

Decision support and vulnerability to interruption in a dynamic multitasking environment

Helen M. Hodgetts^{1,2}, Sébastien Tremblay¹ Benoît R. Vallières¹ and François Vachon¹

¹ École de psychologie, Université Laval, Québec, Québec, Canada

² Department of Applied Psychology, Cardiff Metropolitan University, Cardiff, United Kingdom

Correspondence to: Helen Hodgetts, Department of Applied Psychology, Cardiff Metropolitan University, Cardiff, CF5 2YB, or François Vachon, École de psychologie, Université Laval, Québec (QC), G1V 0A6, Canada. E-mails: hhodgetts@cardiffmet.ac.uk or francois.vachon@psy.ulaval.ca

Abstract

Using a microworld simulation of maritime decision making, we compared two decision support systems (DSS) in their impact upon recovery from interruption. The Temporal Overview Display (TOD) and Change History Table (CHT) – designed to support temporal awareness and change detection, respectively – have previously proven useful in improving situation awareness; however, evaluation of support tools for multitasking environments should not be limited to the specific aspects of the task that they were designed to augment. Using a combination of performance, self-report, and eye-tracking measures, we find that both DSS counter-intuitively have a negative effect on performance. Resumption lags are increased, elevated post-interruption decision-making times persist for longer, and defensive effectiveness is impaired relative to No-DSS. Eye-tracking measures indicate that in the baseline condition, participants tend to encode the visual display more broadly, where as those in the two DSS conditions may have experienced a degree of attentional tunnelling due to high workload. We suggest that for a support tool to be beneficial it should ease the burden on attentional resources so that these can be used for reconstructing a mental model of the post-interruption scene.

Keywords: Interruption; multitasking; microworld; command and control; human performance; decision support systems

1. Introduction

In complex command and control (C2) situations such as crisis management, emergency response, aviation, or military operations, working towards a higher order goal generally involves performance of a number of concurrent subtasks (e.g., monitoring, planning, visual search, decision making). Although multitasking is necessary for many operations, generally a cost is incurred relative to focusing on one task individually (Wickens, 2002). Interruptions are another characteristic of C2 situations, whereby the need to integrate an unexpected task within an already complex activity can further compromise the speed or accuracy of task performance (Hodgetts & Jones, 2006a,b; Trafton, Altmann, Brock, & Mintz, 2003). Given that multitasking is unavoidable in composite tasks, it is important to understand how these costs can be minimized and how support tools might be offered to assist the limited cognitive resources of the human operator. Designing and evaluating such support systems is complex, since a potential solution tested in one domain – for example the Change History Explicit tool (CHEX; see St. John & Smallman, 2008) designed to aid change blindness – may then have a negative impact on a different facet of C2 (e.g., classification accuracy) when multiple subtasks are performed together (Vallières, Hodgetts, Vachon, & Tremblay, 2012). It is therefore critical to take a holistic approach to ensure that any support tool benefitting one aspect of performance does not incur costs to another. Microworlds provide a useful methodology for studying multitasking since they go some way towards preserving the complexities and interdependencies of a real world task, but within a simplified and controlled setting. They allow us to test the effect of support tools across the whole task, and not just the specific isolated variable that they were intended to augment. In the current study we examine two decision support systems (DSSs) – originally designed to improve specific aspects of situation awareness – within a multitasking naval air-warfare microworld task in which interruptions occur.

Decision support and vulnerability to interruption

1.1 Decision support

Support systems are increasingly being developed to assist human operators in dynamic decision making tasks (Gonzalez, 2005). In order to provide support for one aspect of multitasking, it is important not only to understand the underlying cognitive mechanisms implicated in that particular subtask, but also the interaction with the other subtasks involved. Some interface design features can have unintended consequences, reducing error in one domain while exacerbating another (Imbert, Hodgetts, Parise, Dehais, Vachon, & Tremblay, 2014; Trafton & Ratwani, 2014). In the context of maritime C2, we consider temporal awareness and change detection as two elements of a multifaceted task that may benefit from external support tools. In terms of temporal awareness, operators in dynamic and time-pressured environments need to have an overarching internal representation of the timecourse of events – both past and future – and the interactions between them. The temporal overview display (TOD; Rousseau, Tremblay, Lafond, Vachon, & Breton, 2007) provides an on-screen grid-based representation of temporal events that goes beyond a mere snapshot in time, to facilitate the processes of planning and executing complex series of actions (e.g., Potter, Gualtieri, & Elm, 2003). In a naval air-warfare simulation task, TOD was successful in promoting a time-based decision heuristic, improving temporal awareness (Tremblay, Vachon, Rousseau, & Breton, 2012) and situation awareness (Vachon, Lafond, Vallières, Rousseau, & Tremblay, 2011). Dealing with interruptions is a common necessity in C2 situations (e.g., responding to a colleague or a system alert), which must be interleaved with the ongoing task so as to minimize overall disruption. TOD's explicit representation of temporal information may be beneficial to participants experiencing task interruptions by helping them to plan, coordinate and prioritise more efficiently in the post-interruption phase, thus reducing the time needed to get 'back on track'.

Another device designed to augment situation awareness (SA; Endsley, 1995) in C2 tasks is the Change History Explicit tool (CHEX; see St. John & Smallman, 2008) which

Decision support and vulnerability to interruption

automatically detects situational changes in the immediate environment and logs them in a table dynamically linked to the geospatial display. By referring to the text in the table, it eases the burden on the operator to notice changes occurring on the radar which may be subtle and difficult to detect on a visually cluttered and dynamically changing screen. Using a simplified version of this tool, Vachon et al. (2011) found that participants not only demonstrated better SA than those performing the task without support, but also possessed a more accurate perception of their level of SA. This feature may be particularly useful in the case of task interruption when the operator's attention is necessarily directed elsewhere and the visual scene continues to evolve. By automating part of the task, a CHEX-like tool can alert the operator to changes that have occurred, and in turn facilitate the process of regaining SA after an interruption (St. John & Smallman, 2008). Although both the TOD and CHEX tools have been successful in supporting the particular processes they were designed to augment (temporal awareness and change detection, respectively), the question remains as to how useful these tools are in multitasking situations in which interruptions occur. The CHEX tool was developed within a change-detection-only environment that did not require performance of other concurrent subtasks, and TOD has not been tested when interruptions have the potential to break the task's focus. There may be a fine balance between a support tool beneficially automating part of the task on one hand, and detrimentally imposing additional visual information load on the other (Perry, Wiggins, Childs, & Fogarty, 2013). Given that interruptions are known to increase perceived workload (e.g., Kirmeyer, 1988), within an already complex multitasking environment there is the potential that the presence of a support tool could place too high demands on attentional resources and actually impair performance. The current study uses a holistic approach to examine these two support tools beyond their original purpose, by assessing their impact on interruption recovery.

1.2 Theoretical perspective

Decision support and vulnerability to interruption

One influential framework in the domain of task interruption is Memory for Goals (MfG; Almann & Trafton, 2002) which derives from the ACT-R (Adaptive Control of Thought – Rational) cognitive architecture (e.g., Anderson & Lebiere, 1998). It is based on the idea that all memory elements have a fluctuating level of activation and if suspended – because of interruption for example – that level will decay over time. The activation of suspended task goals can be boosted by the processes of strengthening (e.g., rehearsal) and priming (e.g., associative activation from environmental cues that are present both when the goal is suspended and when it is to be resumed). If one assumes that these processes are effortful, then the success or timecourse of interruption recovery might be dependent upon the availability of attentional resources at goal suspension and resumption. In this regard, a DSS that automates some part of the task could be expected to ease interruption recovery by freeing up attentional resources at critical points around the interruption, thus facilitating the processes of strengthening and priming. On the other hand however, any addition to the interface that is too complex or attention-demanding may in fact direct resources away from important associative cues in the task environment, and consequently compromise the strengthening and priming processes.

While the memory-based processes described in the MfG framework are useful for understanding resumption in static tasks, it is unclear the extent to which they can be applied to interruption in C2 task environments. Salvucci (2010) argues that in applied domains in which interruptions are in the order of several seconds/minutes – and the complexity of the task would involve more than a single memory retrieval – memory-based accounts are difficult to reconcile. We would also add that many applied environments are dynamically changing, and so although spatial memory may guide resumption in static tasks (Ratwani & Trafton, 2008, 2010), memory for pre-interruption locations and information may be of little use post-interruption in an evolving task (e.g., Hodgetts, Vachon, & Tremblay, 2014). Salvucci argues that task resumption in applied environments must involve a process of reconstruction, incorporating perceptual,

Decision support and vulnerability to interruption

cognitive and motor behaviors, starting by creating a new problem state to replace the problem state lost. In accordance with this viewpoint, there should perhaps be less of an emphasis on the seconds immediately before and immediately after interruption, and more of a need to assess the processes and strategies at play over a longer post-interruption recovery period. Whether resuming a task after interruption involves memory retrieval or reconstruction, it is likely that the effect of the DSS on workload – whether positive or negative – will play a critical role in this process.

1.3 Microworlds

To study the effect that changing one aspect of a C2 task may have on another requires a task environment that takes into account some of the complexities and interdependencies of the real-world situation, but with a high level of experimental control. Microworlds are interactive computer-based tasks that allow the study of human behaviour – as individuals or teams (e.g., Brehmer & Dörner, 1993; Tremblay, Vachon, Lafond, & Kramer, 2012) – within a controlled scenario. They are cognitively demanding and engage a variety of cognitive functions such as situation assessment, decision making, monitoring, complex problem solving, causal learning and planning (Gonzalez, Vanyukov, & Martin, 2005), and are often characterized by added stressors such as uncertainty, temporal pressure, and limited resources (Gonzalez et al., 2005; Granlund & Johansson, 2004). Unlike many computer-based studies that simply administer a task and display static information, microworlds are better able to capture the complex, dynamic, multitasking type of environment in which interruptions frequently occur. As well as mundane realism, microworlds offer high tractability (Gray, 2002); the researcher can easily control the environment in a way that is only possible with a computer-based task and not in the field (e.g., observations of naturally occurring interruptions). It is this capability to isolate specific variables that allows researchers to establish cause-and-effect relationships at a functional level that can be generalised beyond the original task environment. Microworlds are also ideal for the study of

Decision support and vulnerability to interruption

interruptions because they allow for a range of dependent measures which may be differentially impacted by interruption (e.g., error, time costs), or may provide converging support for the behavioural effects observed (e.g., eye movements, subjective measures). This facilitates the use of a holistic approach (Lafond, Vachon, Rousseau, & Tremblay, 2010), allowing researchers to achieve a comprehensive understanding of performance as a whole, and of interactive effects between variables (e.g., Hodgetts et al., 2014; Vachon et al., 2011).

