EmailValet: Managing Email Overload through Private, Accountable Crowdsourcing

Nicolas Kokkalis, Thomas Köhn, Carl Pfeiffer, Dima Chornyi, Michael S. Bernstein, Scott R. Klemmer
Stanford University HCI Group
Computer Science Department, 353 Serra Mall, Stanford, CA
{nicolas, thomasko, carlp87, chornyi, msb, srk}@cs.stanford.edu

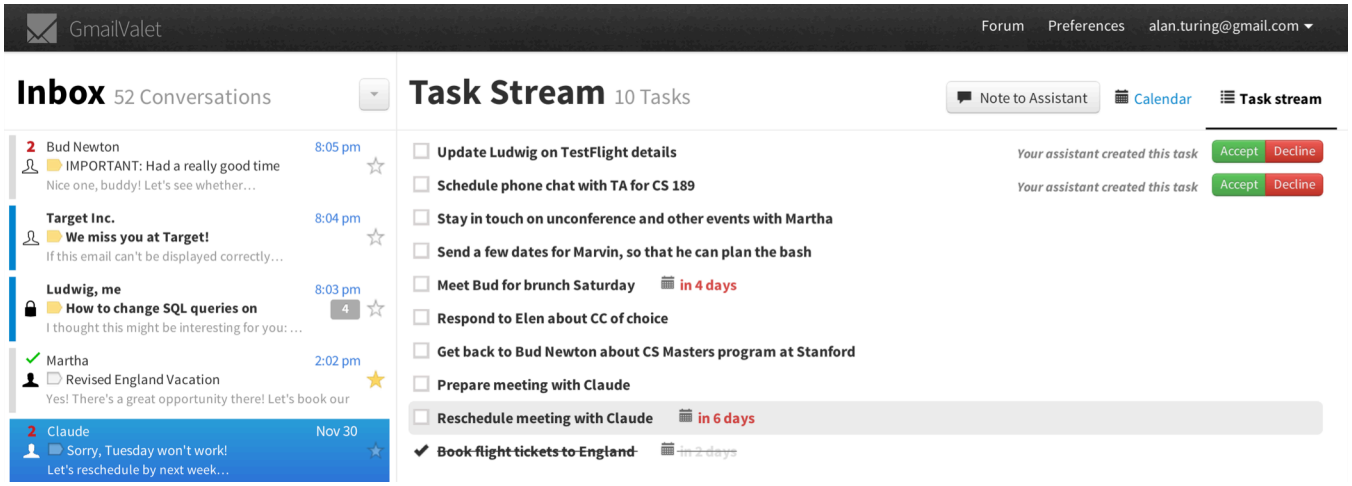

Figure 1. The EmailValet email client draws on crowdsourced expert assistants to transform a cluttered inbox into an organized task stream. Assistants are given limited, accountable access to the user's inbox so that they may extract tasks from each email.

ABSTRACT

This paper introduces privacy and accountability techniques for crowd-powered systems. We focus on email task management: tasks are an implicit part of every inbox, but the overwhelming volume of incoming email can bury important requests. We present *EmailValet*, an email client that recruits remote assistants from an expert crowdsourcing marketplace. By annotating each email with its implicit tasks, EmailValet's assistants create a task list that is automatically populated from emails in the user's inbox. The system is an example of a *valet* approach to crowdsourcing, which aims for parsimony and transparency in access control for the crowd. To maintain privacy, users specify rules that define a sliding-window subset of their inbox that they are willing to share with assistants. To support accountability, EmailValet displays the actions that the assistant has taken on each email. In a weeklong field study, participants completed twice as many of their email-based tasks when they had access to crowdsourced assistants, and they became increasingly comfortable sharing their inbox with assistants over time.

CSCW '13, February 23–27, 2013, San Antonio, Texas, USA.

Author Keywords

Crowdsourcing; Email Overload; Human Assistants; Task Management; Access Control.

ACM Classification Keywords

K.4.3 [Organizational Impacts]: Computer-supported collaborative work.

General Terms

Design; Human Factors.

INTRODUCTION

Email management means triaging a never-ending tide of incoming requests. New messages push important requests out of view, and those requests can be unintentionally missed [26, 38]. To avoid overlooking important messages, people spend large amounts of time carefully processing their inbox or triage by focusing only on high priority messages [6, 18, 35]. However, people often keep unfinished tasks in their inbox [38], and triaging is error-prone [35]. As a result, tasks are often mixed with other emails, get pushed down by new messages, become hard to find, and forgotten.

Current approaches for handling email-based tasks are limited and/or expensive. Integrating task management directly into the email client [3, 8] or asking communicators to structure their requests [39] requires significant manual effort. Automatic techniques have shown some promise in identifying tasks in emails [12, 16, 19, 25], but they are not yet fully reliable [25] and require heavy-handed user inter-

action [12] and training [19]. Finally, while privileged individuals have long relied on personal assistants to help with email, their integration into the email client is limited and the cost can be prohibitive.

In this paper, we combine the accuracy and oversight advantages of personal assistants with the large-scale availability and affordability of crowdsourced experts. We recruit workers from expert crowdsourcing platforms to help extract tasks, and share these assistants across multiple people. This multiplexing increases employment for assistants and affordability for users. To present these task-organized messages, we introduce the *EmailValet* mail client. EmailValet's task list (Figure 1) condenses an inbox to actionable items, makes tasks more prominent and easier to track, increases efficiency through task-oriented interactions (rather than coopting general email primitives, like marking as unread), and focuses attention on the most important emails.

