Email Makes You Sweat: Examining Email Interruptions and Stress with Thermal Imaging

Fatema Akbar[1], Ayse Elvan Bayraktaroglu[1], Pradeep Buddharaju[2], Dennis Rodrigo Da Cunha Silva[3], Ge Gao[1], Ted Grover[1], Ricardo Gutierrez-Osuna[3], Nathan Cooper Jones[2], Gloria Mark[1], Ioannis Pavlidis[2], Kevin Storer[1], Zelun Wang[3], Amanveer Wesley[2], Shaila Zaman[2*]

[1] Dept. of Informatics
University of California, Irvine
Irvine, CA, 92697 USA
{fatemaa, ayseelvan, ge.gao, grovere, gmark, storerk}@uci.edu

[2] Computational Physiology Lab
University of Houston
Houston, TX 77004, USA
Buddharaju@uhcl.edu,
nathancooperjones@gmail.com,
{ipavlidis,awesley4,szaman4}@uh.edu

[3] Dept. of Computer Engineering
Texas A&M University
College Station, TX 77843 USA
{silva.dennis, rgutier, wang14359}@cse.tamu.edu

ABSTRACT

Workplace environments are characterized by frequent interruptions that can lead to stress. However, measures of stress due to interruptions are typically obtained through self-reports, which can be affected by memory and emotional biases. In this paper, we use a thermal imaging system to obtain objective measures of stress and investigate personality differences in contexts of high and low interruptions. Since a major source of workplace interruptions is email, we studied 63 participants while multitasking in a controlled office environment with two different email contexts: managing email in batch mode or with frequent interruptions. We discovered that people who score high in Neuroticism are significantly more stressed in batching environments than those low in Neuroticism. People who are more stressed finish emails faster. Last, using Linguistic Inquiry Word Count on the email text, we find that higher stressed people in multitasking environments use more anger in their emails. These findings help to disambiguate prior conflicting results on email batching and stress.

*All authors contributed equally to this work.

ACM ISBN 978-1-4503-5970-2/19/05…$15.00
https://doi.org/10.1145/3290605.3300898

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**; • **Social and professional topics** → User characteristics.

KEYWORDS

Interruptions; stress; email; multitasking; sensors; personality; thermal imaging; empirical study

ACM Reference Format:
Fatema Akbar[1], Ayse Elvan Bayraktaroglu[1], Pradeep Buddharaju[2], Dennis Rodrigo Da Cunha Silva[3],, Ge Gao[1], Ted Grover[1], Ricardo Gutierrez-Osuna[3], Nathan Cooper Jones[2], Gloria Mark[1], Ioannis Pavlidis[2], Kevin Storer[1], Zelun Wang[3], Amanveer Wesley[2], Shaila Zaman[2]. 2019. Email Makes You Sweat: Examining Email Interruptions and Stress with Thermal Imaging. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland UK*. ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3290605.3300898

1 INTRODUCTION

Workplace environments are characterized by frequent interruptions. This subject has sparked a long interest in the HCI field to understand how interruptions affect productivity, time to resume work, mood, and stress, e.g. [5, 10, 23, 32, 42]. While interruptions can be beneficial, e.g. by providing useful information [23] or social interaction [38], they can also be detrimental, e.g., by lengthening task time [13], disrupting focus [59], increasing errors [42], and creating attention residue that interferes with the current task [29]. In particular, a documented negative effect of interruptions on work is that they increase stress, cf [37]. Workplace stress is important to understand and address as it impacts health [27] and wellbeing [2].

Work interruptions can arise from a variety of sources (phone calls, text messaging, face-to-face interactions, social media), a common form being email. For example, the Radicati group [17] estimates that in 2017, users managed an average of 125 emails at work daily. A study of 40 information workers for 12 days found that they spend an average of 1 hour and 23 minutes per day on email, checking the inbox 77 times a day [33]. While email provides numerous benefits [4, 62], continual switching of attention away from tasks to manage email can increase cognitive workload and consequently stress [28, 33].

To relieve the burdensome effects of email, some organizations have experimented with different strategies for delivering email as a way to reduce task interruptions and increase focus. These strategies have included email-free days and batching email. Batching may decrease stress by avoiding disruptions of task activity and reducing cognitive load [59]. However, results are mixed as to whether individual strategies of checking email are related to stress [5, 28, 33]. Email is often managed while multitasking with other work tasks, and relatively little research has explored how batching email might affect stress with other concurrent task performance. Further, the relationship of email use and stress may vary with other situational and dispositional factors in the workplace.

Our current study expands on understanding the relationship between email use and workplace stress in three ways. First, we designed a controlled experiment to compare the effects on stress of two different patterns of email use. Second, since information workers in high-stress environments are often exposed to constant stressors (e.g. deadlines, upcoming reviews) in addition to emails, we examined how a persistent, or anticipatory, stressor affects task performance. To our knowledge, investigating the effect that a stressor has on email behavior is unique. Third, we examined how the relationship between email use patterns and stress varies with other individual factors, in particular, personality. Although the effect of each of these factors on stress has been discussed [12, 15, 41, 52], to date the findings are not well-connected.

Consequently, our current study examines the interplay between email use patterns, stressors, and task performance, a combination that best models the complexity of a real-world work environment. Further, in contrast to much of the previous research using self-report measures, we use a state-of-the-art thermal imaging system to capture objective and unobtrusive stress signals while people are working at a computer. One unique benefit of deploying this method, as indicated by recent sensor-based research, e.g. [9, 33, 40, 57, 61, 66], is that it enables us to precisely synchronize the stress measure with time and/or tasks. To our knowledge, this is the first study using thermal imaging to measure stress during computer work under different interruption patterns.

Our results show that people higher in Neuroticism are significantly more stressed when managing emails in a batching mode compared to those lower in Neuroticism. We also find that as stress increases, the length of time spent answering emails decreases. Last, using Linguistic Inquiry Word Count (LIWC) analysis we find differences in email content: people experiencing higher stress use more words to convey anger in their emails. We discuss the relationship between personality, email use pattern, stress, and task performance. Finally, we propose how our results inform the future design of personalized systems that can help people better manage workplace interruptions and reduce stress.

2 RELATED WORK

Email Interruptions and Stress

Despite its benefits for communication, task management, and coordination, email has been long recognized as one of the main sources of workplace interruptions. To manage the growing number of incoming emails and expectations of prompt reactions to them–both considered a source of stress [16], most employees have their email clients running all day [8], creating the potential for continual interruptions [49]. Czerwinski et al. [10] found that information workers view email as a task that needs repeated attention throughout the day, one that constitutes on average 23% of the tasks they perform daily. Thus, interruption and the consequent disruption of the task-at-hand are important aspects of email use that can cause stress [24, 25, 56].

Researchers have investigated how email interruptions increase stress [3, 11]. Mano and Mesh [31] found that while the continual flow of emails can contribute to one's work performance by providing necessary information, it can also create stress. Email can disrupt focus: cutting off email for five days caused information workers to increase their task focus and lower their stress [34]. Perceived interruptions from messages are shown to lead to experiencing higher time pressure, which itself is a job stressor [55]. The relationship between email load and workload stress has also been explained due to the time spent on email [56]. In related work, Barley et al. [3] found that time on email is correlated with the number of incoming emails.

In sum, despite the multi-faceted benefits of email [4, 62], studies have repeatedly shown that email is associated with stress. Other proposed explanations are the large volume of incoming email, disruption of notifications, habits of checking and self-interrupting, opportunity costs taking time away from other work, uneven costs of the receiver and sender [11] and, perhaps above all, its psychological association with work [3].