1.4 The current study

In the current study, the task we use is the Simulated Combat Control System (S-CCS) microworld (Lafond et al., 2010; Vachon et al., 2011) which provides a simplified simulation of above-water command and control warfare practiced aboard the Canadian Navy's frigates. It simulates at a functional level the processes involved in monitoring and risk assessment and shares similarities with the Argus Prime radar task (Scholles & Gray, 2001) and the Ballas task (Ballas, Heitmeyer, & Perez-Quinones, 1992). The participant plays the role of a tactical coordinator who must monitor changes in the operational space, and conduct aircraft threat assessments including categorization and prioritization of threats. Decisions are based on criteria very similar to those required in real-world scenarios. A baseline condition with no DSS was compared to two support systems originally designed to support specific aspects of the task (temporal awareness and change detection). The first, a temporal overview display (TOD) was designed to aid planning and execution of activities by explicitly presenting each aircraft across a timeline according to its proximity to the ownship. The second, a Change History Table (CHT), was designed in accordance with the CHEX tool (Smallman & St. John, 2003; St. John, Smallman, & Manes, 2005), and automatically detects and logs changes to the airspace in a table dynamically linked to the geospatial display. Although these two DSS have both shown to provide support in this C2 task by improving SA (Vachon et al., 2011), within a multitasking environment there are many other aspects of performance that should also be taken into account

Decision support and vulnerability to interruption

before concluding that a tool is beneficial. Here we look at whether these DSS may help or hinder the recovery from interruptions, a common occurrence in a C2 environment.

In order to study dynamic task interruption, eye movements were analysed in an event-based manner, using the interruption as an anchor point and comparing periods immediately before and after. Although not providing a full insight into all facets of attention, eye fixations provide a useful proxy for which part of the screen a participant is attending to (e.g., Goldberg & Kotval, 1999), although their interpretation may depend upon context (see, e.g., Poole & Ball, 2006). Quicker fixations are associated with rapid encoding (Gartenberg, McCurry, & Trafton, 2011), while longer fixations are linked to higher cognitive processing loads (Callan, 1998; Recarte & Nunes, 2000) and difficulty in extracting information (e.g., Goldberg & Kotval, 1999; Just & Carpenter, 1976). We expected shorter fixations following interruption to reflect the rapid encoding process as participants attempt to regain SA (Gartenberg et al., 2011; Hodgetts et al., 2014). Previous research has shown that spatial memory can guide task resumption, and participants are generally very good at resuming a static visual search task after brief interruptions (Lleras, Rensink, & Enns, 2005); indeed, in keeping with the MfG theory, participants in interrupted tasks tend to look back to the pre-interruption location for associative cues to prime goal retrieval (e.g., Ratwani & Trafton, 2008; 2010). The role of spatial memory in visual search tasks is illustrated by the finding that initiating a new search display takes longer than resuming an old display (Lleras et al., 2005), and changes to task relevant features of a target (like location) will affect the search process (Lleras, Rensink, & Enns, 2007). In the current study we will examine whether spatial memory guides resumption by assessing the concordance rates between pre and post-interruption fixation points, in each DSS condition. However, in our dynamic microworld task, the visual scene continues to evolve during the interruption and so the scene that participants return to will not be the same as that which they

Decision support and vulnerability to interruption

left; as such, it is unsure the extent to which pre-interruption locations will be useful in reactivating goals associated with the moving aircraft.

We compared interrupted with uninterrupted trials across each DSS condition in terms of behavioural performance measures, self-reported workload and eye movements. It was anticipated that the behavioural performance measures (decision-making times and defensive effectiveness) would be negatively affected by interruption because taking attention away from the primary task would break focus and incur a cost to performance. The effect of the two DSS would depend on the extent to which they ease – or add to – the burden on cognitive processes during task resumption; and self-reports of workload were expected to concur with these findings.

2. Method

2.1 *Participants*

Sixty-two students at Université Laval (32 men; mean age = 23.24 years) participated in the two-hour experiment and received CAD \$20 for their time. All reported normal or corrected-to-normal vision and normal hearing. They were randomly assigned to each DSS condition: No-DSS (n = 21), TOD (n = 21), CHT (n=20).

2.2 *Apparatus/Materials*

The S-CCS microworld (see Vachon et al., 2011), run on a PC computer, provides a functional simulation of threat evaluation and combat power management processes (i.e., planning, execution, and situation monitoring). Participants were required to perform three concurrent tasks: 1) to determine the threat level (hostile, non-hostile, uncertain) of all the aircraft on the radar screen; 2) determine the threat immediacy of hostile contacts (i.e. time until they hit the ship); and 3) engage a missile to neutralize a hostile contact. The visual interface is made up of three parts: a radar screen, a parameters list, and a set of action buttons (see Figure 1). The ownship is represented by a dot in a circle at the center of the screen, while aircraft move

Decision support and vulnerability to interruption

in the vicinity in real time. Each aircraft is represented by a white dot surrounded by a green square, with a line attached that indicates the speed and direction of the aircraft (line length is proportional to the aircraft speed). Each scenario started with five aircraft and involved 27 aircraft in total (maximum of 10 at any one time). Sixteen 4-minute scenarios were created which were equivalent in difficulty (e.g., number of aircraft, number of hostile aircraft) but differed in surface characteristics such as parameter values or different aircraft trajectories.

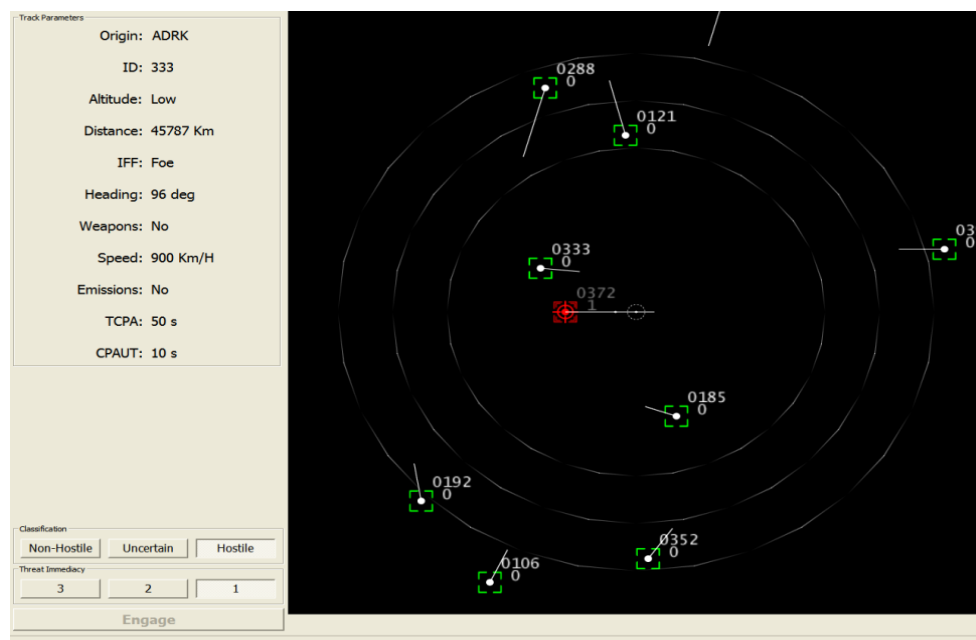


Figure 1. Screenshot of the S-CCS microworld visual interface.

To assess threat level of each aircraft, participants accessed the parameter values by clicking on the aircraft icon with the mouse, at which point the surrounding square would turn red. The parameters list displayed a range of information relating to the aircraft, some of which were not included in the threat assessment task (e.g., heading, distance, speed). Five critical parameters each displayed one of two options, whereby one was classed as threatening and the other was not: country of origin (*ADRK* = threatening), altitude (*low* = threatening), intention friend or foe (IFF; *foe* = threatening), weapons detected (*yes* = threatening), and military emissions (*yes* = threatening). Based upon these, participants were required to classify each

Decision support and vulnerability to interruption

aircraft as either non-hostile (0 or 1 threatening parameters), uncertain (2 or 3 parameters), or hostile (4 or 5 parameters), and click the associated action button. Once classified, the level of threat assigned to the aircraft was indicated by a change in color of the aircraft's white dot to green (non-hostile), yellow (uncertain), or red (hostile). Participants needed to sometimes check back at the parameters of already-classified aircraft in case circumstances changed and the threat level needed reassessing. For aircraft classified as hostile, the participant then had to take further steps to rate the immediacy of that threat (on a scale 1 to 3), based upon the TCPA parameter (Time to Closest Point of Approach; < 15 s, 15–30 s, or > 30 s, respectively). The participant was then required to neutralize that aircraft by launching a defense missile: Clicking on the 'engage' button launched a missile with a 2-second delay, and only one could be airborne at any one time. There were eight hostile aircraft in each scenario that were programmed to hit the ship.

Eye movements were monitored throughout the task using a Tobii T1750 eye tracker (Tobii Technology, 2006), integrated into a 17-in monitor with a resolution of 1024×768 pixels and with a sampling rate of 50 Hz. Participants were seated approximately 50 cm from the eye-tracker computer screen, and eye movements were calibrated after the microworld familiarization session. All eye movement data were analyzed using ClearView software (Tobii Technology, 2006).

2.3 Manipulations

One third of participants completed the task without a DSS. Another third of participants were assigned to the TOD condition: The interface featured a grid on the right-hand side of the screen which presented the same information about speed, distance, and direction as the geospatial display but as a single visual representation (Figure 2a). The grid featured vertical lines corresponding to specific temporal intervals, and a single red vertical line that indicated "now". Each aircraft was represented by a horizontal rectangle that moved from right to left, and when the right end of this rectangle crossed the red line then ship would be hit. Clicking one

Decision support and vulnerability to interruption

of the rectangles in the timeline highlighted the contact location on the radar, and equally every action made on the geospatial display occurred simultaneously in the TOD.