To explore the potential for crowdsourced solutions, we conducted a formative survey finding that people feel a tension recruiting remote assistants for managing complex, personal information: they want help, but have reasonable concerns about giving strangers unfettered access. To address this tension, this paper introduces the *valet* crowdsourcing approach. Like a valet key, valet interfaces seek a dual objective of parsimony and transparency. First, they parsimoniously give assistants just enough access to help with most of the work. Second, they make access boundaries transparent so users have an accurate model of what the assistant can and cannot do, and they cause transgressions to leave fingerprints. Achieving these dual objectives provides peace-of-mind for users and limits liability for assistants.

We illustrate this approach through EmailValet. For parsimony, EmailValet's access control shares a limited window of the inbox (by default, the 100 most recent messages), and limits assistants' actions (to creating tasks). Users author a whitelist and blacklist of messages to be shared, and for transparency, each message presents its access status with glanceable icons. Furthermore, EmailValet provides a viewable log of all assistants' actions.

We hypothesize that EmailValet helps people manage the tasks in their inbox, while maintaining acceptable privacy and accountability practices. A field experiment compared participants' task completion rates when they did not use EmailValet's task functionality (*Control* condition) to when they extracted tasks themselves (*Self* condition) and to when they took advantage of crowdsourced expert task extraction (*Assistance* condition). With crowdsourced task assistance, people completed twice as many email-based tasks than with either standard email or EmailValet's task interface without assistants (Figure 2). Participants grew more comfortable sharing their inbox with assistants as the study progressed.

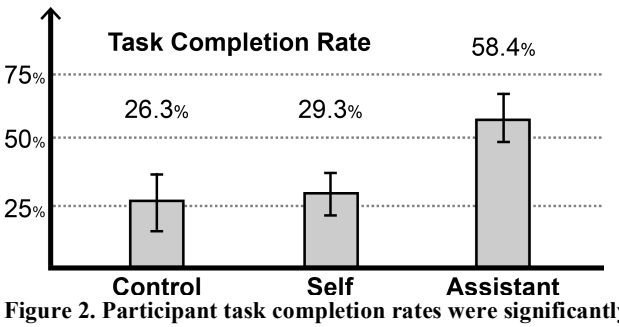

Figure 2. Participant task completion rates were significantly higher when assistants auto-extracted tasks for them.

This work contributes:

- The first use of crowdsourced expert assistants to support personal information management.
- The valet approach for privacy and accountability, which seeks parsimonious, transparent sharing.
- The EmailValet system for email task management, which introduces crowd assistants, richer sharing filters, and icons summarizing assistants' activity.
- A task-integrated feedback interface. This lightweight feedback structure helps users negotiate common ground so assistants can personalize their support. These conversations are shared with all assistants helping that user.
- Empirical results that crowdsourced assistants manage information accurately, enabling EmailValet users to accomplish more tasks.

RELATED WORK

EmailValet extends work on email overload, automated email management, and task extraction. It contributes the valet approach, crowdsourcing these goals to an expert assistant who is multiplexed across many users.

Email overload is an enduring challenge [18, 38]. While email has many positive impacts on organizations [13], frequent email interruptions can decrease productivity and large volumes can create information overload [21, 22, 36].

Data mining, categorization, and metadata extraction can help users organize their email. These approaches ease email filing [2, 7, 11, 23, 31, 32, 33], detect important emails [2, 11, 20, 33], summarize messages [29], forward and reply to messages [21], and filter spam [30]. For example, Gmail's Priority Inbox identifies and highlights emails that the user is likely to want to read and respond to. Other work shares EmailValet's goal of detecting action items and tasks in emails [9, 12, 19, 33] or even automating their execution (*e.g.*, booking a meeting room) [19]. However, these machine learning approaches often suffer from false positives [12] that a professional assistant would not make [15]. Consequently, EmailValet reaches out to human assistants.

Information workers manage many of their tasks using their email clients [25, 38]. Common mail clients (*e.g.*, Outlook, Gmail) allow people to mark a message as a task and set a due date [3, 4]. Taskmaster grouped message threads by task and allowed users to customize these groupings [8]. TimeStore-TaskView is an email interface that is centered

around the relationship between messages, with the focus on managing future tasks [20]. EmailValet, like TaskView and Taskmaster, foregrounds tasks rather than messages. It differs from this prior work by integrating human assistants into the loop to lower users' organization burden. EmailValet's inbox interface also adds iconic cues that reflect assistants' activity.

Executive assistants help stem the tide of email overload, focusing their principal's attention on important messages, shielding them from unimportant ones, and handling simple tasks autonomously [15]. Reflecting this, mail clients such as Microsoft Outlook allow users to delegate limited inbox access to assistants [1, 5]. This paper generalizes and extends these relationships to a large class of new users and assistants. To do so, we introduce techniques for an assistant to support multiple users, and focus on building common ground and maintaining transparency with remote assistants.We hypothesize that the manual effort of user-controlled systems and low accuracy of automated systems have yielded low adoption in practice. Our intuition was that paid crowd assistants could be a high-accuracy, low-effort, and affordable solution, but that user trust would be a main concern. Previous crowd-powered systems (*e.g.*, [10]) have largely sidestepped the challenges of sharing private data with crowd workers.

FORMATIVE SURVEY AND INTERVIEWS: CONCERNS WITH CROWD ASSISTANTS

We combined a large-scale online survey with semi-structured interviews to establish current email privacy and security concerns. We began with an online survey on Amazon Mechanical Turk with 585 U.S. residents (59% female; 36% aged 18–25 and 33% aged 26–35). It offered a $0.01 payment, which reduced the incentive to spam and selected for workers with intrinsic interest in the subject.

We followed up with 48 semi-structured interviews (32 male). We recruited participants through email lists at our university and from our professional networks. Thirty-three participants were MBA students; 15 had technical backgrounds, five of whom were at the manager level, six were students, and four were researchers. Seventeen interviews were conducted in-person; 31 were online. The interviews discussed specific concerns and opportunities with crowd-powered inbox assistants.

