TimeAware: Leveraging Framing Effects to Enhance Personal Productivity

Young-Ho Kim[1] **Jae Ho Jeon**[1] **Eun Kyoung Choe**[2]
Bongshin Lee[3] **KwonHyun Kim**[4] **Jinwook Seo**[1]
[1]HCI Lab, Dept. of Computer Science & Engineering, Seoul National University, Seoul, Korea
[2]College of IST, The Pennsylvania State University, University Park, PA, USA
[3]Microsoft Research, Redmond, WA, USA
[4]Interdisciplinary Program of Cognitive Science, Seoul National University, Seoul, Korea
{yhkim, jhjeon}@hcil.snu.ac.kr, echoe@ist.psu.edu,
bongshin@microsoft.com, {skydome7, jseo}@snu.ac.kr

ABSTRACT
To help people enhance their personal productivity by providing effective feedback, we designed and developed TimeAware, a self-monitoring system for capturing and reflecting on personal computer usage behaviors. TimeAware employs an ambient widget to promote self-awareness and to lower the feedback access burden, and web-based information dashboard to visualize people's detailed computer usage. To examine the effect of framing on individual's productivity, we designed two versions of TimeAware, each with a different framing setting—one emphasizing *productive activities* (positive framing) and the other emphasizing *distracting activities* (negative framing), and conducted an eight-week deployment study (N = 24). We found a significant effect of framing on participants' productivity: only participants in the negative framing condition improved their productivity. The ambient widget seemed to help sustain engagement with data and enhance self-awareness. We discuss how to leverage framing effects to help people enhance their productivity, and how to design successful productivity monitoring tool.

Author Keywords
Productivity tracking; self-monitoring; self-tracking; personal informatics; framing effects; semi-automated journaling; data engagement.

ACM Classification Keywords
H.5.2. Information interfaces and presentation (e.g., HCI): User interfaces.

ACM 978-1-4503-3362-7/16/05$15.00
DOI: http://dx.doi.org/10.1145/2858036.2858428

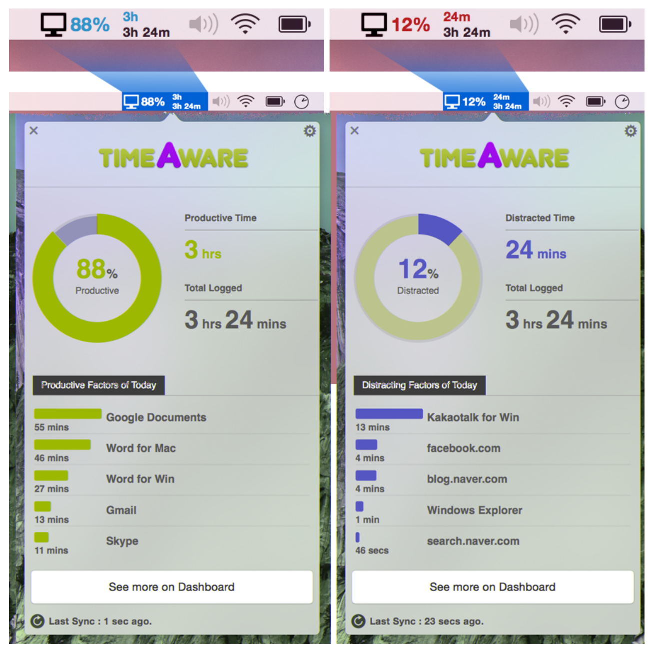


Figure 1. TimeAware widget for OS X with productivity-emphasized setting (left) and distraction-emphasized setting (right). Clicking the widget on menu bar (top), people can see the expanded view with detailed information, which also provides a shortcut to the information dashboard website.

INTRODUCTION
People use multiple applications and switch contexts frequently when they work at the computer. Although this multitasking environment can boost productivity, it could also distract people because they could easily switch into distractors (e.g., browsing the internet, turning the game on using shortcut icons). Thus, we have been witnessing various approaches to help people effectively spend time on computers. They include applications for tracking productivity (e.g., RescueTime [38], SLife [39]), blocking distracting apps during work hours (e.g., Focus [5]), and discouraging multitasking by supporting a main task on minimal interface (e.g., iA Writer [44]).

Given the importance of self-awareness and self-reflection in behavior change [6], a self-monitoring component is commonly embodied in productivity applications (e.g., RescueTime [38], SLife [39], Beeminder [2]). For example, RescueTime, one prominent commercial self-monitoring tool, automatically tracks individuals' computer usage behaviors such as duration of application usage and website browsing. Each application and website (referred to hereinafter as 'activity') is then assigned to one of five productivity levels—very productive to very distracting—to calculate a daily *productivity score*, which indicates how well a person has been working on productive tasks at the computer. RescueTime also provides feedback with a web dashboard and sends weekly reports via email.

Although prior research suggests potential usefulness of such systems (e.g., [34,35]), researchers also discovered a major drawback, that is, extremely low engagement with data. For example, the average duration of RescueTime usage (i.e., accessing the RescueTime website) was only 4.68 seconds per day even though participants received a daily phone call from the researchers [12].

In this work, we aimed to help people enhance personal productivity by providing effective feedback. Toward this goal, we designed and developed TimeAware, a self-monitoring system for capturing and reflecting on personal computer usage behaviors, leveraging RescueTime's API. Inspired by recent research on the effect of providing information on the widget (e.g., [9,13]), we employed an *ambient widget* to provide near real-time productivity (or distraction) score based on individual's computer usage behaviors. We also enabled people to manually edit the RescueTime's automated categorizations of productivity level to enhance the accuracy of the data.

Our work complements and extends the self-monitoring research in the context of productivity tracking by exploring how to leverage framing effects [8,11] to design effective personal productivity feedback. We designed two versions of TimeAware, each with a different framing setting—one emphasizing *productive activities* (positive framing, or PF) and the other emphasizing *distracting activities* (negative framing, or NF). We then studied the effect of framing on individual's productivity through a 3-phased (*Baseline, Intervention*, and *Withdrawal*) between-subjects (positive and negative framing) field deployment study with 24 participants during 8-week period. We found a significant effect of framing on participants' productivity: only NF participants improved their productivity when the feedback was displayed, but this effect disappeared when we removed the feedback. Participants also expressed differing receptiveness and attitudes towards the two versions of TimeAware feedback. The ambient widget seemed to help sustain engagement with data and enhance self-awareness.

The contributions of this work are threefold: (1) the exploration of the effects of framing on productivity, (2) field deployment study for ecologically valid assessment of framing on actual behaviors (prior framing studies were often conducted using a static survey with hypothetical scenario), and (3) the implications of designing productivity monitoring system learned from the study.

In what follows, we summarize related work and illustrate the TimeAware system in detail. We then describe the study design and report on findings from the field deployment study. Based on the lessons we learned, we discuss implications and future opportunities for designing successful personal productivity monitoring systems and propose stepwise guidelines for conducting framing studies in the HCI context.