Email Use Patterns and Stress

Email batching, i.e., limiting email use to certain times of the day, has been proffered as a solution to reduce the frequency of interruptions and counteract stress. However, studies investigating effects of batching behavior on stress report conflicting results. In one study, where participants adopted a once-a-day email checking strategy, no difference was observed in reported stress levels [5]. Similarly, Mark et al. [33] found no evidence that batching behavior would lead to lower stress. In agreement with these results, a field study by Barley et al. [3] found that almost none of the respondents used a batching strategy, probably due to believing that such strategies would have little effect on stress. In fact, Dabbish and Kraut [11] suggest that compared to checking emails at specific times, checking them when they arrive may be a better strategy to reduce email overload. Compressing the answering of emails into a block of time may lower effort expended through multitasking by reducing switch costs [64] and therefore lowering cognitive load. Kushlev and Dunn [28] found that when limiting email checking to three times a day, participants reported significantly less stress. Furthermore, when timing of email interruptions was controlled via an email client, participants' stress responses to high task demands were lower [14]. Gupta et al. [18] found that tending to emails two to four times a day can improve productivity, which in turn may lower stress. In summary, the results on batching email and effects on stress are conflicting, which suggests further investigation is needed.

Stressors in Information Work

In addition to email, other stressors can also be present in information work. Repetitive challenges in everyday life such as high work demand or interpersonal tensions at home or work can accumulate and lead to persistent stress levels [2]. Common to information work, knowledge of deadlines or upcoming commitments can act as anticipatory stressors, as one prepares for the event to occur [39]. Further, in the fast-paced environment of knowledge work, managing email is not performed as an isolated task but is interspersed with work on other tasks, leading to multitasking, which has been shown to increase workload and stress, e.g. [63].

Individual Differences in Stress Responses

Individual differences in the experiences of stress have been attributed to personality traits (e.g. Extraversion) that are associated with responding more or less effectively to affective strategies [39]. More relevant to our work, the personality traits of Openness to experience and need for personal structure were found to be associated with whether a person experiences stress due to interruptions [32]. Another study found that people who score high in Conscientiousness tend to resist email interruptions [51]. While these findings are intriguing, the interrelationship between personality, email, and stress has not been widely explored.

Stress Measurement

Most studies examining email use and stress have used self-reports (e.g. [3, 11, 19, 49, 56]). Though self-report instruments have been well validated, they are subjective, require the full cognitive attention of the user, and can be affected by memory recall as well as emotional biases. Further, self-report instruments are ill-suited for continuous unobtrusive measurement of stress, especially in an environment where people switch among different tasks.

Physiological measures of stress have gained popularity because they are objective and provide high temporal resolution. However, conventional physiological measures have their own set of problems, which relate to usability concerns and motion artifacts. Palmar EDA (electro-dermal activity) is a case in point, as it cannot be used in an office study where the hands are hard at work. Wrist EDA is a more usable alternative to palmar EDA, but suffers from weak signaling in that part of the body [60]. Chest strap sensors for measuring heart and breathing rate are uncomfortable for participants and loose connection to the skin depends on body posturing [58].

A new generation of physiological measurement methods based on facial imaging promises to address some of the problems plaguing conventional physiological measurement methods that use contact probes. Imaging methods are unobtrusive, and thus user-friendly. They are also paired with increasingly sophisticated tracking methods that ameliorate motion problems. The most popular imaging modalities are visual and thermal. McDuff et al. [36] employed a visual camera to capture breathing rate and heart rate variability (HRV), using them as inputs into a model that predicted cognitive stress with 85% accuracy. Pavlidis and colleagues measured stress through facial thermal imaging in varied application scenarios, e.g. [30, 48].

3 RESEARCH QUESTIONS

Our research questions explore the relationship between email use pattern, stress and work performance. We also investigate the effects of batched versus continual incoming emails, which to date are poorly understood [3, 5, 11, 14, 18, 33]. In addition to email as a stressor, we consider the role of an anticipatory stressor common in information work. An upcoming deadline, performance review, presentation, or needing to leave the office on time to pick up children can become an anticipatory stressor [39]. In addition to situational factors such as acute stressors from tasks, and anticipatory stressors, dispositional factors might also affect one's stress experience and performance at work. As discussed, personality traits can contribute to one's experience of stress (e.g., [26, 27]). However, it remains unclear how each trait might interact with email use pattern, stress and performance.

An interesting related question is whether stress becomes manifest in the content of the emails. While feeling stressed, participants may alter their word choice and tone compared to a baseline level of stress. Research has found differences in immediate autonomic nervous system activity and subtle differences in word production in writing and in speech [45]. Accordingly, our study also examines whether, while under stress, people alter the affective or emotional tone of their email responses. In particular, we investigate three affective variables using LIWC: positive emotion, negative emotion, and anger.

We investigate the effects of these pervasive aspects of information work (email use pattern, anticipatory stressors) on three different outcome measures during multitasking: (1) physiological stress, (2) email performance, as measured in time to complete email and email tone, and (3) concurrent task performance, the latter being a relatively unexplored area with email studies. Namely, we set out to investigate whether email use patterns might have different qualitative effects on the performance of another concurrent task that is done while switching to email. Our research questions are:

Effects on Stress

RQ1a: How does email use pattern (batch or continual) affect (task-induced) stress during multitasking?
RQ1b: How does exposure to an anticipatory stressor affect (task-induced) stress during multitasking?
RQ1c: What is the relationship between personality, an anticipatory stressor, and different email use patterns on (task-induced) stress during multitasking?

Effects on email performance: Time, Accuracy, Affect

RQ2a: How does email use pattern (batch or continual) affect email performance during multitasking?
RQ2b: How does exposure to an anticipatory stressor affect email performance during multitasking?
RQ2c: What is the relationship between personality, an anticipatory stressor, and different email use patterns on email performance during multitasking?

Effects on other concurrent task performance: Errors

RQ3a: How does email use pattern (batch or continual) affect other concurrent task performance during multitasking?
RQ3b: How does exposure to an anticipatory stressor affect other concurrent task performance during multitasking?
RQ3c: What is the relationship between personality, an anticipatory stressor, and different email use patterns on other concurrent task performance during multitasking?

4 METHODS

Experimental Design

To address these research questions, we examined the following three variables:

- **Email use pattern.** Participants were presented with eight emails in one of two conditions: in batch mode (**B**) at one time, or continually (**C**) throughout the session.
- **Anticipatory stressor.** Participants were assigned to one of two conditions: a) with, and b) without an anticipatory stressor. An anticipatory stressor was created by having participants do the preparatory phase of the Trier Social Stressor Test (TSST) [26], a reliable manipulation that exploits the human stress response to situations that involve social evaluation and/or judgement. This variable thus had two levels: High (**H**), where participants did the preparatory phase of the TSST, i.e., they were told in advance that they would have to present their work to an audience at the end of the experimental session, which led to anticipatory stress, and Low (**L**), where participants were not forewarned that they would present their work.
- **Personality.** Participants completed the Big 5 personality survey, which measures the five dimensions of Agreeableness, Conscientiousness, Extraversion, Neuroticism, and Openness [35].

We recruited 63 participants (45 females/18 males), ages 18-54, mean=23.75, to participate in this study. Recruitment took place via online and flyer posting in three university campuses in the U.S. west and southwest. The participants needed to be at least 18 years of age, have done all their schooling in English, and have at least a high school education. All participants signed informed consent and the study was approved by the institutional review boards of the participating universities. The experimental session lasted about 90 minutes.

Participants were randomly assigned to four groups in a 2x2 factorial design. The two factors were Email Condition and Anticipatory Stressor (received prior to the main writing task). Each factor had two levels: Email Condition {**B**atch, **C**ontinual} x Anticipatory Stressor {**H**igh, **L**ow} = {**BH (N=15), BL (N=14), CH (N=17), CL (N=17)**}. The **Continual** level was characterized by a pseudo-periodic arrival of emails throughout the main session, involving a writing task, while the **Batch** level was characterized by simultaneous delivery of all the emails at the beginning of the session. The Anticipatory **High** stress level was implemented by forewarning participants about the upcoming presentation of their writing task, described above.

The full experimental protocol featured one baseline and four treatments in the following order:

1) **Resting Baseline (RB):** Participants were asked to take a deep breath, close their eyes, and think of something relaxing for 4 minutes. The purpose of this session was to bring participants' arousal close to their tonic levels, so it could be used as a normalizing anchor for the physiological measures taken during the treatments. In human studies that depend on physiological sensing, the importance of the RB session cannot be overestimated. The tonic level of arousal differs significantly among people, depending also on the physiological variable used to gauge arousal. Hence, absolute physiological measurements during treatments are not reliable stress indicators, because what really matters is how much the 'arousal needle' moved with respect to the tonic level.