We combined the interview and survey data inductively. We clustered themes from the interviews and matched them with quantitative survey results. Several authors then collaborated to code interview notes and transcripts with those themes.

Results

Respondents expressed strong concerns about sharing their inbox with a crowdsourced assistant. However, they also resonated with the goal of better task management in email and raised design opportunities around filters, whitelists, and accountability.

Task Management in Email

Forty-eight percent of respondents reported using emails to manage tasks. Of those, 77% sent email reminders to themselves, 47% used their inbox as a to-do list, and 41% stated that they would use an online service that helps better manage tasks in email. Survey participants offered a glimpse into their day-to-day email triage: for example, "I want to be able to add little notes to every email in my inbox. I would isolate the exact action I need to take on each email." Another participant shared: "The biggest pain point is that for each email for which I need to take action, I have a specific action I want to take, but I cannot record it anywhere for the email. Every time I see the email, I go into it and rethink about what the action is, and then I decide if I have time to do the action. I do this an average of 2-3 times per email. It kills me."

Participants were interested in having crowdsourced assistants support inbox triage. One manager wanted an assistant to filter the inbox and "automatically archive all messages that do not require me specifically to take some action." Another participant, familiar with Priority Inbox, said: "I want an affordable […] human alternative to applying my rules to emails*, i.e.*, what to escalate to me immediately, respond to quickly with a canned response, flag for my review quickly, flag for my response immediately, distill the action I need to take, delete, etc."

Privacy and Security

The responses indicate that privacy concerns were the major roadblock to adopting a system like EmailValet. More than two thirds of respondents cited major concerns: they were only willing to share a few messages manually (35.4%) or share nothing at all (38.1%). Roughly one quarter (26.2%) were comfortable with an automatic solution via email rules; only a few (4.1%) were ready to share their entire inbox.

Participants who were only willing to hand-pick emails that an assistant could see acknowledged that this might neutralize any time savings the assistant might offer: "Unfortunately, the only way I'd feel totally comfortable is if I could pre-screen messages beforehand, but even the act of forwarding certain messages to an assistant […] feels too time-consuming."

Respondents who were willing to share parts of their inbox wanted strict access restrictions. They viewed historic inbox data as unnecessary: "[Historical access] would make me nervous about finding something I didn't remember was in there." Most participants also wanted the inbox automatically filtered to remove personal emails and financial information and passwords; and the ability to revoke permission when a sensitive email slips through.

The more freedom the assistant has to take actions, such as write replies or archive emails, the more participants felt the need for monitoring. A manager said: "I do not mind sharing, as long as I can verify exactly how my information is

used." This phenomenon has been noted previously with shared inboxes [24].

Trust was the critical concern: to share their inbox, respondents needed to feel they could rely on the assistants. Some suggested that they would want to "talk to them personally", get to "know them somewhat well" or "vet them myself". Furthermore, respondents wanted confidentiality assurance: "I'd want some liability insurance for what a rogue assistant might do with my information."

In summary, despite improvements in email clients, information workers still spend significant effort managing tasks in email. Respondents desired human assistance, but listed trust and privacy as core concerns, requiring mechanisms to limit potential damage. Designing for privacy concerns is important. At the same time, people's real-world practices often differ from their pre-usage estimates. Consequently, it is important to observe actual usage and not rely exclusively on survey data.

EMAILVALET

EmailValet (Figure 1) is a task-based email client that enables collaboration with crowdsourced assistants. The current prototype shares a sliding-window history of a user's inbox with assistants who annotate messages with tasks. In principle, this approach could support other delegated tasks such as summarizing messages, negotiating meeting times, or drafting/sending replies. EmailValet makes all assistant actions visible to the user. Icons clearly identify when messages have been processed by assistants or contain tasks (Figure 3a). Actions performed on each email are displayed along with the email headers (Figure 3c). A complete log is also available to establish peace of mind (Figure 5). Our prototype is integrated with Gmail and is available at https://www.gmailvalet.com.

In the EmailValet inbox, the left column displays email threads (Figure 3). On the right, the system presents a stream of all tasks that the user or assistant created from messages. To view an email, a user either clicks on the thread (in the inbox) or the task (in the task stream). When viewing or composing email, the message consumes the right pane; its related tasks are shown at the top (Figure 3b).

Assistants log in to a limited, 'valet' view of each inbox that they have support. They can read emails and create tasks associated with those emails.

Creating and Viewing Tasks

To lower the friction for task creation, assistants and users can create a task by entering its title at the top of the email (Figure 3b). In the other direction, clicking a task shows its originating email. Icons in the inbox indicate whether an assistant can or has viewed the message, and also whether its tasks are completed (Figure 3a).

The right-hand task stream (Figure 1) gives users an at-a-glance overview of their tasks. Our vision is that users can treat this task stream as an action-oriented view of their email, facilitating efficient message handling. By default, tasks are ordered by recency; users can reorder them by priority.

A calendar view provides users with an overview of when tasks are due. If possible, due dates are automatically extracted using Natural Language Processing (NLP); they can also be manually edited.

Accountability and Access Control

Survey respondents were concerned about limiting assistants' access and monitoring their actions. Consequently, EmailValet introduces facilities for accountability and access control.

