sending on-time responses to our emails. The benefit of using a points-based scheme is that it can be used to explicitly define, and then systematically vary, the relative importance of responding to different types of email. The purpose of our method was not to create a game per se, but simply to use points to indicate importance to receivers. We therefore did not include any other gamification elements. There is strong precedent for using point-based rewards in experimental HCI and psychology research to

operationalize the manipulation of value (Farmer et al., 2018; Gould et al., 2016; Schumacher et al., 1999). These studies have shown that points can be explicitly communicated to participants, and that participants adapt their behavior to maximize these rewards. Using this approach, responding promptly to emails that had high importance (as defined by the sender/experimenter) earned participants more points, whereas responding promptly to emails that were of low importance earned fewer points. Using the subject line of the email we were then also able to manipulate the *salience of importance* by either explicitly stating how many points were on offer, or by leaving this information unspecified, and thereby creating an information gap (Wainer et al., 2011). The subject line of the email was also used to communicate how quickly a response was required in order to earn these points (urgency), whereas the effort required to make give a response (cost) was only evident in the email body.

3.1. Method

Participants

Forty-five participants (31 male) with a mean age of 28 (*SD* = 5.81) were recruited via an online recruitment advertisement. Thirty of the participants were in full-time employment, nine were part-time workers, and six were full-time students. All participants were self-assessed 'high' users of email (60% receive more than 25 emails a day), and stated they check their inbox at least once every few hours.

Participants were motivated to take part in the study by the chance to receive one of three rewards. Two £50 rewards were allocated based on the highest number of points obtained through responding to the emails in the study. The third £50 reward was allocated at random to one participant who completed the study, to avoid drop-outs and maintain engagement.

Design

A 3×3×2×2 entirely within-subjects design was used, with the variables of: *Urgency* (20 minutes, 3 hours, and 24 hours), *cost* (low and high), *importance* (10 points, 30 points, and 100 points), and *the salience of importance* (low and high). The dependent variables were the number of on-time responses (response rate) and the response time.

Materials

We sent participants email, and they earned points for giving on-time responses to these messages. Emails contained information in the subject line as well as in the body of the message (Figures 1 and 2). The information included in the email subject line varied in three ways: urgency, importance, and the salience of importance.

(Figure 1 about here)

(Figure 2 about here)

***Urgency*.** Karagiannis and Vojnovic (2009) demonstrated that the majority of emails are responded to within 24 hours. The median response time is one hour. And 20% of emails are responded to within 5 minutes. In our experiment, three levels of urgency were specified in the subject lines, indicating how quickly the sender required a response. The highest level showed three crosses in the subject line "[+++]", which meant participants had 20 minutes to reply to the email in order to receive the points assigned to that message. The medium level showed two crosses "[++]", indicating a 3-hour window to reply, and the low level showed one cross "[+]", indicating a 24-hour window to reply. An on-time response meant that a participant replied to the email within the time window specified by the level of urgency and thereby received the points indicated in the subject line.

***Cost of responding*.** The cost of responding was indicated in the body of the email rather than the subject line. Unless specified, it is generally hard for users to know the cost of responding until a message is opened. In the low-cost condition, responding to an email involved the participant opening the email and copying and pasting a unique random code from the email body (Figure 2) into the subject line of a reply email. In the high-cost condition, participants were required to click on a link contained within the message (Figure 3). This took them to a website where they were required to rate the emotional content of a series of text messages (Figures 4 and 5). This task lasted two minutes. Only when this task was completed were participants provided with a unique random code, which then had to be copied and pasted into the subject line of a response email. Participants again earned the points on offer if this response email was received on-time.

(Figure 3 about here)

(Figure 4 about here)

(Figure 5 about here)

Importance. To manipulate importance, each email in our study was assigned one of three levels of points that would be earned by an on-time response. The highest level meant that an on-time response was worth 100 points, the medium level was worth 30 points, and the lowest level was worth 10 points. The use of points was to operationalize the importance of responding to an email. To earn the points on offer, the response email had to be received within the specified time window (20 minutes, 3 hours or 24 hours, depending on the level of urgency). In our experiment, points also served as a way to keep participants engaged throughout the three-week study as two of the £50 rewards were given to participants who earned the most points during the experiment.

The salience of importance. There were two levels of the salience of importance. When the salience of importance was high, the importance of an email was included in the subject line (see the second item in Figure 1) and the number of points for an on-time response was shown. When the salience of importance was low, the importance of an email did not appear in the subject line (see the first item in Figure 1). For both levels, the number of points a participant could collect by responding always appeared in the main body of the message (Figure 3).

Procedure

Each participant was sent demographic questionnaires to complete on the first day of the study. This was done to gauge their inbox size and their email management style. Questions included how many emails they had received that day and how many they had sent, along with a Likert scale to score how representative their answers were of their normal daily email workload.

Participants received emails in their existing main email account, every day between 9am and 9pm for three weeks, excluding weekends, from a consistent sender. There were 36 types of emails due to the 3 (urgency) × 3 (importance) × 2 (the salience of importance) × 2 (cost of responding) design of the experiment. All participants received

10 emails of each type, meaning 360 emails in total, over the 15-day study. This averaged out to 24 emails per day, with 28 emails being the highest number sent on any one-day and 20 the lowest. Participants responded to our emails using whichever strategies and devices they usually used for handling their email.

3.2. Results

From the initial survey about their usual email behavior, 66.7% of participants reported receiving on average between 10 and 49 emails per day, with 26.6% receiving more than 50 emails a day, and 6.7% reporting less than 10 emails a day. Participants reported that they *often* replied from their laptop (60.5%) or their smartphone (48.9%), followed by desktop PC (35%) and tablets (24.2%). We also asked participants to indicate which factors they considered when deciding to read an email: 57.8% stated that the sender was *very important*, and subject, date-received, and flagged-as-important were considered *important* by 61.4%, 44.2%, and 52.3% of participants, respectively.

Participation in the Study

We sent 16,200 emails to 45 participants, and participants responded to 65% of these emails (10,551 of 16,200). The total number of points available per participant was 16,800. The mean number of points participants earned was 14,073 (*SD* = 1,465). The top-10 scoring participants were all within 1,590 points of the leader, who scored 16,730.

To gain a better understanding of how each individual participant was engaging with the study, we considered participants' response rate over the duration of the study. These data are shown in Figure 6 (range: 0-100%). It can be seen in the figure that while many participants engaged with the study and responded to the majority of emails that were sent to them, some did not. On closer inspection, we found that 16 participants failed to respond to at least 50% of the emails that were sent to them. Moreover, these same participants also failed to respond to at least one email from each of the experimental conditions. This latter point is particularly problematic as it results in missing cells for the statistical analysis of data.

Excluded participants responded to far fewer emails (M = 28%) than those that were included in the main analysis (M = 85%). Here we quantify the level of non-responsiveness of these excluded participants. The experimental design has 36 cells (3×3×2×2 entirely within-subjects) and data from 16 participants was excluded – this makes 576 cells in total. Of these 576 cells, 205 (36%) had zero responses to the emails that were sent. One participant did not respond to any emails at all, and four participants had zero data in half of all cells in the experimental design. In other words, the scale of non-responsiveness amongst excluded participants was extremely high, making it impractical to exclude cases pairwise or impute data. Hence, we chose to exclude these 16 participants from all subsequent analysis of data. We return to this point in the discussion.

Having excluded 16 (of 45) participants, it is possible that the generalisability of the results might be affected if, for example, the participants who were excluded show entirely different patterns of response behavior. To allay this concern, at the end of the results section we report a descriptive analysis of email responses from participants that were excluded. The analysis necessarily focuses on reporting mean performance values; it is not possible to conduct a thorough statistical analysis due to the number of missing data points from these participants.

(Figure 6 about here)

Of the 29 remaining participants, the mean response rate was 85% (*SD* = 10%, range: 60-99%). The mean of participants' average response times (i.e., the mean of means) was 61 minutes (SD = 36 min). Our fastest participant took, on average, 18 minutes to respond to a message. Our slowest participant took an average of 3hrs 1min to respond. Of the 8,777 responses we received, only 228 (2.6%) were made within 60 seconds. It is therefore evident that emails were responded to during both work and non-work time throughout the period over which they were sent (9am to 9pm, Monday to Friday, for a period of three weeks). For these participants we investigate the effects of condition on the response rate and response time to emails.

Response Rate

We used the R statistical programming environment to perform a repeated measures ANOVA with a significance level of .05 to compare the main effects of urgency, cost, importance and salience of importance on number of responses and to understand the interaction of these main effects. We provide a summary of this analysis in Table 1.

