

Understanding Information Preview in Mobile Email Processing

Kimberly A Weaver*, Huahai Yang[†], Shumin Zhai[‡], & Jeff Pierce[†]

kimberly.weaver@gatech.edu, hyang@us.ibm.com, zhai@acm.org, jspierce@us.ibm.com

*GVU Center
School of Interactive Computing
Georgia Institute of Technology
Atlanta, Georgia 30332-0760

[†]IBM Research - Almaden
650 Harry Road
San Jose, CA 95120

[‡]Google Inc
1600 Amphitheatre Parkway
Mountain View, CA 94043

ABSTRACT

Browsing a collection of information on a mobile device is a common task, yet it can be difficult due to the small size of mobile displays. A common trade-off offered by many current mobile interfaces is to allow users to switch between an overview and detailed views of particular items. An open question is how much preview of each item to include in the overview. Using a mobile email processing task, we attempted to answer that question. We investigated participants' email processing behaviors under differing preview conditions in a semi-controlled, naturalistic study. We collected log data of participants' actual behaviors as well as their subjective impressions of different conditions. Our results suggest that a moderate level of two to three lines of preview should be the default. The overall benefit of a moderate amount of preview was supported by both positive subjective ratings and fewer transitions between the overview and individual items.

ACM Classification Keywords

D.2.5 Software Engineering: Testing—*Usability testing*

General Terms

Experimentation, Human Factors

Author Keywords

mobile phone, email, information preview

INTRODUCTION

When presenting a large collection of information items, such as search results, news, email, songs, and podcasts, a common design question is how much, if any, “information preview” an interface should display to the user. An information preview can be in the form of content snippets, a brief

*Kimberly performed this research while a graduate student intern at IBM Research.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to publish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MobileHCI 2011, Aug 30–Sept 2, 2011, Stockholm, Sweden.

Copyright 2011 ACM 978-1-4503-0541-9/11/08-09...\$10.00.

summary, or simply the first lines of text in each information item, typically immediately following an item's title (e.g., headline, subject). Modern search engines almost always display previews with search results, as do news sites. Different email clients make a variety of choices for preview. Some only display the author and subject line (e.g., the Android, Blackberry, and Symbian mail applications), while others display one (e.g., the Gmail mobile web application, the Windows Phone 7 and webOS mail applications) or two lines (e.g., the iOS mail application) of the message body. Some allow users to configure the preview length, while others do not. This lack of agreement on the best approach shows that more research is needed to better guide preview design decisions.

Obviously, the more information preview the interface provides, the more informed the user will be in deciding whether to select, or “drill down”, on specific items. Equally obvious is that the more preview an interface presents for each item, the fewer information items it can display in a fixed screen size, making it difficult to fit the complete list of potentially relevant items in a single screen. Scrolling or paging beyond the first screen to look for more items imposes additional motor, visual (due to discontinuity), and memory costs on the user. The conflicting needs of accommodating for both a quick overview of the entire list and a preview of each item raises an important theoretical and empirical question: what is the optimal, most useful, or most practical amount of information preview to display? This question is even more important in mobile user interface design because mobile screen sizes are much more limited (in this paper we will use mobile to refer specifically to smart phones unless otherwise noted). Screen real estate is precious on mobile devices, so interfaces must use it as effectively as possible.

It is very difficult to devise a general, theoretical yet useful answer to such a question because the impact of preview is likely dependent on many factors including the user's tasks, goals, the type of information presented, and the cost of navigation. Scrolling or paging at the overview level is one type of navigation cost. Drilling down to a specific information item and coming back up is another. Both types have consequences on the relative value of the amount of preview. Consider two audio overview examples. When selecting music, users may be able to identify songs by the artist and title and

need no additional preview. However, when selecting a podcast lecture the artist and title may be the same for all the lectures in a series. Without additional information, users may be forced to start several podcasts and wait until the introduction of the topic or speaker in each before finding the desired one. If more detailed previews of the available podcasts were provided, the chance for the user to select a wrong podcast could be minimized. Interestingly, we note that current podcast interfaces (e.g. on the iPhone) often do not provide such previews.

Since the impact of preview clearly depends on the task and information content, a productive research strategy requires narrowing the question of how much and what preview to provide to a specific domain before developing more general and theoretical answers. We chose to focus on the design of email message previews for mobile phones, in part because there is so much disagreement between platforms. More specifically, we focus on mobile email processing. Mobile email users typically scan accumulated email messages and attempt to quickly determine which can be deleted or ignored, which are important enough that they need to be handled immediately, and which can be deferred until later (similar to how users triage messages on the desktop [21]). That process requires the ability to both quickly navigate a collection of email messages and quickly ascertain the contents and importance of each individual message. Because scanning is such a critical part of mobile email processing (research suggests that mobile email users defer reading in detail and writing responses to a computer [14]), a better understanding of how to design effective previews for email messages could help the design of interfaces that allow users to process their email more quickly and effectively.

We further limit our investigation to the amount of email preview rather than the preview content. Conceivably, an email preview can consist of a summary or snippets algorithmically distilled from the email body [4, 9, 19, 22]. Such types of summary can potentially contain a higher density of information, but their usability in email information preview is a subject of inquiry beyond the scope of the current paper. Instead, we use the industry convention of employing the initial lines of email as the preview content. In addition to being consistent with current practice, this choice is logical because the first lines tend to be the most informative and because users can more easily maintain continuity upon opening a message.

RELATED WORK

Email has been a common area of research in non-mobile settings [3, 24, 25]. In this paper we will focus on the literature most relevant to our target mobile email processing activity, in particular research investigating short interaction times, the display limitations of mobile devices, and behavior modeling.