RELATED WORK

In this section, we cover related work in the areas of (1) self-monitoring and feedback, (2) productivity monitoring systems, and (3) leveraging visual framing in designing effective feedback.

Self-monitoring and Feedback

Increased *self-awareness*—being aware of one's current state—tends to promote changes in the person's performance or behavior [6]. This is referred to as *reactivity* or *reactive effects* [33]. Researchers found that providing feedback could enhance reactivity because it provides a yardstick that enables a person to compare their current state to the ideal state or goal [21]. Because self-monitoring feedback can facilitate this process, it is important to design effective feedback and adequately deliver it.

How people engage in self-monitoring depends on the type of data capture mechanism. When a person manually captures the target behavior, the person naturally becomes aware of the data during the capture. In automatic tracking methods, a person does not engage in the data capture process, and thus, self-awareness and self-reflection decrease unless effective feedback is provided by the system [10]. Therefore, many automated self-tracking systems deliver feedback in various manners, which may impose additional access burden. To access the feedback, people usually have to open a web browser (e.g., RescueTime, Fitbit dashboard website [18]) or launch a mobile app (e.g., Fitbit app, Health app on iPhone [20]). Thus, researchers have proposed ways to lower the access burden and improve awareness such as providing feedback on smartphone's lock screen widget [9] or on the smartphone's wallpaper [14]. Lee and Dey showed that providing self-monitoring feedback on a tablet display was effective for medication taking adherence [23]. In the desktop environment, menu bar (OS X) or taskbar (Windows) is typically used to place application icons that are frequently used, or to show real-time feedback (e.g., battery status, clock). In prior work, researchers also used this space to inform a person of privacy risks and found that showing information on the taskbar improved information awareness [13]. While we share the similar goals of enhancing information awareness, our focus is to project self-monitoring data in the desktop environment with an aim to enhance individuals' productivity.

Productivity Monitoring Systems

Helping people manage personal productivity in a collaborative work environment has been an active area of research (e.g., [17,31]). Recognizing *time* as an important resource in personal and professional life, researchers have made efforts on helping people improve time management to enhance productivity via better support of scheduling (e.g., [4,45]) and understanding the nature of multitasking (e.g., [16,19,29]).

Recently, we see many personal productivity management systems appeared in consumer markets with the idea of tracking and visualizing an individual's work history. Some tools support manual tracking in multiple platforms (e.g., Pomodoro Timer [41], Attainr [1]). Other productivity monitoring tools usually adopt automated tracking methods to record the computer usage duration and help people monitor their productivity by showing duration of application usage (e.g., ManicTime [28], KnowSelf [22]) or a productivity score derived from application usages (e.g., RescueTime [38], Hubstaff [42]). Most of them provide a client application that captures computer usage and a separate website or standalone application that visualizes the captured data.

Prior work on personal informatics explored how people use such systems and showed their potential usefulness. Pammer and colleagues reported that visualizing application usage history helps people gain insights with regards to time management [34,35,36]. In the experiment using RescueTime, Zhou and colleagues also found that providing individuals' usage duration of social network service (SNS) increased their awareness on SNS usage, although their actual SNS usage duration did not change [46]. In their follow-up study, however, the authors found a major drawback—participants' low engagement with RescueTime [12]. This result calls for alternative ways to engage people and to deliver productivity monitoring feedback beyond the current web-based method, which imposes high information access burden.

Leveraging Visual Framing

To identify more effective way of presenting productivity monitoring feedback, we turned to the well-known *framing effects* [43], which refer to the way people differently react to the same information framed differently (e.g., highlighting information in a positive light versus negative light). For example, in the framing of the odds of a surgical operation, many would prefer having an operation of where the outcome is "90 out of 100 are alive after five years" than one where "10 out of 100 are dead after five years [30]." Framing effect was initially studied in *textual* descriptions, and has recently been applied in designing persuasive *visual* feedback. Called "*Visual Framing,*" prior work leveraged framing effects in designing visual representations of self-monitoring feedbacks [11], mobile app's privacy information [8], and health risk visualizations [26]. Inspired by prior work, we set out to identify effective ways to frame personal productivity data that can nudge people toward increased productivity.

TIMEAWARE

The TimeAware system consists of two components—the ambient widget (Figure 1) and the information dashboard (Figure 2). We used RescueTime's API to incorporate the automatically captured data and designed our own feedback. To make up for the lowered attention due to the automated tracking, we provided feedback on the easily accessible ambient widget, which also served as a shortcut to the information dashboard. In what follows, we describe productivity modeling and feedback design, ambient widget, information dashboard, and implementation details.

Framing the Productivity Score

We used two contrasting frames—one emphasizing *productive activities* (positive framing) and the other emphasizing *distracting activities* (negative framing).

Productivity Modeling

Our productivity model builds upon RescueTime's scoring model. The productivity score of RescueTime is a weighted average of all used applications' productivity level; each computer application is automatically classified into one of the following categories—very productive, productive, neutral, distracting, and very distracting [27]. Because distinction between "very productive" and "productive" (as well as "very distracting" and "distracting") was not well-defined and thus confusing, we decided to simplify the model to use three levels of productivity (*productive*, *neutral*, and *distractive*) in TimeAware; we merged RescueTime's "very productive" category into "productive" category and "very distracting" into "distracting."

In dealing with the *neutral* activities, RescueTime seems to assign the *neutral* label to the activities that are (1) unclassifiable to any of predefined category or (2) not strongly implying productivity (e.g., search engines and browsers). Despite the potential ambiguity it could add, we decided to keep the neutral category, because RescueTime assigns the neutral label to unfamiliar applications or websites, which people can manually edit later.

In TimeAware's productivity model, each application's productivity level depends on the context of the application use. For example, messaging applications are considered productive if used for business meetings, but not if used for chattering with friends. However, such contexts cannot be automatically detected accurately. Therefore, we enabled people to change applications' productivity label. For example, a *productive* or *neutral* application can be changed to a *distractive* application and vice versa. The productivity label can be altered for the whole usage of an application during the entire study period (called app-level editing) or just for the selected fraction of a usage (called log-level editing) (see Figure 3).

We defined the ***productive rate*** by revising the formula of the original ***productivity score*** from RescueTime. The

Positive Framing	Negative Framing
$\textit{productive duration} = t_{prd} + 0.5 \times t_{ntr}$	$\textit{distracted duration} = t_{dst} + 0.5 \times t_{ntr}$
$\textit{productive rate} = \dfrac{\textit{productive duration}}{t_{total}}$	$\textit{distractive rate} = \dfrac{\textit{distracted duration}}{t_{total}}$

$t_{\{prd,ntr,dst\}}$: total duration labeled as {productive, neutral, distractive}
t_{total} : total computer usage duration. Equals to $t_{prd} + t_{ntr} + t_{dst}$.