(Table 1 about here)

First, we consider the effect of response urgency (i.e., how long participants had to respond to an email in order to gain points from responding to it). As expected (**H1a**), participants were significantly more likely to miss the opportunity to respond on time to incoming emails that require fast responses than those where a less urgent response is required (20 minute: M = 74%, SD = 22%; 3 hours: M = 88%, SD = 17%; 24 hours: M = 93%, SD = 14%), $F(2, 56) = 57.10$, $p < .001$, $\eta_p^2 = .67$. This presumably reflects the fact that once the response window to earn points from an email has been missed, there is simply no point in responding to it at all.

Second, we consider the effect of the cost (**H2a**) of response (i.e., the amount of effort and time that is required to respond to each email). We found that participants were more likely to respond to incoming emails that require a low cost response(M = 91%, SD = 14%) than to emails that had a high response cost, (M = 79%, SD = 22%) $F(1, 28) = 21.41$, $p < .001$, $\eta_p^2 = .43$.

Third, we consider the effect of the importance (**H3a**; i.e., the number of points that participants earned) for responding to an email on-time on response rate. As expected, participants were generally more likely to respond to emails that are marked as important than those marked as low importance. Reflecting this, there was a significant main effect of importance on response rate, $F(2, 56) = 10.79$, $p < .001$, $\eta_p^2 = .28$.

Fourth, we consider the effect of the salience (**H4a**) of importance on response rate (i.e., whether inbox-level cues revealed the importance of the email). Results showed that there was no significant main effect of the salience of importance on response rate. The hypothesis was therefore not supported.

We now turn our attention to the interactions between the variables. A significant urgency × cost interaction was found, $F(2, 56) = 4.22$, $p < .05$, $\eta_p^2 = .13$. This interaction is shown in Figure 7. To investigate this more thoroughly we conducted tests of the simple main effect of response cost across each of the different levels of urgency (applying Bonferroni corrections). It was found that the effect of response cost was robust across the manipulation of urgency: Participants were significantly more likely to respond to emails that had a low-cost than a high-cost, and this effect occurred when the response window was either 24-hours, $F(1, 28) = 13.91$, $p < .001$, $\eta_p^2 = .33$, 3-hours, $F(1, 28) = 19.46$, $p < .001$, $\eta_p^2 = .41$, or 20-minutes, $F(1, 28) = 20.29$, $p < .001$, $\eta_p^2 = .42$. In other words, the effect of cost of responding is robust across the manipulation of urgency.

(Figure 7 about here)

There was also a significant importance × cost interaction, $F(2, 56) = 6.11$, $p < .01$, $\eta_p^2 = .18$. This interaction can be seen in Figure 8. It shows that the effect of importance on response rates is moderated by the cost of responding. To investigate this interaction, we report the results of a simple main effects test, with Bonferroni corrections. Results show a significant simple effect of importance on response rates to emails that had a high response cost, $F(2, 27) = 8.31$, $p < .01$, $\eta_p^2 = .38$; participants were more likely to respond to high-cost emails that were worth more points. In contrast, there was no such simple effect of importance on response rate to emails that had a low response cost, $F(2, 27) = 3.21$, $p = .056$, $\eta_p^2 = .19$; participants tended to respond to most emails that were easy to respond to regardless of how many points were on offer.

(Figure 8 about here)

There was a significant urgency × salience interaction, $F(2, 56) = 3.22$, $p < .05$, $\eta_p^2 = .10$. This interaction can be seen in Figure 9. A simple main effects analysis of this interaction with Bonferroni corrections showed that when participants had a brief 20-minute window to respond to an email, then responses were more likely when there was of low-salience (i.e. no inbox-level cues indicating an email's importance), compared to

when this information was of high-salience (i.e., clear inbox-level cues were present), $F(1, 28) = 4.88$, $p < .05$, $\eta_p^2 = .15$. However, when participants had longer to respond (3- or 24-hours), there was no effect of the salience of importance on response rate, all F's < 1.

All other interactions were not significant.

(Figure 9 about here)

Response Time

Having analyzed the *number* of emails participants responded to, we next consider how *quickly* participants responded to messages. We used the R statistical programming environment to perform a repeated measures ANOVA with a significance level of .05 to compare the effects of urgency, cost, importance and salience of importance on response time. Table 1 provides a summary of this analysis.

First, in line with our prediction (H1b) we found that the urgency of an email affected how quickly participants responded to it. Responses to emails were significantly faster when there was a shorter response window (20 minute: M = 20 min, SD = 42 min; 3 hours: M = 38 min, SD = 38 min; 24 hours: M = 125 min, SD = 133 min), $F(2, 56) = 45.80$, $p < .001$, $\eta_p^2 = .62$.

Second, we found that participants were faster at responding to emails that had a low response cost (**H2b**; M = 37 min, SD = 56 min) than to emails that had a high response cost (M = 85 min, SD = 118 min), $F(1, 28) = 37.56$, $p < .001$, $\eta_p^2 = .57$.

Third, participants were faster at responding to emails that had greater importance (**H3b**). This was operationalized in the study by varying the number of points that were earned for responding to the email within the required time frame. We found that participants were significantly faster at responding to emails that were worth more points (100-points: M = 57 min, SD = 88 min; 30-points: M = 60 min, SD = 93 min; 10-points: M = 66 min, SD = 104 min), $F(2, 56) = 5.27$, $p < .01$, $\eta_p^2 = .16$.

Fourth, there was no effect of the salience (**H4b)** of importance on performance. Response times were similar regardless of whether the importance was missing (M = 61 min, SD = 97 min) or visible in the subject line of the email (M = 61 min, SD = 94 min), $F < 1$.

We now turn our attention to the interactions between this variables. It can be seen in Figure 10 that the effect of response cost was moderated by the urgency of response required. Indeed, statistical analysis found a significant urgency × cost interaction, $F(2, 56) = 29.61$, $p < .001$, $\eta_p^2 = .51$. A simple main effects analysis of this interaction with Bonferroni corrections show that when participants had at least 3 hours to respond to an email, there was an effect of response cost. That is, participants were faster at responding to emails that had a low-cost response than those that were high-cost in the 3 hour condition, $F(1, 28) = 26.31$, $p < .001$, $\eta_p^2 = .48$, and the 24 hour condition, $F(1, 28) = 37.40$, $p < .001$, $\eta_p^2 = .57$. However, when participants had a relatively brief 20-minute response window they were equally quick to respond to both the high- and low-cost messages. In other words, there is no significant simple effect of response cost when a time sensitive response was required.

(Figure 10 about here)

It can be seen in Figure 11 that the effect of importance was moderated by the urgency of response required. Statistical analysis found a significant urgency × importance interaction, $F(4, 112) = 3.12$, $p < .05$, $\eta_p^2 = .10$. A simple main effects analysis of this interaction with Bonferroni corrections show that when emails were less urgent and could be deferred for up to 24-hours, there was a significant effect of importance on response times, $F(2, 27) = 3.98$, $p < .05$, $\eta_p^2 = .29$. As can be seen in Figure 11, participants were strategic and responded faster to emails that gave more points. In contrast, when there was a relatively short response window (20 minutes or 3 hours), the number of points earned from responding to an email had no effect on response times.

(Figure 11 about here)

All other interactions were also non-significant, all F's < 1.2.

Analysis of the behavior of those excluded from primary analysis

As stated at the start of the results section, we excluded data from 16 participants as non-responses resulted in many missing cells and this meant that statistical analysis could not be conducted. In this section we show that these participants responded to these different cues the same way i.e., the experimental manipulations affect response behaviors in the same way regardless of response rate. The nature of this data precludes significance testing of these results, but descriptive statistics are provided to illustrate our argument.

Response rate: Excluded participants – by definition – responded to far fewer emails (M = 28%) than those that were included in the main analysis (M = 85%). Despite this difference, the same key patterns emerge in the data, but with all values simply being reduced. For example, when we investigate the effect of urgency we see that excluded participants were less likely to respond to messages when given a shorter time window: 18% response rate when 20 mins, 29% response rate when 3-hrs, 37% response rate when 24-hrs. We also replicate the pattern when investigating the impact of cost: excluded participants were far more likely to respond to the low-cost emails (47%) than high response cost emails (9%). This also holds when investigating the impact of importance on response rate: excluded participants were more likely to respond to emails that were worth more points (100-points: M = 33%; 30-points: M = 26%; 10-points: M = 25%). There is also a similar pattern when we look at the interaction between cost and importance: when there is a low-cost of responding, all emails are equally likely to be responded to regardless of points on offer (100-points, low-cost: M = 50%; 30-points, low-cost: M = 46%; 10-points, low-cost: M = 46%). But when there is a high-cost of responding, there was a higher response rate for emails that were worth more points (100-points, high-cost: M = 16%; 30-points, high-cost: M = 6%; 10-points, high-cost: M = 5%).

