Busy Days: Exposing Temporal Metrics, Problems and Elasticities through Diary Studies

William M. Newman
University College London Interaction Centre
31/32 Alfred Place, London WC1 7DP, UK
wmn@pobox.com

Microsoft Research Cambridge Ltd.
7 JJ Thomson Avenue, Cambridge CB3 0FB
v-wnew@microsoft.com

ABSTRACT
The study of temporal properties of people’s activities helps us to understand their needs and to assess the benefits they would gain from interactive systems. This paper describes the Busy Days study project, which employs an augmented diary study method involving elicting a plan before the diary is kept. The plan helps in exposing not only problems, in the form of disruptions to temporal objectives and daily plans, but also the critical parameters by which the performance of supporting technologies can be measured. Analysis of the data reveals several disrupting effects, two of which (forgetfulness, overruns when writing to completion) are briefly discussed. One significant effect, the temporal elasticity of work with a tangible focus (i.e., artefacts or people), is presented.

Author Keywords
Performance metrics, diary studies, temporal structures, human enhancement, critical parameters.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
Studies of temporal structure in work can guide the design of interactive systems in a number of ways. They can expose the structures and mechanisms through which activities are coordinated, and suggest ways to improve this coordination. For example, Dix et al. have highlighted the role of triggers in coordinating academic processes [2], and Reddy and Dourish the coordinated rhythms of intensive medical care [14], as indicators of requirements for coordination resources. A second benefit from studies is to shed light on time itself as a resource – often the major resource required for the performance of activities, besides human effort. In tightly interlocked collaborative settings such as those just mentioned, outcomes may indeed depend just as much on coordination as on simply finding sufficient time. The studies I present here, however, suggest that availability of time is of primary concern to people, particularly those who lead busy lives. Busy people often express an overriding interest in getting certain things done by the end of the day, and value anything that will truly enhance their ability to do so.

I mention human enhancement because this, in the opinion of several distinguished writers, is the central role of technology in society [4, 5, 16]. Technology enhances us physically, equipping us to travel at enormous speeds, transport huge loads and ward off disease. It enhances our senses, enabling us to hear astronauts on the moon and see galaxies near the edge of the universe. It also enhances our intellectual capacity, providing us with rapid access to vast amounts of information and the ability to analyse it at incredible speed. Many of the artefacts we see advertised, such as cars, domestic appliances and electric tools, offer to enhance us in tangible ways; it is significant that the enhancement often lies in their obvious ability to save us time. Interactive systems should do the same, but here the time savings are often far from obvious.

The human-enhancement imperative implies, I believe, that our systems must help people use their time better as well as coordinate better with others. Neither enhancement should be pursued to the exclusion of the other. This means that we should study how they use their time as well as how they manage their collaborations. In order to be sure we have improved both aspects of their work, we must discover how to measure the improvements our systems deliver. This paper describes a programme of diary studies that have been applied to understanding these aspects of work and of other types of time-constrained human activity. It finishes by discussing several types of temporal disruption that have emerged from the study.

DIARY STUDIES
Diary studies have been used by systems researchers for a number of years, primarily as a means of investigating specific aspects of people’s daily activities. An early paper by Riemann described their use in studying exploratory learning [15], and since then they have been used by Palen and Salzman [13], O’Hara et al. [12] and Sellen et al. [18].

In these studies the researchers were often concerned with a particular type of activity, such as web usage by knowledge workers, and with understanding the frequency with which certain activities occurred.

The method applied here has rather different aims from these earlier studies. It focuses on constructing an explicative account of the participant’s day, drawing on a written diary of events and on elaborations provided during subsequent interviews. The participant makes diary entries throughout the day, consisting of just a few words describing each distinct activity, together with start and stop times. An interview is held at the end of the day, or early the following day, during which the participant is invited to elaborate on each activity.