Table 1. Productivity metrics used in TimeAware in the two framing conditions.

productive rate means the ratio of the *productive* duration to the total computer usage duration. In Table 1, we show how we calculate the *productive rate* and *distractive rate*.

We dealt with the *neutral* applications as RescueTime does, assuming that they can be treated either as *productive* or *distractive*, and formulated the equation so that a half of the *neutral* duration contributes to productive duration, and the other half to distracted duration. *Distractive rate* can be easily inferred due to the symmetry in the equation.

Emphasizing the Frames in Feedback Design
We emphasized each framing condition by using different color encodings: we used *blue* for encoding productive elements and *red* for distractive elements. We also emphasized or de-emphasized framed information by varying the color saturation (e.g., see the arcs on the donut charts in Figure 1), and by using different wording for each condition—the phrases *productive rate* and *productive duration* were used in the positive framing condition, while *distractive rate* and *distracted duration* were used in the negative framing condition.

Ambient Widget
To help people quickly access productivity feedback, we showed the productive rate (or distractive rate) and total computer usage duration at the *ambient widget* on the menu bar and more detailed information on the *expanded view* (Figure 1). TimeAware's ambient widget is an automatically updated, always visible application, which displays a brief summary of computer usage. The menu bar acts as an ambient display where the person can check his or her current productivity without much cognitive load [32]. The ambient widget is either embedded in the menu bar (OS X) or is fixed above the system tray (Windows).

The widget on the menu bar shows a basic summary including today's productive rate (or distractive rate), total productive duration (or total distracted duration), and total computer usage duration. When the person clicks on the ambient widget to open up the expanded view, more detailed information is revealed such as top 5 activities that contributed most to today's productive (or distractive) rate. The expanded view also contains a link to the web-based information dashboard.

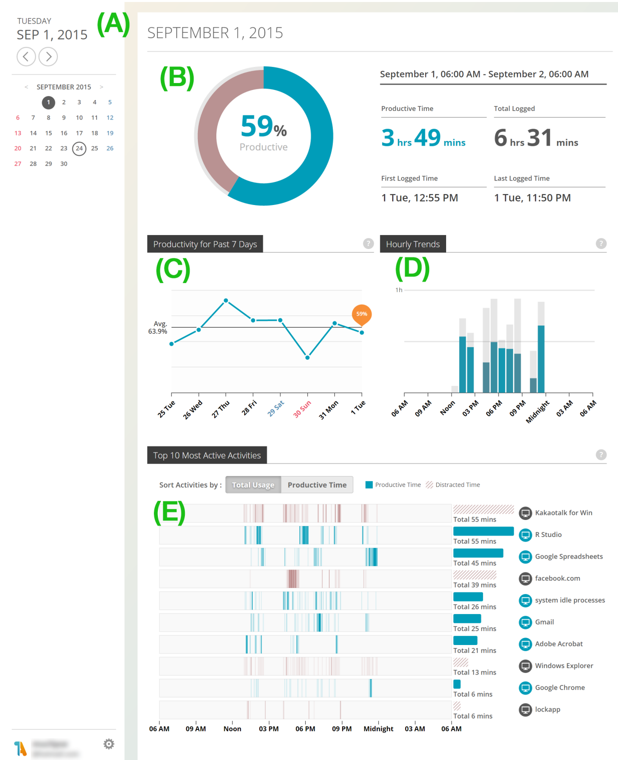


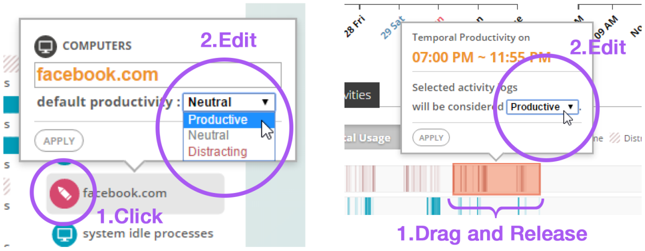
Figure 2. TimeAware information dashboard for the Positive Framing condition: (A) the calendar navigator, (B) the summary panel, (C) the history chart, (D) the hourly trends panel, and (E) the top 10 activities panel.

Information Dashboard
The information dashboard displays detailed daily performance including hourly trends and breakdown, past 7-day trends, and a list of top 10 activities sorted by duration. Figure 2 shows the dashboard for the positive condition (see supplementary document for the negative condition design). The calendar allows people to browse the historical data by date (Figure 2A). We describe the details of information dashboard using the interfaces designed for the positive framing condition.

The **Summary Panel** (Figure 2B) displays the productive rate (e.g., 59%) and productive duration (e.g., 3 hours and 49 minutes), total device usage duration (e.g., 6 hours and 31 minutes), and timestamps of when the person started and ended using the computer.

The **History Chart** (Figure 2C) illustrates the 7-day trend of productive (or distractive) rate with a line chart. The chart is devised to help the person maintain a regular productivity level by comparing the current status with past records [25].

The **Hourly Trends panel** (Figure 2D) shows how the person spent time each hour using a histogram**.** Each bin represents an hour of day, and is a stacked bar with two sections—the colored bar and the gray bar. The height of the stacked bar indicates the total computer usage duration, with the colored bar indicating productive duration and the



(a) App-level Editing (b) Log-level Editing

Figure 3. Stepwise process of the two editing features: (a) editing default productivity of an application (app-level) and (b) assigning temporal productivity on specific time span (log-level).

gray bar indicating distractive duration. When the mouse is hovered over the histogram, the tooltip containing information of the top 5 most-used applications is shown.

We designed the **Top 10 Activities panel** (Figure 2E) to help people be aware of the applications that contribute most to their productive (or distractive) rate. Application usage pattern is visualized by heatmaps, which show not just the duration of each application but also the fragmented nature of the person's computer usage patterns. Next to the heatmap is a bar that displays the total usage duration of the application. Because the panel displays information regardless of whether it is distractive or productive, we emphasized or de-emphasized the representation according to the framing condition.

As we mentioned earlier, an application's productivity label depends on the context of the use (e.g., work, play), which can be hard to automatically detect. Thus, we provided **Manual Editing** features in the information dashboard. TimeAware determines an application's default productivity label based on RescueTime's criteria. However, when the RescueTime's categorization is inaccurate, people can manually edit the productivity label at the application level (See Figure 3a), or at the log level (i.e., applying the change to a specific duration of time; See Figure 3b). Assuming that the explicit labeling on a specific duration is more deliberate than the activity level assignment, a time span labeled with log-level editing is not affected by app-level editing.

Implementation
Using RescueTime's API, we collected the names of active applications (and the domain names if a web browser is active) and the duration of usage. Every activity is accumulated into a 5-minute-sized bin. We periodically dumped RescueTime data into our server and used the data to implement TimeAware-specific functionalities such as editing productivity labels.

FIELD DEPLOYMENT STUDY
We performed a field deployment study with a between-subjects design to evaluate the effects of framing by comparing the two versions of TimeAware interface. The study was conducted in Seoul, South Korea.