Mobile email

One factor that determines a person's decision to perform a task such as reading email on a mobile phone rather than on a laptop or desktop is that the phone requires a lower level of

engagement [7]: people can use a phone without necessarily interrupting another main activity. The disadvantage, however, of this type of intermittent mobile interaction is that periods of continuous attention to a mobile device can be very short [17]. Due to the competition for attention which is inherent in many mobile computing tasks, interfaces should focus on communicating as much useful information as possible to the user in the smallest amount of time.

In an email triage study by Pierce et al., moving from displaying no preview of an email messages body to displaying one line of preview improved the accuracy of participants attempting to determine the purpose of messages from 51% to 80% [18]. However, the addition of more lines resulted in diminishing returns; using five lines of preview only improved accuracy to 88%. This result indicates that there is likely to be a crossover point where adding more lines of preview will not significantly impact a users need to open a message to determine its purpose but will require them to expend significantly more effort scrolling through their mailbox to see all of their messages.

Information navigation on small displays

Sweeney and Crestani investigated the relationship between device display size and the summary length of web search results and found that users were the most effective with shorter summaries on all devices [20]. More generally, conventional wisdom with mobile device interface design has been to reduce the amount of scrolling required [12], but this design guideline may not be as applicable today given the greater ease of scrolling associated with the current generation of touchscreen devices. In contrast, Jones et al. produced guidelines for small screen search engines which included reducing the amount of page to page navigation and providing more information instead of less for each search result due to the high cost of navigating to another page on a small device [11]. They acknowledged the existence of a trade-off between these two guidelines. This trade-off is what we explore in the context of mobile email.

Factors influencing triage behavior

Information foraging theory and the concept of *information scent* from the realm of web searching [5] provide useful insights and a biological metaphor for human information seeking. We can view the questions of how often the value of a message can be gleaned in the first few lines and how providing that extra value in the inbox view impacts the time to process all messages as the scent design questions for more effective foraging. A large body of research exists related to how people choose results based on text preview in web searches [2, 13, 16, 23, 26], but little has been done on the impact of text preview on the need (and decision whether) to open or not open email messages when scanning them. Researchers have shown that sender characteristics and message content influence a person's perceptions of message importance [8]. Sender information is provided by all modern email clients. The question then becomes how much content from the message is necessary to provide users with enough information to make accurate judgments about the value of an individual message without requiring them to open the

message. The focus of our current study is on understanding how the amount of preview changes the pattern of scrolling in the inbox view versus opening full email messages when processing email and how that preview amount impacts task performance time.

METHOD

Application design

In order to investigate the impact of preview length on message viewing and scrolling, we created a custom mobile iOS application. We chose to create a custom application to allow us to manipulate message previews and to log user actions. We built the application on iOS because of the availability of a pool of existing iPhone owners whom we could recruit as participants. Developing for only one operating system also meant that we could more easily compare results between participants because the interaction experience would be the same for all participants.

The initial application screen presents users with the number of new, unread, and read messages in their inbox. Within the inbox view, the application groups email messages into two lists: new messages and read and unread messages. If there are no new messages, the application shows only the header and list for the read and unread messages. The application presents messages in reverse chronological order: newer messages appear at the top of a list and older messages at the bottom. Figures 1(a), 1(b), and 1(c) show how this inbox view appears with different preview length settings. Independent of the preview setting, the inbox view shows a visual indication of read or unread status, the sender, time received, and subject for each message. Depending on the preview setting, the preview of the email body can vary between zero lines (as seen in Figure 1(a)) and five lines in length. A comparison between the zero lines of preview shown in Figure 1(a) and the four lines of preview shown in Figure 1(b) demonstrates how quickly the number of simultaneously visible emails decreases as the number of preview lines increases. With a zero lines preview setting, 10.5 messages are visible at once, while with a four lines setting, only four messages are visible. Users can open an email message and view its entire contents (shown in Figure 1(d)) by tapping on the messages detail disclosure button at the right.

In addition to allowing different choices for the number of default preview lines, we also introduced the ability for users to incrementally add lines to a messages preview by tapping the preview. When adding lines this way, users are not limited to five lines of the message body; they can continue tapping on the preview entry until the entire email is visible. We chose not to provide a direct method to reduce the size of a preview once expanded. The preview settings remain even after open and closing messages, although users can reset the previews by exiting to the home screen and then returning to the inbox view. Figure 1(c) shows how the inbox view might appear after manually adding additional preview lines to some messages. While some existing email clients allow the configuration of preview lines, the tap-to-add email preview is a novel addition to our client. Including two types of preview manipulation (default and tap-to-add) allows users