Response time: Responses were generally slower for excluded participants (M = 82 min, interquartile range 10 – 83 min) than those that were included in the main analysis (M = 61 min, interquartile range 11 – 59 min). Despite this difference, all but one of the same key patterns emerge in the data, but with all values simply being reduced. For

example, when we investigate the effect of urgency, we see that just like the included participants, excluded participants were also faster at responding to emails when that email had a shorter response window (20-minute: M = 29 min; 3-hours: *M* = 46 mins; 24-hours: *M* = 160 min). We also replicate the pattern when investigating the impact of cost: just like the included participants, excluded participants were also faster at responding when there was a low response cost (*M* = 76 min) than when there was a high response cost (*M* = 95 mins). When investigating the impact of importance on response rate we found a difference in the behavior of those who were excluded when compared to the those included: in contrast to the included participants, excluded participants were not faster at responding to emails that were worth more points (100-points: *M* = 88 min; 30-points: *M* = 89 min; 10-points: *M* = 68 min). In addition, excluded participants also showed no evidence of being sensitive to salience of importance: response times were similar when importance was missing (*M* = 85 min) compared to when it was visible (*M* = 80 min).

4. GENERAL DISCUSSION

This paper makes three novel and significant contributions. First, we provide a detailed and empirically grounded understanding of how people prioritize responses to emails in their inbox based on subject line cues given by the sender. Second, we contribute a novel method for conducting field experiments into how people triage their daily situated email. Third, we contribute a set of design implications that can inform the development of future email management systems.

4.1. Email Response Behavior

Response Rate

We hypothesized that the *number* of responses will be dependent on the urgency (**H1a)** of the response required (those that require fast responses will be less likely to be responded toy than those that have a longer response time), the cost of responding (**H2a**) (incoming emails that require a low cost response will be more likely to be responded to than those that have a higher cost), the importance of the response (**H3a)** (incoming emails with a higher marked importance will be more likely to be responded to than those with a lower marked importance), and the salience of importance (**H4b**) (incoming

emails without an indicator of importance will be more likely to be responded to due to the curiosity to discover how many points the message is worth).

We found that participants were more likely to respond to messages that required an urgent, time critical response, that could be responded to quickly and easily and that were marked as important. These results corroborate previous research showing that people use their mobile devices to give quick and simple responses to urgent messages while on the move (Cecchinato et al., 2015), but that messages that require a more complex response are postponed until the user reaches a desktop computer (Matthews et al., 2009). This previous research has focused on the strategies that people use to try and respond to urgent emails in a timely way – but it does not tell us whether these strategies are successful. The results of our field experiment complement these insights from previous qualitative studies, by providing evidence that people actually do consistently prioritize responding to emails that require quick and easy responses over messages that require composing longer and effortful responses.

Our finding that people prefer to do respond to emails that are easy and quick before tasks that are hard and time consuming has parallels in the broader HCI literature. For instance, Payne, Duggan, and Neth (2007) had participants choose between two Scrabble games (i.e., find as many words as possible given two separate sets of letters). Payne et al. found people were sensitive to their own rate of productivity (i.e., rate of finding words within each set of letters) and prioritized the task with the high rate of return (i.e., that was easiest to find letters in). This finding is relevant to us because it shows that people are more likely to choose easier/shorter tasks over harder/longer tasks, in order to maximize the overall rate of return. This idea is broadly consistent with Pirolli and Card's seminal work on Information Foraging theory in HCI (Pirolli, 2007; Pirolli & Card, 1999) and Payne and Howes' (2013) characterization of Adaptive Interaction: both assume that people adapt their behavior to maximize their rate of return towards meeting their ongoing goals. Our novel contribution is to show that these same processes can help explain email triaging behavior.

Our findings put limits on the generalizability of Kraut et al.'s (2002) conclusions about inbox-level cues. In this work, senders paid 'postage' that reflected the importance they placed on the message they were sending – the more important they felt it was, the greater the amount of postage allocated. Surprisingly, they found no effect of importance on prioritization - recipients did not use the postage paid by senders as a signal of importance. In contrast, we found a significant main effect of importance, suggesting that perhaps Kraut et al.'s manipulation of importance was not effective. Our instantiation of importance was both reliable and consistent; responding on-time to important emails earned participants more points. By using this points scheme, we were able to consistently communicate and objectively manipulate the importance of an email. As we shall discuss in the limitations section below, the importance of an email is not always as clearly defined.

We found further differences between our findings and those of Wainer et al. (2011) when we examined the effect of the salience of importance. Wainer et al. argued that *"curiosity drives attention in the email context"* (p. 3446), but we found no effect of the salience of importance on response rate except when messages were marked as urgent. Wainer et al. note that the effect of the salience of importance disappeared when their participants received a high volume of emails to process in a limited time. One explanation is that our participants were also processing a high volume of email (both for our experiment and from their normal email activity). Taken together, these results suggest that sending emails with ambiguous or empty subject lines is not an effective strategy to grab the attention of people that receive and process a higher volume of email – senders should provide explicit and accurate subject line cues.

Response Time

We hypothesized that the *speed* of responses will be dependent on the urgency (**H1b)** of the response required (those that require fast responses will be responded to more quickly than those that have a longer response time), the cost of responding **(H2b)** (incoming emails that require a high cost response will be responded to more slowly than those that have a lower cost), the importance of the response (**H3b)** (incoming emails with a higher marked importance will be responded to more quickly than those with a

lower marked importance), and the salience of importance (**H4b**) (incoming emails without an indicator of importance will be responded to more quickly due to the curiosity to discover how many points the message is worth).

We found that emails that had high importance to the receiver, and were quick and easy to respond to, were dealt with first. Emails that had low importance, and were costly to respond to, were deferred. However, when emails needed a fast response (i.e., were urgent), participants replied to low importance messages as quickly as to high importance messages. Urgency trumped other factors, even when a more pluralistic assessment of message characteristics would have yielded better results.

These findings are consistent with previous research that demonstrates that decisions about urgent events are not always rational (Bogunovich & Salvucci, 2011; Kerstholt, 1994). Given that participants only managed to respond to around 65% of the messages that we sent to them, they would have been better off allocating their attention to the high importance/less urgent messages than to the low importance/urgent messages.

These findings fit with Kahneman's (2011) dual process account of judgement and decision making and with experiments from cognitive psychology, which have shown that people's ability to make rational judgements breakdown when they are under time-pressure (see, e.g., Alter et al., 2007). This is because, under these circumstances, people tend to favor intuitive 'System 1' heuristics when they do not have time to employ more deliberate 'System 2' strategies. In this case, the time-pressured 'urgent' messages led to people responding as quickly as possible, exhibiting behaviour that suggests that they had not given explicit thought as to whether it was the best strategy. The most 'rational' strategy would have been to respond to all the high point emails first, ahead of the low point but urgent emails. To the best of our knowledge, no previous research has demonstrated that people are biased to prioritize urgent unread emails over others in their inbox.

Our results also extend the findings of Karagiannis and Vojnovic (2009) who found that email response times varied depending on the length of the email that was being responded to, and whether it had an attachment or not. Our findings demonstrate that people are not only sensitive to the cost of reading longer emails, but also to the cost of composing replies. Even a two-minute task can influence the speed with which a response is sent. The Email Charter (http://bit.ly/emailcharter1) suggests indicating that emails require action (i.e., are high-cost) by adding information to the start of subject line (e.g., "[Response Required]"). Our findings indicate that such inbox-level cues will be used by receivers to defer attending to such messages, making responses slower. Conversely, indicating that a short quick reply is needed is likely to prompt the recipient to give a speedy response.

The Method

The work described in this paper makes a methodological contribution. We developed a novel approach for conducting field experiments into how people triage their daily situated email. Previous research that has investigated email triage behavior has been of three types: post-hoc analyses of email activity logs (e.g., Kooti et al., 2015), lab-based experiments (e.g., Wainer et al., 2011), and qualitative enquiries (e.g., Cecchinato et al., 2015). Studies of email archives are incredibly valuable for understanding response patterns (e.g., Kalman & Rafaeli, 2005). However, it is difficult to learn about *why* people responded in the way that they did from post-hoc analysis of activity logs. In contrast, both lab-based experiments and qualitative enquiries offer well-established methodologies for uncovering the factors that influence triaging behaviors, but both methods have limitations. Qualitative studies can be subject to unreliable self-reported data as participants do not always have a good understanding of the factors that influence their own behavior. In contrast, lab-based experiments allow for the variables of interest to be systematically controlled and manipulated by the experimenter to see what effect they have on behavior. However, in doing so, lab-based experiments are often criticized for being artificial and lacking ecological-validity. Field experiments potentially offer a middle ground.