Participants
We advertised the study on Facebook, university mailing lists, and a campus recruiting website. Among the 41 people who filled out the screener, 24 people met our inclusion criteria: (1) Windows or Mac users (TimeAware only supports Windows and Mac OS); (2) not an undergraduate student; (3) use computer for more than 3 hours a day; (4) not taking a vacation for more than 3 consecutive days and not for more than 7 days in total during the study period; (5) have no experience using any automated productivity tracking systems; (6) have administrator's right to install software on their work computer; and (7) interested in self-tracking and have motivation for enhancing productivity. Among the 24 participants, 62.5% were male (n = 15), and their ages ranged from 22 to 38 (M = 27.88). Eighteen participants were full-time graduate students, three were full-time office workers, and three were freelancers.

Participants were randomly assigned to one of the two conditions: positive framing (PF) condition (n = 12, 5 females, 8 graduate students) or negative framing (NF) condition (n = 12, 4 females, 10 graduate students). Participants in the PF condition used the productivity-emphasized version of TimeAware interface, and those in the NF condition used the distraction-emphasized version. We offered 80,000 KRW (about 70 USD) to compensate participants in appreciation for their time.

Procedure
We organized our study procedure in three main components: a tutorial session, deployment of TimeAware, and pre- and post-study questionnaires.

Tutorial Session
Participants in a small group (3–5 people) who were assigned to the same condition attended an hour-long tutorial session. They were instructed to install TimeAware and RescueTime clients on their main work computers on which they want to monitor their activities. Most participants brought their own work laptop and installed TimeAware and RescueTime clients during the tutorial session. We stressed that study compensation is not tied to their use of TimeAware nor their productivity improvement. During the deployment, participants were not allowed to visit RescueTime's website to receive separate feedback.

Deployment
Right after the tutorial session, participants were instructed to use TimeAware for 8 weeks. To measure the participants' baseline activity levels, the widget and information dashboard were hidden while TimeAware was running in the background for 2 weeks (***Baseline*** period, 10 working days). For the following 4 weeks, participants were instructed to freely use TimeAware (***Intervention*** period, 20 working days) with the widget and the dashboard activated. Because notifications could affect one's engagement [3], we did not contact participants during this period. After the Intervention period was over (i.e., at the

end of week 6), we removed the feedback, deactivating the widget and dashboard. We, however, continued to track application usages for two more weeks to see the effect of withdrawing the feedback (***Withdrawal*** period, 10 working days).

Pre- and Post-study Questionnaires

Participants were asked to fill out a questionnaire before the Baseline period (**pre-study**) and after the Intervention period (**post-study**). The **pre-study** questionnaire contained short-answer questions about participants' estimates on their productivity level and computer usage duration and patterns. The **post-study** questionnaire contained the same questions from the pre-study questionnaire, but also included additional open-ended questions to gather qualitative feedback on self-reflection and behavior change (e.g., What did you learn using TimeAware?), and ways to improve the TimeAware system.

Dataset and Analysis

Participants installed TimeAware on one or more devices: Fourteen participants installed TimeAware on one device; seven participants on two devices; and three participants on three devices. Over the course of 8 weeks, excluding weekends, we collected 4,874 hours of computer usage data (5.2 hours per user per day) captured from participants' computers. With the original labels before editing, sum of productive, neutral, and distracted duration was 2505, 1383, and 986 hours, respectively. The durations changed after manual editing to 2890, 1092, and 892 hours, respectively. From these logs, we extracted a set of time-series data including trends in *productive rate*, *productive duration*, *distracted duration*, and *computer usage duration*. We used *productive rate* in NF condition to make it comparable with the PF condition. In addition, we also captured participants' interaction with the widget and dashboard (e.g., access count and duration) and editing actions, which we refer to as the *usage logs.* We excluded weekends and holidays from our analysis because the main goal of using TimeAware was to improve the work productivity.

To analyze the change of productive rate over time, we used *mixed-effects models* against time because these models can handle unbalanced data with repeated measures from the same participant [37]. We used each participant's daily *productive rate* as a data point ($N = 890$); we excluded days with no computer usage (e.g., business trip) because that days cannot yield a score. The productive rate was transformed into logit scales following guidelines in [40] to handle the heterogeneity of residual variance and the boundedness of the outcome variable [15]. Data points were weighted by total usage duration of the day because the productive rates from short usage durations are meaningless. After testing various controlling variables—*age*, *gender*, and *elapsed days* were not significant and thus excluded—that could affect productivity, we used intercept as a random effect and period and group as fixed effects.

We used *t*-test for other comparisons (e.g., comparing participants' behaviors in the PF and NF conditions during the ***Intervention*** period). We however note that our small sample size ($N = 24$) raises a possibility of Type II error when using a *t*-test for comparing engagement level.

We digitized all the qualitative feedback from the two questionnaires and analyzed the texts to identify common themes regarding self-awareness and self-reflection, receptiveness to the TimeAware feedback, and recommendations for improving TimeAware. We used Li et al.'s two phases of self-reflection [25]—***Discovery*** and ***Maintenance***—for deductive coding as well as bottom-up thematic analysis to identify emerging themes.

RESULT

We explore the results of our study in three parts: (1) effects of TimeAware on productivity, (2) engagement in TimeAware, and (3) self-awareness and self-reflection.

Effects of TimeAware on Productivity

In this section, we report on changes in participants' productivity over time comparing the PF and NF conditions. Note that we used the productivity data with the participants' manual editing applied for the analysis.

Productivity Change Between Periods

Using mixed-effects models, we found a significant interaction between Framing and Period on the outcome variable (logit-transformed productive rate), $p<0.0001$. We conducted post-hoc comparisons using Holm-Bonferroni correction. Table 2 shows pairwise comparisons among the three periods in each condition. Positive effect size means that productive rate increased after the first period. Participants in the NF condition showed a significant improvement in productive rate during the Intervention (IV) period compared to the Baseline (BL) period. Their productive rate significantly declined after TimeAware's feedback was removed during the Withdrawal (WD) period (Figure 4, right). We did not observe this effect in the PF condition (Figure 4, left).

Passage of time did not have a significant effect as we excluded it from fixed effects in the model ($p = 0.75$ in Maximum-likelihood test). In other words, NF condition's productive rate was not a gradual change over time, but increased immediately after participants started to receive TimeAware feedback (Figure 4, right).

Comparison		Effect Size	*t*-value	*p*-value
PF	BL vs IV	-0.125	-1.00	0.93
	IV vs WD	-0.066	-0.527	0.72
	BL vs WD	-0.191	-1.336	0.93
NF	BL vs IV	0.634	4.522	0.0000***
	IV vs WD	-0.473	-3.433	0.003**
	BL vs WD	0.161	1.013	0.93

***$p<0.001$; **$p<0.01$; *$p<0.05$

Table 2. Summary of statistical differences showing pairwise comparisons among the three periods in each condition.

