The Cost of Interrupted Work: More Speed and Stress

Gloria Mark
Department of Informatics
University of California, Irvine
Irvine, CA, U.S.A. 92697
gmark@ics.uci.edu

Daniela Gudith and Ulrich Klocke
Institute of Psychology
Humboldt University
Berlin, Germany
daniela.gudith@gmail.com, klocke@rz.hu-berlin.de

ABSTRACT
We performed an empirical study to investigate whether the context of interruptions makes a difference. We found that context does not make a difference but surprisingly, people completed interrupted tasks in less time with no difference in quality. Our data suggests that people compensate for interruptions by working faster, but this comes at a price: experiencing more stress, higher frustration, time pressure and effort. Individual differences exist in the management of interruptions: personality measures of openness to experience and need for personal structure predict disruption costs of interruptions. We discuss implications for how system design can support interrupted work.

Author Keywords
Multi-tasking, interruptions, experiment, context

ACM Classification Keywords
H5. Information interfaces and presentation: H.5.2 User Interfaces: Theory and Methods

INTRODUCTION

The role of interruptions in the workplace has begun to receive a lot of attention in HCI in the last few years. Empirical studies have focused on identifying the extent of interruptions and how they affect tasks [6], the recovery of tasks after an interruption [3, 9], and timing of interruptions, e.g. [1]. Spurred on by these field and laboratory studies systems have been developed to help people manage interruptions (e.g. [3]. Yet as more studies in multi-tasking and interruptions emerge, so do conflicting ideas on how interruptions might affect work.

INTERRUPTIONS AND CONTEXT

In a field study of managers, Hudson et al. [8] reported that interruptions might be beneficial. Other lab and diary study results have described them as detrimental [1, 3]. Mark et al. [10] on the other hand discovered that interruption effects might be more nuanced: in a field study their informants reported that interruptions of the same context as the current task were beneficial, whereas interruptions of a different context than the current task were disruptive.

We decided to investigate the different perspectives raised by these studies. Interruptions during the course of the workday might be of the same context as the current task at-hand or they might be random, related to other topics. If indeed interruptions as the same context as the task at-hand are beneficial, then this has important implications for system design. For example, systems might be designed to help colleagues gear their interruptions to others so as match the context of their tasks.

We were interested in measuring the *disruption cost* of interruptions. One type of a disruption cost is the additional time to reorient back to an interrupted task after the interruption is handled. These previous studies introduce conflicting notions as to whether the interruption context is related to a disruption cost. For example, one might be working on a paper and be interrupted by a completely different topic, such as a question about a budget. If an interruption has a different context than the current task at-hand, this could introduce a disruption cost as it involves a cognitive shift of context to attend to the interruption, and then one must reorient back to attend to the interrupted task. On the other hand, one might be interrupted by a question that concerns the same context as the paper one is working on. This might be beneficial but if the context of the interruption and primary task are similar, this could lead to interference with the primary task [5] and in this way may introduce a disruption cost. A third possibility is that the interruption context may not matter. Perhaps *any* discontinuity in the task creates a disruption cost for work.

Disruption costs of interruptions can also involve other factors such as stress. A second question that we asked is how interrupted work might change the state of a user; is interrupted work significantly more stressful or viewed as more effortful than noninterrupted work? Though we might intuitively believe it does, it remains an open question.

Disruption costs might also be mitigated by personality factors. We expected that a) the more open one is to new experiences (and thus better able to handle new tasks), and b) the less need one has for personal structure (and thus is more flexible), the lower the disruption cost would be. We

reasoned that these measures would indicate if some can adapt quicker than others to a new situation (the interruption) and then reorient back to the task. Such individual differences could be used to design customization in a system to adapt to workplace preferences.

The goal of our experiment was to better understand how interruptions affect work: its patterns, user strategies, and disruption costs in order to inform system design to help people manage interrupted work.

EXPERIMENTAL DESIGN AND PROCEDURE

A 3x2 factorial experimental design was used. The within-subject factor was interruption context with three levels: no interruption (**B**, baseline), same-context interruption (**S**), and different-context interruption (**D**). The order of interruption context was fully counter-balanced. Because interruptions can come through various sources, we checked whether media might affect the context of interruptions. The between-subjects factor was media type: subjects were interrupted with telephone or IM.