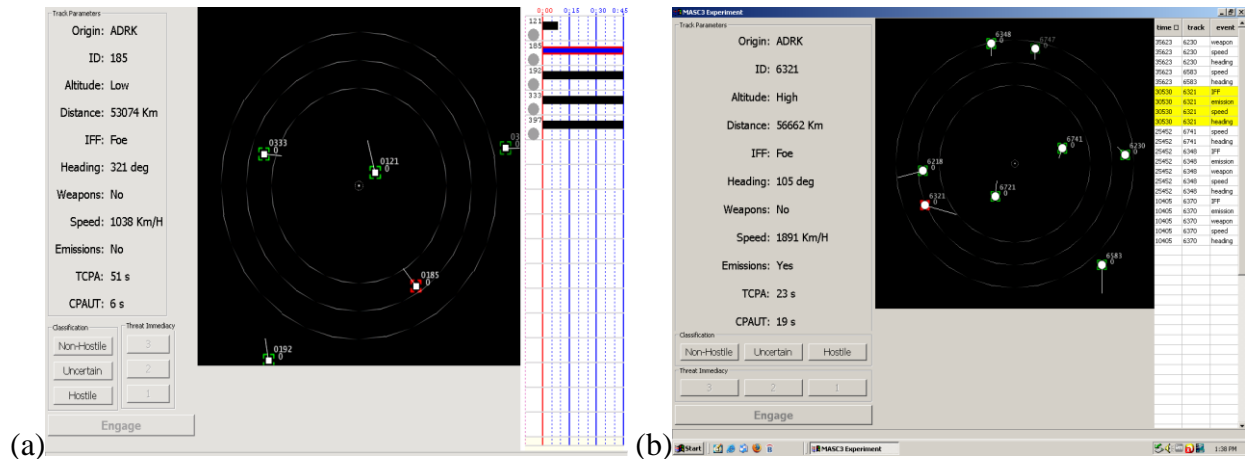


Figure 2. The S-CCS microworld visual interface with a DSS on the right of the screen, (a) the Temporal Overview Display and (b) the Change History Table.

In the CHT condition, a table was added to the right-hand side of the interface which automatically detected and displayed all changes as a permanent record, and in chronological order (Figure 2b). Its three sortable columns allowed the participants to determine whether a particular aircraft made a certain type of change and when this change happened. This table was dynamically linked to the geospatial display so that clicking on an aircraft on the radar highlighted associated entries in the table, and vice versa. As well as a comparison between the different DSS conditions, we also compared interrupted and uninterrupted scenarios. Half of scenarios were interrupted for 24 s at a point between 55 s and 125 s into the 4-min burst. The whole S-CCS interface went blank and displayed three consecutive questions regarding the status of the mission which required a yes/no answer by clicking with the mouse. These questions were intended to simulate requests for information from an external authority. Questions changed automatically every 8 s, until after 24 s when the S-CCS interface was restored. Participants resumed the same scenario that they had been engaged in previously,

Decision support and vulnerability to interruption

although this would have continued to evolve in real time during the interruption interval. Of course in real-world scenarios, being interrupted away from the primary task could impair performance by obscuring a critical event occurring during that time; however, in the current study no aircraft became hostile during the interruption (or during the 20 s immediately preceding or following), thus allowing one to examine the impact of the interruption itself (i.e., a break in cognitive focus of the task), rather than on the consequences of missing a critical event whilst otherwise engaged. Furthermore, there was a 40s hostile-free period at the equivalent point in uninterrupted trials, which ensured the equivalence in task difficulty across all scenarios. This strict control on the occurrence of hostile aircraft around the interruption (or where the interruption would have occurred on control trials) enabled a direct comparison between interrupted/uninterrupted trials, and more specifically, a comparison within trials, in terms of differences in the speed/nature of cognitive processes operating prior to and directly after interruption.

2.4 Measures

For all eye movement analyses, the threshold to detect a fixation was set at 100 ms and the fixation field corresponded to a circle with a 30-pixel radius (equivalent to 1.15 degrees of visual angle when seated at a distance of 50 cm). For each condition, we recorded the duration of individual fixations around the interruption period, the total dwell times on different areas of interest (AOIs) on the screen and the number of transitions between AOIs, as well as the location of first fixations after interruption (and their concordance with the last location previously fixated prior to interruption). Resumption time was the time between the offset of interruption and the first action on the task (mouse click). Decision-cycle time was recorded as the time between selecting one aircraft and selecting the next: This incorporated parameter assessment and classification, possibly an immediacy assessment and weapon engagement (for hostile aircraft), and the time to search and select the next aircraft. One of the participants' tasks was to

Decision support and vulnerability to interruption

identify and neutralize hostile aircraft before they hit the ship, and so a behavioral measure of defensive effectiveness was used which related to how close the ship came to being hit. Finally, self-reports of mental load and time pressure were taken after each scenario by clicking at the appropriate number on a 10-point Likert scale where 1 = low, 10 = high (adapted from the NASA-TLX questionnaire; Hart & Staveland, 1988).

2.5 Procedure

Participants read through a PowerPoint tutorial at their own pace which explained the context of the simulation and the tasks to execute. To check understanding of all the information given, participants were presented with nine static screenshots from the microworld task and asked to perform the threat classification task and the threat immediacy task, if applicable. They then familiarized themselves with the microworld simulation in a 1-min session before completing two training sessions. After a 15-min break, participants completed four blocks, each comprising four scenarios of four min each. Participants were allowed a short rest between each test session, after which a summary of instructions was presented on screen and participants clicked a 'Continue' button to initiate the first scenario. Following each 4-min scenario, subjective workload was measured using the NASA-TLX technique (Hart & Staveland, 1988). Within each block, half of scenarios featured an interruption.

3. Results

The three DSS conditions were compared in relation to their ability to support recovery from interruption using a range of objective and subjective measures. All post hoc tests used the Bonferroni correction for multiple comparisons.

3.1 Decision-Cycle Time

Mean decision-cycle time was first analyzed using a 3 (DSS condition: none, TOD, CHT) \times 2 (interruption or no-interruption) mixed design ANOVA (Figure 3). However, this revealed no significant effects of interruption, $F(1, 59) < 1$, or DSS condition $F(2, 59) < 1$, and

Decision support and vulnerability to interruption

no significant interaction, $F(2, 59) = 1.54, p = .22$. It is possible that by averaging across entire scenarios, any effects of interruption may have been lost: The interruption did not occur until one minute or so into the scenario, and one might presume that any effects following interruption would not persist for the remaining duration of the 4-min burst. Thus it seems wise to consider differences in decision making speed confined to time periods specifically around the interruption.

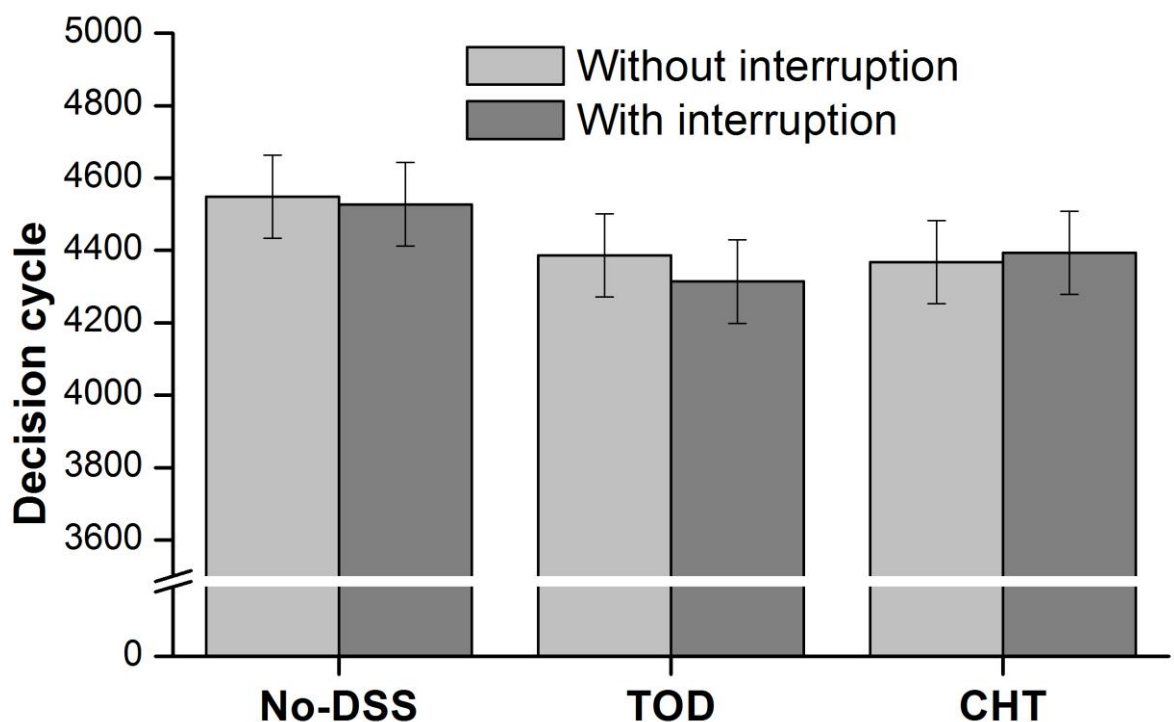


Figure 3. Mean decision-cycle time (ms) according to DSS condition and the presence/absence of interruption. Error bars represent 95% confidence intervals with Masson and Loftus' (2003) method.

We then restricted our analysis to just 80 s of the scenario, comparing four critical time intervals: the 20 s before interruption, and the three consecutive 20s intervals afterwards (i.e., -20s, +20 s, +40 s, +60s around the interruption), for each of the three DSS conditions (Figure 4). These specific epochs were chosen based on the mean duration of a decision cycle (4 to 5 s on

Decision support and vulnerability to interruption

average), allowing sufficient data points to calculate a mean for each time period. The 20s period before interruption and the 20s period after, were equivalent in terms of complexity as no aircraft turned hostile during these times. A 4 (time interval) \times 3 (DSS condition) mixed design ANOVA revealed a significant main effect of time interval, $F(3, 177) = 6.37, p < .01, \eta_p^2 = .10$ with a significant increase in decision cycle time in the 20s immediately following interruption compared to the 20s preceding. Although there was no effect of DSS, $F(2, 59) < 1$, there was a significant interaction, $F(6, 177) = 2.51, p = .02, \eta_p^2 = .08$. Post hoc tests showed that in comparison to pre-interruption decision cycle time, levels in the DSS conditions were still marginally elevated at +40s ($p = .08$), while decision cycle time in the No-DSS condition had returned to pre-interruption decision making speed already by this point (no difference between -20s and +40s).