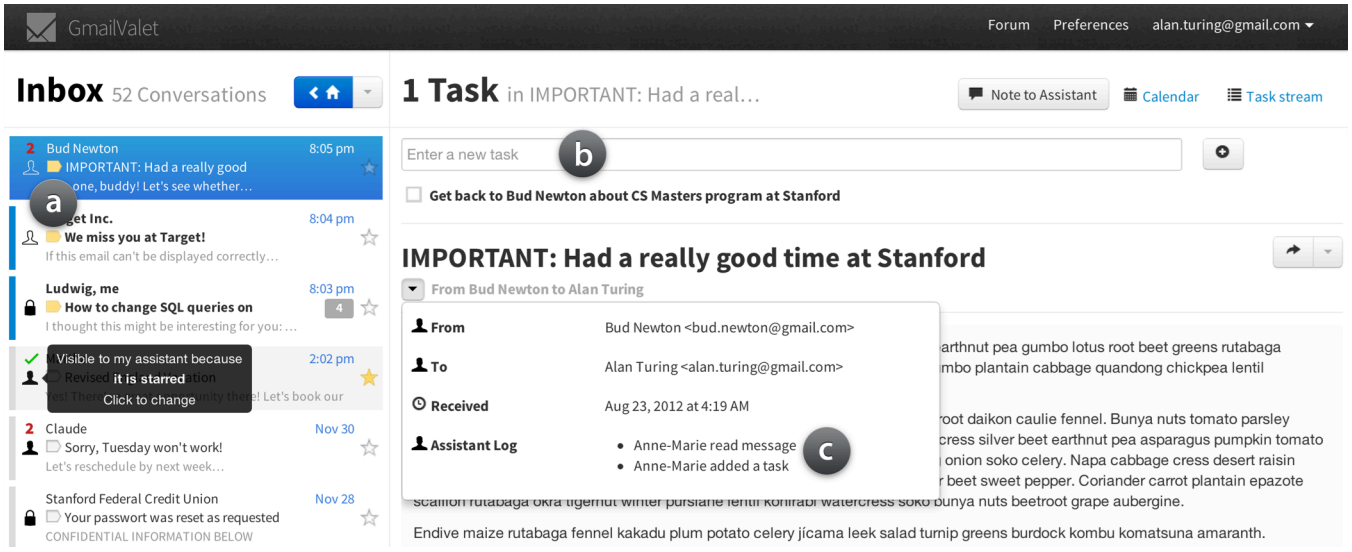

Figure 3. a) Icons to the left of the message summary indicate whether an assistant can or has viewed the message, as well as how many open tasks remain. Hovering reveals the system's reason for the current visibility setting. Clicking opens the email on the right and displays its associated tasks. Users can also add a task (b) and view actions the assistant has taken with this message (c).

Accountability

EmailValet offers three monitoring techniques. First, inbox icons show whether the assistant has processed an email and extracted a task (Figure 3a). Second, the detailed message view lists all logged activities for that email: for example, opening an email, sending a response, or creating a task (Figure 3c). Third, a user can view a complete chronological log of all an assistant's actions (Figure 6). Logging doesn't prevent abuse, but does leave "fingerprints" [37]. We anticipate this log's primary benefit to be deterrence and peace of mind — like a security camera — rather than frequent monitoring by users.

Access Control

Assistants can only view a user's mail through EmailValet. EmailValet, by default, restricts the assistant to a sliding window of the most recent 100 messages and search is disabled. Users can provide whitelist and blacklist filtering for finer-grained control. Whitelist filters restrict the assistant's view to particular labels or folders, such as starred messages, Gmail's Priority Inbox, or messages sent to oneself. Furthermore, users can provide blacklist filters to exclude sensitive messages, such as emails from family, passwords, or financial information (Figure 4).

These restrictions attempt to balance the assistants' need to understand contextual connections with the user's desire not to expose their whole history to the crowd. Assistants' actions are also limited; EmailValet's current policy prevents assistants from deleting messages.

Finally, EmailValet also integrates with automatic approaches. To illustrate this, with the EmailValet prototype, users can restrict assistant-visible emails to Gmail's Priority Inbox.

Feedback and Learning

The tasks that assistants create may not always be the tasks that users want. Tasks mean different things to different users: this is one reason automated approaches often fail [19]. To provide assistants feedback, the simplest way for a user to remove a task is to decline it (Figure 1). EmailValet shows assistants which tasks were accepted or declined, and encourages users to add an explanatory comment to the assistant (Figure 5). To help assistants learn what they missed, they can see tasks that users create themselves.

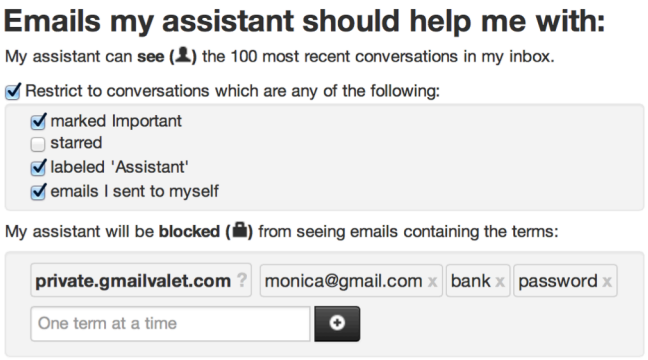

Figure 4. For privacy, users can specify rules for which emails will be visible to their assistant.

To frame discussions, users can leave an introductory note about their preferences. This note is especially helpful for providing new assistants with context and user-specific heuristics. For example, users may want to emphasize certain senders or ask the assistant to apply labels to their tasks (*e.g.*, "put [Event] in front of every event").

Assistants and users can also open a chat window to clarify any confusion. While a few crowdsourcing systems provide limited interactive feedback with requesters [14], we are unaware of other crowd-powered systems that support interactive worker conversation with end users. We have found that chat helps efficiently achieve common ground:

Assistant: *Do you want me to create tasks from [name]?*
User: *Yes, please.*
Assistant: *Ok I will, I made one and it got declined so I just wanted to make sure.*

Multiple Users per Assistant

Assistants can help many users simultaneously, increasing affordability for users and labor opportunities for assistants. To provide easy access to multiple accounts, assistants can switch between users with a drop-down menu. Awareness indicators on this menu show users with unread messages.