2) **Writing Baseline (WB):** Participants were given 5 min to write a short essay expressing their opinions on the subject of competition vs. collaboration. This session allowed participants to warm up for the subsequent writing session, and provided a baseline of writing skills for each participant.

3) **Priming – Stroop OR Relaxing Video:** This session occurred directly before the main writing session in order to reinforce arousal for the High Anticipatory Stressor group, while subduing arousal for the Low Anticipatory Stressor group. Priming for the Anticipatory High stress group was implemented via 5 min of the Stroop color word test, while priming for the Anticipatory Low stress group was implemented via 5 min of viewing a relaxing natural landscape video.

4) **Dual Task (DT):** This was the main writing session. Participants were asked to write an essay on the topic of technological singularity (i.e., when machines overtake human intelligence). We chose this topic as it is complex and we expected it to require careful thought. As noted earlier, participants in the two **High** Anticipatory Stress conditions were told that they would have to present their essay to a panel of judges at the end of the session. Participants were given 50 min to compose the essay, during which they would also have to respond to eight emails. In the **B**atch condition, the eight emails arrived 10 minutes after the start of the DT, and participants had 5 minutes to start replying to them. In the **C**ontinual condition, individual emails arrived every 4 minutes (on average), and participants had 10 seconds to start replying to each email. If participants did not start their reply within the transitional time allotted, the interface shifted into the email page in order to ensure consistency across participants of the same email group (Batch or Continual). Five emails asked for opinion/advice and three were scheduling tasks. The five opinion/advice email prompts were chosen from a pilot study on MTurk where an original selection of 30 emails were presented to 270 workers on the platform. Each email was presented to 9 different Mturk workers who were asked to compose a reply as if they worked for a company.

Table 1: Example of email content during the dual task (DT).

Opinion/advice: To help us collect some information to design an emotional wellbeing program, could you please answer the following questions: If someone you cared about lied to you, would you be angry? Why or why not? Have you ever lied to someone you care about? Is it reasonable to expect other people to be honest with you, in all circumstances? What are the exceptions?
Scheduling: Please find a time for a student, administrator, and professor to have an hour-long meeting, in a conference room that can seat 3 people. Peoples' schedules are presented on the first tab of the spreadsheet attached, and room availability is in the second. People have three types of availability: completely free, possible commitments which can be moved if necessary, and commitments that cannot be moved. You may not schedule a meeting in unavailable time slots. Scheduling decisions should minimize the number of possible commitments rescheduled, and prioritize the schedule of the professor, then the administrator, then, lastly, the student. Reply to this email with the meeting start time and room you select.

Then, we selected the five emails that generated the highest average word count in replies. Examples of email content is shown in Table 1. The order of the emails was randomized.

5) **Presentation (P):** At the end of the DT, all participants were asked to deliver their 3-minute oral presentation in front of a panel of three judges, who were attending remotely (via Skype). In addition to serving as an anticipatory stressor for the **H**igh stress groups, the presentation session also acted as an upper bound of stress for all groups, facilitating additional validation of our measurement methods.

Between the DT and P sessions, participants completed other surveys for exploratory analyses, which are not reported in this paper. Experimental materials are found at [1].

Measures

• *Stress.* In physiological terms, we refer to stress as significant levels of sympathetic arousal. When a person's sympathetic arousal is significantly higher than his/her tonic level, then the person is under some form of stress – mild, moderate, or severe. Arousal can be measured indirectly via peripheral physiological variables. In this case, we use perinasal perspiration (**PP**), a signal extracted through the thermal imaging (detailed in the section on Thermophysiological measurements) to gauge the levels of arousal in participants, and thus have a measure of their ongoing 'stress'. PP stress measures were normalized appropriately for validity testing (Thermophysiological Measurements) and in support of analytical modeling (in Results).

• *Personality.* Based on the 44-item Big 5 inventory [35], which measures the five basic personality traits of Agreeableness, Conscientiousness, Extraversion, Neuroticism, and Openness. All variables were mean-centered and scaled. Due to a technical error, one item was left out of the Neuroticism dimension, leaving the score to be calculated by 7 items instead of 8.

• *Time on Email (measured in seconds).* Time spent answering emails during the DT, as logged by the interface: starting when the email is opened and until 'send' is clicked.

• *Affective content of email.* Using LIWC, which calculates a percentage of words in a sample of text belonging to predefined categories. Past empirical studies have found LIWC to reliably detect themes and meaning from text in a wide variety of contexts, as well as to detect emotional states, intentions, motivations, thinking styles, and individual differences [21, 22, 46]. In our work, we use the latest available LIWC dictionary (LIWC 2015), which provides a more comprehensive range of word categories relating to different social and psychological processes compared to earlier LIWC versions [46].

• *Concurrent task performance.* We graded total errors (grammar, usage error, mechanic error and style errors) of the essays using ETS' Criterion ® Online Writing Evaluation Service [7, 50], a web-based system that uses automated scoring and evaluation of student essays. It uses two complementary applications based on natural language processing methods to a) extract features and run statistical models and b) to evaluate errors and provide feedback.

Experimental Setup

The experimental setup was designed to be realistic, unobtrusive, and replicable. We set up three identical layouts – one in each participating campus – in typical office booths. In these booths, the participants operated on standard Dell desktops featuring 24-inch displays. A thermal camera was located under the participant's computer monitor and a visual camera was set up at the top of the computer monitor (Figure 1). Both cameras were imaging the participant's face, sending their video sequences to the experimenter's laptop, where they were processed and then curated in Amazon Web Services. The experimenter with her/his laptop was sitting in a neighboring booth, physically separated from the booth of the participant to minimize distractions.

The thermal imaging system consisted of a Tau 640 longwave infrared (LWIR) camera (FLIR Commercial Systems, Goleta, GA); it featured a small size (44×44×30 mm) and adequate thermal (<50 mK) and spatial (640×512 pixels) resolution. The Tau 640 camera was outfitted with a LWIR 35 mm lens f/1.2 with an auto-focus mechanism. The visual camera was an HD Pro Webcam C920 (Logitech).

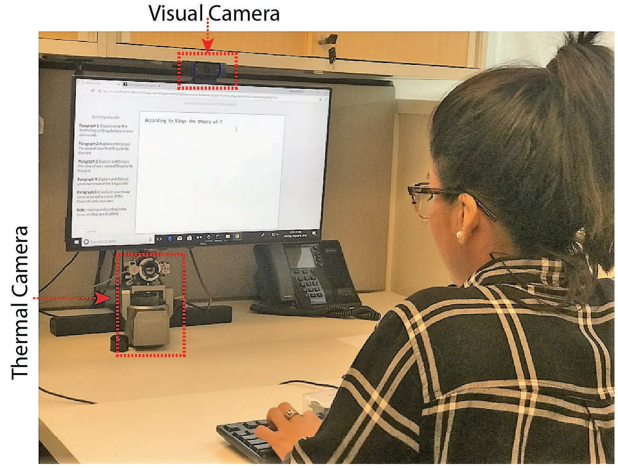

Figure 1: Experimental setup. The thermal imaging camera is located below the computer monitor, while the visual camera is tucked at the top of the monitor. The snapshot shows a participant while writing her essay during the Dual Task (DT).

Hence, the setup did not differ from a typical office setup, the participants were not tethered or restricted in any way, and the sensing was unobtrusive, enabling the experiment to model the look and feel of an office environment.

Study Interface

We used a custom-developed web application to guide the participants throughout the study. The application contained links to the priming tasks and questionnaires, text areas for writing the essays, and a simulated interface similar to Gmail, where participants reply to emails (Fig. 2). The software also logs timestamps for all events. Whenever the participant received an email, the study interface showed a notification on the right-hand side of the essay text area, indicating that a new email had just arrived, shown in Figure 2a. If the participant clicked on the notification, or if they ran out of time, the email interface appeared, where they could select an email in the inbox and reply.