Forty-eight subjects participated. 81% of subjects were German university students with a mean age of 26; all had a German high school degree (roughly equivalent to one year in a U.S. university). Most majored in psychology (27.1%), medicine/sciences (16.7%), or mathematics/engineering/IT (10.4%). Fifteen were males and 33 were females. All but one participant had been using email for at least two years. Half of the subjects had been using IM for at least one year or longer; 29.2% had never used IM before. Subjects were paid 12 Euros for their participation.

Experimental task. We simulated an office environment in the lab and chose an email task, common in information work. Subjects read instructions that they were to play a role as a human resource manager at a medium-sized craft supply company. They had just returned from vacation and were carefully instructed in all conditions (and were given incentives) to answer all emails in their inbox as quickly, correctly and politely as possible. Subjects were given a simple fact sheet to use in answering the emails, e.g. staff job levels and education, overtime hours, etc. For each interruption type condition, subjects had to answer 12 emails. Based on pilot studies, the emails in the folders were equally distributed by topics. We confirmed in the pilot study that the email questions in all conditions were equally demanding. The content of the emails consisted of questions from various people, e.g. "Where do I get more information on internships in your company?" Subjects first performed one trial email with the experimenter to familiarize them with the equipment and the task.

Interruptions. Subjects were told that their "supervisor" (the experimenter who sat in another room) would contact them periodically and ask questions. No interruptions were given in the baseline condition (B). In the S condition participants were interrupted by the supervisor with questions concerning the human resource context (e.g. "How many employees, including you, are in the department today?"). These questions were of the same context as the primary email task, i.e. human resources. In the D condition subjects were interrupted by the supervisor with questions about a different topic not related to the context of human resources. These interruptions were on random topics, such as the upcoming company party (e.g. "How many hot dogs do we need for 240 employees?"). These types of random interruptions were designed to simulate the types of interruptions one might expect in real office work. Pilot studies confirmed that questions in the S and D conditions were judged by subjects to be same and different contexts, respectively, to the human resource scheme.

Subjects were instructed to attend to interruptions immediately, i.e. to pick up the telephone or attend to the IM window. Interruption frequency was set to two minutes, based on observations from [6] and the pilot experiments. During the experiment, the experimenter adjusted the length of time for interruptions to try to make the times as equal as possible across conditions.

Dependent variables: Our primary variable of interest was the total time to perform the task. The total time needed to complete each block of emails and the time spent on interruptions were manually recorded. The time to perform the task was computed as [total time to perform task – time spent on interruptions]. If the time to perform the task was higher with an interruption, then this could indicate that extra time was needed to perform the task after an interruption. We expected less politeness and more errors in email messages that were disrupted. A politeness metric was computed by assigning points for the use of standard greeting/closing phrases and polite words (e.g. Dear Mr.; sincerely yours, please, thank you). Errors were measured as spelling errors, typos or others (e.g. misspelled names).

Subjective Workload was measured by a modified NASA Task Load Index (TLX) [7], used by [1]. We added a stress measure to the six given rating scales: mental and physical demands, performance, and temporal demand, which we changed to time pressure, effort, and frustration. Subjects rated these factors on the standard NASA 20-point scale.

Personality measures. 'Openness to experience' was measured as part of the NEO Personality Inventory Revised using the Openness to Actions subscale of the German translation [11] since these items best fit our topic. Ten items of the Personal Need for Structure (PNS) subscale from the Need for Cognitive Closure/Personal Need for Structure inventory were used [2], (e.g. "I do not like it at all to change my plans at the last minute.").

Procedure. The experiment was conducted at a university lab and took about 1½ hours. Subjects were randomly assigned to media type and to the interruption order (the treatments were counterbalanced with six permutations). Subjects first filled out a questionnaire to obtain personality measure scores and demographic and computer experience

information. Subjects then performed the task for each interruption context. After completing each block of emails, subjects assessed the workload measures in a questionnaire.

RESULTS

Our first research question was whether the context of the interruption matters. We performed repeated measures analyses with media type as a between-subjects factor. Means are shown in Table 1. For Time to Perform Task there was a significant difference between interruption types ($F(2,77.98) = 3.36$, $p<.05$). Surprisingly, paired contrasts showed that subjects took the longest time in the baseline condition to perform the task (B vs. S: $F(1,46=4.22)$, $p<.05$; B vs. D: $F(1,46=4.13)$, $p<.05$). There was no significant difference between S and D contexts.

	Time to perform task* (minutes)	Avg. number errors in emails	Length of email message* (avg. # words)	Politeness of email messages
Baseline (no interruption)	22.77 (7.60)	1.94 (.91)	31.49 (8.1)	28.98 (5.37)
Same context interruption	20.31 (5.94)	1.93 (.88)	29.17 (7.02)	28.69 (5.89)
Different context	20.60 (4.93)	1.84 (.92)	30.16 (7.18)	28.90 (6.30)

Table 1. Mean measures of task performance (s.d.). *=p<.05.

