Trigger Analysis
understanding broken tasks

Alan Dix[1], Devina Ramduny-Ellis [1], Julie Wilkinson[2]
Lancaster University and Huddersfield University

[1]Computing Department, Lancaster University, SECaMS Building,
Lancaster, LA1 4YR, U.K.
Email: alan@hcibook.com, devina@comp.lancs.ac.uk

[2]School of Computing and Mathematics, University of Huddersfield, Queensgate,
Huddersfield, HD1 3DH, U.K.
Email: j.wilkinson@hud.ac.uk

http://www.hcibook.com/alan/topics/trigger-analysis/

Why do things happen when they happen? Trigger analysis exposes the triggers that prompt sub-tasks to occur in the right order at the right time, and assess whether tasks are robust to interruptions, delays and shared responsibility. Trigger analysis starts with task decomposition via any suitable method, proceeding to uncover what trigger prompts each sub-task. The obvious answer: "because the previous sub-task is complete" is often the precondition, not the actual trigger. Previous analysis has uncovered certain primary triggers and potential failure points. The complete analysis produces an ecologically richer picture, with tasks interacting with and prompted by their environment.

Keywords: sub-tasks, triggers, placeholders, artefacts, shared tasks, environmental cues, task timing, interruptions, delays

Introduction

Many task analysis and workflow techniques decompose the overall task into smaller subtasks, processes or activities. The order of these subtasks is then typically specified or described (e.g. plans in Hierarchical Task Analysis (HTA) [Shepard, 1995], links between processes in a workflow, temporal connectives in CTT). For short- term tasks, performed by one person, without interruption, this may be the end of the story:

to photocopy a document:
(a) open copier lid, (b) put original on glass, (c) close lid,
(d) select number of copies, (e) press copy button,
(f) when copying complete remove copies, (g) remove original

But what happens if:

- the number of photocopies is large and so you go for a cup of tea between (e) and (f)?
- someone comes in and interrupts between (f) and (g)?
- instead of a small photocopier this is a large print machine and the person responsible for each stage may be different?

Trigger analysis deals with exactly these issues. Why do things happen when they happen and do they happen at all? By exposing the triggers that prompt activities and sub-tasks to occur in the right order at the right time, trigger analysis allows us to decide whether tasks are robust to interruptions, delays and shared responsibility (even across organisational boundaries).

Trigger analysis starts with a task decomposition obtained by any suitable method (and can therefore be used in combination with many TA and workflow methods). It then proceeds to uncover what trigger causes each subtask to occur. The initial answer is "because the previous sub-task is complete", but this is often merely a precondition, not the actual trigger.

Previous empirical and theoretical analysis has uncovered a small set of primary triggers including the simple "previous task complete", timed events – "every hour I check the mail", and environmental cues – "the document is in the in-tray". For each class of trigger there are a set of subsequent questions, for example "what happens if you are interrupted between tasks?", "how do you know when it is the right time?", "are there several tasks with the same environmental cues?".

Triggers are what make activities happen when they do. A closely related issue is: knowing where in task sequence you are. Often, environmental cues act in both roles: triggers saying that something should happen and placeholder saying what should happen. Typically, the complete analysis produces a highly ecologically rich picture and rather than cognitively-driven tasks acting on the environment, we see tasks interacting with and prompted by their environment. Note that trigger analysis is not an alternative task analysis technique or notation but instead an additional concern that should be grafted to existing analysis methods.

This chapter examines the trigger analysis method in depth, with particular reference to its grounding in empirical work. The starting point is a brief discussion of the theoretical underpinning for the fieldwork undertaken. We then suggest some reasons why prolonged interactions tend to break down in organisational contexts, culminating in an explanation of the five trigger types that have emerged during our investigations. A second, related pattern – the 4Rs framework – is also presented in detail, as we suggest the 4Rs form an important and fundamental unit of work. Finally, we demonstrate a comprehensive application of the 4Rs and the benefits they provided for the analysis of work in a systems development project.

Theoretical background

The roots of trigger analysis lie in two principal theoretical foundations: the study of pace of interaction [Dix, 1994; Dix, 1992] and status–event analysis [Abowd *et al.*, 1994, Dix 1991, Dix *et al.*, 1998]. The primary basis is the issues surrounding pace – that is, the rate at which users interact with computer systems, the physical world and with one another. Thinking about pace makes one concentrate on the timescale over which interaction occurs, both the similarities between interactions of widely different pace and also the differences.

Status–event analysis

Status–event analysis is a collection of formal and semi-formal techniques all focussed on the differences between events (things that happen) and status (things which always have a value). Applications of status-event analysis have included auditory interfaces [Brewster, 1994], formal analysis of shared scrollbars [Abowd *et al.*, 1994] and software architectures for distributed agent interfaces [Wood *et al.*, 1997].

Status–event analysis allows the distinction between actual events – some objective thing that occurs, from perceived events – when an agent (human or machine) notices that the

event has occurred. Sometimes this is virtually instantaneous, but more often there is a lag between the two. Many formal and informal analyses of events assume simultaneity between cause and effect. However, accepting that there is often a gap allows us to investigate what actually causes secondary events to occur when it does. The lag between the actual event and the perceived event can in turn influence the pace of interaction, for example the time it takes to notice that email has arrived.

Status–event analysis also looks at the effect of events on agents. An event may simply be intended to inform, but more often is intended to initiate actions, which in turn may cause further events. The actions of agents may change the status of the agent or the world, but in addition changes in status are themselves events that may trigger further action. Furthermore, most notations in computing focus primarily on events with little, if any, description of status phenomena. However, we shall see that environmental cues (things that are around us) such as piles of papers or notice boards play an essential role in acting as triggers and placeholders.

The analysis of status and events has also allowed us to see some common features between human-human, human-computer and internal computer interactions. For example, it is common to see status mediation whereby one agent communicates an event to another by manipulating a status that will eventually be observed by the second agent. Moreover, polling – the periodic observation of a status phenomenon to detect change, is not just a low-level computational device, but something people do as well. This rich interplay of status and event phenomena is reflected in the ecological perspective that colours the analytic stance of our current study.

Because status–event analysis gives equal weight to status and event phenomena it is able to address interstitial behaviour – the continuous relationships between status phenomena between events (in the interstices). This interstitial behaviour is often what gives the 'feel' to computer systems – you move the mouse and the window is dragged around the screen. Part of the power of the human motor system is its capability for near continuous control and response – as in sports, driving, manipulating objects etc.; and paradigms such as direct manipulation and virtual reality use this power.

This continuous interaction is important in many non-desktop technologies including aspects ubiquitous computing (see Dix 2002b). However, in this chapter we will be concentrating more on office style interactions and the discrete activities commonly dealt with in task analysis. Often the lowest level subtask or activity may involve continuous interaction, but we will look principally at the events that drive the transitions between these

Pace of interaction

The 'pace' of interaction with other people or with a computer system is the rate at which you send messages/commands and then receive a response. It varies from tens of milliseconds in a video game to hours or days when interacting by post.