Moving experimental studies out of the lab and into the field means that participants are processing the experimenter-sent emails alongside all the other email that they normally receive. This approach comes with its own drawbacks, such as uncontrolled user behavior. But by having the experiment run over an extended period of time (three weeks in the field as opposed to one hour in the lab), participants triage their email during their other normal activities and this creates a more naturalistic environment than a lab setting. On balance, this approach offers an improvement on the ecological-validity of a traditional lab-based experiment while at the same time retaining the ability to systematically manipulate variables of interest and draw causal inferences between the factors we manipulated and the behavior we recorded. In doing so, we avoid the subjectivity of qualitative diary studies and interviews with people about their daily email habits.

One might wonder though whether our participants did behave as they would normally when checking their emails. As Cecchinato et al. (2016) describe, email is received and managed across a variety of accounts, with participants in their study reporting having three email accounts on average. Whichever of their accounts our participants used to receive these emails, our findings demonstrate that these accounts were checked frequently. However, perhaps participants might have been too enthusiastic in their engagement with the study, going out of their way to set-up systems and tools to help them respond in a timely fashion. Given that, on average, participants took an hour to respond to emails and that only 2.8% of responses were made within 60 seconds, there is little evidence to suggest this was happening. While we might expect people to change their behavior at the outset of the study, it would be difficult for them to maintain this over the entire three-week period of the study.

One might also wonder about the ecological validity of the points system used within our method. The points associated with each email provide an incentive for answering each of the messages. Varying the points associated with email enables us to systematically manipulate how important each of the messages is to answer within the context of our study. Of course, it is not enabling the investigation of how participants would perceive the importance of responding to one of our emails against the importance

responding to an email from another sender. But it does enable us to see that importance is taken into account when deciding which email to respond to next. Whilst the points system does not directly replicate the importance, or lack thereof, of emails received on a daily basis, it does enable us to approximate how emails received vary in importance.

It is also important to reflect on the external validity of the method. To what extent does the variability in urgency given to particular emails in our study replicate that seen in other contexts? As we state above in section 2.1, Karagiannis and Vojnovic (2009) found very little evidence of email response times being influenced by organizational seniority. Therefore, the fact that all the messages in our study came from a single sender seems unlikely to be a major limitation of the external validity of our method. Instead, we explicitly manipulated the urgency of responses in our study using cues in the subject line and body of the message. Future work could extend our method so that emails come from multiple senders who play different roles within the context of the study and thus implicitly manipulate the urgency and importance of responses.

4.2. Design Implications

Finally, we discuss the implications of results for the design of intelligent email systems and tools. Such systems could use machine learning algorithms to prioritize whether a message is of high importance to a receiver, while also taking into account the fact that, as we have demonstrated, users' email triage behaviors are not optimal when faced with time pressures.

Display cues that indicate the cost and importance of responses.

Our study demonstrates that people use these cues in their email management strategies. Previous research has demonstrated that using time efficient email processing strategies reduces the time spent each day on email (Bradley et al., 2013), and that when people spend less time on email they have reduced feelings of stress and overload (Kushlev & Dunn, 2015). Supporting people in quickly identifying the emails that need responding to, while filtering out the majority that are irrelevant or do not require immediate attention (Buthpitiya et al., 2009), would be helpful to users.

There are a number of ways in which the time-cost and importance of responding could be estimated by a machine learning algorithm. One option is for senders to be asked to provide estimates of both of these cues when composing a message. Measures of the alignment of a sender's estimate of importance and cost with that of the receiver's perception could be calculated by the system by comparing the sender's rating of importance with a similar rating provided by the receiver, and by comparing the sender's estimate of time to reply with the actual time taken to compose the response. Such a process would enable responders to influence the prioritization mechanism by providing additional information for the system to adjust its learning algorithm and better prioritize emails. Such a system does not do 'free' work, though. It would mean moving workload associated with classifying messages from recipients to senders. It would therefore represent a trade-off over where the system demanded effort be expended – by senders or receivers of messages. This trade-off would likely be context-bound, influenced by the context in which given messages are being sent and received.

Provide additional support to help users select valuable urgent emails amongst the urgent trivia that makes its way into their inbox.

Our study demonstrates that people are less strategic when processing their urgent emails than they could be. Previous research in cognitive psychology has shown that decision-making is often negatively impacted when people are under time pressure (Kerstholt, 1994). The problem of choosing between urgent and important problems was famously described by Eisenhower: "*I have two kinds of problems: the urgent and the important. The urgent are not important, and the important are never urgent*" (Wright, 2016). In fact, the Eisenhower urgent-important matrix categorizes tasks into four categories: urgent+important, non-urgent+important, urgent+not-important, and non-urgent+not-important. Our data shows that urgent trivia (e.g., someone asking if you're free for coffee in five minutes) – an example of an urgent+not-important message –steals attention from less urgent but more important messages, in much the same way as multitasking research has shown that 'urgent' events are often prioritized over important but less urgent ones (Bogunovich & Salvucci, 2011; Kerstholt, 1994). People's productivity could be improved by pointing them toward the highest importance messages waiting for their attention, steering them away from being captured by urgent,

low importance messages. The Eisenhower urgent-important matrix (Wright, 2016) may provide inspiration for implementation.

4.3. Limitations and Future Research

Study Disengagement

This was a field experiment, and so a degree of experimental control was sacrificed in return for greater ecological validity. As a result, we had little control over participation rates, and had to exclude participants when evidence suggested that they had ceased engaging with our study.

We acknowledge that, rather than excluding participants, there are a number of alternative ways of dealing in which such missing data can be inferred. For example, pairwise deletion involves ignoring missing values and conducting analysis on the remaining data. Such an approach has the advantage of preserving all collected data. However, when there are many missing values, this approach cannot be adopted. An alternative strategy would be to perform multiple imputation by replacing the missing values with, for example, the mean values of the variables. However, if the missing values are not random then this can lead to biases in the data. Wood et al. (2004) demonstrate that the most common approach (65% of studies) to dealing with missing data is to simply omit those cases with the missing data and analyse the remaining data.

A 35% exclusion rate might be concerning in a traditional laboratory experiment, but we do not think it should raise the same level of concern when evaluating the validity of a field study. Wood et al (2004) conducted a systematic review of how missing data is handled in randomized control trials in the medical field. Such trials are by definition, conducted in the field and are therefore a good comparison point for us. They report the percentage of participants for how data was missing across 70 studies. The median percentage of participants with missing data was 10%, and 6 (9%) trials had more than 30% of participants with missing data. Our rate of engagement is at the high end of this distribution. Arguably people will be more motivated to engage in a randomized control trial in medicine which might have implications for their own health than they are to engage in a field experiment about email. We acknowledge that our filtering process

introduced a degree of selection bias toward those with the time and motivation to participate over the extended three-week period of the study. However, our participants with low response rates still seemed to show similar patterns of behavior to those included in our main analyses suggesting that that low responders might be employing the same strategies as other participants.

We do not know why some of our participants dropped out. It may simply be that some participants decided that our study was not important enough to ever prioritize above their other daily emails. A greater reward for responding to our messages might have induced these participants to prioritize them above other messages. However, it also seems plausible that extremely busy people might still have been engaged, but may have chosen to defer responding to our messages until much later and missed the window for timely responses entirely. In contrast, participants who did complete the experiment were arguably highly motivated to respond to our messages by the incentives offered and thus may have responded more often and more quickly to our messages than to others in their inbox – they may have over-valued our messages. Of those who completed the study, 42% ranked winning £100 as the most important factor for participating in the study suggesting that they were motivated to earn points to get a cash reward. In future work it will be important to understand whether this variation reflects participant attitudes toward the study specifically or to email in general.

Improving Telemetry

Our measure of response time enables us to measure the impact of our manipulations on deferral strategy. We can clearly see the impact of manipulating the cost, importance, and urgency of responding. However, we are unable to determine the relative contributions of reading and responding to processing time. Overall response time in this study comprises the time from receipt to reading and then time from reading to replying. We are only able to infer that the extended response time is as a consequence of a longer time from reading to replying in the high-cost condition.

Our data also lack email metadata, such as the email client used, and the devices used to read and reply to emails. Future research should take the devices used and the activity

carried out on each one into account so as to better understand how multiple devices are used to triage emails. Future work should also record when each message is opened, how many times it is opened before replying, and collect more metadata information. One way of doing this could be by using existing email tracking technology like Streak (https://www.streak.com/).

Exploring Cue Variation

Our indicators of message importance were entirely reliable. All emails marked with a high level of importance delivered the highest number of points if they were responded to on time. Similarly, all emails marked with low importance delivered the lowers number of points. In reality, we have all received messages flagged as 'Important' that turn out to be trivial. As a sender, deciding on the importance message that we are sending can be difficult. For instance, we might lack the contextual information to determine the importance of a message.