There was no significant difference of media type, i.e. whether subjects were interrupted by telephone or IM ($F(1,92)=.20$, $p<.66$). There was no significant interaction of media type and interruption type on time to perform task.

We found there was no significant difference in the number of errors that were made across interruption types ($F(2,92)=1.70$, $p<.19$). We compared our politeness metric and found no significant difference across conditions ($F(2,92)=.10$, $p<.91$). For both variables, we found no interaction with media type.

It was possible that the reason that it took longer to do the task in the uninterrupted condition is because people wrote more. We examined the length of the email messages across conditions. A repeated measures analysis on the average number of words per email (Table 1) showed that there was a significant difference among conditions ($F(2,92)=3.34$, $p<.04$). Email messages were longest in the B condition, with no interruptions (B vs. S: $F(1,46=5.57)$, $p<.05$; B vs. D: $F(1,46=2.20)$, $p<.15$). There was no significant interaction of media type with number of words.

	β	t	p
Openness to experience	-.363	-2.22	.03
Need for personal structure	-.346	-2.13	.04

Table 2. Regression coefficients for personality measures.

As a control, we checked whether the actual time spent on interruptions was different in the same or different contexts as longer interruptions could introduce a higher disruption cost. We found no significant difference in interruption length (difference=.62 min, sd=2.8), $t(47)=1.52$, $p<.14$).

Personality variables. We next examined to what extent the measures of openness to experience, and need for personal structure are predictive of the amount of time one needs to complete a task that is constantly interrupted. Our primary variable of interest was disruption cost and since there was no difference in time to complete the task between same/different interruption contexts, we combined the times for these conditions to form a single dependent measure (time to complete the task - time spent on interruptions). A stepwise regression analysis using the personality measures as predictors showed that both 'openness to actions and 'need for personal structure' are significant predictors of the time to complete an interrupted task ($F(2,47)=3.41$, $p<.04$), $R^2=.14$). There is an inverse relationship: the higher one scores on 'openness to experience' and 'need for personal structure', the quicker it takes to complete an interrupted task (Table 2).

	Mental work-load*	Stress **	Frus-tration **	Time pressure**	Effort **
Baseline (no interruption)	10.02 (3.90)	6.92 (3.85)	4.73 (2.93)	11.02 (4.57)	9.50 (3.38)
Same context interruption	10.83 (3.96)	9.46 (3.97)	6.63 (4.19)	12.69 (4.45)	11.04 (3.78)
Different context	11.50 (3.55)	9.13 (4.10)	6.48 (4.45)	12.17 (4.26)	11.52 (3.31)

Table 3. Mean (s.d.) workload measures across interruption types. Scale is 1(low)-20 (high), *=p<.05, **p<.01.

Measures of workload. We tested the difference of the NASA mental workload measures (Table 3) across interruption type. A repeated measures analysis showed a significant difference ($F(2,92)=3.82$, $p<.03$). Workload was highest for the D context condition (D vs. B: $F(1,46=7.38)$, $p<.01$; D vs. S: $F(1,46=2.09)$, $p<.16$). We also found that stress was rated as significantly different across interruption type ($F(2,92)=12.15$, $p<.001$) and was highest for both interruption conditions (S vs. B: $F(1,46=20.32)$, $p<.001$; D vs. B: $F(1,46=14.94)$, $p<.001$). Level of frustration was also significantly different across interruption types ($F(2,92)$-5.21, $p<.007$), and highest in the interruption conditions (S vs. B: $F(1,46=7.88)$, $p<.01$; D vs. B: $F(1,46=7.55)$, $p<.01$). Time pressure was also rated significantly different across interruption types ($F(2,92)=4.71$, $p<.01$), with highest time pressure rated in the interruption conditions (S vs. B: $F(1,46=10.65)$, $p<.01$; D vs. B: $F(1,46=3.65)$, $p<.10$). The amount of effort invested in the task was also significantly different across interruption types ($F(2,79)=8.50$, $p<.001$), with most effort reported in the interruption conditions (S vs. B: $F(1,46=6.92)$, $p<.05$; D vs. B: $F(1,46=14.60)$, $p<.001$). There was no interaction with media type in any of these measures.