The method includes one relatively unusual feature: before commencing the diary the participant is asked to provide a list of the things they want to achieve during the day, their likely duration if known, and their importance: high (must be completed on the day), medium (may be postponed to the next day) and low (may be postponed further). The purpose of this plan is to permit analyses that set the diary’s events in the context of the participant’s aspirations, and to suggest how those aspirations influenced the day’s activities. Eliciting a plan is controversial, because it can influence the participant’s activities. Several steps are taken to reduce this influence; in particular, the participant is not provided with a copy of the plan to refer to, and data analysis focuses on high-priority activities, which are likely to be motivated less by the stated plan and more by necessity.

The author developed this method jointly with Marge Eldridge at Xerox Research Centre, Cambridge, on the basis of a suggestion from John Hughes of Lancaster University. It was used in a number of Xerox studies [3, 10, 17], and the method has since been refined considerably on the basis of that experience. In this paper I make some use of data from the earlier studies to supplement the body of data gathered recently.

THE BUSY DAYS STUDY PROJECT

The findings reported in this paper have been drawn primarily from a recent programme of studies of people with busy days. A number of participants were recruited, all of whom (a) were so busy that they admitted to failing regularly to achieve all their daily objectives, and (b) were able largely to set their daily agendas. These two criteria were set in order to increase the chance of observing the occurrence and solution of problems. As in [18], each participant was studied for just one or two days, during which they recorded each activity on a form as the day progressed, with start and finish times to the nearest five minutes. They were paid between 30 and 40 pounds sterling for each day. At the time of writing nearly 60 days’ data have been gathered, from participants aged from under 20 to over 70. As an indication of the range of occupations, they include:

- Chloe: project manager in a local authority
- Colin: partner in a candle supply company
- Ellie: mother of a small baby
- Geoff: art school director
- Ian: website developer
- Jonathan: finance officer
- Kenneth: architectural historian
- Loretta: student doctor
- Maria: student and part-time administrator
- Nick: sculptor
- Odette: secondary-school teacher
- Sophie: university director of health and safety
- Ursula: university lecturer
- Violet: furniture maker
- Winifred: pub landlady
- Zack: convenience-store manager

Interviews were audio recorded, the transcript was incorporated into the diary, and the participant checked the result. The participant also signed a consent form. All names in the diary were then anonymized and the original was discarded. Excerpts from a plan and diary are shown in Table 1.

IDENTIFYING RECURRING ACTIVITIES AND THEIR PERFORMANCE CRITERIA

Diary studies provide an effective method of identifying recurring activities that offer opportunities to achieve human enhancement. They involve much less effort than gathering observational data and reconstructing activities’ task structures. Rather, the diary presents a structured record of the day’s events, divided into units of activity that are meaningful to the diary keeper, each furnished with a similarly meaningful description that can be elaborated in the subsequent interview.

Even after elaboration, diaries such as these of course contain less detail than could be extracted from the analysis of contemporaneous video recordings. Consequently one would not rely on diary data to establish precise design requirements. They are adequate, however, for the purpose of identifying recurring activities and structures. For example, the Busy Days data include over 100 instances of phone calls, both placed and received, described in sufficient detail to categorize them according to purpose and outcome. Frequently recurring call types included outgoing calls requesting information (18), outgoing calls to schedule future activities (18), and incoming calls providing information (12). Most of these could be linked to other activities during the day.
Diaries also record the duration of activities, and thus shed some light on their temporal properties. Actual duration is rarely a particularly informative metric on its own, because its variance is often high. The 18 calls requesting information, for example, ranged from 1 minute to 20 minutes in length. Lack of detail can make it difficult to explain such variations.

A more promising source of temporal measures lies in the descriptions of planned activities provided by participants before the start of the working day. These specify, as mentioned above, the activity’s importance and its duration if known. Table 2 is an excerpt from a plan provided by Ian, a website developer, and shows (highlighted) two high-priority activities with quite precise duration estimates:

Expected duration is, I suggest, a useful and valid metric because it draws on the participant’s experience of performing the activity in question. A precisely specified duration, as in the two highlighted activities of Table 2, may indicate that the participant frequently performs this activity and pays particular attention to how long it takes. If so, we could be justified in treating these times as critical performance metrics, i.e.:

- Delivery time to local premises
- Time to find a website copywriter.