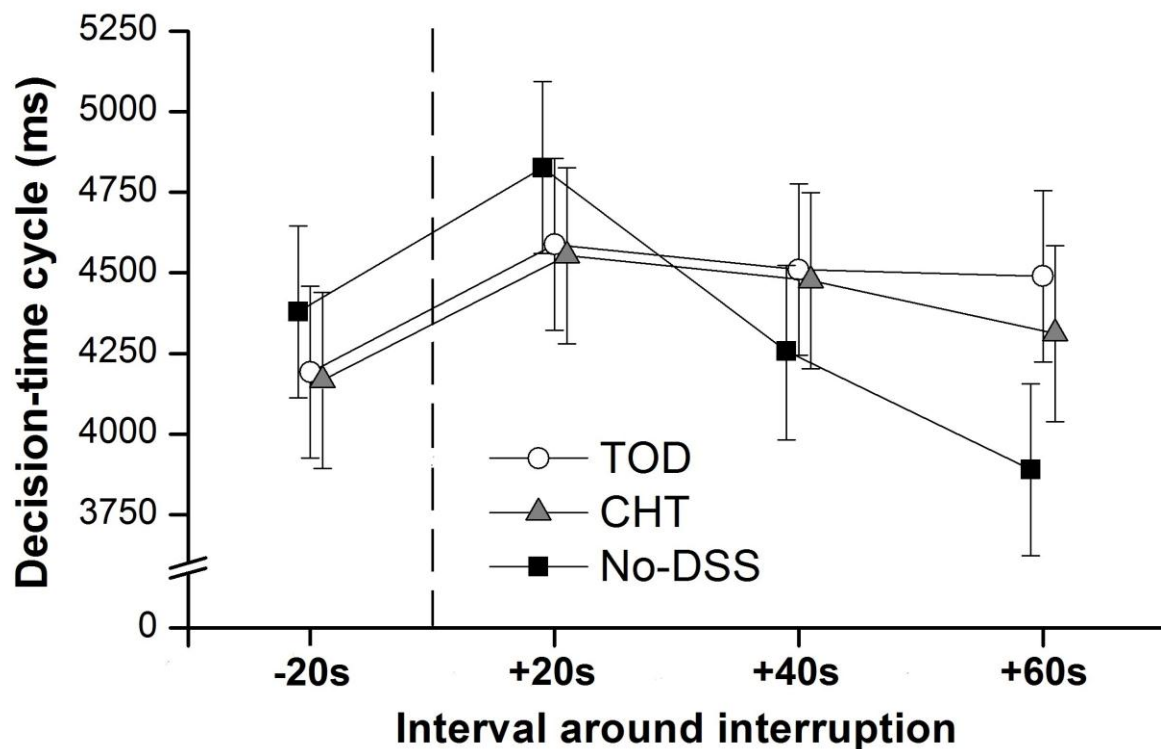


Figure 4. Mean decision-cycle time (ms) according to time interval around the interruption and DSS condition. Interruption line represents 24 s of time in the task. Error bars represent 95% confidence intervals with Masson and Loftus' (2003) method.

3.2 Resumption lag

To take an even more precise measure of task resumption, we calculated the time taken to make the first action on the task (e.g., to select an aircraft or click on a button) following the offset of interruption. Mean resumption lags (ms) were as follows: Control (2148.47, SD = 598.04), TOD (2954.15, SD = 1399.92), CHT (2569.13, SD = 479.57). Due to a violation of Levene's test of equality of variance, the data were log transformed before being submitted to a one-way between subjects ANOVA which showed a significant effect, $F(2, 59) = 4.57, p = .01, \eta_p^2 = .13$. Post hoc tests showed that resumption time in the control condition was significantly quicker than in either the TOD or CHT conditions, which did not differ significantly from each other.

3.3 Eye Movements

Eye-tracking data were collected for each DSS condition and in accordance with the same four time intervals used previously (Figure 5). Shorter fixations are thought to reflect the process of rapidly encoding of a visual scene (Gartenberg et al., 2011). A 4 (time interval) x 3 (condition) mixed design ANOVA showed a main effect of time interval, $F(3, 177) = 7.23, p < .01, \eta_p^2 = .11$, such that fixations were significantly longer in the 20s before interruption than in the next two time periods following interruption (+ 20s and +40s) ($ps < .01$). Fixation duration at +60s was no different to pre-interruption levels. Although Figure 5 suggests a trend for fewer fixations in the TOD condition, the main effect of DSS condition did not reach significance, $F(2, 59) = 2.50, p = .09$. There was no significant interaction between interval and condition, $F(6, 177) < 1$.

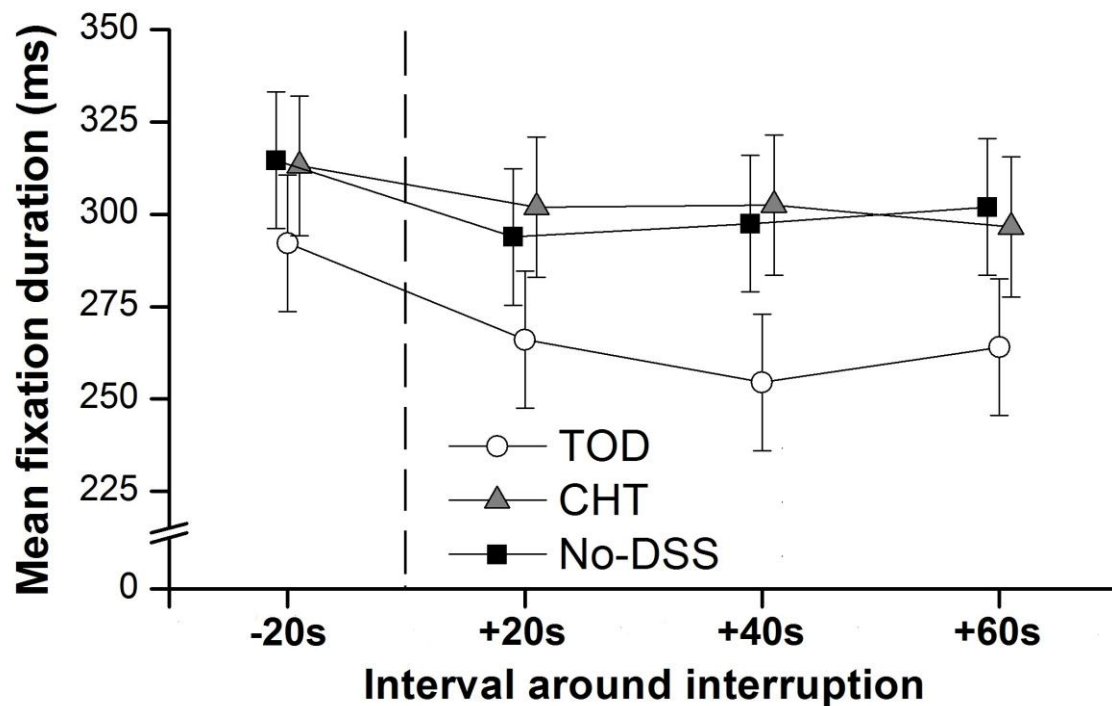


Figure 5. Mean fixation duration (ms) according to time intervals around the interruption and DSS condition. Interruption line represents 24 s of time in the task. Error bars represent 95% confidence intervals with Masson and Loftus' (2003) method.

Given the significant difference between conditions in terms of resumption lag, we took a more detailed look at eye movements and fixations during this post interruption period specifically. Figure 6 defines four areas of interest (AOIs) on the screen: central radar, peripheral radar, parameters, and action buttons. In the case of the TOD and CHT conditions, the support tool formed a fifth AOI.

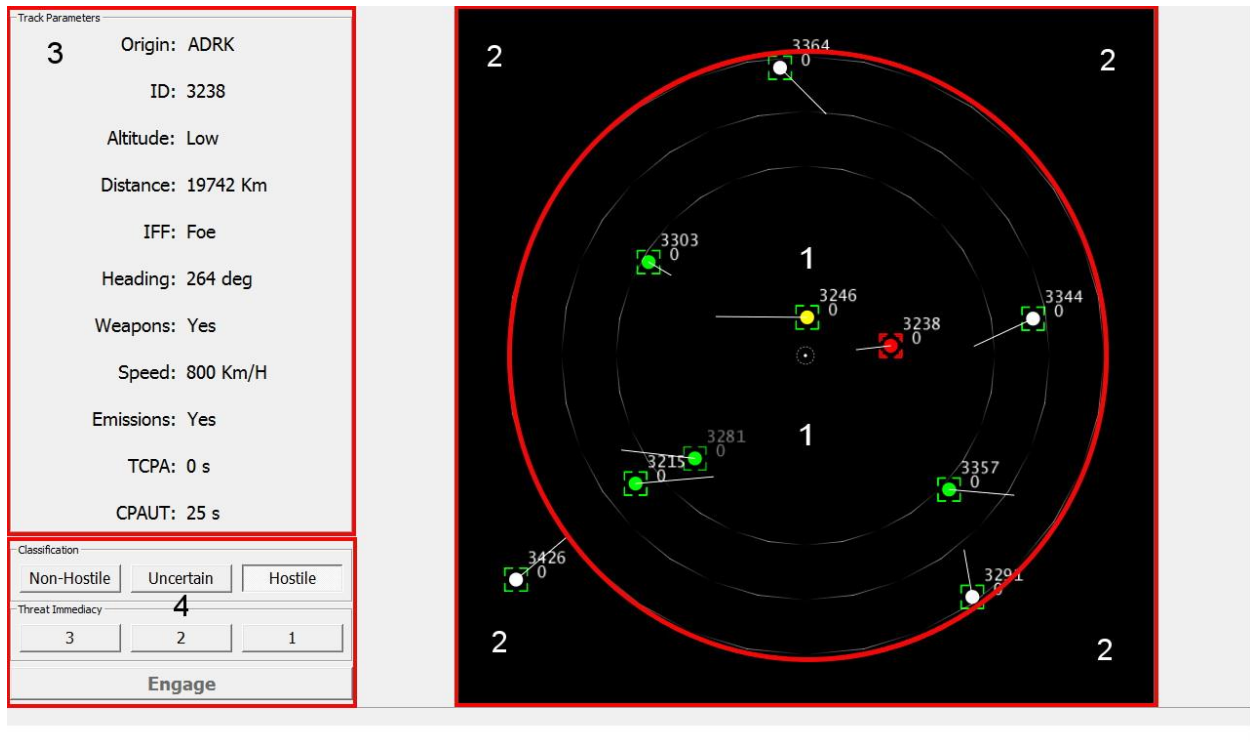


Figure 6. Areas of interest in the control condition: Central radar (1), peripheral radar (2), parameters (3), action buttons (4). In the TOD and CHT conditions, the tool created a fifth AOI (not shown).

We calculated the mean number of transitions between separate AOIs during the first 3s following interruption (to roughly correspond with the resumption lag), as differences in gaze behavior could potentially account for differences in resumption time. In the No-DSS condition the mean number of transitions was 2.67 (SD = 0.59), in TOD this was 2.01 (SD = 0.64) and 1.85 (SD = 0.82) for CHT. This represented a significant effect of DSS condition according to a one-way ANOVA, $F(2, 59) = 8.17, p < .01, \eta_p^2 = .22$. Post hocs showed significantly more transitions between separate AOIs in the No-DSS condition than when a DSS was present ($p < .01$), with no difference between TOD and CHT. We also looked to see if this difference persisted during the longer post-interruption recovery period from 0 – 40s after interruption (to roughly correspond to the point at which decision cycle times return to pre-interruption levels).