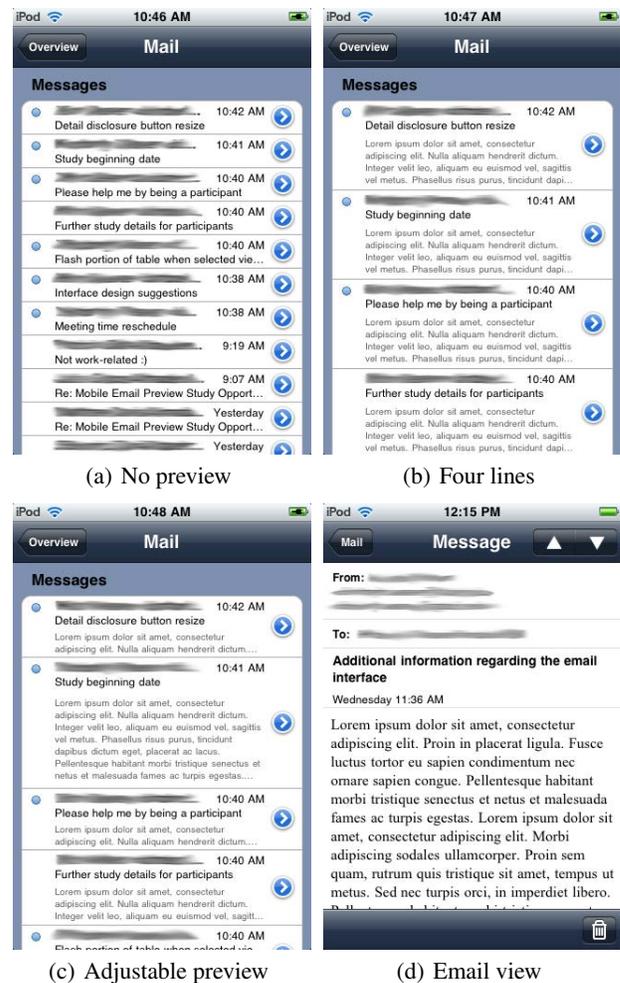


Figure 1. Interface screenshots

to have both gross control over the preview by setting the number of default lines for all messages and more focused control if they wish to see more preview lines for particular messages.

Evaluation methods

There is a wide spectrum of HCI research methods that we could apply for a study, ranging from the experimental to the observational and from objective performance measurements to subjective user preferences. Our initial inclination was to conduct a tightly controlled experimental study that would allow us to quantitatively measure and model the trade-offs introduced by different amounts of information preview with statistical confidence. However, that approach would require that we hold the study conditions consistent across users by introducing artificial constraints and controlled content from an email corpus such as the Enron corpus [6]. The risk in that approach was that the essence of the information previewing activity might actually be lost. A simulated task might not afford the real needs of information preview. The opposite extreme would be to take a completely naturalistic approach and deploy email clients with different information preview settings to a large number of

| Day | Preview Length | Adjustable |
|-----|---------------------------|------------|
| 1 | 0 1 3 5 lines* | No |
| 2 | 1 3 5 0 lines* | No |
| 3 | 5 0 1 3 lines* | No |
| 4 | 3 5 0 1 lines* | No |
| 5 | User's Choice (0-5 lines) | Yes |

Table 1. Presentation of the experimental conditions over the five experiment days. The * symbol indicates that presentation order was varied by participant.

users. The risk with this approach was that people's daily mobile email activities and patterns might be so varied and noisy that no conclusive trends could be easily measured. After repeated deliberation we settled on a middle ground approach that provides elements of control while remaining basically realistic. We asked participants to process their real work emails on their smart phones as they normally would as their first task at work for each of five days, using a different preview setting each day. Our application logged their actions during use and we asked participants to complete a survey questionnaire immediately after each session as well as at the completion of the study after they had experienced all settings.

Having users work with their own email still introduces large variability in the number of emails received, the importance of messages, and the length of messages. However, users are much more likely to be able to determine the contents and context from messages that are from their own contacts and are part of their own conversations than they are with messages they have never seen before from people they do not know. Thus, our use of participants' own email messages allowed us to have more realistic, albeit more messy, study data. While our goal was to have reliable enough data to enable quantitative measurements, we also relied on the subjective ratings from participants to help detect statistically significant differences among preview settings.

Experimental design

We designed the study to last six days. We used the first day to install the custom mobile application on the participant's phone and to collect initial demographic information such as job type, typical email usage patterns, and the mobile and non-mobile clients they employed to check email. For the remaining five days, participants checked their email with our application and we collected usage data. During the first four of these five days, participants had different fixed preview lengths each day (we disabled the ability to increase individual preview lengths during these days). Each day the experiment assigned participants a preview length of 0, 1, 3, or 5 message body lines. All participants experienced each condition, but with the order in which they encountered each condition determined by a partially balanced Latin Square. On the final day, we allowed participants to set their desired level of preview (anywhere from zero to five lines) and enabled the ability to manually increase the previews for individual email messages. A summary of the study parameters is shown in Table 1.

We scheduled the study sessions to occur immediately upon the participant's arrival at work for the day. We instructed participants not to view their email in the morning before using the study application in order to ensure that there would be a sufficient number of unseen email messages from which to gather data. Due to the potential overlap of participants arriving in the morning it was necessary to design the study to minimize the demand on the experimenter so that participants could be run in parallel and so that participation in the study would minimally interfere with the work responsibilities of the participants.

Participant recruitment

We recruited sixteen participants for the study from the employee population of a large US corporation. At the start of the study we asked volunteers to provide some general demographic information and details about their current mobile device usage and email practices. All participants were iPhone users. Participants were between the ages of 28 and 60 years old ($M = 39$, $SD = 11.1$). Eleven were male and five were female. The participants represented a variety of job types: five held management positions, eight were researchers, one participant was an executive assistant, and two were software engineers. The average number of emails participants reported receiving each day was highly variable, with a range of 10 to 100 emails per day ($M = 44.4$, $SD = 31.4$). All participants had owned their iPhone for at least one year ($M = 1.96$, $SD = 0.59$).

Twelve of the participants reported using the native iPhone mail application. The default preview setting for that application is for two lines of body preview [1]. Although few knew it, the native iPhone email application does allow users to adjust the amount of preview in the Settings to anywhere from zero to five preview lines. Five of the participants reported using Gmail in Safari. Gmail uses one line of preview from the body of the message [15]. Three participants reported using Lotus iNotes on their phone, which has no lines of preview from the message [10]. One participant each reported using Hotmail in Safari, the Yahoo! email application and AOL.

Data collection

We collected experiment data from two sources: direct reports from the participants through questionnaires and logs of participant actions within our custom email application.