Metrics such as these can guide the design of enhancing technologies [9, 11, 19]. As one who has devoted a great deal of effort to discovering such metrics by other means, I am finding the use of daily plans to be a particularly effective discovery method. Table 3 shows some examples.

**GENERIC PROBLEMS AFFECTING TEMPORAL STRUCTURE**

Diaries and plans, taken in conjunction, invite analysis in a variety of ways. The earlier Xerox studies focused particularly on analyses of collaboration, made possible by studying entire groups concurrently [10, 17]. The studies also investigated how people dealt with major disruptions (“agenda-benders”) in the workplace [3].

Table 1. Excerpts from (a) Loretta’s plan and (b) her diary for the same period. Her own diary entries are shown in italic, annotated with details gained from interviewing.

<table>
<thead>
<tr>
<th>activity</th>
<th>start</th>
<th>end</th>
<th>priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arrive at Hospital</td>
<td>07:45</td>
<td>08:00</td>
<td>high</td>
</tr>
<tr>
<td>Go into clinical school, check email</td>
<td>08:00</td>
<td></td>
<td>medium</td>
</tr>
<tr>
<td>Drop off letters</td>
<td></td>
<td>08:30</td>
<td>medium</td>
</tr>
<tr>
<td>Get coffee in canteen</td>
<td>08:30</td>
<td></td>
<td>medium</td>
</tr>
<tr>
<td>Read Oxford handbook, 8 pages</td>
<td>09:00</td>
<td></td>
<td>medium</td>
</tr>
</tbody>
</table>

(a)

<table>
<thead>
<tr>
<th>activity</th>
<th>start</th>
<th>end</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arrived Hospital</td>
<td>07:45</td>
<td>07:55</td>
</tr>
<tr>
<td>Left car during conversation mother on mobile phone</td>
<td>07:55</td>
<td>08:00</td>
</tr>
<tr>
<td>Go to library - check email</td>
<td>08:05</td>
<td>08:10</td>
</tr>
<tr>
<td>Bathroom</td>
<td>08:10</td>
<td>08:15</td>
</tr>
<tr>
<td>Go to locker, deposit coat, realise I have forgotten purse and have no money to buy coffee or breakfast</td>
<td>08:15</td>
<td>08:20</td>
</tr>
<tr>
<td>Sat in Hatfield room reading</td>
<td>08:20</td>
<td>08:45</td>
</tr>
</tbody>
</table>

(b)

Table 1. Excerpts from (a) Loretta’s plan and (b) her diary for the same period. Her own diary entries are shown in italic, annotated with details gained from interviewing.

Diaries also record the duration of activities, and thus shed some light on their temporal properties. Actual duration is rarely a particularly informative metric on its own, because its variance is often high. The 18 calls requesting information, for example, ranged from 1 minute to 20 minutes in length. Lack of detail can make it difficult to explain such variations.

A more promising source of temporal measures lies in the descriptions of planned activities provided by participants before the start of the working day. These specify, as mentioned above, the activity’s importance and its duration if known. Table 2 is an excerpt from a plan provided by Ian, a website developer, and shows (highlighted) two high-priority activities with quite precise duration estimates:

Expected duration is, I suggest, a useful and valid metric because it draws on the participant’s experience of performing the activity in question. A precisely specified duration, as in the two highlighted activities of Table 2, may indicate that the participant frequently performs this activity and pays particular attention to how long it takes. If so, we could be justified in treating these times as critical performance metrics, i.e.:

- Delivery time to local premises
- Time to find a website copywriter.

Metrics such as these can guide the design of enhancing technologies [9, 11, 19]. As one who has devoted a great deal of effort to discovering such metrics by other means, I am finding the use of daily plans to be a particularly effective discovery method. Table 3 shows some examples.

**GENERIC PROBLEMS AFFECTING TEMPORAL STRUCTURE**

Diaries and plans, taken in conjunction, invite analysis in a variety of ways. The earlier Xerox studies focused particularly on analyses of collaboration, made possible by studying entire groups concurrently [10, 17]. The studies also investigated how people dealt with major disruptions (“agenda-benders”) in the workplace [3].

Table 2. Plan excerpt, showing two high-priority activities with precise estimates of durations.