Decision support and vulnerability to interruption

The same pattern was true during this longer post interruption period, $F(2, 59) = 6.26$, $p < .01$, $\eta_p^2 = .18$, with significantly more transitions made in the No-DSS group (mean = 4.56, SD = 0.77), than with TOD, (mean = 3.89, SD = 0.82), $p < .05$ or with CHT (mean = 3.63, SD = 1.02), $p < .01$. The difference between TOD and CHT was not significant. Thus during interruption recovery in the No-DSS condition, participants change their gaze towards different areas of the screen more frequently than when a DSS is present.

To further decompose the nature of interruption recovery, we examined the specific areas of the screen that participants looked to in order to resume the task. Specifically, we recorded the location of the first fixation after interruption. Table 1 shows that participants in the No-DSS condition most often looked at the radar periphery, while in the TOD and CHT conditions participants tended to look first at the centre of the radar. Although it was not possible to perform statistical analyses on these data due to the number of empty cells, the differences are compelling.

Table 1. *Frequency of first fixations after interruption on each of the designated AOIs, in accordance with the final fixation made before the interruption occurred. Data includes 8 interruptions per participant in each condition, a) No-DSS (n = 21) b) TOD (n = 21) and c) CHT (n = 20). Any fixations on non-designated AOIs are not included.*

(a)					
No-DSS					
AOI before	AOI after interruption				
	Radar	Periphery	Parameters	Buttons	Total
Radar	2	7	0	0	9
Periphery	24	79	10	5	118
Parameters	4	23	2	0	29

Decision support and vulnerability to interruption

Buttons	0	8	1	0	9
Total	30	117	13	5	165

(b)

TOD		AOI after interruption				
AOI before	Radar	Periphery	Parameters	Buttons	DSS	Total
Radar	133	2	0	1	8	144
Periphery	6	0	0	0	0	6
Parameters	8	1	0	0	0	8
Buttons	3	0	0	0	1	4
DSS	2	0	0	0	0	2
Total	152	3	0	1	9	165

(c)

CHT		AOI after interruption				
AOI before	Radar	Periphery	Parameters	Buttons	DSS	Total
Radar	98	0	8	8	3	117
Periphery	0	0	0	0	0	0
Parameters	24	0	4	0	0	28
Buttons	3	0	1	0	0	4
DSS	0	0	0	0	0	0
Total	125	0	13	8	3	149

We also recorded the last area that participants had fixated before the onset of interruption; Table 1 shows first post interruption fixation in accordance with the last pre-interruption fixation. We were interested to see whether any condition had a higher concordance rate between pre and post interruption fixations as this could indicate a better spatial memory to

Decision support and vulnerability to interruption

guide task resumption. For No-DSS the concordance rate was 49.40% (SD = 18.74), while this increased to 64.38% (SD = 18.26) for CHT and 79.17% (SD = 16.93) for TOD. A between participants ANOVA showed these differences to be significant, $F(2, 59) = 14.37, p < .01, \eta_p^2 = .33$. Pairwise comparisons showed that concordance rates in the No-DSS condition were significantly lower than in the CHT condition ($p < .01$), and CHT concordance rates were significantly lower than in the presence of TOD ($p < .05$). The concordance rates in TOD and CHT conditions were largely accounted for by participants focusing on the centre of the radar both when the task was interrupted and when it was to be resumed.

Table 2. *Mean (and standard deviation) dwell time (ms) on each AOI according to condition (control, TOD, CHT) and post interruption time period (up to 3s and up to 40s post interruption).*

AOI	No-DSS		TOD		CHT	
	0 -3 s	0 – 40s	0 – 3 s	0 – 40 s	0 – 3s	0 – 40 s
Radar centre	407 (200)	3197 (1212)	1117 (409)	9698(3719)	1128 (460)	12158(3825)
Radar periphery	1118 (325)	10365(2396)	88 (113)	891 (672)	9 (23)	1586 (3580)
Parameters	476 (226)	11419(3023)	372 (217)	7879(3299)	563 (327)	8396 (4104)
Action buttons	118 (75)	3387 (1276)	79 (87)	3118(1218)	123 (145)	2689 (1391)
DSS (if available)	--	--	111 (193)	1650(2427)	97 (185)	1016 (1133)

To go beyond first fixations, we then examined which areas of the screen participants fixated during both the short-term and longer-term processes of interruption recovery. Table 2 shows mean dwell time (ms) on each AOI during the resumption lag (0 – 3 s after interruption), and during the 0 – 40s post interruption recovery period. For 0 -3 s, a 4 (AOI, excluding DSS) x 3 (condition) mixed design ANOVA showed main effects of AOI, $F(3, 177) = 98.32, p < .01, \eta_p^2$

Decision support and vulnerability to interruption

= .63, and condition, $F(2, 59) = 4.10, p = .02, \eta_p^2 = .12$, and also a significant interaction between the two, $F(6, 177) = 57.42, p < .01, \eta_p^2 = .66$. Post hoc tests showed that during the resumption lag in the control condition – similar to the first fixations data – participants spent significantly more time looking at the periphery than those in the TOD or CHT conditions, and they fixated on this area significantly more than any other AOI. With TOD and CHT, the radar centre was fixated significantly more than any other area, and significantly more than in the control condition. Given that there was no DSS tool on screen in the control condition, dwell time on the fifth AOI was analyzed separately for the TOD and CHT conditions. Dwell time on the tool was low: 6.3% of time in the TOD condition and 3.9% of time in the CHT condition. An independent t-test showed no difference between TOD and CHT conditions, $t(39) < 1$. During the longer 40s post-interruption period there were also main effects of AOI, $F(3, 177) = 90.64, p < .01, \eta_p^2 = .61$, and DSS condition, $F(2, 59) = 5.38, p < .01, \eta_p^2 = .15$, as well as a significant interaction between the two, $F(6, 177) = 55.85, p < .01, \eta_p^2 = .65$. Pairwise comparisons showed that during this longer recovery period, in the control condition participants spent significantly longer fixating the radar periphery and the parameters, than the centre of the radar or the action buttons. Again, an opposite pattern was found when a DSS was present: Participants in both the TOD and the CHT conditions spent significantly longer fixating the centre of the radar than any other AOI, and significantly less time on the periphery. Dwell time on the DSS tool was again analyzed separately for the TOD and CHT conditions. In the TOD condition, 7.1% of time in the 40s following interruption was spent looking at the tool, and in the CHT condition this was just 5.0% of time. An independent t test revealed no significant differences, $t(39) = 1.08, p = .29$.

3.4 Defensive effectiveness

One part of participants' mission was to identify and neutralize hostile aircraft as soon as possible, and before they hit the ship. A behavioral measure of defensive effectiveness was used, being defined as the sum of the time-to-ship values for all hostile contacts at the point they were

Decision support and vulnerability to interruption

destroyed (a higher score thus indicating that they were destroyed sooner, further away from the own ship). A total of zero would mean that all hostile contacts that attempted to hit the own ship during the period of reference succeeded in doing so. The total was then divided by the number of hostile contacts in order to obtain an average time-to-ship value which is easier to interpret. Greater values indicate a greater defensive effectiveness. This measure was analyzed across the scenario as a whole, because there were too few data points to examine specific epochs around the interruption (Figure 7). A 3 (DSS condition) \times 2 (interruption: present or absent) mixed design ANOVA showed a significant effect of interruption, $F(1, 59) = 18.44, p < .01, \eta_p^2 = .24$, with reduced defensive effectiveness in interrupted scenarios. There was also a main effect of DSS condition, $F(2, 59) = 4.44, p < .02, \eta_p^2 = .13$, with post hoc comparisons indicating that defensive effectiveness was significantly better in the No-DSS than the TOD condition ($p < .02$), while the difference between No-DSS and CHT was approaching significance ($p < .10$). The interaction between DSS and interruption was non-significant, $F(2, 59) < 1$.

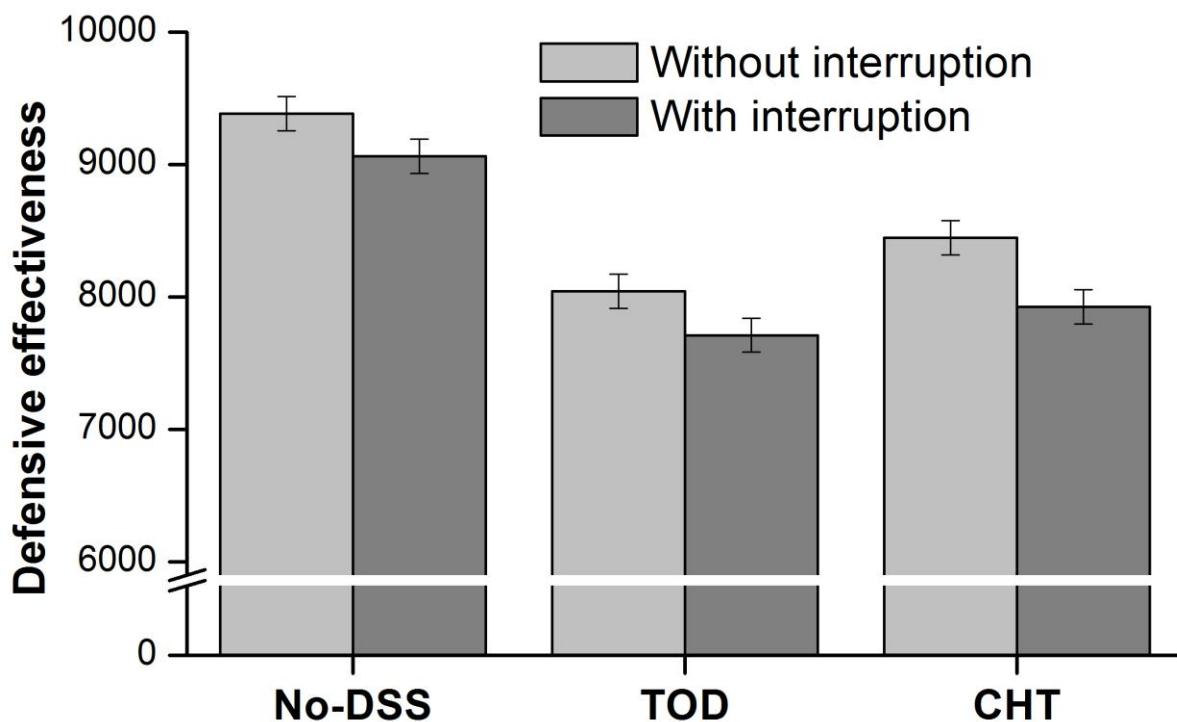


Figure 7. Defensive effectiveness according to DSS condition and presence/absence of interruption. Error bars represent 95% confidence intervals with Masson and Loftus' (2003) method.