FIELD STUDY

A weeklong field study investigated whether EmailValet helps users complete tasks and whether its privacy and accountability features satisfy users' concerns. We found that EmailValet accurately extracted tasks from email, that users found value in the system for their task management, and that users became increasingly comfortable with EmailValet's privacy tradeoffs.

Method

We deployed EmailValet in a one-week field study, recruiting twenty-eight participants using mailing lists at our university. Six were MBA students and twenty-two were university students of other majors, mostly technical. We offered participants a $50 gift certificate. We hired three online assistants through the oDesk crowdsourcing marketplace: one from Illinois, two from California. Two were work-at-home mothers. Assistants were compensated at $8 per hour to process all shared emails during the study.

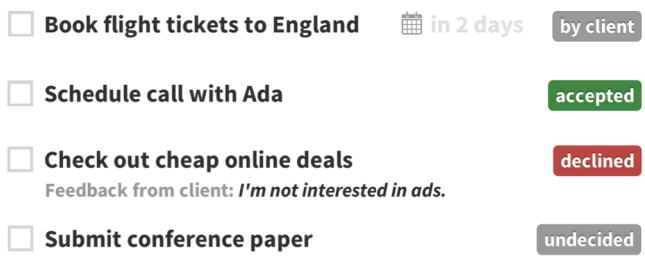

Figure 5. The assistant's view of the task stream. Feedback helps the assistant to learn the user's intentions: accepted and rejected tasks, freeform text, and user-created tasks.

Action	Task	Message	Assistant	Time
Read message	—	[#VWJ-824-43394]: Explanation from Stanford U.	Ludwig	6 min ago
Created task	Set up unconference meeting with Martha	MyGSB Digest for the week of from MyGSB Digest	Ludwig	6 min ago
Read message	—	MyGSB Digest for the week of from MyGSB Digest	Ludwig	16 min ago

Figure 6. The log supports accountability by showing all of the assistant's activities to the user.

Participants began by authoring whitelists and blacklists that determined the assistants' access control. Each participant was instructed to use EmailValet at least twice each day to process their new emails and tasks. At the conclusion of the study, participants completed a survey focused on EmailValet's qualitative usefulness, privacy, and assistant quality.

We ran a within-subjects experiment to investigate whether EmailValet helps users complete the tasks in their email. Participants rotated through three interface conditions. Following a one-day warm-up, each participant spent two consecutive business days in each condition:

- *Control:* Participants could not see assistant-created tasks or create their own tasks.
- *Self:* Participants could not see assistant-created tasks. However, they could create their own tasks.
- *Assistance:* Participants could see assistant-created tasks and create their own tasks. Participants could give task feedback to their assistant.

The order of conditions was randomized for each participant, in a Latin square manner. Participants could always read and write email. The assistant extracted tasks from emails in all three conditions. However, in the Control and Self conditions, the assistant did not receive feedback on tasks from the participant because the participant could not see their extracted tasks. At the end of each condition, participants saw any previously-hidden tasks that the assistant created. To produce ground truth, participants then accepted or rejected these tasks, gave feedback, created any tasks the assistant missed, and marked whether or not they had completed each task.

We measured the percentage of new tasks that the user marked completed while each condition was in effect. This metric combines the tasks that the assistant authored and the user accepted with the tasks that the user authored themselves. If the assistant's tasks were hidden during the condition (e.g., Self or Control), we counted tasks that were accepted during the end-of-condition review. We manually merged any duplicate tasks that the user and assistant both created. Users and assistants may make extraction mistakes, so this metric may not represent every task in the inbox.

We hypothesized that users would complete more email-based tasks with EmailValet's task extraction than when they must extract tasks on their own (Assistance > Self) or cannot extract any tasks (Assistance > Control). We further hypothesized that the discipline of self-managed task management would still produce some benefit: Self > Control.

Results

Of the 28 participants who filled out the final survey, 16 consistently and successfully participated in all three conditions and thus had measurable task completion rates. In this section, we analyze qualitatively and quantitatively: 1) the assistants' accuracy at extracting tasks, 2) users' behavior and feedback regarding EmailValet as a task support tool, and 3) users' feedback to the privacy concerns and functionality in EmailValet. Table 1 displays examples of assistant-extracted tasks.

Assistants' Accuracy

We measured the *precision* and *recall* of assistant-created tasks so that we could better understand how accurate the assistants were at extracting important tasks for users.

Precision measures the percentage of assistant-created tasks that participants accepted. On average, users accepted 71.6% of the tasks extracted by assistants. This ratio was very similar across all three assistants (σ=3.5%). Precision rose throughout the study, starting at 62.1% on the first day and ending with 84.8% on the last day. The rise in precision is most likely due to participants' feedback regarding tasks and to the assistants learning more about their users. One assistant noted, "it has become easier to extract good and accurate tasks from my clients' emails over time. I feel I have gotten to known my clients better and understand the conversations better".

The final survey asked, a free-response question: were the assistants' tasks relevant, or just busywork? About two thirds of participants (19 of 28) found assistant-created tasks to be of value and worth completing. Notably, four of these participants praised the assistants' creations as being on the same level as their own; one thought the assistant was even better than themselves ("they used more detail than I did") and three lauded the assistants for extracting

- *Schedule to meet first week of September with Priya*
- *Complete flyer by the end of the week*
- *Print out flight information*
- *Send International Student Advisor a scan of my I-20*
- *Review Emily's ideas and if possible add onto it*
- *Lunch with Al*

Table 1. Tasks extracted by assistants during the study.

more tasks and events than they would have done ("made the week easier"). However, some participants felt that the assistants were overeager, complaining that assistants created tasks from irrelevant mailing lists, created tasks that were too ambiguous or missed tasks completely. On the other hand, a participant explicitly called their assistant "too conservative" and would have preferred to receive more tasks. Users typically did not mind false positives: as a participant wrote, "Deleting [tasks] was easier than creating."