Methods and Validity of Thermo-physiological Measurements

While participants underwent the baseline and experimental treatments, their faces were imaged visually and thermally. The visual imaging sequences had a secondary role; they were meant to be used, as needed, to disambiguate bouts of eustress vs. distress in the physiological signals [61]. We use the method in [54] to extract perinasal perspiration (PP) from the facial thermal imaging, clinically validated to be

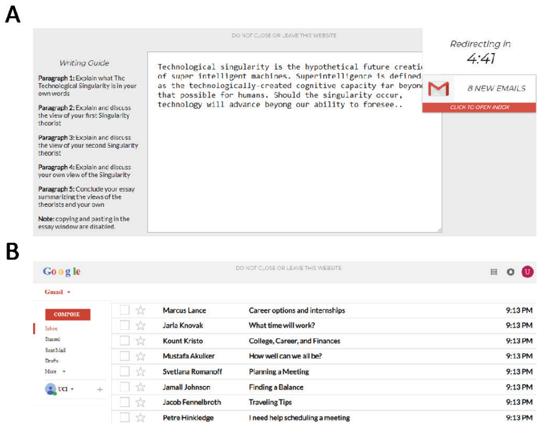

Figure 2: Study interface. A. The essay interface in Batch mode with the time remaining to switch to the email task. B. The Gmail-like user interface.

on par with palm EDA [53, 54], the gold standard in arousal measures. The PP signal extraction method uses facial tracking [65] to nullify head motions. Our methodology has been validated in (1) physical/cognitive/emotional studies (surgeons [44], distractions [43], social categorization [47]) and (2) against heart and breathing rate [58]. Its software implementation is publicly available in figshare [6].

Any significant elevation of arousal beyond the subject's tonic level marks the onset of stress. To track such elevations in the context of the present experiment we computed the within-subject difference in average PP during the treatment and during the Resting Baseline (RB). Figure 3 shows the resulting normalized PP distributions per treatment for participant groups BH, BL, CH, and CL. To ascertain the goodness of the Resting Baseline (RB) as a tonic level proxy and the ability of the thermophysiological method to detect small/moderate sympathetic excitations, we subjected the normalized PP distributions to validity tests (one t-test per distribution). Ideally, these tests should show that:

1) In all participant groups, the mean arousal levels in the stressful treatments (Writing Baseline, Dual Task, and Presentation) were significantly higher than the paired Resting Baseline levels (i.e., above the zero line).

2) In groups BH and CH, the mean arousal levels in the Stroop treatments were significantly higher than the paired Resting Baseline levels (i.e., above the zero line).

3) In groups BL and CL, the mean arousal levels in the Relaxing Video treatments were on par with the paired Resting Baseline levels (i.e., about the zero line).

As the star annotations in Figure 3 show, from the 16 validity tests only one failed, that is, the mean PP value of the Relaxing Video treatment in the CL group was found to be higher than the paired Resting Baseline (RB) – ideally, it should have been on par with it. However, still the Relaxing Video (RV) treatment had significantly lower mean normalized PP with respect to the stressful treatments, thus remaining relatively closer to the Resting Baseline (RB), which was the original aim.

The validity tests associated with each normalized PP distribution are called lower-bound tests, because they verify the existence of an anchoring line as well as the sensitivity of the measurement method. We also carried out upper-bound tests, aimed to examine if the measurement method scales up to strong stimuli, providing a healthy range upon which to base analysis. In the context of the present study, the upper-bound validity tests focused on ascertaining that the Presentation (P) was the most stressful treatment in each participant group. The tests assumed the form of post-hoc analysis of variance among the normalized PP distributions of each participant group. Hence, four such upper-bound tests took place – one for BH, BL, CH, CL; all tests were successful (using ANOVA, $p<0.001$), a result that is visually evident in panels A-D of Figure 3.

The successful outcome of the lower and upper bound validity tests asserts that the experiment was carried out competently, the treatments had the anticipated arousal effects, and the measurement method successfully captured these effects across a healthy range (see Figure 4 and the Results section).

5 RESULTS

Based on visual inspection of the Stress physiological signal (PP) distribution, we eliminated one extreme outlier, and used N=62 in the analyses. Mean PP stress measures during both the Resting Baseline (RB) and Dual Task (DT) were first standardized, and then RB Stress was subtracted (as a baseline) from DT Stress to ameliorate interindividual differences in arousal phenotypes. We refer to this normalized measure of stress used in the subsequent analytical models as PP Stress. For all models reported, we tested for multicollinearity and found that it was not a problem.

RQ1a-c. Effects on Stress

Our first set of research questions investigated how email pattern, presence of an anticipatory stressor and personality might be associated with Stress (as measured by the PP signal during the task). To examine what personality traits might be associated with Stress during the Dual Task, a Pearson correlation of the Big 5 personality traits with Stress (PP signal) revealed that the only personality trait that was significant was Neuroticism. Using Stress as a dependent variable, we ran a general linear model (GLM) in SPSS entering the following terms as independent variables: Email Condition (B/C), Anticipatory Stress (H/L), Neuroticism, and

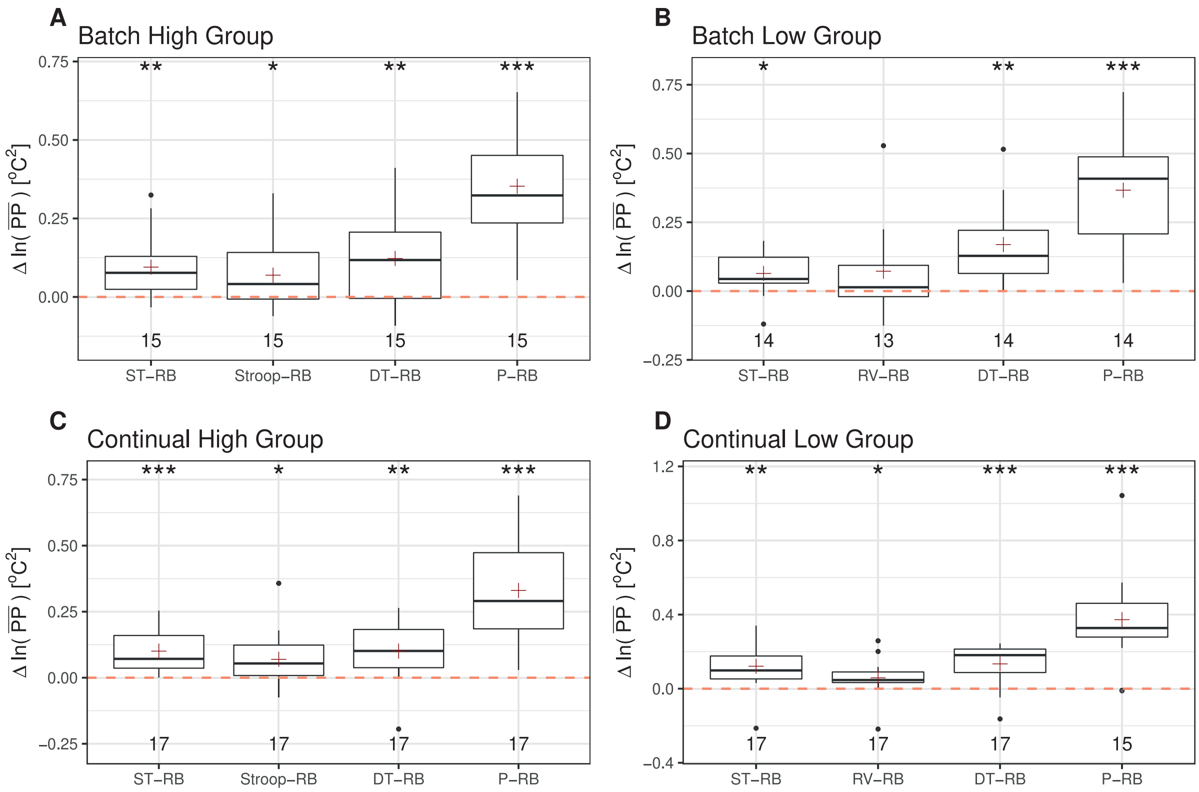

Figure 3: Normalized PP distributions annotated with the results of the lower-bound validity tests. Logarithmic correction was applied to the mean PP signal values to satisfy normality assumptions for the statistical tests. Within-subject normalization was implemented by subtracting (Δ) the Resting Baseline means from the means of treatments. Terminology: WB ≡Writing Baseline; RB ≡Resting Baseline; RV ≡Relaxing Video; DT ≡Dual Task; P ≡Presentation. A. Batch High (BH) subject group B. Batch Low (BL) subject group C. Continual High (CH) subject group D. Continual Low (CL) subject group. We set levels of significance at $\alpha = 0.05$ designated by *, or $\alpha = 0.01$ designated by **, or $\alpha = 0.001$ designated by *. The numbers below the boxplots indicate the respective N sizes.**

all two-way interactions. Table 2 shows a main effect of Neuroticism and a significant Neuroticism × Email Condition interaction. The model is: F(6, 52)=2.16, p<.06, adj R^2=.11. Neuroticism is positively correlated with Stress. Figure 5 shows a plot of the interaction. Interestingly, in the Email Continual condition, the level of stress does not change over low to high Neuroticism. However, in the Batch condition, stress increases with Neuroticism.