Another point of interest was whether participants' defensive effectiveness improved over time, and whether this differed according to DSS condition. That is, we were interested in whether there was a training effect such that performance in the DSS conditions might be lower initially as the complex interface takes longer to learn, but then increase differentially as the DSS tool becomes more familiar and has the potential to facilitate performance. Defensive effectiveness was calculated over time in a 3 (DSS condition: none, TOD, CHT) by 4 (scenario over time: 1st, 2nd, 3rd, 4th) mixed design ANOVA. One participant was removed from the TOD condition for this analysis as they were an outlier (> 2.5 SD from mean) in the second scenario. A Greenhouse-Geisser correction was used due to a violation of sphericity. The main effect of scenario order approached significance, $F(2.03, 117.52) = 2.72, p = .07$; but although there appeared to be a trend for improved performance across time (1st mean = 8143, SE = 206; 2nd mean = 8385, SE = 236; 3rd mean = 8403, SE = 229; 4th mean = 8682, SE = 238), no pairwise comparisons were statistically significant (all $ps > .20$). There was no significant interaction between DSS condition and scenario over time, $F(5.18, 152.90) < 1$.

3.5 Workload

Perceived workload was measured using the mental load and temporal pressure scales of the NASA-TLX, whereby participants indicated a score on a 10-point Likert scale after each burst (Figure 8). A 3 (DSS condition) \times 2 (interruption: present or absent) mixed design ANOVA showed that mental load differed according to DSS condition, $F(2, 59) = 4.03, p < .05$, with post hoc comparisons confirming that mental load was significantly greater in the CHT

Decision support and vulnerability to interruption

condition than control. There was however, no effect of interruption $F(1, 59) < 1$, and no interaction, $F(2, 59) = 2.13, p = .07$.

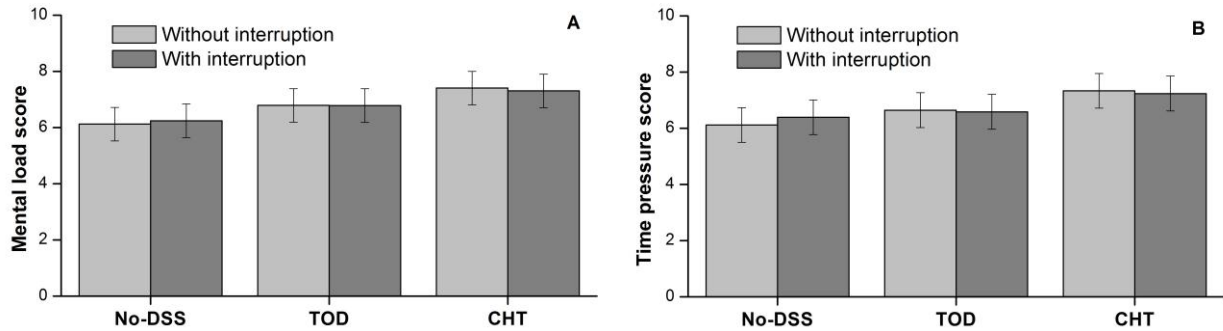


Figure 8. Mean mental load (A) and temporal pressure (B) scores according to DSS condition and the presence/absence of interruption. Error bars represent 95% confidence intervals with Masson and Loftus' (2003) method.

In terms of temporal pressure, there was again a significant effect of DSS condition, $F(2, 59) = 3.16, p < .05$, with post hoc tests demonstrating that pressure was significantly greater in the CHT than control condition. There was no effect of interruption, $F(1, 59) = 1.44, p = .24$, but there was a significant interaction, $F(2, 59) = 5.68, p < .01$. Post hoc analyses showed that interruption significantly increased temporal pressure in the control group, but did not increase levels of pressure for TOD or CHT which were already high.

4. Discussion

We assessed the use of two decision support systems in a multitasking situation in which interruptions occurred: The TOD – which provides temporal information about the situation – and the CHT – which provides information about changes occurring in the environment – have previously shown to be successful in increasing SA in a C2 decision-making task (St John & Smallman, 2003; Vachon et al., 2011), but the question raised in the current study was whether these tools could still augment SA in a multitasking environment in the face of interruption, and facilitate the interruption recovery process. We observed a clear impact of interruption in terms

Decision support and vulnerability to interruption

of eye movements (shorter fixations in the period following interruption) and performance variables (increased decision cycle times). These metrics demonstrated that the impact of interruption is not just limited to the few seconds needed to resume work on the suspended task, but can persist for up to a minute before full SA is restored. In terms of the two DSS, we found that rather than helping, the presence of a support tool on the screen actually prolonged interruption recovery and impaired performance (i.e., in terms of eye movement patterns, increased resumption lags, higher perceived workload, decreased defensive effectiveness). Although it is conceivable that defensive effectiveness with a DSS could improve over time as participants learn to use the more complex interface, this did not appear to be the case over the four experimental scenarios in this study. We highlight the need to combine multiple measures in the assessment of support tools, as those designed to enhance one aspect of multitasking e.g., scheduling (TOD) or change detection (CHT), may incur costs to another (recovery from interruption), and consequently result in a negative impact on the overall operator-interface relationship.

The two DSS – both designed to improve aspects of SA – failed to facilitate the interruption recovery process. Post interruption dwell times showed that participants spent very little time looking at the tools, and so this could explain why their presence did not improve performance. Perhaps participants felt that they could miss critical information on the radar if they devoted time and resources to the support tool, and preferred instead to rely on the naive realism provided by the geo-spatial display (Smallman & St John, 2005). However, it was not the case that the support tools were simply irrelevant, since their mere presence on the screen actively impaired performance relative to the No-DSS condition. Resumption lags were increased and the elevated post interruption decision cycle times persisted for longer. In this complex, dynamic, and time pressured environment, participants in the TOD or CHT conditions struggled to regain SA as quickly as those in the baseline condition. This negative impact of

Decision support and vulnerability to interruption

DSS, specifically during the post-interruption period, suggests a limitation in the availability of – or a conflict in the coordination and allocation of – attentional resources, at critical points in the task when workload is highest.

We looked for differences in eye movement behavior to try to explain these effects. Concordance rates between pre and post interruption fixations were actually higher in the TOD and CHT conditions than with no DSS, thus the delayed recovery cannot be related to worsened spatial memory in the presence of an onscreen support tool. The high concordance rate was largely accounted for by participants focusing on the central area of the radar both before and after interruption; furthermore, participants spent the highest proportion of post-interruption dwell time fixating the centre of the radar. On the other hand, participants in the No-DSS group spent most time fixating the radar periphery. One might presume that the central region of the radar would be the more beneficial area to fixate, but fixations on the peripheral area associated with the No-DSS condition gave rise to better performance. We speculate that fixations on the radar periphery (an area of limited relevance to the task) were actually an attempt to encode information from two areas concurrently (e.g., the central radar and the parameters). The idea that participants in the No-DSS condition were perhaps encoding a greater area of the interface during resumption is supported by the finding that these participants also made significantly more transitions between separate AOIs. If participants in the No-DSS group were allocating attention over wider areas of the screen then they may regain SA more quickly, explaining why participants in this condition appeared to experience more efficient interruption recovery.

Why then, might participants in the No-DSS condition allocate attention over wider areas of the screen compared to when a support system forms part of the interface? We suggest that the mere presence of a support tool on the screen can alter the way a task is performed and can add additional load (McCrickard, Catrambone, Chewar, & Stasko, 2003; Rousseau et al., 2007). Subjective reports of workload indicated higher mental load and temporal pressure in the CHT

Decision support and vulnerability to interruption

condition; furthermore, the decrease in defensive effectiveness in the TOD (and tendency in the CHT) condition relative to control also suggests an issue of higher load. Although participants were not fixating the tools for long periods, the decision of when, or even if, to switch attention towards an automated tool can impose an overhead cost. When operators are overloaded they can experience attentional tunneling (Chan & Courtney, 1993; Wickens & Alexander, 2009) whereby their attention focuses on one part of the task, often to the detriment of other aspects. Attentional tunnelling can be problematic if other parts of the task are neglected, and especially if warning signals are missed due to excessive focus (Beringer & Harris, 1999; Dehais, Causse, Vachon, & Tremblay, 2012; Dehais, Tessier, Christophe, & Reuzeau, 2010). Given the high workload demands of C2 tasks, it is important that a DSS designed to ease cognitive burden does not inadvertently compound these effects.

The MfG framework (Altmann & Trafton, 2001) emphasises that the efficiency of interruption recovery is dependent upon the opportunity for task encoding and priming processes at critical points before and after interruption (Hodgetts & Jones, 2006a; Ratwani, Andrews, Sousk, & Trafton, 2008). Our findings would seem compatible with this theory if we assume that the TOD and CHT – rather than easing the burden of cognitive processing – were actually more resource-demanding than the condition with No-DSS at all. These demands could have led to a diminished opportunity for encoding associative cues at goal suspension, and consequently a longer time was required to regain SA. However, given the dynamic nature of our task and the fact that a previously fixated target would have been substantially displaced during the course of the 24s interruption, it is questionable whether memory for associative cues encoded before interruption could account for post-interruption recovery. The visual search literature shows that participants rely on memory for the relative locations of display items to guide search resumption; although a shift in the absolute location of a target does not significantly impair search, the global configuration of other items must remain the same (Shen & Jiang, 2006). In

Decision support and vulnerability to interruption

our dynamic visual display, the locations of all aircraft change in real time, altering the global spatial arrangement of items on the screen and making it more effortful to rapidly regain SA. Moreover, the concordance rate between pre and post interruption fixation areas in the No-DSS group was below 50% (and less than in the other two groups), further weakening the idea that spatial memory could account for the differences. Instead our findings suggest more of a reconstructive strategy (Salvucci, 2010), whereby the No-DSS participants who appeared to be encoding a wider area of the post-interruption display (as shown by dwell areas and a greater number of transitions between AOIs), recovered from interruption quicker than those who encoded less of the screen, and may have experienced attentional tunnelling.