Participant responses

Every day, after processing their email with a particular preview setting, we asked participants to rate their impression of that preview setting on a 7 point scale where 1 was the worst and 7 was the best. We asked them to explain their rating and to list any advantages or disadvantages of that particular setting. Finally, we asked participants to report any strategies they used to check their email while using that setting. On the final day of the study, we asked participants to rank the settings from best to worst and to again explain their rankings. We also asked them to indicate their preference for fixed or adjustable previews and to explain their choice.

Logging participant actions

The types of activities recorded by the application can be divided into four different categories: email metadata, email actions, preview settings, and scrolling behaviors.

- *Email metadata*: Traits of an email which may contribute to the likelihood of being read, independent of preview size.
 - Number of recipients (e.g., sole, few, many, broadcast)
 - Size of email message represented by the number of vertical pixels the email occupies when rendered.
- *Email actions*: Actions performed on a specific email
 - The email was opened
 - The email was deleted
- *Preview settings*: Actions related to preview size in the email overview.
 - Default settings including the number of preview lines and whether or not the size of individual email previews can be increased inside the application
 - Whether the user has increased the number of lines for a specific message preview
- *Scrolling behaviors*: Actions determining which emails are visible at any point in time
 - Items visible when mail is first loaded
 - Scrolling has begun
 - Items appear in view due to scrolling
 - Scrolling has ended, includes direction of scrolling and how far the user scrolled (in pixels)

RESULTS

We present our results in three sections. We first present a summary of participants' comments on their experiences with the different preview settings. This summary provides a useful background for understanding the subsequent two sections. The second section presents results drawn from the logged user actions. The last section presents the quantitative experience ratings participants gave to the preview settings.

Participant comment analysis

In this section we present participants' subjective responses (their written comments) on the different experimental conditions. We collected the written responses from all participants and categorized them into positive, negative, and neutral comments about the number of visible preview lines. Table 2 shows the percentage of positive, neutral and negative comments out of all of the comments for each preview line condition. For the zero lines preview condition, the percentage of negative comments was greater than the percentage of positive comments, indicating that participants were largely unsatisfied with this preview setting. Responses were equally distributed between positive and negative comments for the one line condition. The three lines condition had the

| Preview Length | Positive | Neutral | Negative |
|----------------|----------|---------|----------|
| 0 lines | 31.3% | 12.5% | 56.3% |
| 1 line | 48.3% | 3.5% | 48.3% |
| 3 lines | 75.0% | 14.3% | 10.7% |
| 5 lines | 63.3% | 3.3% | 33.3% |

Table 2. Percentage of positive, neutral and negative comments about each of the conditions.

| Comment | Preview Length | | | |
|---------------------------------|----------------|---|----|----|
| | 0 | 1 | 3 | 5 |
| <i>Opening messages</i> | | | | |
| Had to open a lot of messages | 4 | 5 | | |
| Didn't have to open messages | | | 9 | 7 |
| <i>Level of detail</i> | | | | |
| Not enough detail | 10 | 9 | | |
| Too many lines | | | 3 | 7 |
| <i>Scrolling</i> | | | | |
| Scrolling was easy | | | | 1 |
| Scrolled too much | | | | 3 |
| <i>Other</i> | | | | |
| Provided useful information | 2 | 7 | 12 | 11 |
| Overwhelming number of messages | 4 | | | |
| Fast and concise | 8 | 7 | | |

Table 3. Common themes emerging from participant comments

most positive comments and the least negative comments, indicating the highest level of satisfaction. This positive to negative ratio began to shift back for the five lines condition, indicating it is not necessarily true that users are always more satisfied with more preview.

In order to better understand the advantages and disadvantages of the different preview settings, we then examined the positive and negative comments to determine common themes. These themes are shown in Table 3. With regard to opening messages, we see a switch in participants' responses between one and three lines of preview. Participants felt that they had to open a lot of messages for zero or one preview lines, possibly because at this level they could not make decisions about how to handle a message without opening it to see the actual message body. By contrast, opening messages often became unnecessary with three and five preview lines. As a corollary, participants felt like there was not enough detail provided by zero or one line, while they made comments about there being too many lines for the three and five lines settings. There were not many comments about scrolling overall, but some participants did feel they had to scroll too much in the five lines preview setting, and one participant said that scrolling was easy in the five lines setting. Participants reported that the preview did provide useful information at each of the preview levels, however this type of comment was more prevalent in the three and five preview lines settings. One quarter of the participants said that when using zero lines, the number of visible messages was overwhelming. Half of the participants said that the zero and one line preview settings were fast and concise.