<table>
<thead>
<tr>
<th>activity</th>
<th>duration</th>
<th>priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read email</td>
<td>15-30m</td>
<td>high</td>
</tr>
<tr>
<td>Take son’s music (forgotten) and drop off at school</td>
<td>20m</td>
<td>high</td>
</tr>
<tr>
<td>Start new project</td>
<td>?</td>
<td>high</td>
</tr>
<tr>
<td>Find someone to prepare website copy</td>
<td>1 hr</td>
<td>high</td>
</tr>
<tr>
<td>Write proposal, 2-3pp, not too long or technical, maybe send it today</td>
<td>2-3h</td>
<td>high</td>
</tr>
</tbody>
</table>

Table 3. Examples of metrics based on planned activities.

When plans and diaries are inspected side by side, they almost always exhibit differences that suggest lines of inquiry. Planned activities may not be mentioned in the diary, usually because they didn’t get done, or actual activities may have been left out of the plan because they were forgotten at the time or were unexpected. Also the duration or outcome of an activity may differ from what was planned. These discrepancies can be spotted and followed up during interviews. By taking this approach I have identified a number of generic problems that impact the temporal structure of the day. They include:
1. Forgetfulness
2. Overruns when writing to completion
3. Last-minute iterative schedule preparation – the “dinner seating plan” problem
4. Inaccessibility and unreliability of office technology
5. Re-jigging the day in the face of major unanticipated problems requiring urgent resolution.

Each of these occurred at least three times in the data, causing significant temporal disruptions to the participants’ days. I cannot do justice to all six problems in the space available here, but will briefly discuss the first two.

### Forgetfulness

Memory problems have for some years been regarded as a potential target for Ubiquitous Computing and other computer architectures [7, 8]. One of the initial motivations of the Busy Days study project was to gain some field data on the nature, frequency and cost of people’s memory lapses. Participants could not, of course, be relied upon to note down all such lapses, but those that caused significant disruptions to the day could be detected during the interview if not in the diary itself.

A total of 13 memory lapses occurred during the 57 days studied. Four of these involved forgetting the diary form, and so should probably be considered atypical. Three other lapses also involved leaving things behind. Only one lapse occurred among the ten participants aged over 60. Five of the lapses were incurred by one participant (Loretta), who started by forgetting her purse and ended up completely reorganising her day in order to retrieve it and buy something to eat. Apart from Loretta, however, nobody suffered much from their lapses – at most the participant spent an extra 10 minutes retrieving the situation.

These results, admittedly from a small survey, suggest that memory lapses may not be a particularly serious and frequent source of disruption for busy people, and advocates of ubiquitous memory aids may need to consider this. Older people, on the other hand, often have serious memory problems, some of which can be alleviated by specialized mobile systems, and these developments could be informed by diaries, possibly kept by carers.

### Overruns when writing to completion

Many of the Busy Days participants engaged in writing activities, as did those taking part in earlier Xerox studies. In some cases the document in question was completed during the day, while in others it was left incomplete, to be worked on at a later date. The diary data expose an interesting temporal phenomenon here.

As can be seen, writers who did not complete the document always spent less time on it than they planned; they put off getting started, or gave excuses such as “I’d been looking at it too long,” or “I like to stop after 45 minutes’ writing.” In contrast, completing a document almost always took longer than planned, sometimes more than twice as long. There were often serious costs associated with over-runs and failures to complete: deadlines were missed, weekend plans were cancelled, other planned work was postponed.

It appears possible that many writers are in fact fairly accurate in their estimates of how long actual text creation will take, but fail to allow for the possibility that they will need to produce evidence to back up what they eventually write. This was certainly true of one of Chloe’s planned tasks, finishing a report, which she had estimated would take her an hour and a half, but which took her nearly four hours. In fact it involved her in only 86 minutes of text creation, but she also had to spend over two hours preparing data for a table to include in the report. Comparison of estimated and actual completion times detected other activities that tend to take longer than expected, including fixing software problems and printing.