The importance of a message is likely to be determined dynamically by a recipient according to their moment-to-moment context. As a result, the points assigned by us may not reflect the overall or relative importance of other emails that our participants receive in their daily lives. A tardy reply to a vital work-related email is different to responding late to a high-point email experiment: failing the former has a much higher personal cost compared to the latter. There were no social or work-related costs in this experiment which are often a critical part of one's communication with others. Nevertheless, our results do show that in situations where the importance of messages can be appropriately inferred by a sender, importance cues can influence response behavior.

As well as cue reliability, there is also the issue of cue salience. The content of the emails used in our study is artificial. Participants could quickly look for the cues in the email that reliably indicated urgency, importance, and cost. Extracting this information from most of the other messages in their inbox would likely have been more complicated and time consuming. This puts limits on how far we can generalize the response rate and actual response times in our experiment.

5. CONCLUSIONS

The results of our field study show that when emails were less time-sensitive, participants deferred responses to a later time. However, when they did respond they gave greater priority to messages that were easier to respond to (lower cost) and those that carried the greatest importance. In contrast, when presented with emails that required an urgent, time-sensitive response, participants prioritized these and disregarded any other cues, even when a more nuanced assessment of message characteristics would have been more efficient.

Our results have implications for the users of email who desire timely responses. Users should consider ways in which they can convey the amount of time it will take the responder to reply, the urgency with which the response is needed, and the importance of the recipient's response. Composing emails that require short responses is the best way to ensure that the receiver will deal with it promptly.

ACKNOWLEDGEMENTS

<The first author> supervised the design of the study and the data collection, lead the writing, and provided funding and insight to the project as a whole. The other authors (alphabetical order) contributed to the whole project, with <2nd author> designing the field study and constructing the materials to run it and co-supervising the data collection, <3rd> and <4th> contributing actively to the data analysis process, and <5th> contributing to the writing of the manuscript. The data was collected by <student>. This work was supported by the EPSRC Digital Epiphanies Project (EP/K025392/1).

REFERENCES

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology-General*, *136*(4), 569–576. https://doi.org/10.1037/0096-3445.136.4.569

Bogunovich, P., & Salvucci, D. D. (2011). The effects of time constraints on user behavior for deferrable interruptions. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 3123–3126. https://doi.org/10.1145/1978942.1979404

Bradley, A., Brumby, D. P., Cox, A. L., & Bird, J. (2013). How to Manage Your Inbox: Is a Once a Day Strategy Best? *Proceedings of the 27th International BCS Human Computer Interaction Conference*, 20:1–20:6. http://dl.acm.org/citation.cfm?id=2578048.2578077

Buthpitiya, S., Madamanchi, D., Kommaraju, S., & Griss, M. (2009). Mobile Context-Aware Personal Messaging Assistant. In T. Phan, R. Montanari, & P. Zerfos (Eds.), *Mobile Computing, Applications, and Services* (pp. 254–272). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-12607-9_17

Cecchinato, M. E., Cox, A. L., & Bird, J. (2015). Working 9-5?: Professional Differences in Email and Boundary Management Practices. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 3989–3998. https://doi.org/10.1145/2702123.2702537

Cecchinato, M. E., Sellen, A., Shokouhi, M., & Smyth, G. (2016). Finding Email in a Multi-Account, Multi-Device World. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 1200–1210. https://doi.org/10.1145/2858036.2858473

Collins, E. I. M., Cox, A. L., & Wootton, R. (2015). Out of Work, Out of Mind?: Smartphone Use and Work-Life Boundaries. *Int. J. Mob. Hum. Comput. Interact.*, *7*(3), 67–77. https://doi.org/10.4018/ijmhci.2015070105

Dabbish, L. A., & Kraut, R. E. (2006). Email Overload at Work: An Analysis of Factors Associated with Email Strain. *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work*, 431–440. https://doi.org/10.1145/1180875.1180941

Di Castro, D., Karnin, Z., Lewin-Eytan, L., & Maarek, Y. (2016). You'Ve Got Mail, and Here is What You Could Do With It!: Analyzing and Predicting Actions on Email Messages. *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, 307–316. https://doi.org/10.1145/2835776.2835811

Eppler, M. J., & Mengis, J. (2004). The Concept of Information Overload: A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines. *The Information Society*, *20*(5), 325–344. https://doi.org/10.1080/01972240490507974

Farmer, G. D., Janssen, C. P., Nguyen, A. T., & Brumby, D. P. (2018). Dividing Attention Between Tasks: Testing Whether Explicit Payoff Functions Elicit

Optimal Dual-Task Performance. *Cognitive Science*, *42*(3), 820–849. https://doi.org/10.1111/cogs.12513

Gould, S. J. J., Cox, A. L., Brumby, D. P., & Wiseman, S. (2016). Short links and tiny keyboards: A systematic exploration of design trade-offs in link shortening services. *International Journal of Human-Computer Studies*, *96*, 38–53.

Grandhi, S. A., & Lanagan-Leitzel, L. K. (2016). To Reply or To Reply All: Understanding Replying Behavior in Group Email Communication. *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 560–569. https://doi.org/10.1145/2818048.2819981

Jarvstad, A., Rushton, S. K., Warren, P. A., & Hahn, U. (2012). Knowing When to Move On Cognitive and Perceptual Decisions in Time. *Psychological Science*, *23*(6), 589–597. https://doi.org/10.1177/0956797611426579

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kalman, Y. M., & Rafaeli, S. (2005). Email Chronemics: Unobtrusive Profiling of Response Times. *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, 108b–108b. https://doi.org/10.1109/HICSS.2005.231

Karagiannis, T., & Vojnovic, M. (2009). Behavioral Profiles for Advanced Email Features. *Proceedings of the 18th International Conference on World Wide Web*, 711–720. https://doi.org/10.1145/1526709.1526805

Kerstholt, J. H. (1994). The effect of time pressure on decision-making behaviour in a dynamic task environment. *Acta Psychologica*, *86*(1), 89–104. https://doi.org/10.1016/0001-6918(94)90013-2

Kooti, F., Aiello, L. M., Grbovic, M., Lerman, K., & Mantrach, A. (2015). Evolution of Conversations in the Age of Email Overload. *Proceedings of the 24th International Conference on World Wide Web*, 603–613. https://doi.org/10.1145/2736277.2741130

Kraut, R. E., Morris, J., Telang, R., Filer, D., Cronin, M., & Sunder, S. (2002). Markets for Attention: Will Postage for Email Help? *Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work*, 206–215. https://doi.org/10.1145/587078.587108

Kushlev, K., & Dunn, E. W. (2015). Checking email less frequently reduces stress. *Computers in Human Behavior*, *43*, 220–228. https://doi.org/10.1016/j.chb.2014.11.005

Matthews, T., Pierce, J., & Tang, J. (2009). No smart phone is an island: The impact of places, situations, and other devices on smart phone use. *IBM RJ10452*.

Neustaedter, C., Brush, A. J. B., & Smith, M. A. (2005). Beyond 'from' and 'Received': Exploring the Dynamics of Email Triage. *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, 1977–1980. https://doi.org/10.1145/1056808.1057071

Payne, S. J., Duggan, G. B., & Neth, H. (2007). Discretionary task interleaving: Heuristics for time allocation in cognitive foraging. *Journal of Experimental Psychology: General*, *136*(3), 370–388. https://doi.org/10.1037/0096-3445.136.3.370

Payne, S. J., & Howes, A. (2013). Adaptive Interaction: A utility maximization approach to understanding human interaction with technology. *Synthesis Lectures on Human-Centered Informatics*, *6*(1), 1–111. https://doi.org/10.2200/S00479ED1V01Y201302HCI016

Pirolli, P. (2007). *Information Foraging Theory:Adaptive Interaction with Information*. Oxford University Press, USA.

Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review*, *106*(4), 643–675. https://doi.org/10.1037/0033-295X.106.4.643

Porter, S. R., & Whitcomb, M. E. (2005). E-mail Subject Lines and Their Effect on Web Survey Viewing and Response. *Social Science Computer Review*, *23*(3), 380–387. https://doi.org/10.1177/0894439305275912

Sappleton, N., & Lourenço, F. (2016). Email subject lines and response rates to invitations to participate in a web survey and a face-to-face interview: The sound of silence. *International Journal of Social Research Methodology*, *19*(5), 611–622. https://doi.org/10.1080/13645579.2015.1078596

Schumacher, E. H., Lauber, E. J., Glass, J. M., Zurbriggen, E. L., Gmeindl, L., Kieras, D. E., & Meyer, D. E. (1999). Concurrent response-selection processes in dual-task performance: Evidence for adaptive executive control of task scheduling. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(3), 791.

