PaintShop: A MICROWORLD EXPERIMENT INVESTIGATING TEMPORAL DECISIONS IN A SUPERVISORY CONTROL TASK

Michael Hildebrandt and Michael Harrison
{hilde, mdh}@cs.york.ac.uk
Department of Computer Science
University of York, York YO10 5DD, UK

This paper explores temporal aspects of control behavior in order to support the design of systems where functions can be allocated flexibly in time. Dynamic Function Scheduling, an extension of Dynamic Function Allocation approaches, highlights the role of temporal information and temporal reasoning in supervisory control decisions. The microworld experiment presented 30 participants with a supervisory control task where they had to monitor production in a simulated paint station, make strategic decisions about automatic or manual production, and handle faults. Independent variables were event rate, the knowledge of event rate information, the availability of an online progress indicator, and the cost of fault servicing. Results showed that knowledge of event rate information improved performance, but availability of an online progress indicator had no additional effect. Implications for the investigation of temporal control behavior are discussed.

INTRODUCTION

Temporal aspects rank highly among the defining characteristics of today's socio-technical systems: these systems are usually subject to real-time requirements, they involve adaptation to dynamic changes, and they often operate under multi-task conditions. One approach for designing such systems is to consider joint human-automation control, and distribution of workload through Dynamic Function Allocation (DFA). This solution provides adaptivity over time, but in a way that is best described as "snapshot allocation" (Hildebrandt and Harrison, 2002, 2003): DFA decisions are made on the human-automation resource dimension, but fail to take account of the additional degree of freedom provided by allocating functions along a joint human-automation timeline. This second type of allocation decision, which we refer to as Dynamic Function Scheduling (DFS), is defined as *any adaptive control decision in a dynamic system where time is the input* (e.g. in terms of temporal information, temporal knowledge) *and/or the output* (e.g. scheduling decision) *of a decision process*. Analysis of DFS decisions is supported by formal modelling (e.g. queuing models or task representations or model checking; c.f. Loer et al., in print), but an understanding of the trade-offs and benefits involved in time design choices requires richer notions of time (Hildebrandt et al., 2004), where time is seen as

- *a property of the automation or interface:* e.g. service rate, responsiveness, display of temporal information, temporal validity, interface support for temporal awareness
- *an aspects of user behaviour:* e.g. perceptual/physiological timing issues, temporal orientation, anticipative/reactive control mode, temporal reasoning, temporal memory, reaction to time stress, pace of interaction, personal/social attitudes towards time
- *a property of the task*: e.g. interleavability, pre-emptability
- *a property of the environment:* predictability/regularity of task arrival, self-paced/system-paced interaction, deadlines

This is the first in a series of experiments designed to investigate temporal control behaviour, use of temporal knowledge, and biases in temporal decision-making. Participants were faced with a simulated production plant where time was an aspect of the system (service rates), the task (event rate) and the control decisions (manual or automatic processing; trade-offs in fault servicing between time and financial expenses). The aim of these experiments is to explore the strategies and information sources used to achieve and maintain temporal awareness; the use of flexible scheduling in adaptive control decisions; biases in decisions involving temporal and monetary costs; and the effect of workload levels on these measures.

METHOD

The experiment simulated an industrial paint station with two parallel production lines (Fig 1). The participant's task was to supervise the production process and intervene in case of problems. Unpainted items arrived on a conveyor belt at a rate of one item every 2, 3, or 4 seconds (within-subjects factor *EventRate*). When items approached the paint station, they were distributed by a "lift" to one of the two parallel production lines. By default, items would be allocated to whatever station was available. However, the participant was able to override this automation by setting the lift to *up* or *down* (all items to upper or lower station).

Each 3-minute trial was divided into three 1-minute blocks of high, medium or low event rate (i.e. one item every 2, 3 or 4 seconds). Combinations of these blocks resulted in nine types of trials: three trials with identical event rates over the three blocks (*Low-low-low; Med-med-med; High-high-high*), and six mixed trials (*Low-med-high, Low-high-med*, etc.).

Each type of trial was presented twice, under each of the levels of the within-subjects factor *AvailabilityProgressIndicator*: during half of the trials, participants could request an online progress indicator, during the other half of the trials this feature was unavailable.

Half the participants were informed of the event rate distribution prior to the trial, the other half did not know the block structure (between-subjects factor *AvailabilityEventRateInf*). For the first group, this information was also displayed within the progress indicator display.

For each paint station, two painting modes were available, *automatic* (4 seconds) or *manual* (2 seconds). Each painted item "earned" the participant 1 pence.