The pace of interaction is influenced by and influences three principal factors (see figure 1):

(i) The intrinsic pace of the channels through which you communicate.
(ii) The pace of the collaborative tasks.
(iii) The users' own natural pace for different forms of mental activity.

Problems may occur when there is any significant mismatch between any of these and the resulting pace of interaction [Dix, 1992, Dix *et al.*, 1998b].

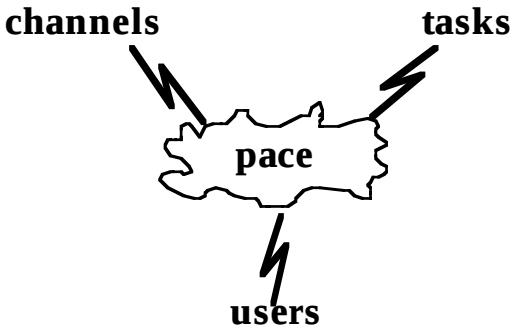

Figure1. Factors affecting pace

Prolonged Interaction

The study of pace helps us to understand interactivity in a wider context. A system or collaborative process is not interactive because it is fast or it has instant feedback. Instead, interactivity is about the appropriate pace of interaction in relation to the task at hand. This is certainly the case in many collaborative situations where the pace of communication may be over days or weeks.

The reason for the prolonged nature of these interactions varies: it may be due to the communication medium (e.g. normal postal delays), or due to the nature of the task (e.g. a doctor waiting for X-ray results). One of the key points is that models of interaction that concentrate on a tight cycle between action and feedback break down [Dix, 1992]. This is typified by Norman's execution–evaluation cycle [Norman, 1986, 1988]: a user has a goal, formulates actions that will further that goal, executes the actions, and then evaluates the results of those actions against the expected outcome and the goal (see figure 2a). This model effectively assumes that the results of the user's actions are immediately available. If the delay between executing actions and observing the results is greater than short-term memory times, then the evaluation becomes far more difficult. This problem has been called the 'broken loop of interaction'.

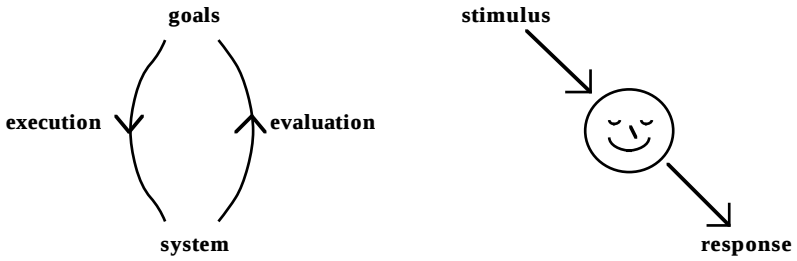

Figure 2. (a) Norman's execution–evaluation cycle (b) stimulus–response model

Another model of interaction used in more industrial settings is to treat the worker in a stimulus–response manner (see figure 2b). Commands and alarms act as stimuli and the effective worker responds to these in the appropriate manner. However, in a pure form, this model does not allow for any long-term plans or goals on the part of the worker; the worker is treated in a mechanistic manner, a cog in the machine.

To incorporate both these perspectives we need to stretch out the interaction and consider the interplay between the user and the environment over a protracted timescale. We use the term environment to include interactions with other users, computer systems or the physical environment. Such interaction is typically of a turn-taking fashion: the user acts on the environment, the environment 'responds', the user sees the effects, and then acts again.

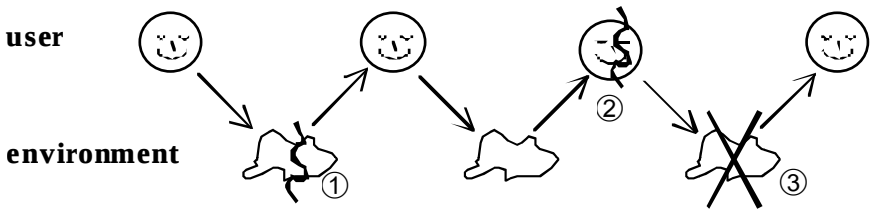

Figure 3. Problems for prolonged interaction

This process is illustrated in figure 3. Notice how the Norman loop concentrates on the user–environment–user part of the interaction whereas the stimulus–response model concentrates on the environment–user–environment part. We can see various ways in which long-term interaction affects this picture:

① **action–effect gap** – The user performs an action, but there is a long delay before the effects of that action occur, or become apparent to the user. For example, you send an email and some days later get a reply. The problem here is loss of context; how do you recall the context when you eventually receive the feedback. When the reply comes you have to remember the reason why the original message was sent and what your expectations of the reply were. The way in which email systems include the sender's message in the reply is an attempt to address this problem. In paper communications the use of 'my ref./your ref.' fulfils a similar purpose.

② **stimulus–response gap** – Something happens to which the user must respond, but for some reason cannot do so immediately. For example, someone asks you to do something when you meet in the corridor. The problem here is that you may forget. Hence the need for to-do lists or other forms of reminder. In the psychological literature this has been called prospective memory [Payne, 1993].

③ **missing stimulus** – The user performs an action, but something goes wrong and there is never a response. For example, you send someone a letter, but never get a reply. For short-term interactions this is immediately obvious, you are waiting for the response and when nothing happens, you know something is wrong. However, for long-term interactions you cannot afford to do nothing for several days waiting for a reply to a letter! In this case you need a reminder that someone else needs to do something – a to-be-done-to list!

Trigger analysis focuses on problem ②, the stimulus-response gap, but when looking at the robustness of the process all three problems are important.

Fundamental concepts of trigger analysis

Triggers – why things happen when they happen

Workflows and process diagrams decompose processes into smaller activities and then give the order between them. Similarly in Hierarchical Task Analysis (HTA) [Shepherd, 1995], plans give some specification of the order of sub-tasks and in Conquer Task Trees (CTT) these temporal orders are made more specific using operators derived from LOTOS [Paternó 1999].

Figure 4. Simple work process

Figure 4 shows a simple example, perhaps the normal pattern of activity for an office worker dealing with daily post. Notice the simple dependency that the post must be collected from the pigeonhole before it can be brought to the desk and before it can be

opened. However, look again at the activity "open post" – when does it actually happen? The work process says it doesn't happen before the "bring post to desk" activity is complete, but does it happen straight away after this or some time later?

In previous work [Dix *et al.*, 1996, 1998b], we have looked in detail at the triggers that cause activities to happen when they happen. In the case of opening post this could easily be something like "at coffee time" rather than straight away. In our work we identified a number of common triggers:

- *immediate*: straight after previous task
- *temporal*: at a particular time or after a particular delay
- *sporadic*: when someone thinks about it
- *external event*: some event occurs such as a phone call or the receipt of a message
- *environmental cue*: something in the environment prompts action

We will look at a more detailed breakdown later.

Now we can augment the work process with triggers for each activity (figure 5).