Siu, N., Iverson, L., & Tang, A. (2006). Going with the Flow: Email Awareness and Task Management. *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work*, 441–450. https://doi.org/10.1145/1180875.1180942

The Radicati Group. (2015). *Email Statistics Report, 2015-2019*. http://www.radicati.com/?p=12964

Tyler, J. R., & Tang, J. C. (2003). When Can I Expect an Email Response? A Study of Rhythms in Email Usage. In K. Kuutti, E. H. Karsten, G. Fitzpatrick, P. Dourish, & K. Schmidt (Eds.), *ECSCW 2003* (pp. 239–258). Springer Netherlands. https://doi.org/10.1007/978-94-010-0068-0_13

Wainer, J., Dabbish, L., & Kraut, R. (2011). Should I open this email?: Inbox-level cues, curiosity and attention to email. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 3439–3448. https://doi.org/10.1145/1979442.1979456

Whittaker, S., Bellotti, V., & Moody, P. (2005). Introduction to This Special Issue on Revisiting and Reinventing e-Mail. *Hum.-Comput. Interact.*, *20*(1), 1–9. https://doi.org/10.1207/s15327051hci2001&2_1

Wood, A. M., White, I. R., & Thompson, S. G. (2004). Are missing outcome data adequately handled? A review of published randomized controlled trials in major

medical journals. *Clinical Trials*, *1*(4), 368–376. https://doi.org/10.1191/1740774504cn032oa

Wright, C. (2016, May 12). *More 'Prioritization' and Less Speed | SciTech Connect*. http://scitechconnect.elsevier.com/more-prioritization-less-speed/

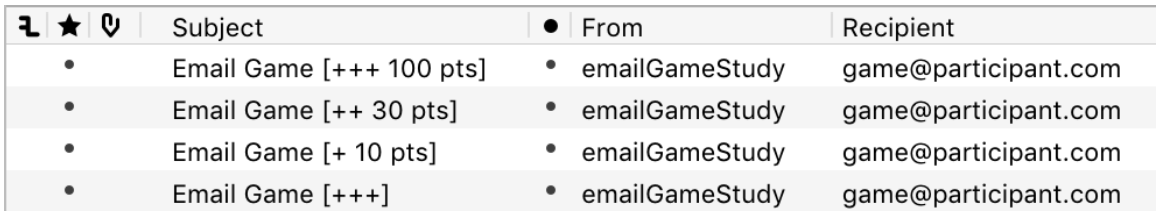
Table 1. Summary of ANOVA results for response rate and response time to emails. Effect sizes are reported for only significant results. Significant interactions are presented in Figures 7–11.

	Response rate		Response time	
Source	*F*	η_p^2	*F*	η_p^2
Urgency (U)	57.10[b] ***	.67	45.80[b] ***	.62
Cost (C)	21.41[a] ***	.43	37.56***	.57
Importance (I)	10.79[b] ***	.28	5.27[b] **	.16
Salience (S)	2.55[a]		0.02[a]	
U × C	4.22[b] *	.13	29.61[b] ***	.51
U × I	0.24[c]		3.12[c] *	.10
U × S	3.22[b] *	.10	0.44[b]	
C × I	6.11[b] **	.18	0.02[b]	
C × S	0.08[a]		0.27[a]	
I × S	1.10[b]		0.30[b]	
U × C × I	0.87[c]		0.08[c]	
U × C × S	0.76[b]		0.13[b]	
U × I × S	0.92[c]		1.18[c]	
C × I × S	0.13[b]		0.30[b]	
U × C × I × S	0.09[c]		0.12[c]	

[a] $df = 1, 28$, [b] $df = 2, 56$, [c] $df = 4, 112$

* $p < .05$, ** $p < .01$, *** $p < .001$

Figure 1. Example subject lines



	Subject	From	Recipient
•	Email Game [+++ 100 pts]	• emailGameStudy	game@participant.com
•	Email Game [++ 30 pts]	• emailGameStudy	game@participant.com
•	Email Game [+ 10 pts]	• emailGameStudy	game@participant.com
•	Email Game [+++]	• emailGameStudy	game@participant.com