Faults developed every 8-14 paint cycles as the nozzles become increasingly blocked. It took 6 items to progress from the initial blockage to a full blockage (indicated by the pressure meter moving towards the "critical" area). Once this state was reached, the station would break and be unavailable for further processing. This would lead to newly arriving items piling up in front of the lift, unless the items could be painted on the other station quickly enough. Once more than four items piled up, each newly arriving item was removed.

Faults could be serviced either by *repairing* the station (unavailable for 24 seconds) or by *replacing* the blocked nozzle (available immediately, but cost of 8p, 6p or 12p, depending on between-subjects factor *ReplaceCost*). Repair and replace decisions could be made at any time, even before a station had broken.

During training, participants were given information about the different strategies for controlling the PaintShop. They were made aware that the different event rates could be handled thus:

- **High**: both stations *auto*, or a single station *manual* mode,
- **Medium**: both stations in *auto* mode, or a single station *manual*, with slack time (which could be used e.g. for reducing a backlog).
- **Low**: one station in *auto* mode is sufficient (any other combination creates slack time).

Participants were also made aware of the relative costs of replace vs. repair (assuming both stations are unavailable and there is a risk of items being removed):

- **6p:** *replace* is as good a choice as *repair* during low, and better under medium and high event rate.

- **8p:** *replace* is a worse choice than *repair* during low, equally good during medium, and better during high event rate.
- **12p:** *replace* is a worse choice than *repair* during low and medium, and equally good during high event rate.

Participants were also reminded to take information about event rate boundaries into account. For instance, if a switch to a high event rate is imminent, participants might want to opt for *replace* even though the current event rate is low.

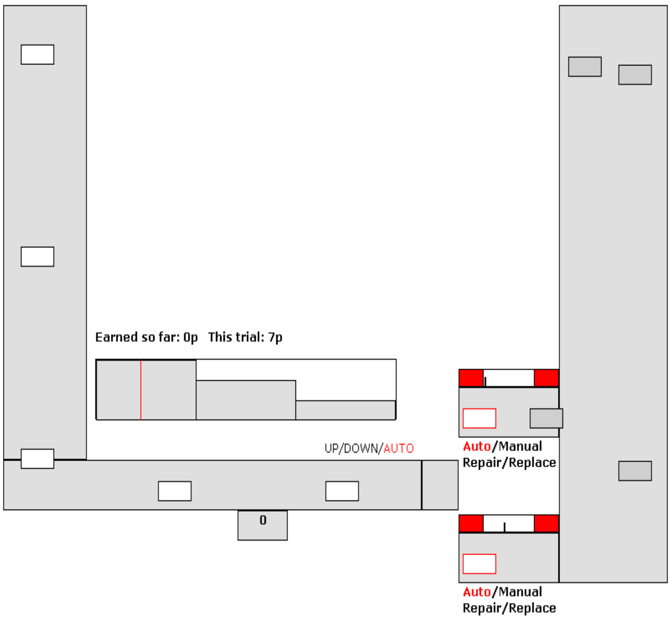

Figure 1. Screenshot of the PaintShop micro-world simulation

RESULTS AND DISCUSSION

Dependent variables were: number of painted items, money earned per block, and the types of control decisions. After each trial NASA-TLX ratings were recorded and participants were asked to provide any comments on the control strategies they used. Data was analysed using a repeated-measures 3x2x3x2 ANOVA (within-subjects factors *EventRate*, *AvailabilityEventRateInf*; between-subjects factors *ReplaceCost*, *AvailabilityProgressIndicator*). Presented below are two selected results (a publication with the full set of results is in preparation).

Event rate information

Results showed a significant reduction in the number of removed items when event rate information was available (Fig. 2). Closer analysis revealed that this effect was specific to the 6p and 8p *ReplaceCost* condition (Fig. 3). Note that the effect of *ReplaceCost* is due pre-trial event rate information, as the availability of the online progress indicator had no additional effect.

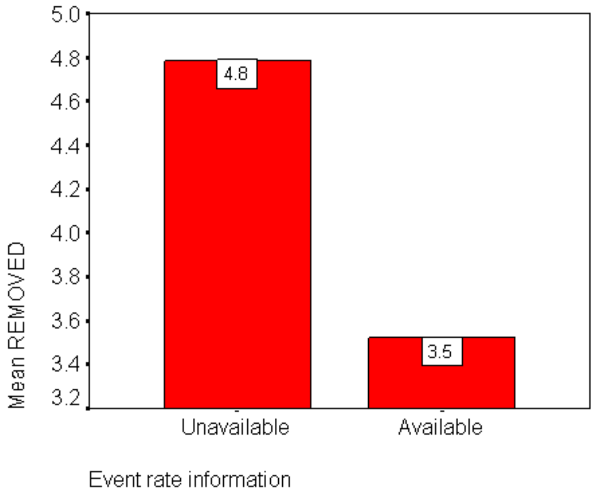

Figure 2. Main effect of *AvailabilityEventRateInf* on removed items.