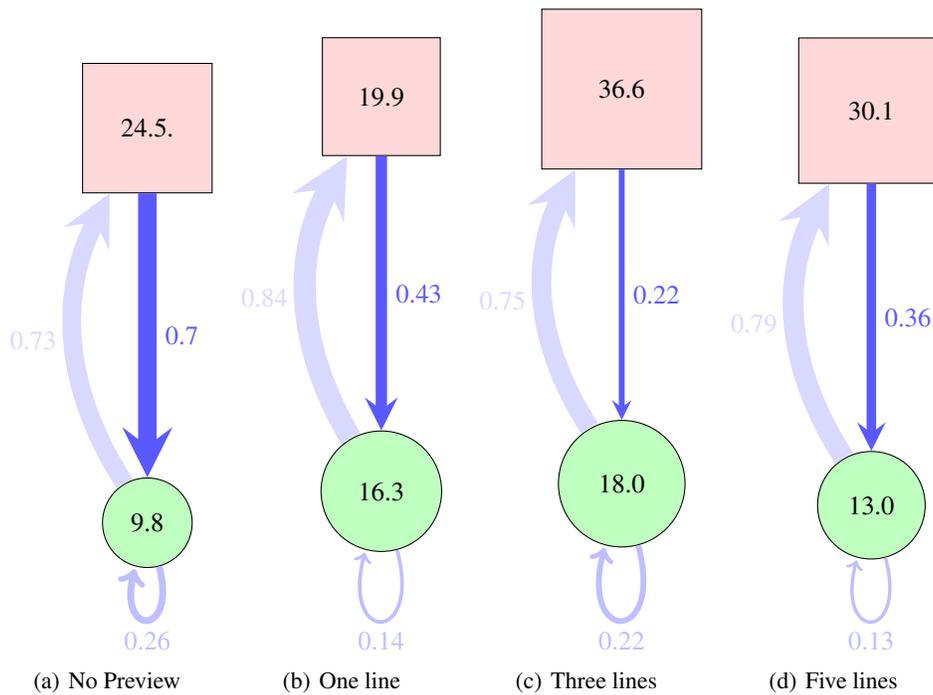


Figure 2. UI State Transitions diagrams. A box represents the *PreviewList* state, a circle represents the *FullMessage* state, and the arrows represent transition likelihood. The size of a box or a circle represents the amount of average time spent in that state before transitioning to another state.

Logging data results

On average, participants had 17.1 new messages at the beginning of each study session, with a range of 0 to 45 and a standard deviation of 10.2. During study sessions, the average duration of application use was 313 seconds. However, the range of duration times was quite large, between 100 and 1388 seconds, with a standard deviation of 276 seconds. In addition, there is no significant correlation between the number of messages received and the work duration ($r = 0.16$, $p > 0.2$). Participants clearly have very different styles when working with email messages on their phones.

UI state transitions

The detailed event logging allowed us to reconstruct the user interface state transitions. We were interested in two UI states. One is viewing the inbox (the list of messages). We denote this state type as *PreviewList*. The second is the full view of an individual mail message, denoted as *FullMessage*. Our application provides three possible transitions among the two states: two transitions going between the two different states (one in each direction), and one looping within the *FullMessage* state itself. Self-looping is either a result of deleting the currently viewed message (which views the next message automatically), or tapping the “previous message” or “next message” interface elements to navigate directly between messages. For each state we calculated the average time participants spent before transitioning to another state, which we encoded as the areas of the state nodes in Figure 2. The areas of the squares are the average time (in seconds) spent in the *PreviewList*; the areas of the circles are the average time spent viewing the *FullMessage*. As might be expected, the duration spent viewing an opened

email message was correlated with the length of the message, $r = 0.39$, $p < 0.005$.

We also counted the number of *PreviewList* to *FullMessage* transitions per session. By dividing those values by the number of mails loaded for that session, we obtained a measure of the likelihood of a participant to open up the full view of a message from the inbox (preview list) view. We calculated the same probabilities for the *FullMessage*’s self-looping and the *FullMessage* to *PreviewList* transitions. We encoded these state transition likelihood measures as the thickness of the arrowed lines in Figure 2. Clearly, participants are the least likely to open up full views of messages when previewing three lines. With no preview, participants are most likely to open up full views of messages.

The state transition models and visualizations in Figure 2 effectively portray the different use patterns in different preview settings. The three lines condition, Figure 2(c), clearly stands out. It has a large box (a 36.6 second *PreviewList* state), a large circle (an 18 second *FullMessage* state), and a thin line from the box to the circle (22% messages opened). This result is quite plausible, because three lines of preview might have enabled users to process many messages (skipping, ignoring, or deferring messages based on sufficiently understood email content) in the *PreviewList* state before drilling down to the *FullMessage* state. Once opening up a message (the *FullMessage* state), a user spends more time reading it because it is likely to be a substantive, relevant, or long message given that the user made a decision to open it after reading its first three lines of preview. In contrast, the no preview condition, Figure 2(a) features a smaller box, a

smaller circle, and a much thicker line from the box to the circle. These characteristics indicate that in this condition users could do much less processing in the *PreviewList* state before having to drill down to the *FullMessage* state. Users might even need to open a message to ensure it is unimportant before deleting it. Upon opening a message, the likelihood for it to be worth spending time reading is much less. Without the ability to preview before opening users are more likely to open irrelevant messages. Recall that we used realistic tasks, so the message number, email content and email length fluctuated naturally. Given these noisy factors, the sharp contrast between the multi-line preview conditions and the zero preview condition is still remarkably clear.

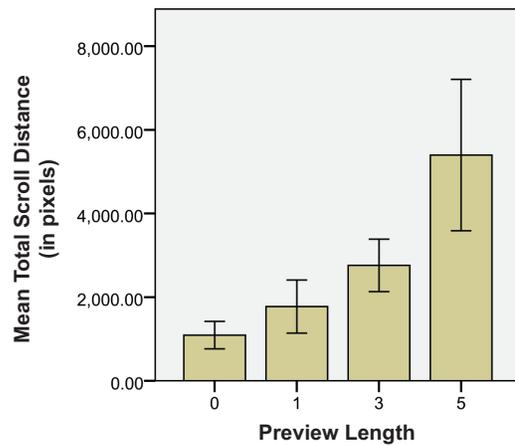
The likelihood of transitioning between the *FullMessage* and the *PreviewList* is relatively uniform, ranging from 0.73 to 0.84. This indicates that participants for the most part deleted emails while in the *PreviewList* view. They also rarely used the “previous” and “next” message buttons. The two conditions with fewer preview lines have the highest probability of transition between the *PreviewList* and the *FullMessage*. This is consistent with participants needing to read the email more often in order to decide on its importance. The lower probability of transition from *PreviewList* to *FullMessage* for the two longer preview lengths paired with the longer time spent in the *PreviewList* indicates that the three and five lines settings allowed participants to get sufficient information from the preview to make decisions on less important messages without opening them.