Figure 2. Example email sent to a participant by the system

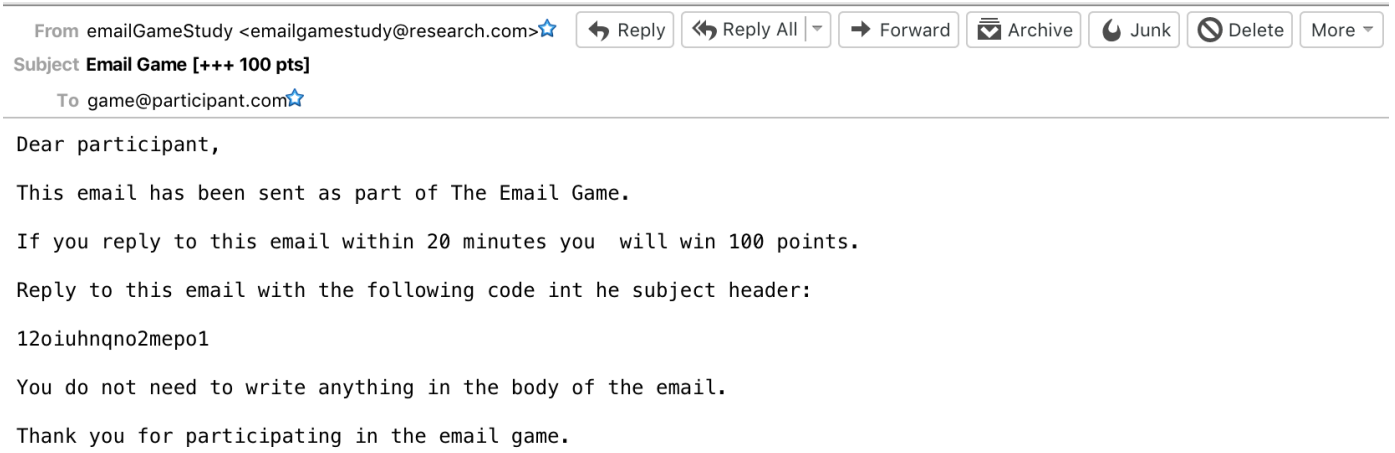


Figure 3. Example of high-cost email

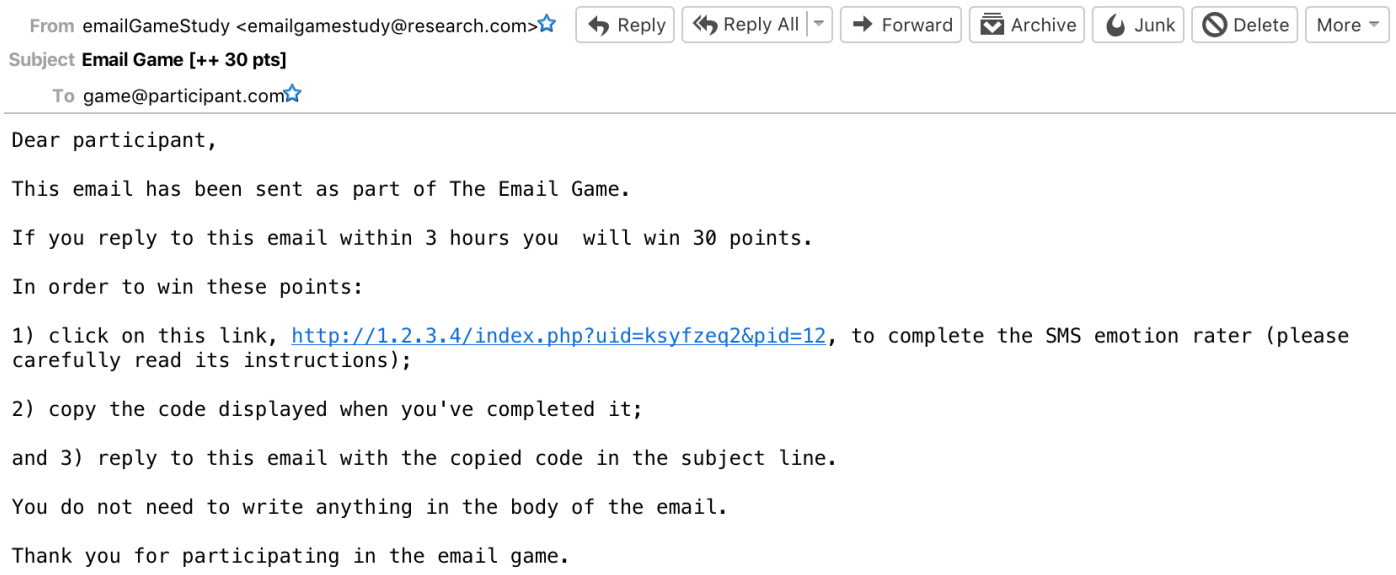


Figure 4. Screenshot of instructions for the high-cost emails

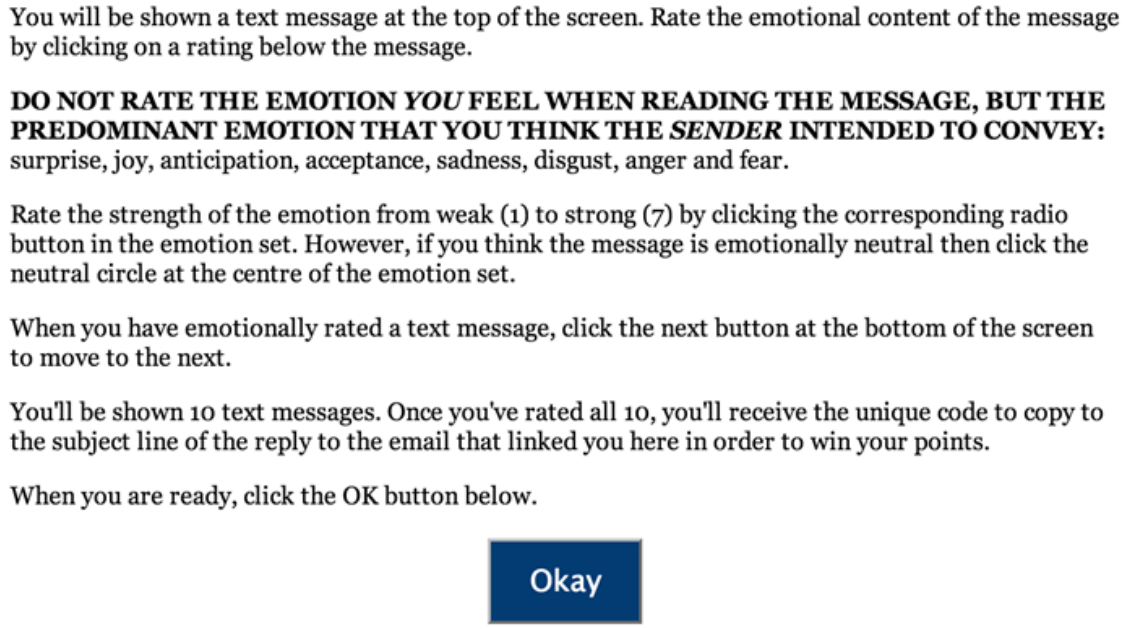


Figure 5. Screenshot of the rating task given to participants when in the high-cost condition. (Question asks whether the recipient is traveling by car or by train.)

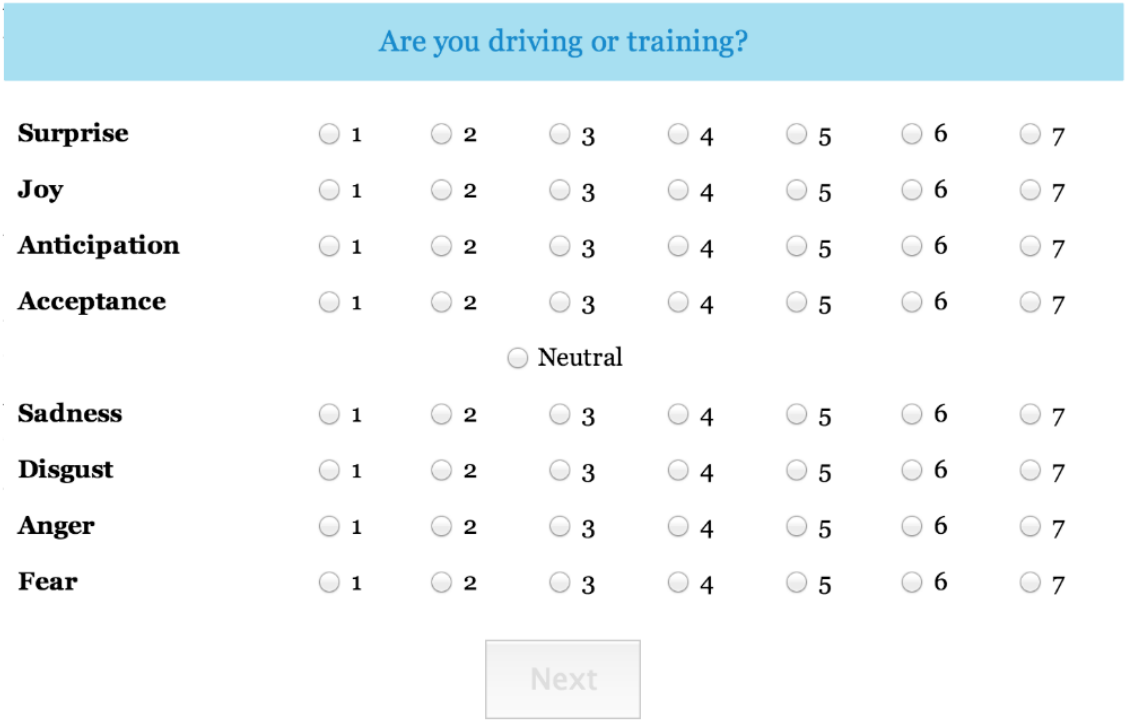


Figure 6. Histogram showing the distribution of participants' mean email response rate across all conditions.

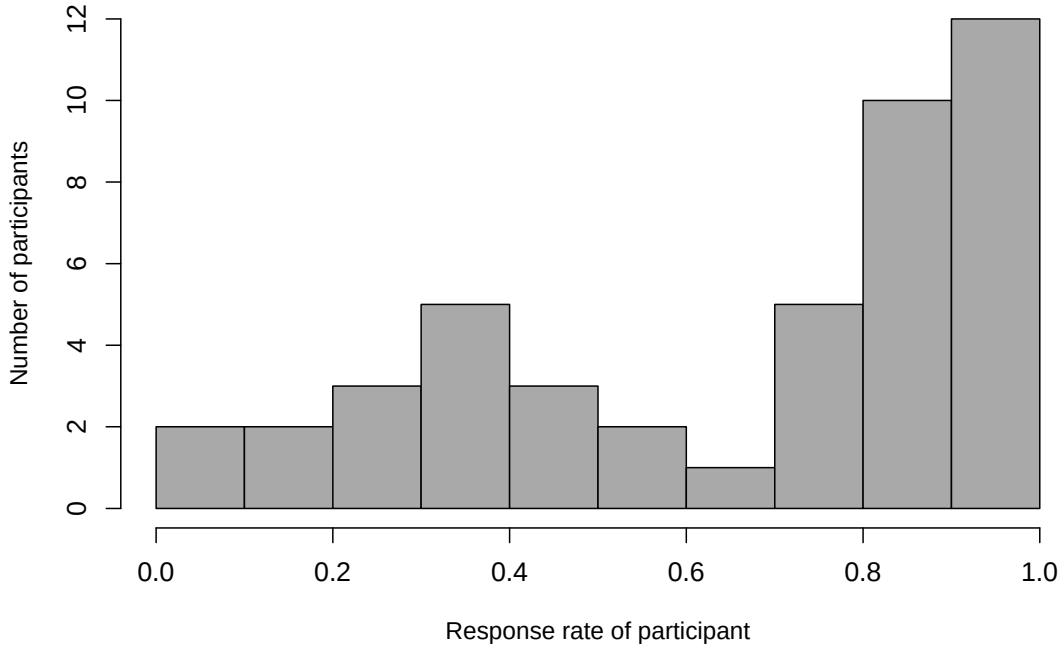


Figure 7. Urgency by cost (U × C) interaction on response rate. Errors bars represent standard errors of the mean. There was a significant effect of cost on response rates, and this effect occurred at all levels of urgency: 24-hours, 3-hours, and 20-minutes.