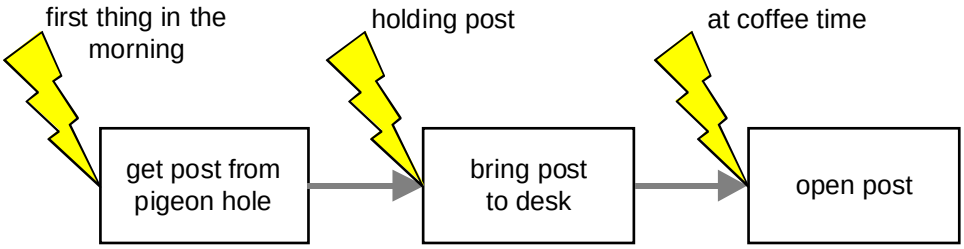

Figure 5. Triggers for activities

Notice how we have examples of several types of trigger, two temporal and one environmental (letters in the office worker's hand prompting her to carry them to her desk).

Triggers are important not only for understanding the temporal behaviour of the task, but also because they tell us about potential failure modes. If two environmental triggers are similar, one might do parts of the task out of sequence. If a trigger may not occur or be missed (likely for sporadic triggers), activities may be omitted entirely. Triggers also help us assess the likelihood of problems due to interruptions – for example, immediate "just after" sequences are disrupted badly but environmental cues tend to be robust (because they are still there).

Sometimes triggers are seen in the plans of HTAs and sometimes 'waiting' subtasks are included for external events, but these are both the exception. The normal assumption is that tasks are uninterrupted. However, it is straightforward to add a trigger analysis stage to most task analysis methods.

In terms of the ecology of interaction, triggers remind us that tasks are not always performed to completion. In the television quiz programme, "Mastermind", the time limit would sometimes buzz when the quizmaster, Magnus Magnusson, was in the middle of a question. He would calmly announce "I've started so I'll finish" and then continue the question. In the heat of the real world it is rarely so easy to ignore interruptions and distractions. In practice, tasks are interleaved with other unrelated tasks or, potentially more confusing, different instances of the same tasks and may be interrupted and disrupted by other activities and events. Furthermore, the performance of the tasks is dependent on a host of, sometimes fragile, interactions with the environment and apparently unconnected events.

Artefacts – things we act on and act with

Notice that one of the trigger types is an environmental cue – things in the environment that prompt us to action. Some years ago one of the authors received a telephone call reminding him to respond to a letter. He could not recall receiving it at all, but searching through a pile on his desk he found it and several other letters over a period of several weeks unopened and unread. What had happened? His practice was to bring the post upstairs to his desk, but not always read it straightaway. Not being a coffee drinker it was not coffee time that prompted him to open the post but just the fact that there was unopened post lying on his desk. This process had worked perfectly well until there was a new office cleaner. The new cleaner did not move things around on the desk, but did straighten higgledy-piggledy piles of paper. However, he had unconsciously been aware that unopened post was not tidy on the pile. This had effectively been acting as a reminder that post needed dealing with – this is a trigger. But with the new cleaner, post that for some reason got missed one day would then look as if it was tidily 'filed' in a pile on the desk.

This story is not unique. The ethnographic literature is full of accounts of artefacts being used to manage personal work and mediating collaborative work. Some of that purpose is to do with the content of the artefacts – what is written on the paper, but much by the physical disposition – by orienting a piece of paper towards you I say 'please read it'. In the case of his desk the cue that said, "post needs to be opened" was purely in the physical orientation (not even the position).

Of course, artefacts do carry information and are often the inputs or products of intellectual work. Furthermore, in physical processes the transformation of artefacts is the purpose of work.

One example that has been studied in detail in the ethnographic literature is air traffic control and all these uses of artefacts are apparent (Hughes *et al*. 1995). Flight strips are central (figure 6) – small slips of card for each aircraft recording information about the aircraft (flight number, current height, heading etc.). This information is important for the controller managing the aircraft, but is also an at-a-glance representation of the state of the airspace for other controllers. In addition, the strips are in a rack and the controllers slightly pull out strips corresponding to aircraft that have some issue or problem. This acts partly as a reminder and partly as an implicit communication with nearby controllers. Finally, the strips in some way represent the aircraft for the controllers, but of course, the real purpose of the process is the movement of the aircraft themselves.

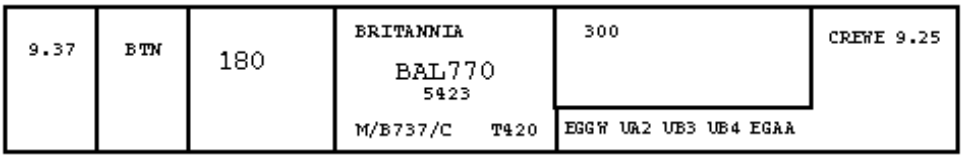

Figure 6. Air traffic control flight strip

Task models often talk about objects, either implicitly in the description of subtasks or explicitly in the task model. However, the objects are always 'second class' – users act on them, but they are not 'part of' the task. CTT and most work process notations do talk about automated tasks, but not the artefacts, whether electronic or physical included within the interaction.

In Unified Modelling Language [e.g. Bennett *et al*., 2001] and other Object Oriented design methods [e.g. Jacobsen, 1992], it is common to give a lifecycle description of 'objects'. However, this is usually because we are intending to store and automate the object electronically. Also workflow analysts study document lifecycles – again largely because of the intention to automate.

The task analysis chapter of Dix *et al*. (1998) treats physical objects as 'first class' within an example of entity-relationship style task analysis. This was based largely on the ATOM method [Walsh, 1989] but, to our knowledge, this style of method has not gained widespread acceptance.

There is no reason why most task analysis methods should not adopt some form of artefact tracking. This may be as simple as recording which artefacts are triggers for, used by, modified by, or produced by any particular sub-task. For tasks where artefacts are particularly central more sophisticated artefact lifecycles could sit alongside the task description. These lifecycles may be mundane (letter closed → letter open), but this is the point: users recruit their everyday knowledge and physical properties of the world to coordinate their activity.

Placeholders – knowing what happens next

It is half past five in the evening. The busy office building is beginning to become quiet as people pack up to go home. One or two employees work late in their offices but as the evening wears on, they too go home. Soon there is only the hum of vacuum cleaners and the clatter of waste bins as the office cleaners do their work, until eventually, the last light goes out and the building sleeps. A few have taken papers and laptops home and continue to work, but eventually they too put aside their work and sleep.

It is three o'clock in the morning, in the darkness and silence of the office and the deep sleep of all the employees, where is the *memory* of the organisation? The next morning at nine o'clock the office is a flurry of activity, it has not forgotten and has restarted its activities, but how?
We have already discussed two aspects of this memory: information required to perform tasks and triggers that remind us that something needs to happen. However, there is one last piece of this puzzle that we have hinted at several times already. As well as knowing *that* we need to do something we need to know *what* to do next. In the complex web of tasks and subtasks that comprise our job – *where* are we?

In fact, when looking at triggers we have already seen examples of this. The post being untidy on the desk said both "something needs to happen", but the fact that it was also unopened said, "it needs to be opened". In that discussion we already noted that similar triggers could cause sub-tasks to be performed out of sequence. If we only have a small number of dissimilar tasks this is unlikely to happen as we can remember where we are in each task. However as the number of tasks increases, especially if we are performing the same task on different things, it becomes harder to remember where we are.