Impact of differing amounts of preview lines

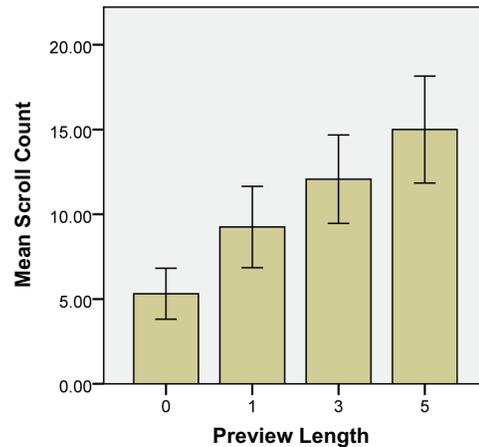
The number of preview lines should directly influence the time that participants spend in the *PreviewList* view. Simplistically, the time spent in the *PreviewList* consists of the time for reading the previews (*ReadingTime*), and the time for scrolling the list (*ScrollingTime*). Ideally, we would like to maximize the former while minimizing the later.

We applied statistical analyses on the user actions from the log data. Due to the nature of the realistic task, we had to remove two outlier data points, probably due to interruption, from the statistical analysis. We recorded the distance (in pixels) for each scrolling action. To see the effect of the number of preview lines on scrolling behavior, we fitted a repeated measure general linear model to the total scrolling distances in the study sessions, with the number of preview lines as the within-subjects factor. The main effect of preview lines is significant, $F_{3,39} = 3.68, p < 0.02$. Figure 3(a) depicts the scroll distance by the number of preview lines.

A similar trend is present for the number of scroll actions performed by participants. Figure 3(b) depicts scroll action counts by the number of preview lines. Clearly, participants scrolled more with more preview lines. A repeated measure general linear model suggests that the effect of preview lines is significant $F_{3,39} = 3.4, p < 0.03$. The no preview condition induces a significantly lower number of scroll actions than the three and five lines conditions, with p values of 0.04 and 0.02 respectively. Other pairwise comparisons are not



(a) Scroll distance



(b) Scroll count

Figure 3. The effect of preview lines on scrolling behavior

significant but are close, with p values ranging from 0.14 to 0.17, except for the difference between one and three preview lines ($p = 0.5$).

Dividing the total scroll distance by the standard mean iPhone scrolling speed of 30 pixels per second, we obtain the total time the participants spent scrolling through the interface. Normalizing it by the number of mails loaded, we obtained the average time spent scrolling per message while viewing the inbox, *ScrollingTime*. We have already calculated the average time spent viewing the inbox (encoded as the size of the squares in Figure 2); subtracting the average *ScrollingTime* from the average time spent viewing the inbox gives us the amount of time on average a person spent reading individual messages in the the inbox *ReadingTime*. Figure 4 plots these two average times (*ScrollingTime* and *ReadingTime*) against the number of preview lines.

Using the number of preview lines as the within-subjects independent factor, we tested the fit of both the average *ScrollingTime* and *ReadingTime* to a repeated measure general linear model. The overall effect of preview lines is significant for the multivariate test, $F_{3,39} = 3.1, p < 0.04$. A

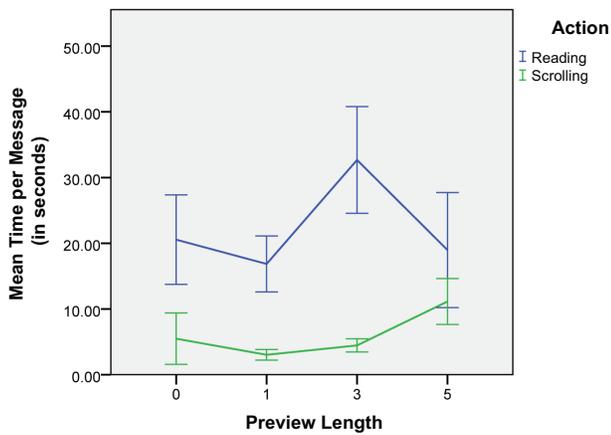


Figure 4. The effect of preview lines on reading and scrolling time

| Preview Length | Mean | Std. Deviation | N |
|----------------|---------|----------------|----|
| 0 Line | 3.58125 | 1.657496 | 16 |
| 1 Line | 4.40625 | 1.189800 | 16 |
| 3 lines | 5.12500 | 0.944281 | 16 |
| 5 lines | 5.06875 | 1.445899 | 16 |
| user choice | 6.06563 | 1.107056 | 16 |

Table 4. User rating and rankings of preview conditions

univariate test for *ReadingTime* was not statistically significant, $F_{3,39} = 1.1$, $p = 0.36$, but the test for *ScrollingTime* was close to significance, $F_{3,39} = 2.5$, $p = 0.08$. For *ReadingTime*, the pairwise comparison between one and three preview lines produced a significant difference ($p < 0.05$). Other pairwise differences were not statistically significant. For *ScrollingTime*, the pairwise comparison between one and five preview lines was significant ($p < 0.05$), and the five preview lines condition's difference from the zero and three preview lines conditions were both close to being significant, with p value of 0.08 and 0.06 respectively. The difference between one and three preview lines was also close to being significant ($p = 0.1$).

Overall, although we chose a naturalistic task and thus fully expected the performance data to be very noisy, we still observed several statistically significant or nearly significant impacts on user behavior from different preview amounts. More preview lines clearly led to more scrolling. The three lines preview condition clearly enabled more reading (and information processing) from the inbox view.