### THE TEMPORAL ELASTICITY OF WORK WITH PEOPLE AND ARTEFACTS

Writing is just one instance of working with information, but it is not the only one that suffers from overruns. Other types of information-focused work arise frequently in the data, including email sessions, updating records, work-related reading and information searches. There were 50 occurrences of such activities in the data that were not only completed by the participants, but for which they also recorded estimated and actual durations. On average these activities overran by nearly 15 minutes.
These activities stand in sharp contrast to those whose main focus was not information, but people or physical artefacts. People-related activities included decision-making meetings (see the example in Table 4), social meetings, teaching and tutoring, hosting visitors, and interviewing. Activities with physical artefacts included metalwork, shopping, tidying up and packing. Far from overrunning, participants were frequently able to complete these types of activities more quickly than estimated, by 5 and 7 minutes respectively on average (see Table 5). The difference from information work was statistically significant, p<0.05 in both cases.

Overall, these results suggest that there is a degree of temporal elasticity in work with artefacts and people that enables people to compress it into less time. In contrast, information-based work is intrinsically resistant to completion in less time, and likely to overrun. There is no obvious explanation for this. It is interesting, however, that certain information-based activities such as writing have been shown to impose heavy cognitive loads [6]. This could explain why they are hard to speed up. People working with artefacts or with other people may tend normally to leave greater reserves of capacity. Further investigations are needed in this area.

**DISCUSSION AND CONCLUSION**

In this paper I have reported on the ongoing diary-based Busy Days study and have presented some interim results concerning temporal structures in work. To conclude, I will comment on the main issues raised: identification of performance metrics, and understanding temporal structures in human activity.

Successful technical innovation involves us in discovering how the technology can deliver enhancements to those who use it, and this involves knowing how to measure an enhancement. It is encouraging to see performance metrics emerging from the plans of participants, but there is more work to be done to validate these metrics and ensure they are usable in design. It is not enough, for example, to inform those trying to improve the design of a word processor that “time to complete the document” is what matters most. A major contributor to this parameter may be the time spent producing evidence; if so, the temporal properties of evidence-production, both during and prior to writing, will need to be investigated. We might conclude that writers themselves need to learn the importance of gathering evidence in advance, and that this should be the focus of innovation, just as Bhavnani and John concluded that users of CAD systems needed better work strategies, not better technology [1].

As regards temporal structures, this study suggests there is much more lying hidden, ready to be unearthed. In particular, the Busy Days study is reaching out to people who have rarely, if ever, been studied by HCI researchers. It is capturing the daily activities of people who interact almost entirely with other people and/or with physical artefacts. It is discovering the skills and social systems that enable them to achieve their daily objectives. Evidence of the relative elasticity of these people’s work, compared with “knowledge work,” is a reminder that we can learn a lot from them. There are almost certainly other desirable properties besides temporal elasticity that we need to take into account in designing interactive systems.

**ACKNOWLEDGMENTS**

I wish particularly to thank Marge Eldridge for her considerable help in developing the diary study method at Xerox, and John Hughes for suggesting the method in the first place. Abigail Sellen and Richard Harper assisted Marge and myself in some of the Xerox studies, and Abi has provided useful advice since then. I am grateful to Hewlett Packard Labs, Palo Alto, for their sponsorship of the initial Busy Days studies, and to Mik Lamming for comments on an earlier draft of this paper. I am also grateful for diary data supplied by Alistair Beresford and Huseyin Dogan, and for Tim Regan’s help with statistics.

**REFERENCES**


<table>
<thead>
<tr>
<th>focus</th>
<th>N</th>
<th>mean overrun</th>
<th>standard deviation</th>
<th>p &lt; n</th>
</tr>
</thead>
<tbody>
<tr>
<td>info</td>
<td>50</td>
<td>14.44</td>
<td>52.05</td>
<td></td>
</tr>
<tr>
<td>artefact</td>
<td>56</td>
<td>-5.05</td>
<td>21.81</td>
<td>0.05</td>
</tr>
<tr>
<td>people</td>
<td>47</td>
<td>-7.34</td>
<td>23.44</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Table 5. Overruns in activities focused on information, physical artefacts and people. N = number of occurrences.