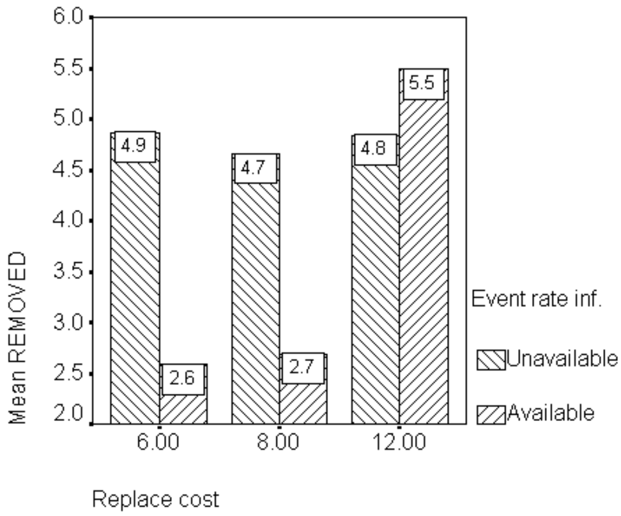

Figure 3. Interaction effect of *ReplaceCost* and *AvailabilityEventRateInf* on removed items.

However, the use of the progress indicator itself is affected by *EventRate*. Participants' use of the progress indicator significantly increased from high to low workload periods (Fig 4).

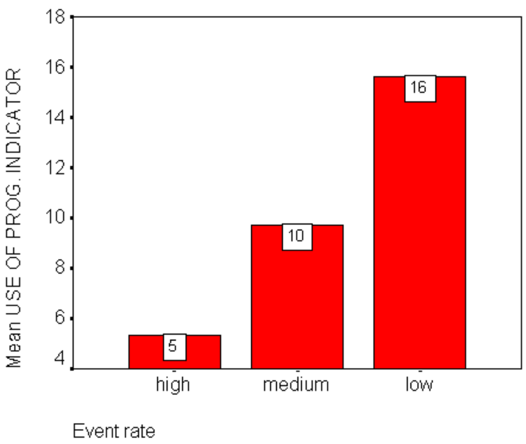

Figure 4. Effect of *EventRate* on use of progress indicator.

Temporal control strategies

To illustrate the richness of temporal control strategies used on the task, selected comments from the feedback forms participants filled in after each trial are quoted below:

a) 8p repair cost,event rate unavailable, timescale unavailable
"I tried to anticipate the best time when to repair/replace the nozzles. However, just when the workload was very high and I decided to replace them, the workload decreased and I felt a certain level of frustration because of the loss of points."

b) 6p repair cost, event rate available, timescale unavailable
"I wanted to set both machines to repair half way through the medium section so they were both ready for the heavy period, but not having the timeline made that very difficult to judge."

c) 8p repair cost,event rate unavailable, timescale unavailable
"Some strategies have developed: (1) I route all traffic onto one station, when this is broke I repair it and use the other. (2) Depending on frequency, I change to manual paint. (3) Replace must be avoided. [*Next trial, timescale unavailable*] I think I have to revise my strategy of not using replace. If the frequency is high […], replace seems to be necessary as backlog increases too quickly. [*Next trial, timescale available*] This time I used the timeline to anticipate changes in frequency. I also repaired stations before breakdown."

d) 12p repair cost, event rate unavailable, timescale available
"The timeline is very useful to check how far through you are and also if the event rate is about to change so you can plan ahead. […] But it is only really useful in medium or low periods when you have time to look at it."

e) 8p repair cost,event rate unavailable, timescale unavailable
"Definitely works best having one machine being repaired all the time and using the other in manual mode. Usually seem to lose about 6 or 7 during a run so better than replacing the nozzles. [*Later trial, timescale unavailable*] Much harder than previous ones. I lost 17 so a couple of replaces would have been useful but it is hard to know when to use replace when the next batch may come at a slower rate."

f) 12p repair cost, event rate available, timescale unavailable
"Just alternate between the two stations, one in action while the other is repaired. Repair pre-emptively. [*Later trial, time scale unavailable*] Attempted same strategy […] However it didn't work, this was due to the zero slack time available, meaning that every time I switched machines and hit the repair button, a build-up was occurring. 15 lost and 2 replacements."

g) 12p repair cost, event rate and timescale unavailable
"Lack of time indicator makes it more difficult – as you cannot tell if it is worth repairing a nozzle (e.g. just in case the next interval will be a faster one)."