Rating analysis

We show the subjective ratings of the five conditions, including the user-choice condition on the last day, in Table 4. Variance analysis shows a significant main effect of preview conditions (0 lines, 1 line, 3 lines, 5 lines, user choice) on participants subjective ratings: $F(4, 60) = 9.722$, $p < 0.0001$. Sphericity corrections (such as Greenhouse-Geisser) did not have meaningful impact on the statistics. LSD pairwise comparisons show that the three lines condition was significantly more highly rated than the no preview ($p = 0.004$) or one line ($p = 0.035$) conditions. From three lines

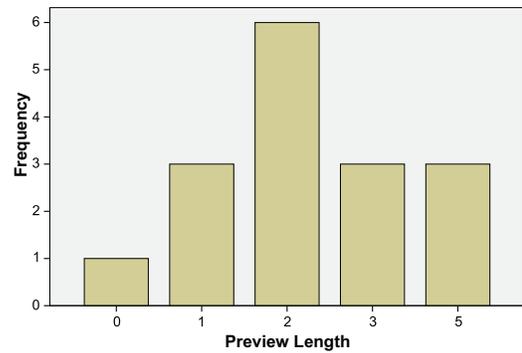


Figure 5. Histogram of participant preview choice.

to five lines there was no significant increase in mean rating ($p = 0.815$). Participants rated the user-choice condition significantly higher than the no preview ($p < 0.0001$), one line ($p = 0.001$), and most notably the three lines ($p = 0.034$) conditions. The user-choice condition was also almost significantly more highly rated than the five lines condition ($p = 0.053$). Additionally, the one line condition was rated almost significantly higher than the no preview condition ($p = 0.056$).

Friedman's nonparametric rank test shows a significant main effect among the preview condition rankings (no preview, one line, three lines, five lines). Pairwise comparisons show that the significantly different pairs in ranking were: three lines was ranked better than zero lines ($p < 0.0001$) and one line was also better than zero lines ($p = 0.006$).

Taking the rating and ranking results together, we can draw the following conclusions about participants' overall experience with each preview condition. First, the one line preview condition is likely to be preferable to the zero lines, no preview condition. This conclusion was strongly supported by their significantly different ranks. It was also supported, although weakly, by the rating data: the one line rating was almost significantly higher than zero line rating.

Second, the three lines condition provides an optimal combination, accommodating different factors that influenced participants opinions. It was significantly higher rated than both the zero lines and one line settings. It was also ranked higher than the zero line setting. There was no significance increase in rating from three lines to five lines. In fact the five lines settings mean rating was slightly lower than the three lines settings rating.

User choice analysis

On the final day of the study, participants picked the preview setting that they preferred. Figure 5 shows a histogram of the participants' choices. At least one participant chose each of the possible preview settings. Three people each chose the one line, three lines and five line settings respectively. The two lines preview setting was the most popular, with six people choosing it. We also asked participants to give their opinion on whether or not they wanted to set the same preview

amount for all of the messages or if they wanted the ability to increase the preview for individual messages. Eleven participants (68.75%) desired the ability to manually adjust previews. Only five participants (31.25%) wished to have fixed preview settings. Of those five participants, three participants may have misunderstood the question. These participants focused more on the tap interaction method we chose to activate the increased preview. They would have preferred the tap action to open up the full view of the message. It is unknown whether these three participants felt there was no value to the adjustable preview, or if they were objecting to the particular interaction. Of the remaining two participants, one felt that adjustable preview was “annoying.” The final participant chose the five lines preview and thus felt that not much extra information would be gained from tapping to increase the preview still further.

Twice as many participants chose to use two lines of preview than any other preview setting. This suggests that the optimal preview length may be two lines rather than three. The absence of overlapping comments in participants’ subjective responses between the one and three lines preview settings may serve as further evidence of the utility of a two lines preview setting. We note that while we explicitly chose not to include the two lines preview setting in the first four days of the study (in order to achieve a distribution of preview lengths), it is the default setting for the iPhone native email application. This fact may have had some impact on participants’ choices. Overall, a moderate amount preview, in two or three lines of text for the iPhone dimensions, appears to be an optimal trade-off range.

DISCUSSION AND CONCLUSIONS

Through an experiment studying realistic email use on mobile phones, we examined the impact of different amounts of information preview on email processing behavior. Combining results drawn from analyses of subjective user comments, UI state transition modeling, logged user actions, overall experience ratings, rankings, and user choices, we can now draw conclusions about the impact of different amounts of message body previews.

Considering all of these factors, providing a preview is clearly preferable to not providing any preview at all. This conclusion is supported by almost all of the measures in our study. On the other hand, it is also clear that providing more preview is not always better. We saw that as the number of preview lines increased, the mean scrolling distance and the mean number of scroll actions also increased. Both the three and five line preview conditions involved significantly more scroll actions and required participants to scroll farther. That scrolling adds an interaction cost.

A moderate amount of preview, incorporating two or three lines from the body of the email, provides the best balance between competing factors. Although a moderate preview requires more scrolling than no preview in the inbox view, at the three lines preview setting that increase in scrolling is balanced by a reduced number of transitions between the inbox view and views of individual email messages. Our

state transition analysis clearly shows that users could process more messages from the inbox view with a moderate information preview than with a shorter preview. User ratings, rankings and settings choices also clearly favored moderate (two or three lines) preview lengths. We note that this result suggests that the iOS mail application is providing sufficient preview information, but that the Android, Blackberry, Symbian, and webOS mail applications all provide too little.

Our results also show that if possible, designers should allow users to adjust the default preview setting. Users rated an adjustable preview significantly higher than a fixed moderate (three line) preview. This conclusion is also validated by the fact that our participants, when given the choice, were divided on which preview setting they wanted to use. Our results also suggest that designers should consider allowing users to increase the amount of preview for individual messages as a middle ground between scanning just the default preview and viewing the entire message.