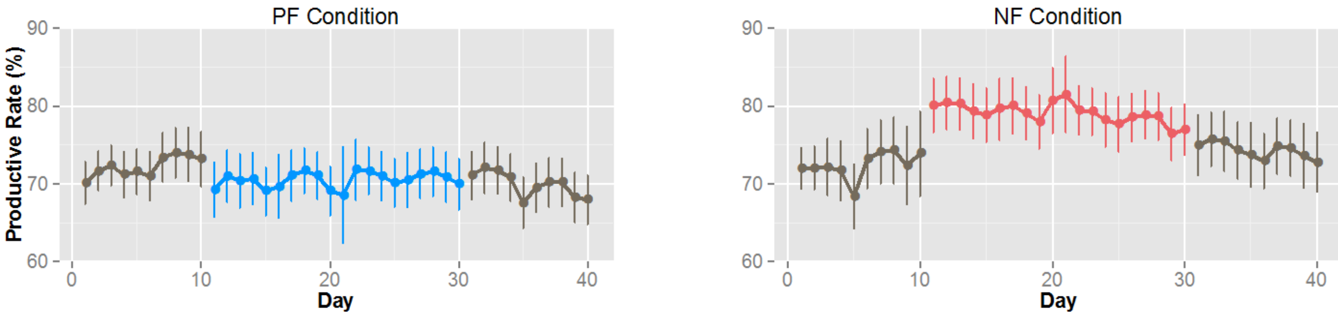


Figure 4. Changes in productive rate (%) as predicted by the mixed model for the positive (left) and negative (right) conditions. The colored lines represent the *Intervention* period of TimeAware. Only NF participants (right) show significant improvement during the *Intervention* period. Note that the model used here contains additional predictors (elapsed days and days of week) to reveal daily temporal trends in the figure.

Engagement in TimeAware

To assess people's engagement with ambient widget and information dashboard, we analyzed the *usage logs* and qualitative feedback from the questionnaire.

Widget and Dashboard Usage Behavior

We examined participants' daily TimeAware usage with three measurements: the number of times a person clicked on the widget to see the expanded view (i.e., widget expansion count), dashboard access count, and dashboard usage duration. We used the widget expansion count as a proxy for the number of eye fixations on the ambient widget because it was not feasible to accurately measure how often participants glanced at the ambient widget.

The daily usage did not differ between the two conditions (Table 3). Not surprisingly, the engagement peaked during the first few days after the intervention started and plateaued over time. However, participants in both conditions checked on the expanded view more than twice per day, and accessed the information dashboard at least once per day, longer than a minute throughout the study period. Given the low RescueTime website usage duration reported in prior work (M = 4.68 seconds / day, SD = 12.03) [12], our result shows the ambient widget's promising effect for enhancing engagement with information.

In the post-study questionnaire, we asked how many times participants glanced at the widget per day. Among the 23 participants who answered, sixteen (69.5%) reported 6–20 times per day, five (21.7%) reported 21–50 times per day, and the remaining two (8.7%) reported at most five times per day. Participants in both conditions mentioned that they consistently checked the ambient widget to maintain a certain level of productivity. PF-3 mentioned, "*When the productivity dropped below my threshold, I became alert and tried harder to stay focused.*" Similarly, NF-11 mentioned, "*Checking on my distracted duration helped me move away from distractions and carry on with my productive work.*"

Engagement	*t*-test		Mean	SD
Widget Expansion Count (times)	$t(21.99) = -.42, p = .68$	PF	2.31	1.73
		NF	2.60	1.76
Dashboard Access Count (times)	$t(11.63) = -.83, p = .42$	PF	1.14	0.74
		NF	2.20	4.39
Dashboard Usage Duration (sec)	$t(19.44) = -.27, p = .79$	PF	61.59	37.05
		NF	66.69	54.21

Table 3. Summary of quantitative engagement measures.

Productivity Editing Behavior

In TimeAware's information dashboard, we offered two editing features—app-level productivity editing and log-level productivity editing. We examined how often participants edit the initial productivity categorization.

In terms of app-level editing, participants in the NF condition (M = 10.92, SD = 8.08) edited more than participants in the PF condition (M = 6.67, SD = 7.06), but the difference was not significant, $t(21.61) = -1.37$, $p = .18$. Similarly with the log-level editing, participants in the NF condition (M = 7.33, SD = 11.06) edited more than participants in the PF condition (M = 1.75, SD = 7.06), but the difference was not significant, $t(16.55) = -1.55$, $p = .13$.

Next, we examined the directional changes of the edits in terms of the edit count. In Figure 5, we visualized the changes in app-level productivity for each group. It shows that participants in both conditions actively edited the *neutral* activities. The majority of the neutral activities were changed to productive activities, which was more often observed in the NF condition. This observation implies that participants might have abused the editing feature to intentionally raise their productivity. Thus, to assess the impact of editing on productive rate during the Intervention period, we conducted a paired *t*-test between the average

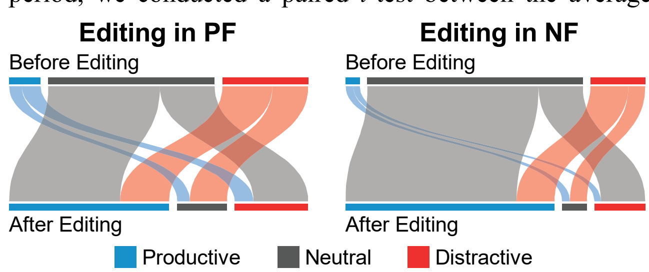


Figure 5. Directional changes of productivity labels with the edited activities by group with app-level editing feature. The length of each bar indicates the number of edited activities with corresponding productivity.

Stage of Reflection	Topics	Example Quotes
Discovery	Application usage patterns	"*I always thought that I used Matlab the most, but the dashboard showed that I spent most of the time reading papers using Acrobat Reader or using Microsoft Word, which was quite new to me.*" [PF-5] "*I found it striking to see applications I thought I only looked in a few times were highly ranked, and those applications accounted for a sizeable chunk of the whole usage duration.*" [PF-6] "*I was curious how much music I listen to, but it* [music player] *wasn't captured because it was in the background* [inactive window]. *So I turned it on another laptop* [with TimeAware] *and tracked it* [music player] *there*." [PF-3]
	Factors affecting productivity	"*Productivity rate was high on the days that I really needed to get my work done, but on the days that I go out, it was generally low.*" [PF-10] "*Working in multiple places, I realized that my productivity differed from place to place.*" [PF-2]
Maintenance	Strategies for improving productive activities	"*I concentrated on my task and ignored Kakaotalk* [Messenger] *and phone calls when working*." [PF-12] "*I was consistently checking my productivity and was stressed at the fact that I had to increase my productivity.*" [NF-9]
	Strategies for reducing distracting activities	"*When the distraction gets too high, I stopped using distracting apps.*" [NF-1] "*I made a habit of turning off the computer when I don't use it. Consequently, I was able to spend my time more constructively.*" [NF-2]. "*Removed a game on my computer.*" [PF-1]

Table 4. Categories of self-reflection and example quotes.

productive rate with the editing applied and without the editing applied (pure RescueTime's automatic ratings). For the PF condition, average productive rate significantly decreased after the editing, $t(11) = -2.27$, $p = .03$. For the NF group, average productive rate did not differ, $t(11) = -.97$, $p = .35$. This result suggests that participants did not abuse the editing feature.