Let us look again at air traffic control. One of the controller's tasks is to manage the flight level of aircraft. A much-simplified model of this activity is shown in figure 7. Because this is a shared task between the controller and the pilot, each box is labelled with the main actor (although tasks 2 and 3 are both communications). Recalling earlier sections, we might ask what information is required at each stage, for example task 1 would depend on radar, locations of other planes, planned take-offs and landings, new planes expected to enter airspace.

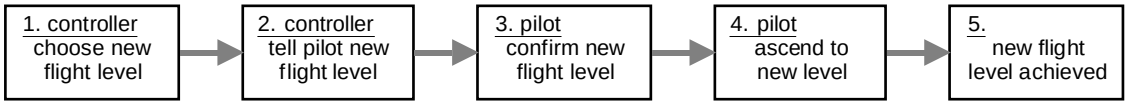

Figure 7. Flight level management task

Note that box 5 is not really a task but more a 'state of the world' that signifies task completion. However, it is important, as the controller will need to take alternative action if it does not happen. Of course, without appropriate placeholders, the controller might forget that a plane has not achieved its target level. This may either cause trouble later, as the old level will not be clear, or even provoke potential conflicts between aircraft.

In fact, the flight strips do encode just such a placeholder (see figure 8). When the controller informs the pilot of the new height he writes the new level on the flight strip (i). When the pilot confirms she has understood the request the pilot crosses out the old level (ii). Finally when the new level has actually been reached the new level is ticked (iii).

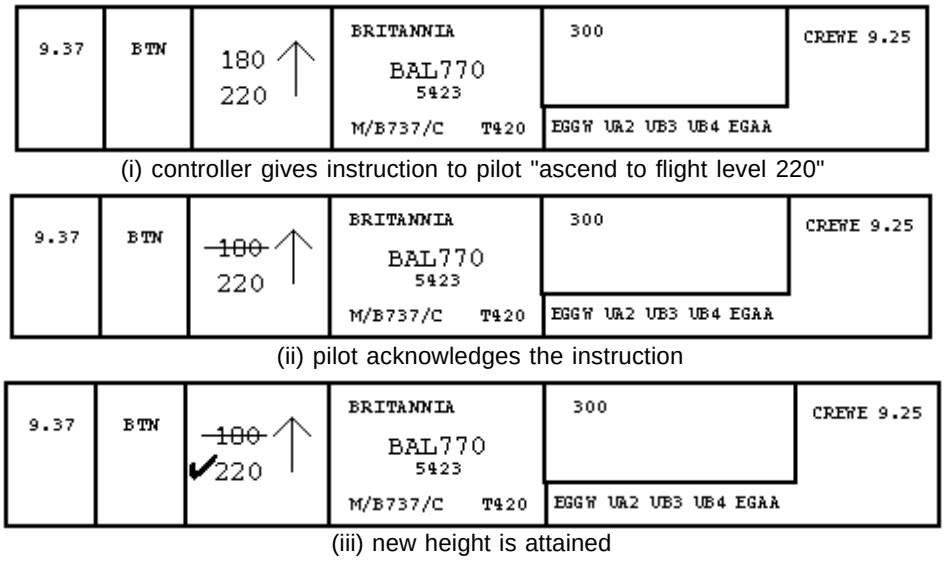

Figure 8. Flight strip annotated during task

Virtually all task-modelling notations treat the placeholder as implicit. The sequence of actions is recorded, but not why the user should do things in the way proposed. Of course, one purpose of task analysis has been to produce training – that is to help people learn what appropriate processes are, but this does not help to actually remember where you are. Just like other forms of information, placeholders may be stored in different ways:

(a) in peoples' heads – remembering what to do next
(b) explicitly in the environment – to-do-lists, planning charts, flight-strips, workflow system
(c) implicitly in the environment – is the letter open yet?

Although often forgotten, placeholders are crucial in ensuring that tasks are carried out effectively and in full. At a fine scale it is rare to find explicit records, as the overhead would be too high. Instead (a) and (c) predominate. As users' memory may be unreliable when faced with multiple tasks and interruptions, it is not surprising to find that various forms of environmental cues are common in the workplace. However, electronic environments do not have the same affordances to allow informal annotations or fine 'tweaking' of artefacts' disposition.

Finding triggers

This chapter is focussed on triggers and associated placeholders. However, these arise from the more general problem of missing stimuli and are linked to issues such as interruptions. The techniques we use are designed to expose all these problems as well. Part of the data we collect is on what is done. In traditional workflow fashion, we catalogue the various activities performed and the dependencies between the activities. However, this is only intended as the superstructure of the analysis, not the focus. Instead, our focus is on when activities are performed and whether they happen at all. The central and distinguishing feature of our work is therefore the way we look explicitly for the triggers that initiate activities.

Standard data collection

Because of the similarity of our study to traditional task analysis, we can use many of the same sources for data collection: documentation, observation or interviews. However, trigger analysis gives us an additional set of concerns.

Documentation

Documentation of long-term processes is likely to be relatively accurate, although it may omit the activities beyond organisational boundaries, and also most of the triggers. However, we can use it as an initial framework, which can be filled out by observation or during subsequent interviews.

Observation

Direct observation is a very effective technique, widely used in ethnographic studies or similar sort of analysis. In many office situations, there are several instances of the same process at different stages of completion. For example, in an insurance office, many claims are processed, each at a different stage. In these cases a 'day-in-the-life' observation may be sufficient. So long as we can see each activity during the study period, we can piece them together afterwards. Even if we never see a process run from end to end we can reconstruct it from its parts. This is similar to observing a natural forest. The complete life cycle of a tree might be hundreds of years long, but by looking at trees at different stages of growth, we can build up a full picture over a much shorter period.

However, direct observation poses special problems when many of the processes of interest extend beyond the time frame of observation and are geographically dispersed. This technique may miss rare but potentially significant events. Observation usually fails when either:

(a) activities are sporadic and long term,
(b) processes are in lock-step and long term, or
(c) unusual events occur.

Interviews

Where direct observation is impractical interviewing can be used. Interviewing allows both prospective (ask what is currently going on) and retrospective (ask what has happened) actions. However, interviewing is often regarded as problematic since the accounts people give of their actions are frequently at odds with what they actually do.

But we are in a strong position as we approach such interviews. Our analytic focus – the structure imposed by task analysis and the specific interest in triggers – allows us to trace omissions and inconsistencies and enables us to obtain reliable results from interviews. This is important as, although we would normally expect some additional direct observation, practical design must rely principally on more directed and less intrusive techniques.

Types of triggers

Based principally on our theoretical analysis and refined by the results of our previous case studies [Dix *et al.*, 1996, Dix *et al.*, 1998b], we have classified the different kinds of triggers that occur. These include the following 5 broad categories:

- *immediate*: occurs when one activity begins immediately after the previous activity reaches completion.
- *temporal*: includes periodic actions that happen at regular intervals or actions that occur after a particular delay (e.g. the expectation of receiving a response by a certain date or the generic task of reminding people based on some time interval).
- *sporadic*: arises when the responsible individual remembers that something must be done.