DISCUSSION AND CONCLUSIONS

Our results showed that *any* interruption introduces a change in work pattern and is not related to context per se. Our results differ from [5] who found similarity of cognitive processes of interruptions to a task were disruptive. We looked instead at similarity of the content of interruptions and a task. Also, [4] found that interruptions extremely consistent with the task were facilitating. Our interruption context shared the same topic as the main (email) task but unlike [4] the operations and details differed. Together, our study along with [10], who report participants' subjective views, show that interruptions that share a context with the main task may be *perceived* as being beneficial but the actual disruption cost is the same as with a different context.

Surprisingly our results show that interrupted work is performed faster. We offer an interpretation. When people are constantly interrupted, they develop a mode of working faster (and writing less) to compensate for the time they know they will lose by being interrupted. Yet working faster with interruptions has its cost: people in the interrupted conditions experienced a higher workload, more stress, higher frustration, more time pressure, and effort. So interrupted work may be done faster, but at a price.

Our results suggest that interruptions lead people to change not only work rhythms but also strategies and mental states. Another possibility is that interruptions do in fact lengthen the time to perform a task but that this extra time only occurs directly after the interruption when reorienting back to the task, and it can be compensated for by a faster and more stressful working style. More sophisticated measurements of working speed directly after an interruption must be done to test this. We found that the more open one is to experiences, the quicker one handled interrupted work and surprisingly, we found the same relation for those who score high on needing personal structure. Perhaps those who need personal structure are better able to manage their time when interrupted.

While laboratory studies are always subject to criticism of ecological validity, our task design was based on real fieldwork. We simulated office conditions using an email task, which is different from many laboratory studies of attention switching that use abstract tasks. The lab environment enabled us to isolate variables of interest.

Our results have implications for system design. A certain amount of interruptions may be tolerable because people can compensate with a higher working speed. However, technology could be used to keep track of and control interruptions over a long period of time so as not to overload people (as our mental workload measures suggest). After only 20 minutes of interrupted performance people reported significantly higher stress, frustration, workload, effort, and pressure. We cannot say whether people would cope over time or if these measures would only increase. Our results confirm experimentally the anecdotal reports of informants in field studies who describe high stress when interrupted in real work situations [10]. Our data also contributed to finding individual differences in interruptions. Our personality measures suggest the need for customization for systems to fit people's preferred interruption tolerance. We hope that our study will spark continued interest in this area.

REFERENCES

1. Adamczyk P. D. & Bailey B. P. If not now, when? The effects of interruption at different moments within task execution. *Proceedings of CHI'04*, (2004), 271-278.
2. Collani, G. Eine deutsche Skala zum Konstrukt Bedürfnis nach kognitiver Geschlossenheit (NFCC)/Persönliches Strukturbedürfnis (PNS). In A. Glöckner-Rist (Ed.), *ZUMA-Informationssystem*. Mannheim: Zentrum für Umfragen, Methoden und Analysen (2004).
3. Czerwinski, M., Horvitz, E., and Wilhite, S. A diary study of task switching and interruptions. *Proceedings of CHI'04*, (2004), 175-182.
4. Czerwinski, M., Cutrell, E. and Horvitz, E., Instant messaging: Effects of relevance and timing. *Proc. of HCI'00*, British Computer Society (2000), 71-76.
5. Gillie, T. & Broadbent, D. What makes interruptions disruptive? A study of length, similarity and complexity. *Psych Research*, 50, (1989), 243–250.
6. Gonzáles, V. & Mark, G. Constant, constant, multi-tasking craziness: managing multiple working spheres. *Proceedings of CHI '04,* (2004), 113-120.
7. Hart, S. G., & Staveland, L. E. Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), Human mental workload**,** Amsterdam, The Netherlands: Elsevier, (1988), 139-183.
8. Hudson, J.M., Christensen, J., Kellogg, W.A. and Erickson, T. "I'd be overwhelmed, but it's just one more thing to do:" Availability and interruption in research management. *Proc. of CHI 2002*, (2002), 97-104.
9. Iqbal, S.T. and E. Horvitz, Disruption and recovery of computing tasks: Field study, analysis, and directions, Proceedings of *CHI 2007*, 677-686.
10. Mark, G., Gonzalez, V., and Harris, J. No Task Left Behind? Examining the Nature of Fragmented Work. *Proceedings of CHI'05*, (2005), 113-120.
11. Ostendorf, F & Angleitner, A. *NEO-Persönlichkeitsinventar nach Costa und McCrae, Revidierte Fassung (NEO-PI-R). Manual.* Göttingen: Hogrefe, (2003).