Recall measures how many tasks the assistant missed: the number of accepted tasks that the assistant authored divided by the number of accepted tasks that either the user or the assistant authored. This measure only includes tasks created during the Assistant condition because this is the only time that users could see the assistants' tasks in real time and add any missing tasks. By this calculation, recall was 68.6%. Often, the user logged in before the assistant had time to process the new email, so eventual recall may be higher.

We asked users a free-response question: were they confident that their assistants would not miss important tasks? We coded their responses and found that more than half of participants (17 of 28) felt they could fully or almost fully rely on their assistant, five participants had mixed feelings, and six had negative impressions. Participants praised assistants' consistency ("after the first few days they established that they could keep track of impt [sic] details"). Missing contextual knowledge was the most common cause of missed tasks: "Many important tasks (that are not obvious) are not extracted."

EmailValet's Usefulness

Users found the assistants to be generally accurate, but did the system help those users manage their tasks? We compare the percentage of tasks completed in each condition and supplement the results with qualitative feedback.

Of the sixteen participants who consistently used EmailValet through all study conditions, users completed a significantly higher percentage of tasks in the Assistance condition (M=58.4%, SD=34.4%) than in the Self condition (M=29.3%, SD=28.7%) or Control condition (M=26.3%, SD=34.5%). We performed a repeated-measures ANOVA with condition as independent variable and task completion rate as the dependent variable. We found a significant effect of condition, $F(2,30)=4.483$, $p<.05$. Posthoc pairwise t-tests with false discovery rate (FDR) correction found a significant difference between Assistance and Control, $t(15) = 2.89$, $p<.05$, and between Assistance and Self , $t(15)=2.49$, $p<.05$. No difference was found between Self and Control $t(15)=0.24$, *n.s.* So, participants completed more tasks in the Assistance condition than in the Self or Control condition.

Participants were enthusiastic about the positive impact EmailValet had on their task management. When asked if they would like to continue using EmailValet, one participant replied, "any help in making sure everything gets done would be greatly appreciated." Another insisted that they would like to use it only with the assistant: "What I need is an extra pair of eyes." A third participant praised the assistant's work because their tasks were "like magic": "very convenient and much easier than doing it myself." For users who didn't feel that assistants supported their task management, they still found that the task-centric interface kept them aware of incomplete tasks. Roughly 40% of the participants wanted to use EmailValet as their main task list following the study (12 of 29), and several continued using it voluntarily the week after the study concluded. Those who were uninterested in continuing felt that it needed deeper integration with other PIM tools for them to fully integrate it into their workflow.

Privacy Concerns and Trust

We asked our participants which emails they made available to assistants and which they blocked. Only two participants opted to manually whitelist individual emails for the assistant to see. Most participants shared everything or whitelisted a set of filters (*e.g.,* priority messages, starred messages, school, work). Those who filtered for privacy blacklisted patterns such as passwords, banking information and more intimate contacts. As one participant put it: "I thought the only way for the service to be most helpful would be to ensure they could look at everything." Some participants loosened their settings over time: as one put it, "Originally I did not share emails from my boyfriend [name], but I changed that after realizing he sent me emails with tasks for me to do".

A majority of participants (18 of 28) initially felt somewhat uncomfortable entrusting an assistant with their emails. However, over half of those with concerns (10 of 18) and over a third of all participants (11 of 28) reported that they felt more comfortable with the service over time, while no one reported a decrease in comfort. Reasons for the increased trust were diverse: One participant named the first task extractions ("I felt more optimistic and was pleasantly surprised […] it was surprisingly useful and effective"). Another one found the assistant biography helpful ("my assistant was a mother"). A third simply got accustomed to what she called a "slight breech [sic] of privacy": "In the beginning, it felt weird that my assistant was reading personal emails from my family and then creating a task for me to do based on it, but I didn't really feel strongly enough about it to change my privacy settings... I just kind of went with it."

Assistants shared the discomfort, as they were initially not at ease with going through their users' emails and occasionally stumbling upon sensitive information: "Some emails I felt like I was invading their privacy because it was emails between my clients and their family, but I didn't react anyway because it's a job I have to do". Similar to users, assistants became more comfortable over time.

Despite of this increased level of comfort, almost two thirds of participants (19 of 28) said the assistant did not care for

them personally but was simply "doing their job," and only one participant felt the assistant personally cared. This was not necessarily bad: four participants explicitly stated that they did not want to get to know their assistant. However, seven other participants suggested more communication with the assistant so that they could feel more like the assistant personally cared. Similarly, opinions were split on whether knowing the assistant made it more or less comfortable share your inbox.

Assistant Economics

The assistants recorded 70 hours of work during the study. They processed 12,321 messages—8,679 for the 28 participants and the rest for 10 existing users of the system who did not take part in the study—and created 779 tasks. On average, each assistant processed 2.94 messages per minute. This rate would extrapolate to 1,408 messages in an eight-hour workday if the assistant were fully occupied. On average, assistants created 6.32 tasks per 100 emails they read. Participants received 38.8 messages per day on average, fewer than recent studies have reported [18].