DISCUSSION

The results indicate that temporal information in the form of prior knowledge of the event rate distribution did indeed improve performance. Surprisingly, this effect cannot be attributed to the availability of the online progress indicator, which, for the group of participants who received event rate information, displayed this information *plus* an indication of the current stage of the trial (and of the point in time when the next event rate change was to occur). In the light of this data pattern it is difficult to speculate about the interpretation of the effect of event rate information. One could assume that the provision of event rate information simply makes the participant more aware of possible workload changes, or focuses the participant's attention in some other way. However this explanation cannot account for the interaction of event rate availability with replace cost, which shows that availability of event rate information *only* reduces the number of removed items in the 6p and 8p condition, whereas in the 12p condition the number of removed items is similar to the condition where event rate information is unavailable. Moreover, the number of removed items is virtually identical across replace cost conditions for the group of participants who received *no* event rate information. As the number of removed items is inversely proportional to the number of *replace* decisions, it could be hypothesized that the availability of event rate information enables or encourages a more appropriate consideration of the trade-offs involved in the *repair* vs. *replace* decision. Participants' comments suggest that this may indeed be the case (c.f. comments *a*, *c*, *d*, *e* and *g*). However, in the absence of a significant effect of event rate information on replace decisions, this explanation remains speculative. Future studies should investigate the determinants and purposes of use of event rate information in more detail. Such studies should address the relation between online and offline temporal information, memory for and awareness of offline temporal information, resolution of conflicts between temporal information, and the effects of unreliable, inaccurate or untimely temporal information.

The decreasing use of the progress indicator at higher event rates (Fig. 4, comment *d*) highlights a common scheduling problem: High workload periods, when accurate control decisions are most necessary, are exactly the periods when fewer attentional resources are available for taking them.

CONCLUSION

The contributions of this line of work to a Human Factors-centric approach to work design are twofold. First, it helps to outline the characteristics, limitations, determinants and biases of temporal control behavior and temporal error. It thereby contributes to a research program in the Human Factors community that has recently begun to recognise time as an important but under-researched domain (e.g. De Keyser, 1995). Such work goes beyond the traditional view of time as merely an external constraint (deadline view), as merely a descriptive property of behaviour (epiphenomenal view), or of decision-making under time pressure (time stress view). Instead it aims to identifiy the various ways in which time is used as information, internalised, externalised, reasoned about, attended to or fading into the background, how time and other decision parameters are traded off against each other, and in what ways time is an aspect of the outcome of a control decision. The crucial question in this domain is whether these decisions can be modelled with general-purpose models, or whether human temporal cognition has properties that distinguish it from other decision domains. A growing literature on temporal factors in judgement and decision-making (e.g. Varey and Kahneman, 1992) suggests this might indeed be the case.

The second contribution of this work relates to the understanding of Dynamic Function Scheduling design choices, or adaptive control choices in general. The long-term aim of this work is to provide a structured method for charting a design space that identifies the various dimensions of temporality (e.g. constraint, information, decision outcome) and allows the assessment of temporal design solutions alongside other options such as adaptive automation. A first step towards this shift in perspective has been achieved by adding a temporal dimension to DFA approaches. The next step is to construct suitable representation tools that allow the expression of rich temporal properties, i.e. the temporal task structure, the workload requirements faced by the system, and the temporal properties of the agents (e.g. service rate, temporal reasoning and planning abilities).

REFERENCES

De Keyser, V. (1995). Time in ergonomics research. *Ergonomics, 38*, 1639-1660.

Hildebrandt, M., Dix, A. & Meyer, H.A. (2004). Time design. In E. Dykstra-Erickson & M. Tscheligi (Eds.), *CHI 2004 Conference on Human Factors in Computing Systems* (pp. 1737-1738).

Hildebrandt, M. & Harrison, M.D. (2003). Putting time (back) into Dynamic Function Allocation. *Proceedings of HFES 2003* (pp. 488-492).

Hildebrandt, M. & Harrison, M.D. (2002). The temporal dimension of Dynamic Function Allocation. *Proceedings of 11th European Conference on Cognitive Ergonomics* (pp. 283-292).

Loer, K.F., Hildebrandt, M. & Harrison, M.D. (in print). Analysing dynamic function scheduling decisions. To appear in *Proceedings of IFIP 13.5 Working Conference on Human Error, Safety and Systems Development*.

Varey, C. & Kahneman, D. (1992). Experiences extended across time: Evaluation of moments and episodes. *Journal of Behavioral Decision Making, 5*, 169-185.