RQ2a-c. Effects on Email Performance: Time on Email

Our next set of research questions addressed email performance. First, we examine how stress affects the amount of time spent responding to email. To address whether personality was related to Time on Email we first looked at a Pearson correlation of all Big 5 traits with Time on Email. None were found to be significant as predictors of Time on Email, and thus were not entered into the model.

For this analysis, we now used Stress (PP signal) as an independent variable in the model in order to investigate how it might contribute to time spent answering email. Using Time on Email as the dependent variable, we ran a GLM with Email Condition, Anticipatory Stress, and Stress (PP signal) as main effects, and all 2-way interactions. We included a variable of Total Word Count of the emails as a control, since time spent answering email could be explained by word count.

The results in Table 3 show a significant main effect of Stress (PP signal), and an Email Condition × Stress (PP signal) interaction. The lower the Stress, the longer time was spent on answering the emails. The control variable of Total Word Count was not significant, indicating that amount of text in the email does not affect duration of time on email. There was a trend for an Anticipatory Stress × Stress (PP signal) interaction, model: F(7, 53)=2.42, p<.03, adj R^2=.14. Figure 6 shows the interaction of Email Condition × Stress (PP).

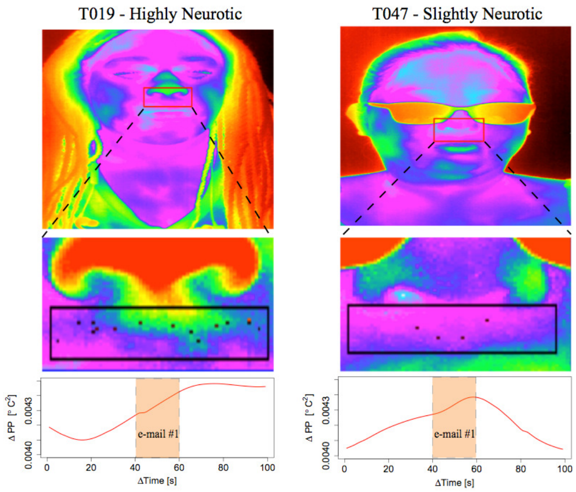

Figure 4: Characteristic samples from the dataset. Participants T019 and T047 belonged to Batch groups, with the first scoring high in the Neuroticism scale while the second scoring low. Black dots in the mid-row images are the thermal imprints of activated perspiration pores in the perinasal region. During email/task handling in DT, these different spatial patterns tended to persist within-subjects, translating to more sustained elevation of PP signals for T019. The shaded areas in the bottom graphs denote the relative times the participants started working on email.

Table 2: Model of RQ1: Stress (PP signal)

	Coeff (SE)	t	p
Intercept	-.01 (.08)	-.12	.91
Email Condition-C	.025 (.11)	.23	.82
Email Condition-B	0		
Neuroticism	.05 (.01)	3.23	.002
Anticipatory Stress-H	.02 (.11)	.18	.86
Anticipatory Stress-L	0		
Email Cond (I) x Neuroticism	-.04 (.02)	-2.50	.02
Email Cond x Anticip Stress-H	-.10 (.15)	-.67	.50
Neuroticism x Anticip Stress-H	-.02 (.02)	-1.08	.29

The trend shows that for Batching, as Stress increases, email is answered in less time. Alternatively, for Continual email use, as Stress increases, participants spent longer time in answering email.

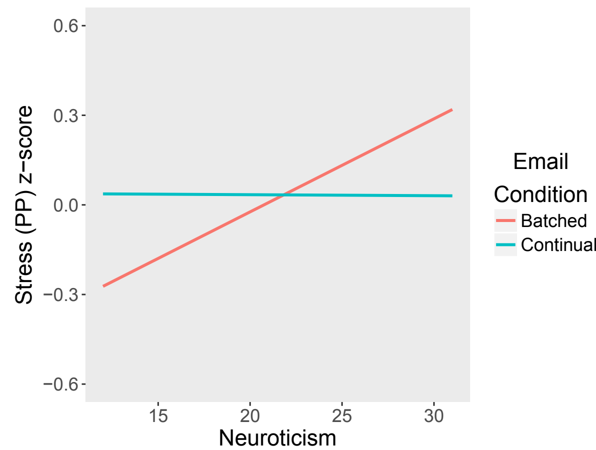

Figure 5: Interaction between Neuroticism and Email Condition on Stress. Stress is measured by $\Delta(\overline{PP})$ $[^oC^2]$, standardized as z-scores. Neuroticism ranges from 7 to 35.

Table 3: Model of RQ2: Time on Email

DV: Time on Email	Coeff (SE)	t	p
Intercept	661.59 (40.71)	16.25	.0001
Email Condition-C	3.21 (31.30)	.10	.92
Email Condition-B	0		
Stress (PP signal)	-171.38 (67.05)	-2.56	.01
Anticipatory Stress-H	46.50 (32.64)	1.43	.16
Anticipatory Stress-L	0		
Total Email Word Count	.44 (.26)	1.66	.10
Email Cond-C x Stress (PP)	162.31 (76.50)	2.12	.04
Email Cond C x Anticipatory Stress-H	12.91 (44.00)	.29	.77
Stress (PP) x Anticipatory Stress-H	139.05 (76.40)	1.82	.07

Effects on Email Performance: Email Scheduling Tasks

Three of the eight emails were scheduling tasks, which we reasoned were cognitively more difficult than the other five emails, which involved text responses. The scheduling tasks

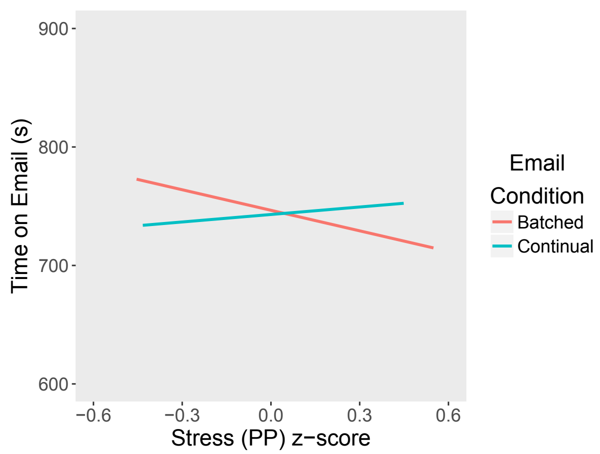

Figure 6: Interaction between Stress and Email Condition on Time on Email. Stress is measured by $\Delta(\overline{PP})$ [$^oC^2$], standardized as z-scores.

had an objective scoring measure, based on an optimal outcome. The scores of the three emails were averaged. A GLM with Email Condition, Anticipatory Stress, and Stress (PP signal) as independent variables, and Total Score as the dependent variable showed no significant effects: F(6,47)=.713, p<.64.