- *external event*: may be due to an alarm or signal (e.g. a wristwatch or automatic calendar set to give a reminder at a specific time) or a specific event (e.g. a telephone call, face-to-face request, receipt of a message, the completion of an automatic activity or even an event in the world).
- *environmental cue*: things in our environment that remind us that something ought to be done, whether explicitly recorded (e.g. a diary entry, to-do-lists) or implicit (e.g., unanswered e-mails or a half-written letter in the typewriter).

Some of these may be evident from observation. For example, if the telephone rings and someone does something (external event). However, others are less clear and this is where direct questioning helps "why did you do that then?". In particular, it may be difficult to tell if someone just remembers something (sporadic) or whether something reminded them (environmental cue). The reason that these are difficult is that they both relate to the users perception and psychological state rather than observable actions.

After having identified the primary triggers, we need to assess their robustness by asking follow-up questions. Sometimes if the primary trigger fails there is some secondary trigger that stops the process from failing.. For example, if someone says that they always do two things one after the other (immediate), we can ask what would happen if there were an interruption. The answer may be that they would need to remember (sporadic) or that the half finished task would remind them (environmental cue).

Sometimes also the primary trigger itself is sufficiently complex that it may need its own secondary trigger. So, depending on the answers, we may have several levels of questions:

1. "why did you send the reminder letter?"
   "because it is two weeks since I sent the original letter" (temporal)
2. "how do you know it is two weeks?"
   "because I write it in the diary as soon as I send the letter" (environmental cue)
3. "what makes you check your diary?"
   "I always do first thing each morning" (temporal)
4. "when you send the first letter, what happens if you are interrupted before entering the expected date in the diary?"
   "I always remember" (sporadic secondary to immediate primary)

Notice that both questions 3 and 4 are suggested by the answer to question 2. Question 3 is about chasing the chain of events that lead to a particular trigger. In contrast, question 4 arises because the answer to 2 was an environmental cue and things in the environment need to have been put there by a previous activity. Note that this should also lead one to re-examine the task associated with sending the original letter as this should include the "write expected date in diary" activity.

We could continue asking follow-up questions indefinitely, but at some point we must stop. We can either believe that a trigger does always occur as specified, or, if not, look at the whole process and assess the consequences should the activity fail to trigger at all and perhaps any delays associated with noticing it.

Table 1. lists some of the main ways in which different types of trigger can fail and the sort of follow up questions that can be asked. It is not complete and is being amended as we learn more. An evolving and more detailed table with examples can be found on the trigger analysis web pages.

trigger type	sub-type	failure mode	follow-up questions
immediate		interruption	does the second activity always proceed immediately? if there is any possibility of a gap, then look for secondary or back-up trigger(s).
temporal	periodic action	may forget	how do you remember to perform the action at the relevant period?
	delay	poor memory	how do you remember the delay? this may be part of a routine, like consulting a diary every morning, but if it is an hourly activity then how do you know when it is the hour? perhaps the clock strikes, an external event – another trigger.
sporadic	memory	may never happen	how do you remember to do something? if a request is made verbally, the recipient has to remember that the request is outstanding until either it can be performed or some record is made of the commitment, a reminder – another trigger.
external event	automatic	reliability	how reliable is the medium of communication?
	communication	delays	are there likely to be any communication delays?
environmental cue	explicit	is moved	can the cue be disrupted?
	implicit	fail to notice	how do you remember to examine the cue? perhaps a periodic activity or in diary – a temporal trigger. can the cue be missed?
		ambiguity	can the cue be confused with others leading to skipped or repeated activities?

Table 1: Summary of triggers, possible failure modes and follow-up questions see trigger analysis web page for more detailed list

Studying artefacts – transect analysis

As we have seen artefacts, both physical and electronic, are an inseparable part of an ecologically valid understanding of work and leisure. Tasks may be initiated to create an artefact (write a chapter abstract), tasks may occur because of artefacts (the memo requesting an action, the broken machine); artefacts may mediate tasks split between several people (patient records, whiteboard), artefacts may record where you are in a task

(document in in-tray, office planner), electronic artefacts may even control tasks (workflow systems, Coordinator).

Studying artefacts can therefore give us a rich understanding of the tasks that they are part of, especially when the tasks are complex, long-lived, or involve different people.

One artefact-centred approach that is particularly useful for uncovering triggers and placeholders is *transect analysis*. This takes an ecologically rich approach looking at the artefacts in their physical context – physical disposition is as important as the artefact itself. It focuses on uncovering the task in praxis as performed in the actual work environment. Transect analysis looks at a snapshot of a work environment (desk, office, or potentially organisation), either at a particular time (noon on Tuesday) or over a relatively short period (day in the life).

This is inspired by the transect, as used in environmental studies, which looks at a cross-section of an environment on a particular day. If the ecology is diverse and non-seasonal it is possible to build up a picture of the life cycles of particular organisms and the interrelations between them even though one is not studying them over time. Similarly we can look at each artefact in the work environment and ask: "who is using it?", "what for?", "why is it here?", "what would happen if it were somewhere else?", "how does it relate to other artefacts?". In particular, we look for instances of the same kind of artefact in different places (e.g. several invoices in different stages of processing) within one person's immediate environment (in-tray, centre desk, out-tray, at an angle on a pile) or between people. By piecing together these snippets of human–artefact interaction we can create models of task processes and artefact life cycles.

This analysis can be done initially by the analyst, but at some stage should be used as a prompt for the actual users. If asked away from the actual environment they may not volunteer many of the unofficial procedures that make the workplace work. However, when shown a document in location and asked “why is it there” they will usually be able not only to knit together the scattered threads of processes, but also often add details beyond the immediate workplace.

If interviews are carried out in the workplace then transect analysis can be woven seamlessly into the normal interview process.

Larger structures

The trigger analysis can be used in the context of virtually any task analysis. In the case studies that we have addressed, we have used a simple process diagram as a method of recording the task analysis and asked specific questions about the overall robustness of the process. However, we believe that the lessons and the broad techniques would be applicable to any process-oriented task analysis, such as HTA.

Processes and activities

We record the processes as a series of circles or bubbles, one for each activity. Each bubble names the activity and the person or persons who perform it. Lines between the bubbles record dependencies and arrows at the beginning of each bubble record the trigger for the activity (see figure 9).

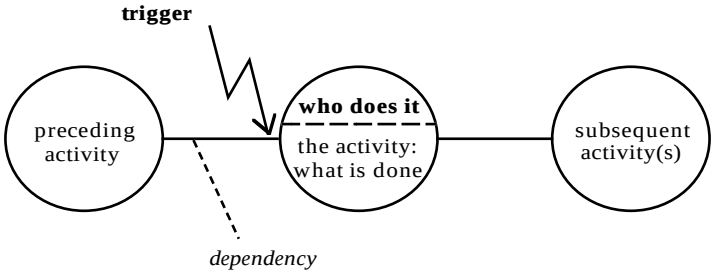

Figure 9. Recording processes

There are plenty of methods for recording processes, and this is not the focus of our work, so we take a minimalist approach. We do not attempt to record all the complexities of real processes in a single diagram. Instead, we use many separate diagrams, often concentrating on specific scenarios. The crucial thing is that for each activity we look for the corresponding trigger.