In closing, we note that our results and conclusions focus on mobile email processing and that we studied previews consisting of the initial lines of message bodies. Our results clearly illustrate the theoretical trade-off between the amount of information and the cost of interaction. There is considerable room for additional experimentation to determine whether our results transfer to other types of tasks users perform on mobile phones. In addition, other types of email message previews (e.g., synthesized summaries or extracted previews) may allow users to determine the contents of messages with shorter previews. Finally, further work is necessary to determine whether mobile devices with different physical dimensions or form factors (e.g., tablets such as the iPad) may have different sweet spots to balance the trade-offs inherent in designing information previews.

REFERENCES

1. Apple, I. Apple - support - iPhone - setting up mail assistant.
<http://www.apple.com/support/iphone/assistant/mail/>, 2010.
2. Bates, M. J. Information search tactics. *Journal of the American Society for Information Science* 30, 4 (1979), 205–214.
3. Bellotti, V., Ducheneaut, N., Howard, M., and Smith, I. Taking email to task: the design and evaluation of a task management centered email tool. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM (Ft. Lauderdale, Florida, USA, 2003), 345–352.
4. Carenini, G., Ng, R. T., and Zhou, X. Summarizing email conversations with clue words. In *Proceedings of the 16th international conference on World Wide Web*, ACM (Banff, Alberta, Canada, 2007), 91100.
5. Chi, E. H., Pirolli, P., Chen, K., and Pitkow, J. Using information scent to model user information needs and actions and the web. In *Proceedings of the SIGCHI conference on Human factors in computing systems*,

- ACM (Seattle, Washington, United States, 2001), 490–497.
6. Cohen, W. M. Enron email dataset. <http://www.cs.cmu.edu/~enron/>, Aug. 2009.
 7. Cui, Y., and Roto, V. How people use the web on mobile devices. In *Proceeding of the 17th international conference on World Wide Web*, ACM (Beijing, China, 2008), 905–914.
 8. Dabbish, L. A., Kraut, R. E., Fussell, S., and Kiesler, S. Understanding email use: predicting action on a message. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM (Portland, Oregon, USA, 2005), 691–700.
 9. Dredze, M., Wallach, H. M., Puller, D., and Pereira, F. Generating summary keywords for emails using topics. In *Proceedings of the 13th international conference on Intelligent user interfaces*, ACM (Gran Canaria, Spain, 2008), 199206.
 10. Jacob, B. Ultralight for iPhone. http://www-10.lotus.com/ldd/dominowiki.nsf/dx/Ultralight_for_iPhone, Nov. 2009.
 11. Jones, M., Buchanan, G., and Thimbleby, H. Sorting out searching on small screen devices. In *Human Computer Interaction with Mobile Devices*, F. Patern, Ed., vol. 2411 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2002, 555–567.
 12. Jones, M., Marsden, G., Mohd-Nasir, N., Boone, K., and Buchanan, G. Improving web interaction on small displays. *Computer Networks* 31, 11-16 (May 1999), 1129–1137.
 13. Khan, R., Mease, D., and Patel, R. The impact of result abstracts on task completion time. In *WWW2009* (Madrid, Spain, Apr. 2009).
 14. Matthews, T., Pierce, J., and Tang, J. No smart phone is an island: The impact of places, situations, and other devices on smart phone use. Technical Report RJ10452, IBM Almaden Research Center, San Jose, CA, USA, Sept. 2009.
 15. McKinley, J. Gmail gets a new engine for iPhone and android-powered devices. <http://googlemobile.blogspot.com/2009/04/gmail-gets-new-engine-for-iphone-and.html>, Apr. 2009.
 16. Oulasvirta, A., Hukkinen, J. P., and Schwartz, B. When more is less: the paradox of choice in search engine use. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, ACM (Boston, MA, USA, 2009), 516523.
 17. Oulasvirta, A., Tamminen, S., Roto, V., and Kuorelahti, J. Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM (Portland, Oregon, USA, 2005), 919–928.
 18. Pierce, J. S., Bunde-Pedersen, J., and Ford, D. A. Triage and capture: Rethinking mobile email. Technical Report RJ10458, IBM Almaden Research Center, San Jose, CA, USA, Jan. 2010.
 19. Rambow, O., Shrestha, L., Chen, J., and Lauridsen, C. Summarizing email threads. In *Proceedings of HLT-NAACL 2004: Short Papers on XX*, Association for Computational Linguistics (Boston, Massachusetts, 2004), 105108.
 20. Sweeney, S., and Crestani, F. Effective search results summary size and device screen size: Is there a relationship? *Information Processing & Management* 42, 4 (July 2006), 1056–1074.
 21. Venolia, G., Dabbish, L., Cadiz, J. J., and Gupta, A. Supporting email workflow. Technical Report MSR-TR-2001-88, Microsoft Research, Redmond, WA, Dec. 2001.
 22. Wan, S., and McKeown, K. Generating overview summaries of ongoing email thread discussions. In *Proceedings of the 20th International Conference on Computational Linguistics* (2004).
 23. White, R. W., Ruthven, I., and Jose, J. M. Finding relevant documents using top ranking sentences: an evaluation of two alternative schemes. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, ACM (Tampere, Finland, 2002), 57–64.
 24. Whittaker, S., Bellotti, V., and Gwizdka, J. Everything through email. In *Personal Information Management*, W. Jones and J. Teevan, Eds. University of Washington Press, Seattle, 2007, 167–189.
 25. Whittaker, S., and Sidner, C. Email overload: exploring personal information management of email. In *Proceedings of the SIGCHI conference on Human factors in computing systems: common ground*, ACM (Vancouver, British Columbia, Canada, 1996), 276–283.
 26. Wilson, M. L., Schraefel, M. C., and White, R. W. Evaluating advanced search interfaces using established information-seeking models. *J. Am. Soc. Inf. Sci. Technol.* 60, 7 (2009), 1407–1422.