Effects on Email Performance: Affect in Email

To test how stress affects the tone of email, we tested three affective variables (using LIWC) as dependent variables in separate GLM models, with Email Condition, Anticipatory Stress, and Stress (PP signal) as independent variables. For Positive Emotion, no significant effects were found: F(6,55)=.38, p<.89. For Negative Emotion, we found a significant Anticipatory Stressor × Stress (PP signal) interaction (b=-.01, (.002), t=12.48, p<.02). Negative emotion rises with Stress when Anticipatory Stress is low. However, the whole model explains very little variance: F(6,55)=1.19; p<.33, R^2=.02. For Anger, we found a significant effect of Email Condition (the Batch mode showed more anger words), a significant effect of PP Stress (the higher the PP Stress, the more anger words expressed), and a significant Email Condition × PP Stress interaction F(6, 55)=3.00, p<.01, R^2=.16 (Table 4). This interaction is shown in Figure 7.

RQ3a-c. Concurrent task performance

In this set of research questions, we examined how performance of the concurrent essay task was affected by Email Use pattern, Anticipatory Stress, and Stress (PP). As the essays were different time lengths, we divided the errors by minute for both the WB (baseline) and the DT essays. We then subtracted [DT errors/minute - WB errors/minute]. A Pearson correlation showed no Big 5 personality variable correlated with this measure, so none were entered into the model. We then ran a GLM with normalized error scores [DT-WB Errors/Minute] as the dependent variable, and entered Email Condition, Anticipatory Stressor, Stress (PP) and all 2-way interactions. There were no significant effects: F(6, 54)=1.74, p<.13, adj. R2=.07. Thus, neither type of email use pattern, nor an Anticipatory Stressor, nor PP Stress (task-induced) affected performance on the concurrent task.

Table 4: Model of RQ2: Affect in email text

DV: Anger (email text)	Coeff (SE)	t	p
Intercept	.003 (.001)	4.61	.0001
Email Cond-C	-.002 (.001)	-1.87	.07
Email Cond-B	0		
Stress (PP signal)	.006 (.002)	3.43	.001
Anticipatory Stress-H	-.001 (.001)	-.95	.34
Anticipatory Stress-L	0		
Email Cond-C x Stress (PP)	-.005 (.002)	-2.76	.008
Email-C x Anticip Stress-H	.001 (.001)	.73	.47
Stress (PP) x Anticip Stress-H	-.003 (.002)	-1.29	.20

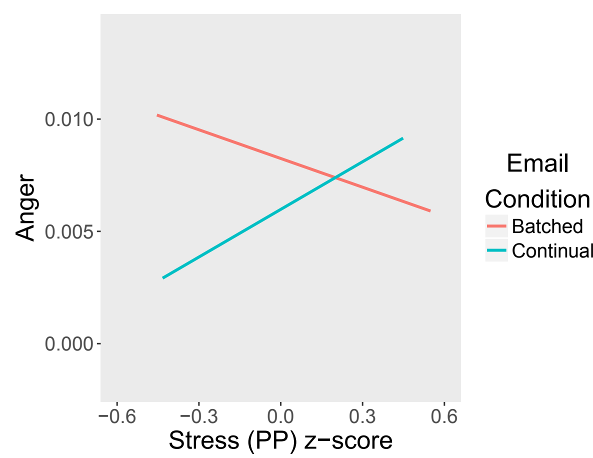

Figure 7: Interaction between Stress and Email Condition on Anger words (from LIWC). Stress is measured by $\Delta(\overline{PP})$ [$^oC^2$], standardized as z-scores. Anger is measured as the proportion of all words that are anger words using LIWC.

6 DISCUSSION

In this study, we set out to disambiguate conflicting results on the relationship of email patterns of batching and continual use with stress using thermal imaging, a proven precision method of detecting stress, to overcome potential biases from self-reports. We also introduced an additional stressor into the experiment, to model anticipatory stress that information workers commonly experience. To address the limited studies on effects of personality on interruptions and more specifically email, we examined personality traits.

Batching, when considered as a sole variable, as previous studies have done [14, 18, 28, 33, 64], shows conflicting results with stress. But its interaction with other factors may explain such discrepant results. Stress increases with Neuroticism when email is batched, but not when email is intermittently managed (i.e. with more email interruptions). What might explain this? Neurotics are more susceptible to stress in general [35]. As handling emails in a batch requires a more sustained focus duration than addressing emails intermittently with breaks in between, and as sustained focus can cause stress [20], it is understandable why the batching task could be more stressful for more neurotic individuals. Given that participants had the same email content, this redirects the attention from email interruptions as stressors to the email itself as a stressor.

Also, when email is batched, as stress increased, people answered emails quicker. This result is consistent with a study showing that when participants worked in an environment with interruptions from phone and instant messaging, their stress increased (measured by self-report), and they finished emails quicker [32]. The current study builds on this work as it shows that the stress-speed relationship does not hold for all types of interruption patterns–only when email is batched. Perhaps when email is batched, people work faster as they can focus for longer durations and switch tasks less often. However, focus is also associated with higher stress responses [20].

Last, we found that when email is batched, it results in the use of less anger words as stress increases. But when email is presented intermittently (with more interruptions), the use of anger words in email increases as stress rises. This result suggests a strong argument for batching emails in organizations: it could potentially reduce the threat of angry emails being circulated and impacting organizational mood.

Our results of an inverse relationship of stress and time on email at first glance seem contrary to findings that show a positive relation with stress and time on email, e.g. [33, 56]. But the key to this puzzle is that in our study we found that participants answered emails quicker when stressed. In an office environment, when people are stressed, they may actually spend a shorter amount of time per email, but may in fact be handling a higher volume. This notion is like a conveyer belt: as incoming email keeps arriving, people handle them faster, presumably to be able to resume other work. Yet email keeps flowing in continually, so speeding up the answering does not reduce email: it rather enables people to handle more email. This result is consistent with Barley's [3] finding that time on email is correlated with number of incoming emails. Responding faster to emails enables people to deal with more email and thus more information, increasing mental load and consequent stress. As time on email is not related to word count, this suggests that it is not fewer words that decreases time on email, but likely spending less time thinking about what one writes. Tying this result to our finding of a positive relationship with stress and anger words included in emails, it suggests that spending less time composing emails can potentially result in more angry emails. Of course, we cannot imply causality.

The fact that we found no significant effect in predicting the total scheduling scores on emails that involved a more cognitively complex process suggests that stress may not be related to the cognitive complexity of the email per se. We would have expected that an anticipatory stressor may have led to worse performance, e.g. through distraction, but it was not the case. It may actually be that it is the nature of the emails themselves that induce stress, e.g. perhaps through their association with work, as Barley [3] suggests.

Last, we found no effect of a batch or intermittent email strategy on the concurrent task performance. Together with the fact that the objective scores of the scheduling task emails showed no difference either suggests that the email strategy one uses may have effects on other aspects of email: speed, and tone of text, rather than accuracy.

Managing workplace stress

In this study we manipulated anticipatory stress, common in information work. We found interactions between anticipatory and acute task-focused stress that intensified the effects on task performance, in terms of time spent on email. This result suggests that to design for lowering workplace stress, different sources of stress need to be considered. The personality trait of Neuroticsm can play an important role in how various stressors (e.g., anticipatory stress, acute stress) affect individuals. This suggests that one-size-fits-all solutions to managing workplace stress are unlikely to work. Instead, a system to ameliorate workplace stress must be able to adapt to (1) the characteristics of each person, their ability to manage distractions, multiple demands, time constraints and commitments, and (2) their current state, as measured by physiological sensors and other sources (e.g., task managers, calendars). Such a system would learn individual characteristics (e.g., based on personality profiles) and work conditions (e.g., looming deadlines, stress levels) for which a particular type of email delivery (e.g., batching) would be beneficial.

Limitations

Our study had several limitations. Except for Anticipatory stress which we manipulated, we cannot infer causality from our Stress (PP) measure. Further experiments would be needed to test this. The thermal imaging technique requires subjects' directional attention. However, moderate natural head motions (±30°) are handled well by the tracker and the method fit well for our study where subjects focused on the computer screen. We lacked one (out of 8 items) on the Neuroticism scale due to a technical error. However, since the other 7 Neuroticism items on the Big 5 are highly correlated, we feel we were able to measure an accurate level of Neuroticism. Participants were mainly university students, and these results may not generalize to information workers. Last, as with any experimental setting, our simulated model of an information work environment enabled us to control variables, and albeit realistic, it may not have the ecological validity of an in situ study.