The level of analysis is also governed by this focus. In general, we place activity boundaries wherever there is the likelihood of a delay or gap. The most obvious such break occurs when people at different sites perform subsequent activities in a process. However, there are often distinct activities performed sequentially by an individual, as in the letter-reading example above. In principle such analysis could go down to the full detail found in HTA. This would be reasonable if, for example, interruptions were possible in the middle of typing a letter.

We deliberately use the term activity rather than action to emphasise that the lowest level of our analysis is far from atomic. Activities may be shared between individuals. For example, having a meeting or dictating a letter would be regarded as a single activity involving several people. Again, one could dissect such an interaction, but this would be the remit of conversational analysis. We may also ignore details of an activity because it is uninteresting or we do not have sufficient knowledge about it. For example, if we issue an order to an external organisation and then wait for the goods to arrive, we may not be interested in the internal processes of that firm. Finally, we include some activities that would normally be omitted in a traditional process model. In particular, we often include the receipt of a message as a distinct activity. This is deliberately to emphasise the gap that may occur between receipt and response.

Process integrity

We can assess the reliability of the work process by asking questions about the triggers for activities. However, nothing is ever 100% correct and it is inevitable that triggers will fail for some reason, activities may be missed, and perhaps the whole process fails to continue because something goes wrong. The combination of a process model together with a well-founded assessment of the reliability of each activity can allow us to assess the robustness of the whole process. If someone fails to complete some activity, and hence quite probably the next activity is never triggered, what happens? Does the whole process seize up, or will the failure eventually be noticed?

Note that this is not simply an *ad hoc* procedure. Following our approach, one can *systematically* go to each trigger and ask – what happens to the entire process if the trigger fails? Furthermore, by looking at the process as a whole we can improve our assessment of the reliability of any trigger. For example, if the trigger for an activity is that a report is in someone's in-tray, we can examine the wider context and assess the likelihood of whether the report will indeed be there when required.

The 4Rs

Although our initial focus with the trigger analysis was on the individual triggers, we began to notice an emerging pattern as we recorded the processes during our case studies. We call this pattern the 4Rs: Request, Receipt, Response, and Release.

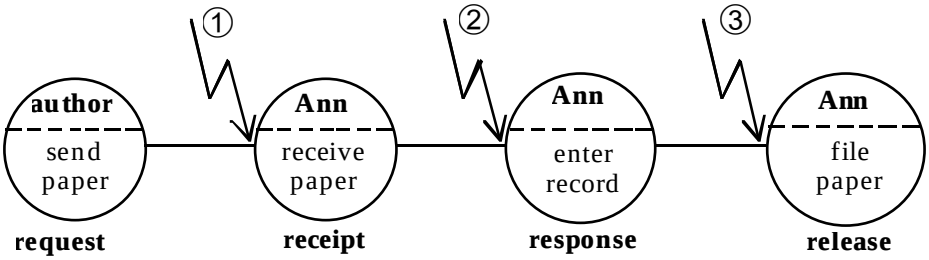

Figure 10. The 4Rs

Figure 10 shows a simplified version of one of the case studies – the operation of the conference office [Dix *et al.*, 1996], which exemplifies the 4Rs. We can see a general structure emerging: request – someone sends a message (or implicitly passes an object) requiring your action, receipt – you receive it via a communication channel, response – you perform some necessary action, and release – you file or dispose of the things used during the process. At this point, if the functional goal has been achieved then the process can be considered to have reached completion.

The papers process in figure 10 is very similar to the process that one of the authors follows when dealing with email. When the mail arrives, he reads it (or at least notes its arrival), but does not deal with it immediately – it stays in his 'in-tray' until he has replied or otherwise dealt with it. Only at that stage does he file it in a folder or discard it. If interrupted after replying, the original message is still in the in-tray (secondary trigger). Once, whilst in the middle of replying to a message, the machine crashed (interruption). When some time later he again read his email, he mistakenly (and unconsciously!) took the continued presence of the email in the in-tray as signifying an interruption before filing (secondary trigger) and hence filed the message without replying.

We believe that the 4Rs are a fundamental unit of long-term work. Not only is the pattern of activities common between different processes, but we also see a similar pattern of triggers. ① is always simply some sort of communication mode and can be assessed for reliability and timeliness. The response activity is typically triggered by ②, the presence of a document or other object. The release activity triggered by ③, which is of the 'immediately follows' kind, removes that cue, but also relies on its existence as a secondary trigger. The problems with the author's email will occur elsewhere! The existence of generic patterns makes it easier to uncover problem situations quickly and to take solutions found in one situation and adapt them to another.

Our case studies show that the 4Rs are normal – the same pattern recurs with similar triggers and similar failure modes. We have also seen that it is normative – if the 4Rs pattern is nearly followed, but with some deviation, this has been seen to be an indication of possible problems. It is also frequently the case that the response of one 4Rs pattern forms the request activity initiating a new 4Rs pattern. A chain of such 4Rs patterns constitutes a sort of long-term conversation (see figure 11). The 4Rs appear to be a pervasive, generic pattern, at a lower level than those identified in speech-act theory [Winograd and Flores, 1986], and perhaps the equivalent of adjacency pairs found in conversational analysis.

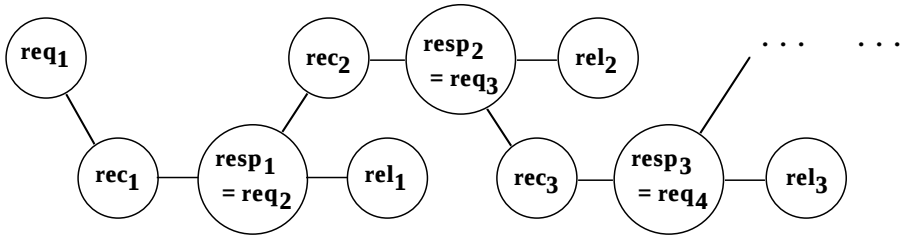

Figure 11. 4Rs chain

Example: The student placement unit

A key aspect of validating the trigger analysis was to apply the framework to prolonged interactions in an extremely busy office environment – the Placement Unit at the School of Computing & Mathematics at The University of Huddersfield. The Unit is staffed on a full- and part-time basis by administrative and academic staff respectively; responsible for helping some 200+ sandwich course students secure one-year placements in industry every year. Besides dealing with students seeking placements (predominantly recruitment and skills-building processes), the Unit also supports those already on placement (including processes in assessment and monitoring).

Also, The MaPPiT Project[1] had just been launched at the School and part of the project remit was to develop a process support system in Notes for the Placement Unit, thereby automating many placement activities. Contact with companies occurs via all media, the telephone being the most common, causing frequent interruptions. Outright winners in the interruption stakes, however, are the students for whom the Unit has an 'open door' policy between 10am and 4pm.

The diagram in Figures 12–14 exemplify the potential for activity triggers to be seriously delayed – sometimes indefinitely – had they been left to be resolved by outside companies.