At this rate, one assistant could support 36 users simultaneously in an eight-hour workday. Such support would cost each user $1.78 per day. However, in our deployment, each assistant was assigned 12 users on average. We asked the assistants to tell us how many users they felt they could support. They generally felt that they had additional bandwidth. For example, one said, "[I have] 10 current clients, it's pretty easy to keep track of the clients, […] so ideal would be probably 20 to keep me busy more".

Limitations

This study has several limitations. First, the sample is not fully representative: the participants are largely young, technology-literate, and self-selecting to participate in a study that shares their private emails with assistants. In addition, the study's short duration capures the first days of usage, but usage patterns may evolve. A longitudinal evaluation with a broad population is a clear next step.

DISCUSSION

EmailValet extends crowdsourcing to tasks that require access to private or sensitive data. In this section, we reflect on the implications of this decision, the importance of context and common ground, and on a future vision of massively multiplexed crowd assistants.

Adaptation to the Privacy Breach

Fully 70% of the formative survey respondents reported they would be unwilling to share significant fractions of their inbox. Yet, after a week of use, only 10% of participants cited privacy as a reason that someone might abandon EmailValet. Even accounting for sampling differences, this is a major shift. Furthermore, participants rarely used EmailValet's accountability features: the features were more useful as a security blanket than as functionality. While engineers and designers often describe privacy as a property—a system either has privacy or it does not—in practice, privacy is dynamic [28]. When the benefit exceeds the invasion, we share information that we previously considered private.

The Assistants' Lack of Context

Traditional crowdsourcing systems focus on context-free micro-tasks (*e.g.*, [10]). However, supporting personal and private tasks requires rich context [15]. There were several situations where EmailValet assistants lacked this context and produced incorrect tasks. The assistants recognized this challenge: in the future, they asked for a short fact file about each user, including their profession and place of residence. We are eager to let assistants flag messages that they find useful for understanding each user, and to allow them to write or share notes between each other. These notes institutionalize assistant knowledge, which would enable a smooth "changing of the guard" as assistants begin and end their work. Companies can also lessen the context and privacy problems by adapting a model midway between EmailValet's distributed experts and in-house administrative assistants.

Tasks and Busywork

Users completed more tasks in the Assistance condition, but is this because EmailValet focused them on unimportant busywork? There is some evidence that this may be happening: one participant explained that they accepted "tasks or events I didn't necessarily plan to do [...] because it was still good to have it there [as] a reminder, whereas I wouldn't have bothered to make note of that task myself." These tasks were likely middle priority: not as unimportant as busywork, but not important enough to be the top priority. In general, however, it is up to the user to decide whether an assistant-created task is important or busywork. The framing for this decision can be communicated to the assistant, helping them adapt to each user's needs.

From Personal to Massively Multiplexed Assistants

Online assistants may be affordable enough for most information workers to use in small amounts. Likewise, assistants can multiplex their services across multiple users. Are we seeing the beginnings of democratized access to executive assistants? Not yet: several issues remain and demand workable solutions. First, as our formative work revealed, people are reluctant to entrust online assistants with access to their sensitive data. As systems like EmailValet establish trust and produce high-quality work, these concerns may lessen. Second, the quality of an online assistant may not compare to that of an executive assistant entirely dedicated to one person. Users of EmailValet-style systems thus face a trade-off between cost and quality. However, collective intelligence of multiple assistants may produce higher quality than any dedicated assistant. Third, as we saw, missing context remains a major struggle for many assistants.

CONCLUSION & FUTURE WORK

This paper introduces *valet* techniques that preserve privacy and accountability for sharing private data with crowdsourced assistants. These techniques take shape in EmailValet, an email client that recruits crowdsourced hu-

man assistants to extract tasks from each email in the inbox. As a valet interface, EmailValet focuses on: 1) parsimony, providing just enough access for assistants to extract tasks and giving users rich controls to limit assistants' permissions; and 2) visibility, surfacing all assistant actions within the interface. In a one-week evaluation, EmailValet users completed about twice as many tasks when they had access to the assistants' help.

Crowdsourced assistants could support many additional personal information management activities. For example, email assistants could crate action plans for tasks [24], identify meeting requests, link references to other threads, mark priorities and write personal reminders. These extensions create an opportunity to design alternative consumption methods for the inbox, for example a compact, glanceable stream of events. The assistant could also support other personal information management tools, curating the user's calendar, maintaining a contacts list or organizing appointments. More ambitiously, assistants could execute tasks such as booking a flight or reserving a conference room. Finally, by formalizing the assistant's actions, we can train machine learning systems and improve the automatic approaches to task extraction.

Additional materials about this research are available at http://hci.st/valet.

ACKNOWLEDGEMENTS

The authors thank Greg Little for thoughtful discussions; oDesk for sponsoring the oDesk workers; Ryan Globus, Thomas Bridges-Lyman, and Arun Prasad for their contributions during their Summer internship; and Ludwig Schubert for work on the figures and video. This research was sponsored in part by NSF POMI 2020 Grant No. CNS-0832820.