4.1 Methodological and practical implications

Microworlds are particularly well suited to the study of multitasking because of the possibility to assess performance on multiple subtasks that together comprise a higher order goal. Although basic cognitive research can inform us of ways to improve particular aspects of cognition, it must be considered that complex real-world situations rarely involve performance of a single isolated task. Microworlds preserve some of the complexities and interdependencies of actual C2 situations, allowing one to examine the impact that modifying one aspect of a subtask may have on another. Although sharing certain key characteristics with their real-world counterparts (being dynamic, complex, opaque; Brehmer, 1992), they are not simply a scaled down version of real life and instead are built to be conceptually relevant and generalizable in terms of the theoretical idea (rather than the setting or sample; Mook, 1983). Because our simulation recreates at a functional level the processes involved in monitoring, planning, and risk assessment, it is possible to extrapolate findings beyond the surface characteristics of maritime decision making, to apply to performance in other complex tasks that engage similar processes. Thus system designers must be alert to the finding that a support tool that is too demanding of attentional resources may overload cognitive processes in high workload,

Decision support and vulnerability to interruption

interruption-prone environments (e.g., aviation, crisis management, emergency response, or military operations).

The variables presented in the current experiment can provide some useful recommendations for the study of interruptions. Measures at a macro level that average across the whole scenario (e.g., comparison between interruption and no-interruption conditions), are likely to dilute the relatively short-lived effects of interruption; as such, concentrating on specific epochs around the interruption is advisable. Eye movements are a rich source of data that can show where participants are looking at precise moments in time; however, we must also be aware that fixations do not necessarily provide a direct window into cognitive processes, and attention may be allocated elsewhere. In the current study, self-reports of workload revealed a mismatch between the subjective and objective measures. Participants reported that interruption did not affect workload, yet defensive effectiveness was reduced in the interrupted conditions. Furthermore, participants believed that only CHT increased workload, yet there was a tendency for defensive effectiveness to be impaired in both DSS conditions. One might question the worth of such subjective measures given that they are retrospective in nature, and that participants are notoriously poor at assessing their own cognitive abilities and limitations (e.g., Levin, Momen, Drivdahl, & Simons, 2000). However, the TLX taps directly into the participant's experience which in itself can be informative. For example, a method or system that increases perceived workload may be less desirable than one that does not, even if performance is equivalent. Furthermore, if operators experience a drop in performance under a particular condition – but no difference in self-reported difficulty – it may be prudent to bring to their attention cognitive fallibilities that may not be apparent to them to avoid dangerous overconfidence (e.g., Levin et al., 2000).

The resumption lag – thought to reflect the time needed to reactivate primary task goals – is highly sensitive to interruption effects and is often taken as the key measure of the degree of

Decision support and vulnerability to interruption

interruption disruption (Hodgetts & Jones, 2006a, b; Trafton, Altmann, Brock, & Mintz, 2003). Of course a certain amount of ‘recovery’ must have occurred in order to decide on an immediate first action following interruption, but relying solely on resumption lag as a measure of interruption recovery may overlook the fact that this process persists beyond the first post-interruption action on a task (Altmann & Trafton, 2007), and may take up to a minute for work to resume at the same rate at which it was left (Hodgetts et al., 2014; Jackson, Dawson, & Wilson, 2003). Given the prolonged nature of interruption recovery in such complex and dynamic tasks, we question whether these processes can reflect the memory-based retrievals proposed in MfG (Altmann & Trafton, 2003). Instead recovery may require reconstruction of a new problem state and situational model (Salvucci, 2010), processes which are dependent upon the availability and adequate allocation of attentional resources.

From a practical point of view, the question is how we can support this process of reconstruction in a high workload task and what modifications of the two DSS would be necessary. The TOD interface had the potential to aid the recovery of SA following interruption, by supporting the reconstruction of a temporal plan to reassess, prioritise, and coordinate subsequent actions. The CHT also had the potential to help reconstruct a mental model of aircraft locations and properties (e.g., speed, trajectory), by highlighting changes that had occurred during the interruption interval. However for both tools the workload was too high to be of benefit within a multitasking environment, and in the face of interruption. Given that these DSS were problematic during the post-interruption period, one possibility could be to remove them at this point of highest workload and gradually introduce them once recovery processes are underway.

Modifications could be made to the support tools themselves so that they may fit more easily within the operator’s workload capacity in a multitasking and interruption-prone environment. In designing a DSS to minimize the burden on attentional resources, it may be

Decision support and vulnerability to interruption

relevant to consider the NSEEV model of attention behavior (NSEEV: noticing – salience, effort, expectancy, value; Steelman, McCarley, & Wickens, 2011). *Salience* of information is key if it is to be attended to with ease, and this is something that would need to be addressed with the two DSS studied. CHT lists all changes occurring (critical or otherwise), which can lead to a cluttered table with information that is difficult to extract. It would benefit from a filter mechanism that prioritises the most important changes and gives them greater salience at the top of the table. The information in TOD may also be difficult to interpret as it is not necessarily intuitive, e.g., on TOD the speed of an aircraft relates to the size of the rectangle and not the speed at which it travels across the screen. Modifications to the TOD interface would make the temporal information it provides more salient and easier to extract. This may have been particularly the case for novice users, as a previous study found that participants were able to use TOD to better advantage only when they had more experience (Rousseau et al., 2007). NSEEV also highlights the *effort* required to shift attention across the screen. The burden on attentional resources could therefore be reduced if critical information from the automated system was integrated within the geospatial display in order to reduce the need for long saccades between the radar and the support tool. Furthermore, the tools are unlikely to be used to optimal advantage in their current form because they are low in *expectancy* and *value*. For example, because the CHT also lists all irrelevant changes amongst critical changes, there is less expectation that entries in the table will convey useful information and a greater value is placed on the naïve realism of the geospatial display.

Summary

Given the costs and stakes associated with introducing new technology to a task, it is essential that any support system is thoroughly evaluated before implementation. A microworld provides a useful platform for this type of testing as it is dynamic, complex and time-pressured, and takes into account interactions rather than isolated variables. Furthermore, it allowed us to

Decision support and vulnerability to interruption

combine various measures that – whether concurring or contradictory – can provide useful insights into user behaviour and experience. Although some performance measures can be robust enough to demonstrate effects when comparing across whole scenarios, the effects of other performance or event-based measures may be diluted when comparing at a more general level. Eye movements and time-based measures allow for more specific assessment of effects directly attributable to the interruption. The current results are difficult to account for in terms of memory-based retrieval, and the long period of interruption recovery is more compatible with an explanation of reconstruction. Thus the practical emphasis of support systems should perhaps not be on how we can facilitate the encoding and remembering of pre-interruption state, but rather on how we can support the reconstruction of a new mental model during the post interruption period.

We demonstrate that any system evaluation needs to be holistic in nature to ensure that benefits to one facet of performance do not incur costs to another. The two DSS examined in the current study – although able to augment SA – reduced defensive effectiveness and prolonged interruption recovery. By looking at variables in combination rather than in isolation we have achieved a broader picture of the processes operating, which in turn can inform theory and practice. To conclude, even with a support system automating some part of a task, multifaceted tasks of risk assessment and dynamic decision making are still, if not more, vulnerable to interruptions.

References

- Altmann, E. M. & Trafton, J. G. (2002). Memory for goals: An activation-based model. *Cognitive Science*, 26, 39–83.
- Altmann, E. M., & Trafton, J. G. (2007). Timecourse of recovery from task interruption: Data and a model. *Psychonomics Bulletin & Review*, 14, 1079–1084.
doi:10.3758/BF03193094
- Anderson, J. R. & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Erlbaum.
- Ballas, J. A., Heitmeyer, C. L., & Pérez-Quñones, M. A. (1992). Evaluating two aspects of direct manipulation in advanced cockpits. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* pp. 127-134. ACM Press. Monterey, CA, United States.
- Beringer, D. B., & Harris, H. C., Jr. (1999). Automation in general aviation: Two studies of pilot responses to autopilot malfunctions. *International Journal of Aviation Psychology*, 9, 155–174.
- Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. *Acta Psychologica*, 81, 211–241. Retrived from http://ac.els-cdn.com/000169189290019A/1-s2.0-000169189290019A-main.pdf?_tid=c2fabe80-a0af-11e3-a760-00000aacb360&acdnat=1393616430_1483bc71630f441aba73fb264d17ee1c
- Brehmer, B., & Dörner, D. (1993). Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field study. *Computers in Human Behavior*, 9, 171–184.
- Callan, D. (1998). Eye movement relationships to excessive performance error in aviation. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting* (pp.

Decision support and vulnerability to interruption

1132–1136). Santa Monica, CA: Human Factors and Ergonomics Society.

doi:10.1177/154193129804201516

Chan, H. S., & Courtney, A. J. (1993). Effects of cognitive foveal load on a peripheral single-target detection task. *Perceptual and Motor Skills*, 77, 515–533.

Dehais, F., Causse, M., Vachon, F., & Tremblay, S. (2012). Cognitive conflict in human-automation interactions: A psychophysiological study. *Applied Ergonomics*, 43, 588-595.

Dehais, F., Tessier, C., Christophe, L., & Reuzeau, F. (2010). The perseveration syndrome in the pilot's activity: Guidelines and cognitive countermeasures. *Human Error, Safety and Systems Development. Lecture Notes in Computer Science*, 5962, 68-80.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic system. *Human Factors*, 37, pp. 32–64.

Gartenberg, D., McCurry, J. M., & Trafton, J. G. (2011). Situation awareness reacquisition in a supervisory control task. *Proceedings of the Human Factors and Ergonomics Society 55th Annual Meeting* (pp. 355–359). Santa Monica, CA: Human Factors and Ergonomics Society. doi:10.1177/1071181311551073

Goldberg, J. H., & Kotval, X. P. (1999). Computer interface evaluation using eye movements: methods and constructs. *International Journal of Industrial Ergonomics*, 24, 631–645. doi:10.1016/S0169-8141(98)00068-7

Gonzalez, C. (2005). Decision support for real-time, dynamic decision-making tasks. *Organizational Behavior and Human Decision Processes*, 96, 142 – 154. doi: 10.1016/j.obhdp.2004.11.002

Gonzalez, C., Vanyukov, P., & Martin, M. K. (2005). The use of microworlds to study dynamic decision making. *Computers in Human Behavior*, 21, 273–286.