Although we provided sophisticated ways to assign productivity level for specific time spans, participants did not actively use the log-level editing feature. Among the 24 participants, only 7 participants (2 in PF and 5 in NF) used the log-level editing feature.

In the post-study questionnaire, participants described reasons for the low usage of the editing features—***web access burden*** (e.g., "*The stepwise process of expanding the widget, clicking the button* [dashboard shortcut], *and checking the dashboard seemed somewhat tedious to me.*" [PF-9]), ***subjective nature of the productivity labeling*** (e.g., "*Although TimeAware can accurately track application usage duration, it's quite subjective to determine if an application is productive or distractive. After noticing this* [subjectiveness], *I found myself not using it as much as before*" [PF-2]), and the ***mental burden*** due to retrospective editing (e.g., "*I couldn't remember the exact time span for log-level editing because the text 'PuTTY'* [activity name] *was the only available information, so I couldn't figure out what I used it for*" [NF-10]).

Self-Awareness and Self-Reflection

To assess the effect of TimeAware on self-awareness, we analyzed the pre-study and post-study questionnaires on self-estimation of productive rate (i.e., "How much portion of your computer usage time was spent on productive [distractive] tasks?") and compared to TimeAware's productive rate. Comparing the pre-study self-estimation with TimeAware Baseline data, participants in both conditions *underestimated* their productive rate before the Intervention period—the difference was 11.71% for the PF condition (SD = 7.66%) and 11.73% for the NF condition (SD = 9.80%). However, the difference between self-estimation and TimeAware data significantly decreased after the Intervention period for both conditions—although participants still underestimated their productive rate, the difference was 5.53% for the PF condition (SD = 3.78%), $t(11) = 2.74$, $p = .01$; and 4.73% for the NF condition (SD = 3.09%), $t(10) = 2.83$, $p = .01$. The significant reduced gaps imply the increased awareness of individuals' personal productivity.

We analyzed the qualitative feedback to examine what participants learned during the self-reflection with TimeAware. Due to the qualitative nature of the data, we did not seek measurable difference between the conditions. We categorized our data according to Li et al.'s two phases of self-reflection [25]—***Discovery*** and ***Maintenance***—and then further identified emerging themes via deductive coding (Table 4). Self-reflection regarding *Discovery* of new findings included *application usage pattern* and *factors affecting productivity*. Self-reflection regarding *Maintenance* behaviors included *strategies for improving productive activities* and *strategies for reducing distracting activities.*

Using TimeAware, participants gained new knowledge on their computer usage. The majority of *discovery* topics include surprisingly low total computer usage duration, fragmentation of productive hours, and sizable use of distractive applications. Some participants discovered that factors that were not captured by TimeAware are actually very important in analyzing their productivity data. For example, PF-2 remarked, "*Working in multiple places, I realized that my productivity differed from place to place.*"

From the ambient widget, participants could see their total computer usage duration along with productive duration (or distracted duration) (Figure 1). It appears that this feedback served as a guide to decide how much more distraction is tolerable for the rest of the day. PF-7 mentioned, "*I checked my usage duration of the web browser and tried to stop using it if my score* [productive rate] *got low.*" NF-10 also remarked, "*I also checked my computer usage duration. If I thought I didn't work much today, I overworked till night.*" Participants' responses suggest that they had an implicit standard of productivity goal either in terms of productive duration or productive rate.

Although the distraction-emphasized feedback was more effective in increasing productivity, many participants in the NF condition reported much more stress than those in the PF condition. NF participants reported that being aware of their distractive rate caused much stress. NF-5 remarked, "*I felt very uncomfortable and was stressed out when my distractive rate got higher than 50%. I then started working to make the score* [distractive rate] *go down.*" Similarly, NF-8 mentioned, "*It feels like something was suppressing me that my distracted duration should be low.*"

DISCUSSION

In this section, we describe the lessons we learned and implications of productivity monitoring system design.

Framing Effects in Productivity Monitoring

Although the results showed that the engagement (measured in terms of the feedback access frequency) of the two conditions were similar, the productivity was improved only in the NF condition. We suspect that framing was effective, making people perceive the two types of feedback differently and establish a different level of personal threshold. In other words, 70% productive rate may be perceived as a decent level of productivity although 30% distractive rate may be perceived as not as productive even though the two are semantically equivalent. The perception of low level of achievement might have pushed the participants in the NF condition work harder such that they can improve their productivity; however, we should consider cultural differences as our study sample—South Korean workers—limits the generalizability of the findings.

In addition, the improved productivity of NF participants dropped immediately after the feedback was withdrawn. From the result, we can conclude that the distraction-emphasized feedback can help people improve their productive rate, but this behavior change might not be sustained when the feedback is withdrawn. Therefore, showing one's distraction level on the ambient widget could be a good way to boost their productivity at the beginning, but is not enough to make behavior changes that last. We envision that actionable guidance on *how* to improve one's productivity can be provided in addition to TimeAware's current feedback on the person's productivity status. For example, the next generation of TimeAware might be able to deliver more specific suggestions on individuals' ideal working environment including when and how they can boost the productivity based on their historical data, thereby helping them create and identify productive working environment and healthy working habits.

Backfire: Dark Side of Productivity Monitoring

As participants stated in the post-study questionnaire, projecting a person's productivity data—especially when the data is negatively framed—on the menu bar or taskbar could backfire, stressing out the person. We suspect that participants perceived the distractive rate as a *punishment,* and thus were stressed (as opposed to perceiving productive rate as a *reward*). To make matters worse, annoying feedback could interrupt the flow of work and negatively impact the person's productivity. A compromised way to project this effective yet stressful feedback would be to reduce the frequency of feedback exposure. For example, push notifications can be shown every hour, or only when the person's productivity level goes below their personal threshold. In addition, the push notification popups can automatically disappear after a few seconds. Designing the feedback that maximizes performance and induces lower stress warrants further research.