Case 1 – Job adverts

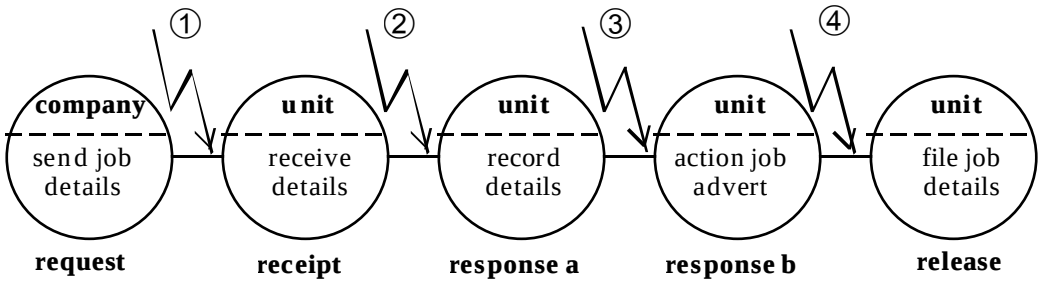

Figure 12. Initial job advert

The establishment of a new placement starts with the initial request from a company for a placement student (see figure 12). This is a 4Rs pattern, but with a two stage response. One of the dangers of multi-stage responses is that the triggers at ② and ③ are often similar, leading to problems.

In the above process the company initiates the request. Many placement providers recruit annually, diarising to send the Unit their requirements. They 'drive' the process by setting

---

[1] MaPPiT – Mapping the Placement Process with Information Technology, a HEFCE-supported 2-year project through the Fund for the Development of Teaching & Learning. Details available at: <http://www.hud.ac.uk/scom/mappit/>

closing dates for applications. Another set of companies is triggered into providing placement details by a standard letter from the Unit, sent out fortnightly via a diary system when deemed most appropriate to make the contact. Looking at ① we asked the administrative staff how they would know if details failed to arrive, thus breaking the chain of activities. At present the only back up is the diary so a time delay occurs between the failure occurring and the next fortnightly, and sometimes monthly, check for responses from the previous month's companies. A follow-on question would be – how do you remember to look in the diary? At present the answer is that the paper-based diary remains highly visible on the Placement Officer's desk. However, as the year progressed, we noticed the diary being checked less and less.

The next activity of 'Record Details' should, ideally, follow on directly from the first activity. This class of trigger is insecure and liable to interruptions – a common occurrence in the Unit. The staff member then has to remember what to do next. Usually, there is the environmental cue of paper on the desk – jotted down from a phone call, a fax copy or a letter, maybe an open email message on screen. The follow-on question was – what if you do not record the receipt of details in the diary? Another staff member could check the diary, see the assumed non-receipt of details and annoy a company by chasing for details already sent in, unless the person checked the Job Adverts Log (Response b) first or checked the company file for the ad. All this checking seemed unnecessarily complicated.

With Notes in mind the project team accepted these current problems as needing resolution. The diary could easily become electronic with built-in 'navigators' (agents) that automatically trigger reminders to execute activities. Gone is the need to remember to check the diary, as reminders appear in individuals' To-Do lists. Even To-Be-Done-To lists can be constructed. Receipt of a job description need only be recorded in one place so there is only one checking activity before chasing a company. Furthermore, the electronic record means that any inconsistency between recorded details can be displayed and thus act as a trigger at ③.

Case 2 – Submission of CVs

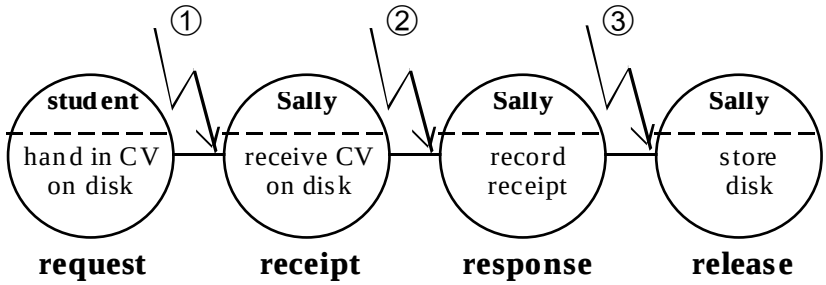

Figure 13. Students submit CVs

A large proportion of the Unit's placement providers happily accept standard CVs from students so CVs are lodged with the Unit and checked by placement tutors very early in the year. Figure 13 shows this process and again we see a straightforward 4Rs pattern.

At ʊ these two activities usually happen face-to-face so there is little risk of breakdown. We noticed that some students, unfortunately, ignored the office hours of the Unit and so 'posted' CV disks under the door after staff had left for the day, thus risking the disk subsequently being disposed of, damaged, lost or misplaced. Similarly the Unit's activities here are all exposed to interruptions and therefore incompletion. Also, staff changes in the Unit brought some new faces unaccustomed to the sub-process of receiving and recording students' disks. The circumstance arose where a student made alterations to the CV, returned the disk and the new member of staff promptly lost the update having been interrupted several times to do other more complex tasks. Furthermore, disks were sometimes found to be corrupt when the CV was required, leading to another set of interactions with the student and further delays.

The planned Notes implementation for this completely bypasses the current error-prone process. Students would fill in a CV template using a web browser. The CV is then automatically submitted to a Notes database that logs the receipt and sends an email to the Unit to confirm that the CV has been submitted on time. Students can update the CV at any time without bothering the administrative staff, in the full knowledge that the latest version will be sent to potential placement employers.

Case 3 – Company decisions

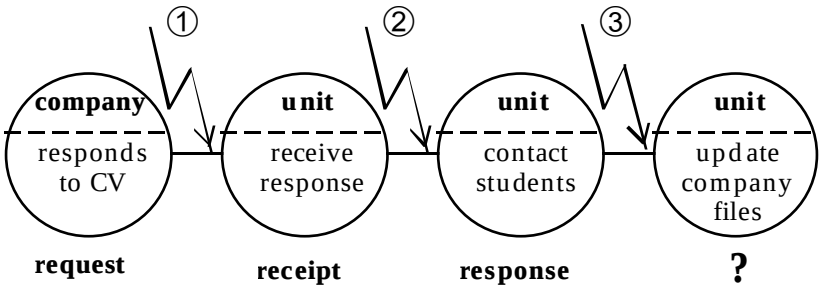

Figure 14. Company decision

After seeing the students CVs, companies decide on students to shortlist for interviewing. Figure 14 demonstrates how the pace of interaction can really slow down when pursuing students to arrange interviews or to provide feedback if rejected.

At ϖ the pace slows considerably once the students are on vacation and hard to track down. Much time can be spent trying to contact students and assuming this is successful, the next activity can be stalled by a phone call or face-to-face enquiry. We rely on an individual's good short-term memory and/or an environmental cue to ensure the sequence is fulfilled!

Although the process appears at first to be a simple 4Rs pattern, we have put a question mark against the last activity. The release usually consumes or destroys the environmental cues that have prompted previous activity. It is not clear that this is the case for this process – what are the environmental cues? The company's decision will arrive in a letter or be recorded on paper, but the slow pace of the response means that the cue may be lost or grow 'stale', ceasing to be salient because it is around too long.