Decision support and vulnerability to interruption

- Granlund, R., & Johansson, B. (2004). Monitoring distributed collaboration in the C3fire microworld. In S. G. Schiflett et al. (Eds.), *Scaled worlds: Development, validation, and applications* (pp. 37-48). Burlington, VT: Ashgate Publishing Limited
- Gray, W. D. (2002). Simulated task environments: The role of high-fidelity simulations, scaled worlds, synthetic environments, and laboratory tasks in basic and applied cognitive research. *Cognitive Science Quarterly*, 2, 205–227. Retrieved from <http://psycnet.apa.org/psycinfo/2002-06806-004>
- Hart, S. G., & Staveland, L. E. (1988). Development of the NASA-TLX (Task Load Index): Results of the experimental and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 139–183). Amsterdam, Netherlands: North Holland Press.
- Hodgetts, H. M., & Jones, D. M. (2006a). Contextual cues aid recovery from interruption: The role of associative activation. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 35, 1120–1132. doi:10.1037/0278-7393.32.5.1120
- Hodgetts, H. M. & Jones, D. M. (2006b). Interruption of the Tower of London task: Support for a goal activation approach. *Journal of Experimental Psychology: General*, 135, 103–115. doi:10.1037/0096-3445.135.1.103
- Hodgetts, H.M., Vachon, F., & Tremblay, S. (2014). Background sound impairs interruption recovery in dynamic tasks: Procedural conflict? *Applied Cognitive Psychology*, 28, 10 – 21.
- Imbert, J.-P., Hodgetts, H. M., Parise, R., Vachon, F., Dehais, F., & Tremblay, S. (2014). Attentional costs and failures in air traffic control notifications. *Ergonomics*. doi:10.1080/00140139.2014.952680
- Jackson, T., Dawson, R., and Wilson, D. (2003), Reducing the effect of email interruptions on employees. *International Journal of Information Management* 23, 55–65. doi:[10.1016/S0268-4012\(02\)00068-3](https://doi.org/10.1016/S0268-4012(02)00068-3)

Decision support and vulnerability to interruption

- Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, 8, 441–480. doi:10.1016/0010-0285(76)90015-3
- Kirmeyer S. L. (1988). Coping with competing demands: Interruptions and the Type A pattern. *Journal of Applied Psychology*, 73), 621-629.
- Lafond, D., Vachon, F., Rousseau, R., & Tremblay, S. (2010). A cognitive and holistic approach to developing metrics for decision support in command and control. In D. B. Kaber & G. Boy (Eds.), *Advances in Cognitive Ergonomics* (pp. 65–73). Danvers, MA: CRC Press.
- Levin, D. T., Momen, N., Drivdahl, S. B., & Simons, D. J. (2000). Change blindness: The metacognitive error of overestimation change-detection ability. *Visual Cognition*, 7, 397–412.
- Mack, A., & Rock, I. (1998). *Inattention blindness*. Cambridge, MA: The MIT Press.
- Lleras, A., Rensink, R. A., & Enns, J. T. (2005). Rapid resumption of interrupted visual search: new insights on the interaction between vision and memory. *Psychological Science*, 16, 684-688.
- Lleras, A., Rensink, R. A., & Enns, J. T. (2007). Consequences of display changes during interrupted visual search: rapid resumption is target specific. *Perception & Psychophysics*, 69, 980-993.
- Masson, M. E. J., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology*, 57, 203-220.
- McCrickard D. S., Catrambone R., Chewar C. M., & Stasko J. T. (2003). Establishing tradeoffs that leverage attention for utility: Empirically evaluating information display in notification systems. *International Journal of Human-Computer Studies*, 58, 547-582.
- Mook, D. G. (1983). In defense of external invalidity. *American Psychologist*, 38(4), 379.

Decision support and vulnerability to interruption

- Perry, N. C., Wiggins, M. W., Childs, M., & Fogarty, G. (2013). The application of reduced-processing decision support systems to facilitate the acquisition of decision-making skills. *Human Factors*, 55, 535-544. doi:10.1177/0018720812467367
- Poole, A., & Ball, L. J. (2006). Eye tracking in human-computer interaction and usability research: Current status and future prospects. In C. Ghaoui (Ed.), *Encyclopedia of Human Computer Interaction* (pp. 211–219). Hershey, PA: Idea Group.
- Potter, S. S., Gualtieri, J. W., & Elm, W. C. (2003). Case studies: Applied cognitive work analysis in the design of innovative decision support. In Elm et al. (Eds), *Applied cognitive work analysis: a pragmatic methodology for designing revolutionary cognitive affordances* (pp. 357-382). Boca Raton, FL: CRC Press
- Ratwani, R. M., Andrews, A. E. , Sousk, J. D. , 7 Trafton, J. G. (2008). The effect of interruption modality on primary task resumption. In *Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting*.
- Ratwani, R. M. & Trafton, J. G. (2008). Spatial memory guides task resumption. *Visual Cognition*. 16, 1001-1010.
- Ratwani, R. M. & Trafton, J. G. (2010). An eye movement analysis of the effect of interruption modality on primary task performance. *Human Factors*. 52, 370-380.
- Rousseau, R., Tremblay, S., Lafond, D., & Vachon, F., & Breton, R. (2007). Assessing temporal support for dynamic decision making in C2. *Proceedings of the 51st annual meeting of the Human Factors and Ergonomics Society* (pp. 1259-1262). Santa Monica, CA: Human Factors and Ergonomics Society. doi:10.1177/154193120705101844
- Salvucci, D. D. (2010). On reconstruction of task context after interruption. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: CHI 2010* (pp. 89-92). New York: ACM Press.

Decision support and vulnerability to interruption

- Schoelles, M. J. & Gray, W. D. (2001). Argus: A suite of tools for research in complex cognition. *Behavior Research Methods, Instruments, & Computers*, 33(2), 130-140.
- Shen, Y. J. & Jiang, Y. V. (2006). Interrupted visual searches reveal volatile search memory. *Journal of Experimental Psychology: Human Perception and Performance*, 32, 1208–1220.
- Smallman, H. S., & St. John, M. (2003). CHEX (Change History EXplicit): New HCI concepts for change awareness. *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting* (pp. 528–532). Santa Monica, CA: Human Factors and Ergonomics Society. doi:10.1177/154193120304700358
- Smallman, H. S., & St. John, M. (2005). Naive realism: Misplaced faith in realistic displays. *Ergonomics in Design: The Quarterly of Human Factors Applications*, 13, 6–13. doi:10.1177/106480460501300303
- Steelman, K. S., McCarley, J. S., & Wickens, C. D. (2011). Modeling the control of attention in visual workspaces. *Human Factors*, 53, 142-153. doi:10.1177/0018720811404026
- St. John, M., & Smallman, H. S. (2008). Staying up to speed: Four design principles for maintaining and recovering situation awareness. *Journal of Cognitive Engineering and Decision Making*, 2, 118–139. doi:10.1518/155534308X284408
- St. John, M., Smallman, H. S., & Manes, D. I. (2005). Recovery from interruptions to a dynamic monitoring task: The beguiling utility of instant replay. *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting* (pp. 473–477). Santa Monica, CA: Human Factors and Ergonomics Society.
- Trafton, J. G., Altmann, E. M., Brock, D. P., & Mintz, F. E. (2003). Preparing to resume an interrupted task: Effects of prospective goal encoding and retrospective rehearsal. *International Journal of Human Computer Studies*, 58, 582–602. doi:10.1016/S1071-5819(03)00023-5

Decision support and vulnerability to interruption

- Trafton, J. G., & Ratwani, R. M. (2014). The law of unintended consequences: the case of external subgoal support. In *CHI 2014 Proceedings*: 1767-1776.
- Tremblay, S., Vachon, F., Lafond, D., & Kramer, C. (2012). Dealing with task interruptions in complex dynamic environments: Are two heads better than one? *Human Factors*, 54, 70–83. doi:10.1177/0018720811424896
- Tremblay, S., Vachon, F., Rousseau, R., & Breton, R. (2012). Promoting temporal awareness for dynamic decision making in command and control. In M. Kay & K. S. Hale (Eds), *Advances in cognitive engineering and neuroergonomics* (pp. 188-198). Boca Raton, FL: CRC Press
- Vachon, F., Lafond, D., Vallières, B. R., Rousseau, R., & Tremblay, S. (2011). Supporting situation awareness: A tradeoff between benefits and overhead. *Proceedings of the 1st IEEE International Conference on Cognitive Methods in Situation Awareness and Decision Support* (pp. 282–289). Miami Beach, FL: IEEE Conference Publications. doi:10.1109/COGSIMA.2011.5753460
- Vallières, B., Hodgetts, H. M., Vachon, F., & Tremblay, S. (2012). Supporting Change Detection in Complex Dynamic Situations: Does the CHEX Serve his Purpose? *Proceedings of the 56th Annual Meeting of the Human Factors and Ergonomics Society*. (pp. 1708-1712). Santa Monica, CA: Human Factors and Ergonomics Society.
- Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, 3, 159-177. doi:10.1080/14639220210123806
- Wickens, C. D. & Alexander, A. L. (2009). Attentional tunneling and task management in synthetic vision displays. *The International Journal of Aviation Psychology*, 19, 182-199.

Acknowledgements

This work was supported by a research and development partnership grant from the National Sciences and Engineering Research Council of Canada with Defence Research and Development Canada—Valcartier and Thales Canada, Systems Division, awarded to Sébastien Tremblay.

Bios

Helen M. Hodgetts is a Research Associate at the Department of Applied Psychology, Cardiff Metropolitan University, UK and affiliated Research Fellow at the School of Psychology, Université Laval. She received her Ph.D in cognitive psychology from Cardiff University in 2005 and was a postdoctoral researcher at Cardiff University (2005 – 2009; 2011 – 2012) and Université Laval (2009 – 2010).

Sébastien Tremblay is currently a professor at the School of Psychology at Université Laval and Head of the Co-DOT lab (www.co-dot.ulaval.ca). His main research interests relate to human cognition and performance. He has expertise in a wide range of cognitive human factors issues. Prior to his appointment at Laval, he held a postdoctoral fellowship at Cardiff University in Cardiff, United Kingdom, funded by the Defence Evaluation and Research Agency. He earned his PhD in psychology in 1999 from Cardiff University, UK.

Benoît R. Vallières is a Ph.D. student in applied cognitive psychology at Université Laval. He completed a research internship at Advanced Distributed Learning Co-Laboratories in Orlando in 2013. He received his bachelor degree in Psychology from Université Laval in 2009. His research interests include change blindness, decision support system, and cognitive training.

François Vachon is an Assistant Professor at the School of Psychology, Université Laval. His main research interests include the basic and applied cognitive psychology of attention and multitasking. He received his Ph.D. in cognitive psychology from Université Laval in 2007. He then completed postdoctoral fellowships in cognitive psychology at Cardiff University (2007), in cognitive neuroscience at Université de Montréal, Canada (2008-2009), and in applied cognitive psychology at Université Laval (2009-2011).