Challenges in Productivity Editing

The two manual editing features—app-level and log-level editing—were not frequently used because of web access burden and subjective nature of productivity. In addition, participants used the editing features in a retrospective manner, making it more challenging to remember the exact purposes of application usages. We also found that the applications that need the most editing are the ones that were used for multiple purposes (e.g., messenger, PuTTY, Google) as NF-10 described: "*I mainly use PuTTY* [terminal emulator] *on my computer, both for programming* [work] *and chatting* [leisure]. *I had set the productivity level of PuTTY as 'neutral' because it was tiring to assign temporal productivity* [log-level editing] *every time I use it. However, because the productivity level of PuTTY was set to 'neutral,' my productivity converged to 50% as I work harder and it was quite depressing.*"

To ease the burden associated with the editing, it might be helpful to provide a cue to remember the nature of the activities when showing the application usage pattern. For example, the system could show a screen capture of the application as a thumbnail on demand, providing the context as a hint.

Configuring When to Track

We tracked people's entire computer usage (i.e., as long as the computer was turned on). However, we learned that people usually use the same device both for work and leisure and that they often have more than one device used for work purposes (Ten participants installed TimeAware on more than one device). Participants did not like that taking a purposeful break from work and having fun (e.g., playing games, web browsing) was captured as unproductive activities and thought that it should not be included in the productive rate calculation. Similarly, they did not like to track their weekend computer usage, which would often score low productive rate (although we took the weekend data out from our analysis). Although NF participants were more sensitive to this issue, regardless of the framing condition, participants reported that they would prefer to track application usage only during a specific period (e.g., working hours) of the day or specific days of the week (e.g., weekdays). For example, PF-3 stated, "*Because TimeAware tracked my activities even on weekends, my weekly average productivity was always lower than my expectations. It was disappointing.*"

One approach to address this issue would be to enable people to configure when to track—let them turn on and off the tracking system or configure the tracking period (e.g., from 9am to 5pm, weekdays only). Another approach would be to allow people to erase or ignore tracking data during their reflection. Investigating the effectiveness of different approaches is an open research question.

Integrating Multiple Devices

All of our study participants owned mobile devices as well as computers, and they often used both devices together, frequently switching between the two. We observed an interesting context switching behavior. Some participants used their mobile phone for distracting activities to avoid them being tracked by TimeAware: NF-3 stated, "*Until now, I frequently surfed the web and read online comics to take a break, but ever since I became aware of my distractive rate, I purposely tried to conceal the distracted duration by using my other devices which were not tracked* [by TimeAware]." Given that mobile devices are prevalent and source of distraction [7], it would be important for future productivity monitoring systems to track data from various devices—such as smartphones, desktops, laptops, and smart watches—providing a more comprehensive view of a person's productivity. We note that many participants expressed the need to track mobile usage[1] in addition to the desktop or laptop usage, and that it would be interesting to explore ways to improve people's productivity more holistically in the context of multi-device ecosystem. Tracking multiple devices would allow us to apply machine learning techniques to the multi-faceted data collected from multiple devices such that we can understand people's true computer usage behaviors.

[1] RescueTime recently began mobile app tracking service for Android 5. However, due to various errors and instability, we did not employ this service in our study.

Conducting Framing Studies in the HCI Context

One of the contributions of this study is the exploration of the framing effects in a field deployment study. Most framing studies including early psychological research as well as HCI research [8,11] were conducted using a hypothetical scenario ([24] provides an extensive review on early psychological framing studies). However, the effect of framing in a real-world situation could be different from that of participants' imagination, thus raises the importance of conducting ecologically valid framing studies. As such, we propose the following steps to conduct an ecologically valid framing study based on our experience: (1) ethical considerations should precede regarding when and where to test framing; (2) identify the framing type (e.g., valence: positive vs. negative) and dependent variables; (3) find semantically equivalent (visual, non-visual) representations; and (4) conduct a field deployment to identify more effective framing.

CONCLUSION

In this paper, we presented the design and evaluations of TimeAware, a self-monitoring system for capturing and reflecting on personal computer usage behavior. Our goal was to help people enhance personal productivity by providing effective feedback. We designed two versions of TimeAware, each with a different framing setting—one emphasizing productive activities (positive framing, PF) and the other emphasizing distracting activities (negative framing, NF), and conducted an eight-week deployment study (N = 24). We found a significant effect of framing on participants' productivity: only participants in the NF condition improved their productivity when the feedback was displayed, but this effect disappeared when we removed the feedback. However, participants in the NF condition reported that looking at the negatively-framed data was stressful. Specific areas for future research include designing for lasting behavior change, making the productivity editing easy, letting people configure when they want to track, and enabling productivity tracking in a multi-device ecosystem. We also demonstrated the importance of running an ecologically valid framing study and proposed guidelines for conducting such a study. Our work contributes to the growing body of literature in personal informatics and self-monitoring with the focus on improving personal productivity. We hope this study can help others working in the field get insights on ways to better design and deliver personal feedback.

ACKNOWLEDGMENTS

We thank our participants for their time and candor. Our colleagues also deserve special thanks for active discussion and constructive feedbacks. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. NRF-2014R1A2A2A03006998). The ICT at Seoul National University provided the research facilities for this study.