REFERENCES

1. Allow Someone Else to Manage Your Mail and Calendar. http://office.microsoft.com/en-us/outlook-help/allow-someone-else-to-manage-your-mail-and-calendar-HA010355554.aspx.
2. ClearContext tames Outlook. http://news.cnet.com/8301-17939_109-9947065-2.html.
3. Create Tasks and To-Do Items. http://office.microsoft.com/en-us/outlook-help/create-tasks-and-to-do-items-HA001229302.aspx.
4. New in Labs: Tasks. http://gmailblog.blogspot.com/2008/12/new-in-labs-tasks.html.
5. Set up Mail Delegation. http://support.google.com/mail/bin/answer.py?hl=en&answer=138350.
6. Barley, S.R., D. Meyerson, and S. Grodal. E-mail as a Source and Symbol of Stress. *Organization Science 22*, 4. 2011.
7. Bekkerman, R., A. McCallum, and G. Huang. Automatic Categorization of Email into Folders: Benchmark Experiments on Enron and SRI Corpora. *Technical Report IR-418*. 2004.
8. Bellotti, V., N. Ducheneaut, M. Howard, and I. Smith. Taking Email to Task: The Design and Evaluation of a Task Management Centered Email Tool. *Proc. CHI 2003*. 2003.
9. Bennett, P.N. and J. Carbonell. Detecting Action-Items in E-mail. *Proc. SIGIR 2005*. 2005.
10. Bernstein, M.S., G. Little, R.C. Miller, B. Hartmann, M.S. Ackerman, D.R. Karger, D. Crowell, and K. Panovich. Soylent: A Word Processor with a Crowd Inside. *Proc. UIST 2010*. 2010.
11. Boone, G. Concept Features in Re:Agent, an Intelligent Email Agent. *Proc. AAMAS 1998*. 1998.
12. Corston-Oliver, S., E. Ringger, M. Gamon and R. Campbell. Task-focused summarization of email. *ACL 2004 Workshop: Text Summarization Branches Out*. 2004.
13. Crawford, A.B. Corporate Electronic Mail — A Communication-Intensive Application of Information Technology. *MIS Quarterly 6*, 3. 1982.
14. Dow. S., A. Kulkarni, S.R. Klemmer, and B. Hartmann. Shepherding the Crowd Yields Better Work. *Proc. CSCW 2012*. 2012.
15. Erickson, T., C. Davis, W. Kellogg, and M. Helander. Assistance: The Work Practices of Human Administrative Assistants and their Implications for IT and Organizations. *Proc. CSCW 2008.* 2008.
16. Faulring, A., B. Myers, K. Mohnkern, B. Schmerl, A. Steinfeld, J. Zimmerman, A. Smailagic, J. Hansen, and D. Siewiorek. Agent-assisted task management that reduces email overload. *Proc. IUI 2010.* 2010.
17. Feldman, M.S. Constraints on Communication and Electronic Mail. *Proc. CSCW 1986.* 1986.
18. Fisher, D., A.J. Brush, E. Gleave, M.A. Smith. Revisiting Whittaker & Sidner's "email overload" ten years later. *Proc. CSCW 2006.* 2006.
19. Freed, M., J. Carbonell, G. Gordon, J. Hayes, B. Myers, D. Siewiorek, S. Smith, A. Steinfeld, and A. Tomasic. RADAR: A Personal Assistant that Learns to Reduce Email Overload. *In Proc. AAAI 2008*. 2008.
20. Gwizdka, J. TaskView – Design and Evaluation of a Task-based Email Interface. *Proc. CASCON 2002.* 2002.
21. Horvitz, E., A. Jacobs, D. Hovel. Attention-sensitive alerting. *Proc. UAI 1999*. 1999.
22. Jackson, T., R. Dawson, and D. Wilson. The Cost of Email Interruption. *Journal of Systems and Information Technology 5*, 1. 2001.
23. Kiritchenko, S. and Matwin, S. Email Classification with Co-Training. *Proc. CASCON 2001*. 2001.

24. Kokkalis, N., T. Köhn, J. Huebner, M. Lee, F. Schulze, S.R. Klemmer. TaskGenies: Automatically Providing Action Plans Helps People Complete Tasks. *Technical Report* CSTR 2012-04. 2012.
25. Lampert, A., R. Dale, C. Paris. Detecting Emails Containing Requests for Action. *Proc. ACL 2010*. 2010.
26. Mackay, W. E. More than Just a Communication System: Diversity in the Use of Electronic Mail. *Proc. CSCW 1988*. 1988.
27. Muller, M. J., and Gruen, D. M. Working Together Inside an Emailbox. *Proc. ECSCW 2005*. 2005.
28. Palen, L., and P. Dourish. Unpacking "Privacy" for a Networked World. *Proc. CHI 2003.* 2003.
29. Rambow, O., L. Shrestha, J. Chen, and C. Lauridsen. Summarizing Email Threads. *Proc. HLT-NAACL 2004*. 2004.
30. Sahami, M., S. Dumais, D. Heckerman, and E. Horvitz. A Bayesian Approach to Filtering Junk E-mail. *AAAI Workshop: Learning for Text Categorization*. 1998.
31. Samiei, M., J. Dill, and A. Kirkpatrick. EzMail: Using Information Visualization Techniques to Help Manage Email. *Proc. IV 2004.* 2004.
32. Segal, R.B. and Kephart, J.O. MailCat: An Intelligent Assistant for Organizing E-Mail. *Proc. AAMAS 1999*. 1999.
33. Shen, J., L. Li, T.G. Dietterich, and J.L. Herlocker. A Hybrid Learning System for Recognizing User Tasks from Desktop Activities and Email Messages. *Proc. IUI 2006*. 2006.
34. Takkinen, J. and N. Shahmehri. CAFE: A Conceptual Model for Managing Information in Electronic Mail. *Proc. HICSS 1998*. 1998.
35. Venolia, G., L. Dabbish, J.J. Cadiz, and A. Goopta. Supporting email workflow. *Technical Report MSR-TR-2001–88*. 2001.
36. Weber, R. The Grim Reaper: The Curse of E-Mail. *MIS Quarterly 28*, 3. 2004.
37. Weiser, M. The Computer for the 21st Century. *Scientific American*. 1991.
38. Whittaker, S. and C. Sidner. Email Overload: Exploring Personal Information Management of Email. *Proc. CHI 1996*. 1996.
39. Winograd, T., and F. Flores. *Understanding Computers and Cognition: A New Foundation for Design*. Ablex, 1986.