7 CONCLUSIONS

Email use and workplace interruptions have been documented as a contributing factor to stress among information workers. In this study, we investigated email strategies to reduce stress, which previously showed mixed results. We used a state-of-the-art thermal imaging technique to measure stress accurately. Our results suggest that batching requires more sustained focus on email, and can increase stress among Neurotics, people susceptible to stress. Our findings suggest that strategies of batching or continually answering email affect processes other than accuracy of task performance. Rather, batching leads people to work on emails faster, and continual interruptions is associated with more angry words in emails. We propose that email batching could benefit organizations by reducing angry text in emails for employees under stress. Email only continues to increase, and we hope that our results will contribute to solutions to ameliorate stress from managing email.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under grants #1704889, 1704682, and 1704636.

REFERENCES

[1] Fatema Akbar, Ayse Elvan Bayraktaroglu, Ge Gao, Ted Grover, Gloria Mark, Kevin Storer, Dennis Rodrigo Da Cunha Silva, Ricardo Gutierrez-Osuna, Zelun Wang, Pradeep Buddharaju, Nathan Cooper Jones, Ioannis Pavlidis, Amanveer Wesley, and Shaila Zaman. 2018. Email Makes You Sweat - CHI'19 : Dataset and Methods. https://osf.io/5surx/

[2] David M Almeida. 2005. Resilience and vulnerability to daily stressors assessed via diary methods. *Current Directions in Psychological Science* 14, 2 (2005), 64–68.

[3] Stephen R Barley, Debra E Meyerson, and Stine Grodal. 2011. E-mail as a source and symbol of stress. *Organization Science* 22, 4 (2011), 887–906.

[4] Victoria Bellotti, Nicolas Ducheneaut, Mark Howard, Ian Smith, and Rebecca E Grinter. 2005. Quality versus quantity: E-mail-centric task management and its relation with overload. *Human–Computer Interaction* 20, 1-2 (2005), 89–138.

[5] Adam Bradley, Duncan P Brumby, Anna L Cox, and Jon Bird. 2013. How to manage your inbox: is a once a day strategy best?. In *Proceedings of the 27th International BCS Human Computer Interaction Conference*. British Computer Society, 20.

[6] Pradeep Buddharaju, Ashik Khatri, and Ioannis Pavlidis. 2017. S-Interface (Previously OTACS). (4 2017). https://doi.org/10.6084/m9.figshare.4244273.v4

[7] Yigal Attali Jill Burstein and Y Attali. 2006. Automated Essay Scoring With E-rater V. 2.0. *The Journal of Technology, Learning, and Assessment* 4, 3 (2006), 31.

[8] Future Work Centre. 2015. "You've got mail!". http://www.futureworkcentre.com/wp-content/uploads/2015/07/FWC-Youve-got-mail-research-report.pdf

[9] Daniel Chen and Roel Vertegaal. 2004. Using mental load for managing interruptions in physiologically attentive user interfaces. In *CHI'04 extended abstracts on Human factors in computing systems*. ACM, 1513–1516.

[10] Mary Czerwinski, Eric Horvitz, and Susan Wilhite. 2004. A diary study of task switching and interruptions. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 175–182.

[11] Laura A Dabbish and Robert E Kraut. 2006. Email overload at work: an analysis of factors associated with email strain. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*. ACM, 431–440.

[12] Jeanette Frost Ebstrup, Lene Falgaard Eplov, Charlotta Pisinger, and Torben Jørgensen. 2011. Association between the Five Factor personality traits and perceived stress: is the effect mediated by general self-efficacy? *Anxiety, Stress & Coping* 24, 4 (2011), 407–419.

[13] Hélène Eyrolle and Jean-Marie Cellier. 2000. The effects of interruptions in work activity: Field and laboratory results. *Applied ergonomics* 31, 5 (2000), 537–543.

[14] Pamela S Galluch, Varun Grover, and Jason Bennett Thatcher. 2015. Interrupting the workplace: Examining stressors in an information technology context. *Journal of the Association for Information Systems* 16, 1 (2015), 1.

[15] Gloria Garcia-Banda, Mateu Servera, Karin Chellew, Victoria Meisel, Joana Fornes, Esther Cardo, Gerardo Perez, Maria Riesco, and Ronald M Doctor. 2011. Prosocial personality traits and adaptation to stress. *Social Behavior and Personality: an international journal* 39, 10 (2011), 1337–1348.

[16] Nicole A Gillespie, MHWA Walsh, Anthony Harold Winefield, Jagdish Dua, and Con Stough. 2001. Occupational stress in universities: Staff perceptions of the causes, consequences and moderators of stress. *Work & stress* 15, 1 (2001), 53–72.

[17] Radicati group. 2015. Email Statistics Report. https://www.radicati.com/wp/wp-content/uploads/2015/02/Email-Statistics-Report-2015-2019-Executive-Summary.pdf.

[18] Ashish Gupta, Ramesh Sharda, and Robert A Greve. 2011. You've got email! Does it really matter to process emails now or later? *Information Systems Frontiers* 13, 5 (2011), 637–653.

[19] Mario Hair, Karen V Renaud, and Judith Ramsay. 2007. The influence of self-esteem and locus of control on perceived email-related stress. *Computers in Human Behavior* 23, 6 (2007), 2791–2803.

[20] Peter A Hancock. 1989. A dynamic model of stress and sustained attention. *Human factors* 31, 5 (1989), 519–537.

[21] Jacob B Hirsh and Jordan B Peterson. 2009. Personality and language use in self-narratives. *Journal of research in personality* 43, 3 (2009), 524–527.

[22] Thomas Holtgraves. 2011. Text messaging, personality, and the social context. *Journal of research in personality* 45, 1 (2011), 92–99.

[23] James M Hudson, Jim Christensen, Wendy A Kellogg, and Thomas Erickson. 2002. I'd be overwhelmed, but it's just one more thing to do: Availability and interruption in research management. In *Proceedings of the SIGCHI Conference on Human factors in computing systems*. ACM, 97–104.

[24] Thomas Jackson, Ray Dawson, and Darren Wilson. 2001. The cost of email interruption. *Journal of Systems and Information Technology* 5, 1 (2001), 81–92.

[25] Thomas W Jackson, Ray Dawson, and Darren Wilson. 2003. Understanding email interaction increases organizational productivity. *Commun. ACM* 46, 8 (2003), 80–84.

[26] Clemens Kirschbaum, Karl-Martin Pirke, and Dirk H Hellhammer. 1993. The 'Trier Social Stress Test'–a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology* 28, 1-2 (1993), 76–81.

[27] Mika Kivimäki, Marianna Virtanen, Marko Elovainio, Anne Kouvonen, Ari Väänänen, and Jussi Vahtera. 2006. Work stress in the etiology of coronary heart disease–a meta-analysis. *Scandinavian journal of work, environment & health* (2006), 431–442.

[28] Kostadin Kushlev and Elizabeth W Dunn. 2015. Checking email less frequently reduces stress. *Computers in Human Behavior* 43 (2015), 220–228.

[29] Sophie Leroy. 2009. Why is it so hard to do my work? The challenge of attention residue when switching between work tasks. *Organizational Behavior and Human Decision Processes* 109, 2 (2009), 168–181.

[30] James A Levine, Ioannis T Pavlidis, Leslie MacBride, Zhen Zhu, and Panagiotis Tsiamyrtzis. 2009. Description and clinical studies of a device for the instantaneous detection of office-place stress. *Work* 34, 3 (2009), 359–364.

[31] Rita S Mano and Gustavo S Mesch. 2010. E-mail characteristics, work performance and distress. *Computers in Human Behavior* 26, 1 (2010), 61–69.

[32] Gloria Mark, Daniela Gudith, and Ulrich Klocke. 2008. The cost of interrupted work: more speed and stress. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 107–110.