Redesigning this sequence in Notes, it was decided that it would be better to record the receipt of the company's contact electronically when the contact actually happens – invariably a phone call, fax or letter. All the information can therefore be displayed on screen and so, if incomplete because of an interruption, it cannot be discarded without a prompt. Note how this action has established an environmental cue within the electronic environment. In the revised process this cue is removed when the contact details are complete, thus making the pattern a true 4Rs with robust triggers throughout. Moreover, if the response stage becomes drawn out and relies on chasing student responses, there is the possibility of automatically signalling if the expected reply is not forthcoming, thus supplying a to-be-done-to facility.

From the three cases discussed above, we can see how different levels of automation have been suggested by the trigger analysis and the 4Rs pattern. At one extreme this has involved the complete bypassing of the human process, but in the others only parts are automated. More importantly, the analysis has ensured that the Notes implementation does not hide existing triggers, as often is the case with electronic filing, but instead is explicitly designed to enhance the triggers with automatic reminders and electronic environmental cues.

Concluding Remarks

This chapter has considered the analysis of work processes with a view to understanding more clearly why things happen when they do. There is a strong indication that work activity is more prone to breakdowns and failures when the interaction is prolonged, or when the pace of interaction is slow, or when the activity is of a collaborative nature that often crossing organisational boundaries. When all three conditions are in place, such failure becomes almost a certainty.

We recommend trigger analysis as a useful method for uncovering what prompts a task to happen. One particular advantage of trigger analysis is that weaker triggers can be identified swiftly and action can be taken to render the tasks more robust, as we saw with the student placement unit example. Moreover, the classification of the triggers by type allows a simpler analysis. A further benefit of this method is that individuals may conduct investigations with little or no expertise in knowledge elicitation techniques – the follow-up questions keep it simple and the analysis can be used with virtually any other forms of task analysis.

Trigger analysis does not only stop at triggers – artefacts and placeholders also have their roles to play. Any activity in a workplace setting involves artefacts of work and we recommend that current task analysis methods should accommodate some form of tracking for artefacts, placing them on a more equal footing with the user in the task context than previously acknowledged. Just as triggers remind us that some thing needs to be done, placeholders must be robust so that we know clearly what needs to be done.

The importance of environmental cues gives us another rich source of information – the work environment itself. We look at an office: there are papers and files on the desk, post-it notes, an in-tray, a wall calendar. Why is that file on the desk? What will happen to it? What would happen if it were not there? We know that environmental cues can be triggers for activities and so we take each item in the environment and look for the activity it triggers, or the coordinating role it fulfils. At the very least a piece of paper left on the desk is saying, “File me please”.

The 4Rs framework was found to be as a fundamental unit of interaction and work - a handy ‘template’ for relatively straightforward work analysis. A key feature of the persistent pattern of 4Rs is its ability to highlight weaknesses in prolonged interaction. We can establish whether or not the 4Rs are complete, as any incomplete 4Rs have always revealed a weak trigger, environmental cue and/or placeholder.

Finally, the study of work situations should not be isolated from the real context in which that work is conducted. Trigger analysis fits easily into such environments, as it needs to examine and understand that environment. Most importantly, it can complement other task analysis methods that also suit field investigations.

In summary, trigger analysis can augment many task decomposition methods. In addition to the description of what should happen found in most such methods it captures why they occur they occur when they do and whether they are likely to be missed or mis-ordered. The strong emphasis on all types of triggers including environmental cues leads to an ecologically rich analytic approach.

Further Information

The most complete theoretically rooted account of triggers and the 4Rs is the Interacting with Computers journal article [Dix *et al.*, 1998b]. For an analysis of how rich ecological features including triggers and placeholders can augment more formal task analysis models and notations see [Dix, 2002]. Finally, for new information and hyperlinked material consult the trigger analysis web pages:

http://www.hcibook.com/alan/topics/trigger-analysis/

References

Abowd, G. and A. Dix, 1994. *Integrating status and event phenomena in formal specifications of interactive systems*. in *SIGSOFT'94*. 44–52. ACM Press.

Bennett, S., Skelton, J. and Lunn, K. (2001). *Schaum's Outline of UML*. McGraw-Hill.

Brewster, S.A., Wright, P.C., Edwards, A.D.N. (1994). *The design and evaluation of an auditory-enhanced scrollbar*. In *Proceedings of CHI'94*. 173-179. ACM Press, Addison-Wesley.

Dix, A.J. (1991). *Formal Methods for Interactive Systems*. Academic Press.

Dix, A.J. (1992). *Pace and interaction*. In *Proceedings of HCI'92: People and Computers VII*. 193-207. Cambridge University Press.

Dix, A. (1994) *Que sera sera – The problem of the future perfect in open and cooperative systems.* In *Proceedings of HCI'94: People and Computers IX*. 397-408. Cambridge University Press.

Dix, A., Ramduny, D. and Wilkinson, J. (1996). *Long-Term Interaction: Learning the 4Rs*. in *CHI'96 Conference Companion*. 169–170. ACM Press.

Dix, A., Finlay, J., Abowd, G. and Beale, R. (1998). *Human–Computer Interaction*, 2nd ed. Prentice-Hall.

Dix, A., Ramduny, D. and Wilkinson, J. (1998b). *Interaction in the Large*. Interacting with Computers - Special Issue on Temporal Aspects of Usability, Fabre, J. and Howard, S. (Eds.) 11 (1), 9-32.

Hughes, J., O'Brien, J., Rouncefield, M., Sommerville, I. and Rodden, T. (1995). *Presenting ethnography in the requirements process.* In *Proc. IEEE Conf. On Requirements Engineering, RE'95*. 27–34. IEEE Press.

Jacobsen, I. (1992). *Object-Oriented Software Engineering.* Addison-Wesley.

Norman, D.A., (1986). *New views of information processing: Implications for intelligent decision support systems.* In Hollingel E. et al. (Eds.) *Intelligent Decision Support in Process Environments,* Springer-Verlag, Berlin.

Norman, D.A., (1988). *The Psychology of Everyday Things*. Basic Books.

Payne, S.J., (1993). *Understanding Calendar Use.* Human-Computer Interaction. **8**(2): pp 83-100.

Paternó, F. (1999). *Model-based design and evaluation of interactive applications.* Springer-Verlag.

Shepherd, A. (1995). *Task analysis as a framework for examining HCI tasks,* In Monk, A. and Gilbert N. (Eds.) *Perspectives on HCI: Diverse Approaches.* 145–174. Academic Press.

Walsh, P. (1989). *Analysis for task object modelling (ATOM): towards a method of integrating task analysis with Jackson system development for user interface software design*. In D. Diaper (Ed.) *Task Analysis for Human-Computer Interaction*. 186-209. Ellis Horwood

Winograd, T. and F. Flores (1986). *Understanding computers and cognition: a new foundation for design*. Addison-Wesley.

Wood, A., Dey, A.K., Abowd, G.D. (1997) *CyberDesk: automated integration of desktop and network services.* In *Proceedings of the 1997 conference on Human Factors in Computing Systems*, CHI'97, pp 552-553.