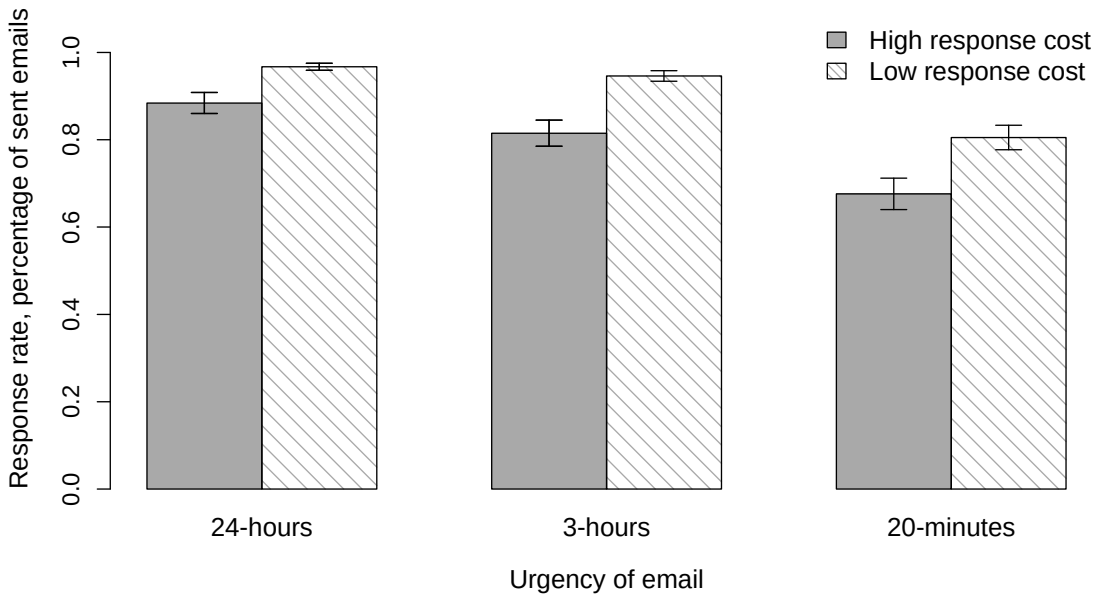


Figure 8. Cost by importance (C × I) interaction on response rate. Errors bars represent standard errors of the mean. There was a significant effect of importance on response rates in the low-cost condition; the effect of importance was not significant in the high high-cost condition.

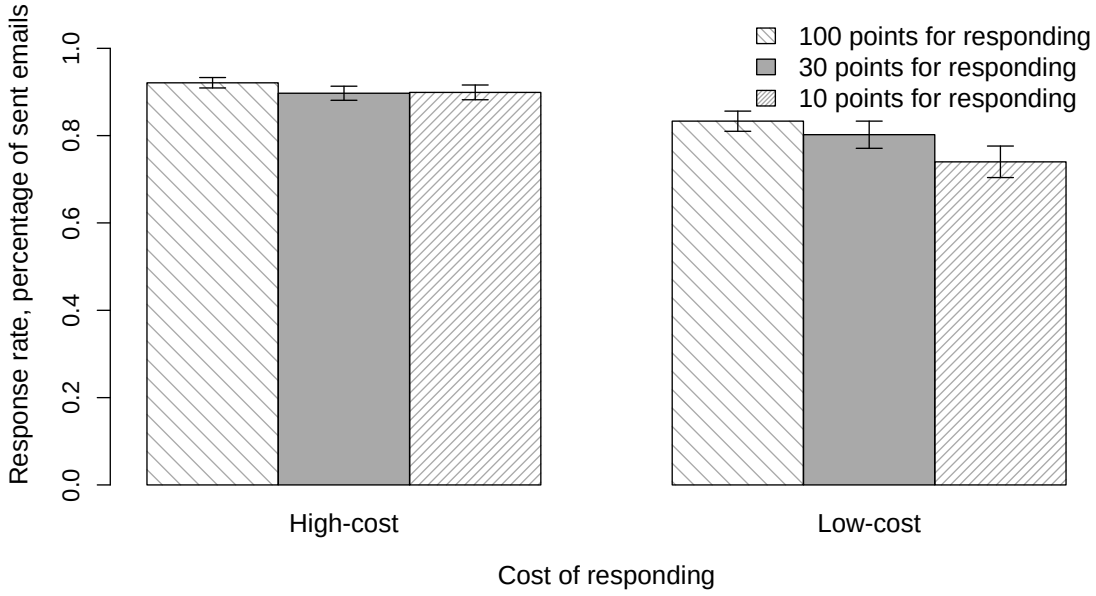


Figure 9. Urgency by salience (U × S) interaction on response rate. Errors bars represent standard errors of the mean. There was a significant effect of salience on response rates in the 20-minute condition; the effect of salience was not significant in either the 24-hour or 3-hour condition.

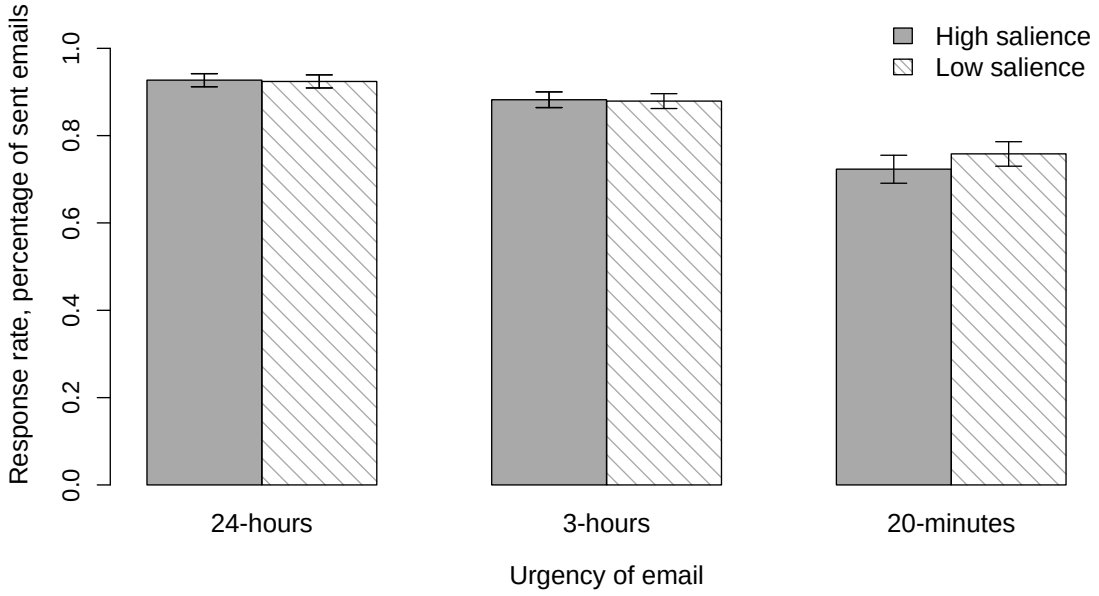


Figure 10. Urgency by cost (U × C) interaction on time to reply to an email. Errors bars represent standard errors of the mean. There was a significant effect of cost on response rates in both the 24-hour and 3-hour condition; the effect of cost was not significant 20-minute condition.

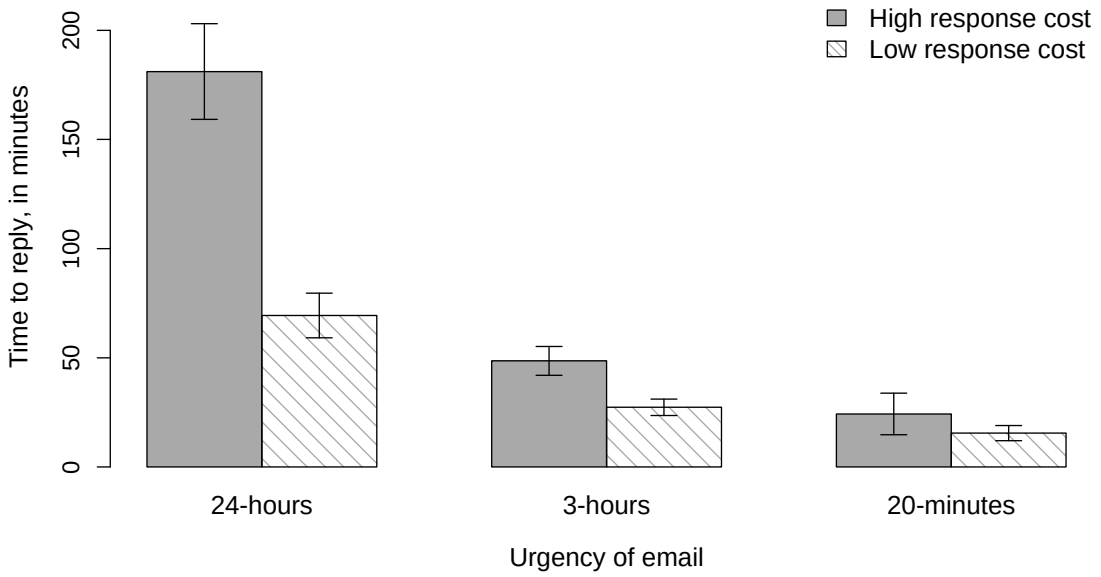


Figure 11. Urgency by importance (U × I) interaction on time to reply to an email. Errors bars represent standard errors of the mean. There was a significant effect of importance on response rates in the 24-hour condition; the effect of importance was not significant in either the 3-hour or the 20-minute condition.