[33] Gloria Mark, Shamsi T Iqbal, Mary Czerwinski, Paul Johns, Akane Sano, and Yuliya Lutchyn. 2016. Email duration, batching and self-interruption: Patterns of email use on productivity and stress. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 1717–1728.

[34] Gloria Mark, Stephen Voida, and Armand Cardello. 2012. A pace not dictated by electrons: an empirical study of work without email. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 555–564.

[35] Robert R McCrae and Paul T Costa Jr. 1999. A five-factor theory of personality. *Handbook of personality: Theory and research* 2, 1999 (1999), 139–153.

[36] Daniel McDuff, Sarah Gontarek, and Rosalind Picard. 2014. Remote measurement of cognitive stress via heart rate variability. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*. IEEE, 2957–2960.

[37] Daniel C McFarlane and Kara A Latorella. 2002. The scope and importance of human interruption in human-computer interaction design. *Human-Computer Interaction* 17, 1 (2002), 1–61.

[38] Brid O'Conaill and David Frohlich. 1995. Timespace in the workplace: Dealing with interruptions. In *Conference companion on Human factors in computing systems*. ACM, 262–263.

[39] Peter O'Connor, Jessica Nguyen, and Jeromy Anglim. 2017. Effectively Coping with Task Stress: A Study of the Validity of the Trait Emotional Intelligence Questionnaire–Short Form (TEIQue–SF). *Journal of personality assessment* 99, 3 (2017), 304–314.

[40] Jussi Okkonen, Tomi Heimonen, Reijo Savolainen, and Markku Turunen. 2017. Assessing Information Ergonomics in Work by Logging and Heart Rate Variability. In *International Conference on Applied Human Factors and Ergonomics*. Springer, 425–436.

[41] Lynn M Oswald, Peter Zandi, Gerald Nestadt, James B Potash, Amanda E Kalaydjian, and Gary S Wand. 2006. Relationship between cortisol responses to stress and personality. *Neuropsychopharmacology* 31, 7 (2006), 1583.

[42] Antti Oulasvirta and Pertti Saariluoma. 2004. Long-term working memory and interrupting messages in human-computer interaction. *Behaviour & Information Technology* 23, 1 (2004), 53–64.

[43] I Pavlidis, M Dcosta, S Taamneh, M Manser, T Ferris, R Wunderlich, E Akleman, and P Tsiamyrtzis. 2016. Dissecting driver behaviors under cognitive, emotional, sensorimotor, and mixed stressors. *Scientific reports* 6 (2016), 25651.

[44] Ioannis Pavlidis, Panagiotis Tsiamyrtzis, Dvijesh Shastri, Avinash Wesley, Yan Zhou, Peggy Lindner, Pradeep Buddharaju, Rohan Joseph, Anitha Mandapati, Brian Dunkin, and others. 2012. Fast by nature-how stress patterns define human experience and performance in dexterous tasks. *Scientific Reports* 2 (2012), 305.

[45] James W Pennebaker. 1993. Putting stress into words: Health, linguistic, and therapeutic implications. *Behaviour research and therapy* 31, 6 (1993), 539–548.

[46] James W Pennebaker, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. *The development and psychometric properties of LIWC2015*. Technical Report.

[47] Giorgia Ponsi, Maria Serena Panasiti, Giulia Rizza, and Salvatore Maria Aglioti. 2017. Thermal facial reactivity patterns predict social categorization bias triggered by unconscious and conscious emotional stimuli. *Proc. R. Soc. B* 284, 1861 (2017), 20170908.

[48] Colin Puri, Leslie Olson, Ioannis Pavlidis, James Levine, and Justin Starren. 2005. StressCam: non-contact measurement of users' emotional states through thermal imaging. In *CHI'05 extended abstracts on Human factors in computing systems*. ACM, 1725–1728.

[49] Karen Renaud, Judith Ramsay, and Mario Hair. 2006. " You've got e-mail!"... shall I deal with it now? Electronic mail from the recipient's perspective. *International Journal of Human-Computer Interaction* 21, 3 (2006), 313–332.

[50] Lawrence M Rudner and Tahung Liang. 2002. Automated essay scoring using Bayes' theorem. *The Journal of Technology, Learning and Assessment* 1, 2 (2002).

[51] Emma Russell, Stephen A Woods, and Adrian P Banks. 2017. Examining conscientiousness as a key resource in resisting email interruptions: Implications for volatile resources and goal achievement. *Journal of occupational and organizational psychology* 90, 3 (2017), 407–435.

[52] Tamera R Schneider, Tara A Rench, Joseph B Lyons, and Rebecca R Riffle. 2012. The influence of neuroticism, extraversion and openness on stress responses. *Stress and Health* 28, 2 (2012), 102–110.

[53] Dvijesh Shastri, Arcangelo Merla, Panagiotis Tsiamyrtzis, and Ioannis Pavlidis. 2009. Imaging facial signs of neurophysiological responses. *IEEE Transactions on Biomedical Engineering* 56, 2 (2009), 477–484.

[54] Dvijesh Shastri, Manos Papadakis, Panagiotis Tsiamyrtzis, Barbara Bass, Ioannis T Pavlidis, and others. 2012. Perinasal imaging of physiological stress and its affective potential. *IEEE Trans. Affective Computing* 3, 3 (2012), 366–378.

[55] Sabine Sonnentag, Leonard Reinecke, Jutta Mata, and Peter Vorderer. 2018. Feeling interrupted–Being responsive: How online messages relate to affect at work. *Journal of Organizational Behavior* 39, 3 (2018), 369–383.

[56] Jean-Francois Stich, Monideepa Tarafdar, Patrick Stacey, and Cary L Cooper. 2018. E-mail load, workload stress and desired e-mail load: a cybernetic approach. *Information Technology & People* (2018).

[57] Feng-Tso Sun, Cynthia Kuo, Heng-Tze Cheng, Senaka Buthpitiya, Patricia Collins, and Martin Griss. 2010. Activity-aware mental stress detection using physiological sensors. In *International Conference on Mobile Computing, Applications, and Services*. Springer, 282–301.

[58] Salah Taamneh, Panagiotis Tsiamyrtzis, Malcolm Dcosta, Pradeep Buddharaju, Ashik Khatri, Michael Manser, Thomas Ferris, Robert Wunderlich, and Ioannis Pavlidis. 2017. A multimodal dataset for various forms of distracted driving. *Scientific data* 4 (2017), 170110.

[59] J Gregory Trafton and Christopher A Monk. 2007. Task interruptions. *Reviews of human factors and ergonomics* 3, 1 (2007), 111–126.

[60] Panagiotis Tsiamyrtzis, Malcolm Dcosta, Dvijesh Shastri, Eswar Prasad, and Ioannis Pavlidis. 2016. Delineating the operational envelope of mobile and conventional EDA sensing on key body locations. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 5665–5674.

[61] Avinash Wesley, Peggy Lindner, and Ioannis Pavlidis. 2012. Eustressed or distressed?: combining physiology with observation in user studies. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*. ACM, 327–330.

[62] Steve Whittaker, Victoria Bellotti, and Jacek Gwizdka. 2006. Email in personal information management. *Commun. ACM* 49, 1 (2006), 68–73.

[63] Christopher D Wickens. 2008. Multiple resources and mental workload. *Human factors* 50, 3 (2008), 449–455.

[64] Glenn Wylie and Alan Allport. 2000. Task switching and the measurement of "switch costs". *Psychological research* 63, 3-4 (2000), 212–233.

[65] Yan Zhou, Panagiotis Tsiamyrtzis, Peggy Lindner, Ilya Timofeyev, and Ioannis Pavlidis. 2013. Spatiotemporal smoothing as a basis for facial tissue tracking in thermal imaging. *IEEE Transactions on Biomedical Engineering* 60, 5 (2013), 1280–1289.

[66] Manuela Züger, Sebastian C Müller, André N Meyer, and Thomas Fritz. 2018. Sensing Interruptibility in the Office: A Field Study on the Use of Biometric and Computer Interaction Sensors. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 591.