REFERENCES

1. Attainr. 2015. Retrieved September 23, 2015 from https://attainr.com.
2. Beeminder. 2015. Retrieved September 23, 2015 from https://www.beeminder.com.
3. Frank Bentley and Konrad Tollmar. 2013. The power of mobile notifications to increase wellbeing logging behavior. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '13), 1095–1098.
4. Ann E. Blandford and Thomas R. G. Green. 2001. Group and individual time management tools: what you get is not what you need. *Personal and Ubiquitous Computing* 5, 4: 213–230.
5. Block distracting websites and apps on your Mac—Focus. 2015. Retrieved September 23, 2015 from https://heyfocus.com.
6. Rafael A. Calvo and Dorian Peters. 2014. Self-awareness and self-compassion. In *Positive Computing: Technology for Well-Being and Human Potential*. MIT Press, 155–172.
7. Luca Chittaro and Luca De Marco. 2004. Driver distraction caused by mobile devices: studying and reducing safety risks. In *Proceedings of the 1st International Workshop Mobile Technologies and Health: Benefits and Risks*, 1–19.
8. Eun Kyoung Choe, Jaeyeon Jung, Bongshin Lee, and Kristie Fisher. 2013. Nudging people away from privacy-invasive mobile apps through visual framing. In *Proceedings of INTERACT 2013 International Conferences* (INTERACT '13), 74–91.
9. Eun Kyoung Choe, Bongshin Lee, Matthew Kay, Wanda Pratt, and Julie A. Kientz. 2015. SleepTight: low-burden, self-monitoring technology for capturing and reflecting on sleep behaviors. In *Proceedings of the 17th international conference on Ubiquitous computing* (UbiComp '15), 121–132.
10. Eun Kyoung Choe, Nicole B. Lee, Bongshin Lee, Wanda Pratt, and Julie A. Kientz. 2014. Understanding quantified-selfers' practices in collecting and exploring personal data. In *Proceedings of the Annual ACM Conference on Human Factors in Computing Systems* (CHI '14), 1143–1152.
11. Eun Kyoung Choe, Bongshin Lee, Sean Munson, Wanda Pratt, and Julie A. Kientz. 2013. Persuasive performance feedback: the effect of framing on self-efficacy. In *Proceedings of AMIA Annual Symposium 2013* (AMIA '13), 825–33.
12. Emily I. M. Collins, Jon Bird, Anna L. Cox, and Daniel Harrison. 2014. Social networking use and RescueTime: the issue of engagement. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct Publication* (UbiComp '14 Adjunct), 687–690.
13. Sunny Consolvo, Jaeyeon Jung, Ben Greenstein, Pauline Powledge, Gabriel Maganis, and Daniel Avrahami. 2010. The Wi-Fi privacy ticker: improving awareness & control of personal information exposure on Wi-Fi. In *Proceedings of the 12th international conference on Ubiquitous computing* (UbiComp '10), 321–330.
14. Sunny Consolvo, Predrag Klasnja, David W. McDonald, Daniel Avrahami, Jon Froehlich, Louis LeGrand, Ryan Libby, Keith Mosher, and James A. Landay. 2008. Flowers or a robot army? Encouraging awareness & activity with personal, mobile displays. In *Proceedings of the 10th international conference on Ubiquitous computing* (UbiComp '08), 54–63.
15. Douglas Curran-Everett. 2013. Explorations in statistics: the analysis of ratios and normalized data. *Advances in physiology education* 37, 3: 213–9.
16. Anton N. Dragunov, Thomas G. Dietterich, Kevin Johnsrude, Matthew McLaughlin, Lida Li, and Jonathan L. Herlocker. 2005. TaskTracer: a desktop environment to support multi-tasking knowledge workers. In *Proceedings of the 10th international conference on Intelligent user interfaces* (IUI '05), 75–82.
17. Casey Dugan, Werner Geyer, Michael Muller, Abel N. Valente, Katherine James, Steve Levy, Li-Te Cheng, Elizabeth Daly, and Beth Brownholtz. 2012. "I'd never get out of this !?$%# office": redesigning time management for the enterprise. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '12), 1755–1764.
18. Fitbit. Retrieved September 23, 2015 from https://www.fitbit.com/app.
19. Victor M. González and Gloria Mark. 2004. "Constant, constant, multi-tasking craziness": managing multiple working spheres. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '04), 113–120.
20. Health App. Retrieved September 23, 2015 from http://www.apple.com/ios/health.
21. Alen E. Kazdin. 1974. Reactive self-monitoring: the effects of response desirability, goal setting, and feedback. *Journal of consulting and clinical psychology* 42, 5: 704–716.
22. KnowSelf. Retrieved September 23, 2015 from http://know-center.tugraz.at/en/knowself.
23. Matthew L. Lee and Anind K. Dey. 2014. Real-time feedback for improving medication taking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '14), 2259–2268.

24. Irwin P. Levin, Sandra L. Schneider, and Gary J. Gaeth. 1998. All frames are not created equal: A typology and critical analysis of framing effects. *Organisational Behaviour and Human Decision Processes* 76, 2: 149–188.
25. Ian Li, Anind K. Dey, and Jodi Forlizzi. 2011. Understanding my data, myself: supporting self-reflection with ubicomp technologies. In *Proceedings of the 13th international conference on Ubiquitous computing* (UbiComp '11), 405-414.
26. Isaac M. Lipkus and J. G. Hollands. 1999. The visual communication of risk. *Journal of the National Cancer Institute* 1999, 25: 149–163.
27. Robby MacDonell. Breaking down the RescueTime productivity score. 2013. Retrieved September 23, 2015 from http://blog.rescuetime.com/2013/03/09/breaking-down-the-rescuetime-productivity-score.
28. ManicTime. Retrieved September 23, 2015 from http://www.manictime.com.
29. Gloria Mark, Victor M. Gonzalez, and Justin Harris. 2005. No task left behind? Examining the nature of fragmented work. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '05) 321–330.
30. Theresa M. Marteau. 1989. Framing of information: its influence upon decisions of doctors and patients. *British Journal of Social Psychology* 28, 1: 89–94.
31. Akhil Mathur, Marc Van Den Broeck, Geert Vanderhulst, Afra Mashhadi, Fahim Kawsar, and Bell Laboratories. 2015. Tiny habits in the giant enterprise: understanding the dynamics of a quantified workplace. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (UbiComp '15), 577–588.
32. Tara Lynn Matthews, Jodi Forlizzi, and Stacie Rohrbach. 2006. Designing glanceable peripheral displays. UC Berkeley. Retrieved September 23, 2015 from http://www.eecs.berkeley.edu/Pubs/TechRpts/2006/EECS-2006-113.html.
33. Rosemery O. Nelson and Steven C. Hayes. 1981. Theoretical explanations for reactivity in self-monitoring. *Behavior Modification* 5, 1: 3–14.
34. Viktoria Pammer and Marina Bratic. 2013. Surprise, surprise: activity log based time analytics for time management. In *Extended Abstracts on Human Factors in Computing Systems* (CHI '13), 211-216.
35. Viktoria Pammer, Marina Bratic, Sandra Feyertag, and Nils Faltin. 2015. The Value of Self-tracking and the Added Value of Coaching in the Case of Improving Time Management. In *Proceedings of the 10th European Conference on Technology Enhanced Learning* (ECTEL '15), 467–472.
36. Viktoria Pammer, Stefan Edler, and Hermann Stern. 2012. Visualising the fragmentation of knowledge work. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design* (NordiCHI '12), 779–780.
37. Jose Pinheiro and Douglas Bates. 2006. Mixed-effects models in S and S-PLUS. *Springer Science &Business Media*.
38. RescueTime. Retrieved September 23, 2015 from http://rescuetime.com.
39. SLife. Retrieved September 23, 2015 from http://www.slifeweb.com.
40. Michael Smithson and Jay Verkuilen. 2006. A Better Lemon Squeezer? Maximum-likelihood Regression with Beta-distributed Dependent Variables. *Psychological methods* 11, 1: 54–71.
41. TeamViz. Retrieved September 23, 2015 from http://www.teamviz.com.
42. Time Tracking Software with Screenshots and Activity Levels. Retrieved September 23, 2015 from https://hubstaff.com.
43. Amos Tversky and Daniel Kahneman. 1981. The framing of decisions and the psychology of choice. *Science* 211, 4481: 453–458.
44. Writer for iOS | iA. Retrieved September 23, 2015 from https://ia.net/writer/ios.
45. Dezhi Wu and Marilyn Tremaine. 2004. Knowledge worker adoption of time management tools: satisfaction and perceived effectiveness. In *Proceedings of the 10th Americas Conference on Information Systems* (AMCIS '04), Paper 433.
46. Yan Zhou, Jon Bird, Anna L. Cox, and Duncan Brumby. 2013. Estimating usage can reduce the stress of social networking. *Personal Informatics in the Wild Workshop*, CHI'13. http://discovery.ucl.ac.uk/1412160/1/zhouBirdCoxBrumbySubmittededited.pdf.