Proceedings of the Human Factors and Ergonomics Society 58th Annual Meeting - 2014

Workload overload modeling: An experiment with MATB II to inform a computational model of task management

Robert S. Gutzwiller[1], Christopher D. Wickens[1,2], & Benjamin A. Clegg[1]
[1]Colorado State University, [2]Alion Science and Technology

Task switching choice was examined building from a model of task overload management. An experiment using the Multi-Attribute Task Battery (MATB) was undertaken to explore the influence of two parameters of the model, task priority and task difficulty. Participants were free to switch between the four component tasks, with the number of switches and task choice for conflicting events observed. A unique post-experiment survey measured subjective ratings of task attributes. We found that task difficulty, by reducing switching, and task priority, which determined whether increased task difficulty increased time in task, significantly influenced task switching predominantly in line with our predictions. The specific role of priority in multi-task management, and future directions including time-on-task related effects and the role of operator fatigue, are discussed.

INTRODUCTION

A post automation failure period of workload transition (Wickens, Laux & Hutchins, 2014; Sebok, Wickens & Clegg, 2014) often requires multi-tasking. Operators might time share between diagnosis, failure management, communications, and other necessary tasks (e.g., keeping a space capsule on a navigational path). Multi-tasking is itself multifaceted (Wickens, 2013), sometimes supporting concurrent task performance, but often forcing sequential task operations. The former has been well modeled by multiple resource theory (Wickens, 2008) and threaded cognition (Salvucci & Taatgen, 2011). Efforts to model the latter have adopted two generic approaches.

First, an **interruption management (IM)** approach (e.g., Trafton & Monk, 2007) focuses on only two generic tasks, an ongoing task (OT) and a single interrupting task (IT). Plenty of empirical data exist from this paradigm (see Wickens et al. 2013 for a summary), and the major effects are captured in the memory-for-goals model (Altmann & Trafton, 2002), based loosely on ACT-R. Such models examine the time and fluency of a single switch between the OT and the IT, and back again. Unfortunately much of these data do not generalize easily to the heterogeneous task environment of post workload transition in crisis, as described above.

Second, and more relevant to the current workload transition issue, are models of **task management (TM).** These focus on the **decision** of whether to switch (if at all), and, if a switch is chosen, which of a set of alternative tasks may be chosen. As reported by Wickens, Santamaria and Sebok, (2013) this paradigm is far less populated with empirical data, particularly for realistic tasks. Freed (2000) developed a highly appropriate model of task switching, proposing attributes of tasks that drive the switch. But no data were presented to support this. Also, models of visual scanning adopt this approach, at least in predicting switches of visual fixation (the eyeball) - but not in predicting switches of task attention (the "brain ball"; Sheridan, 1970; Wickens, 2014).

Last year we presented a model, now called STOM (Strategic Task Overload Management; Wickens et al., 2013). In Wickens, Laux, Hutchins and Sebok (2014) we describe STOM in some detail, but in brief it is as follows. In deciding what tasks to perform, two decisions are made. First, the operator may decide whether or not to switch at all, from an ongoing task (OT) to an alternative task (AT). If there is a decision to switch, s/he must then decide which alternative to switch to when more than one AT is queued to be performed. Each decision is based on multiple attributes of the tasks. The first decision has an inherent bias to stay in the OT, rather than switch, a kind of "task inertia" or switch avoidance. The second decision is guided by specific task attributes. All tasks (both OT and the set of ATs) are characterized by their difficulty, priority, and their interest or "engagement" value.

Based on the analysis performed by Wickens et al. (2013), weightings in the model increase the likelihood of staying (for the OT) or switching to an AT (if a switch is triggered) to the extent that the task is **easier, higher priority** and **more engaging**. In addition, all ATs have an attribute of **task salience,** defining the extent to which its arrival, or presence is signaled by salient reminders (e.g., a tone, or visual text). The most reliable finding is an "easy task preference", thus illustrating a principle of effort-avoidance in task choice (Kool et al., 2010, Kahneman, 2011). The model is "OT-centric" in the sense that once a switch is made from a given OT to an AT, the latter now becomes the OT, with its inherent bias to remain (rather than switch). Our meta-analysis revealed that there is a tendency for switch avoidance about 60% of the time (95% CI [58-62]; Wickens et al., 2014).

The purpose of the current study was to better populate the STOM attribute weightings based on a controlled experiment within a multi-task task switching battery (MATB, described below). A two-pronged approach was adopted: (1) establish estimated weightings of task attributes based upon participant ratings and assess how those ratings influenced actual switching behavior; and (2) manipulate two of the parameters

Copyright 2014 Human Factors and Ergonomics Society. DOI 10.1177/1541931214581179

(priority and difficulty) for one of the tasks (tracking), to determine the effects on switching choices. In particular, we evaluated decision preferences to switch from tracking to one of two *different* ATs.

MATB II Overview

MATB (Multi Attribute Task Battery) II (Santiago-Espada, Myer, Latorella, & Comstock, 2011) is a multi-tasking research platform designed to assess performance on four main, concurrent tasks (tracking, monitoring, resource management, and communications), and is an updated version of the original MATB (Comstock & Arnegard, 1992). All information about tasks was present visually on screen, with the exception of the communications task that relied on participants to listen to simulated air traffic control messages and respond to some of these.

MATB Tasks

The **tracking task (Trk)** was a 2 dimensional random input compensatory task. Participants attempted to keep a target reticle within a small square box. The tracking task was active for the entire trial in the *easy* (low bandwidth) and *difficult* (high bandwidth) conditions. The **monitoring task (Mon)** had two components: operators responded to red and green lights, and to scales that go out of range by registering either too high or too low. The **resource management task (Rman)** represented fuel management aboard an aircraft. Operators maintained fuel in two tanks that continually deplete below target levels. Tanks are connected by pumps, which direct resource flow into or out of each of the tanks, and are controlled by the operator to regulate tank levels. Events in resource management are failures of interconnecting pumps. In each test trial, pumps failed once in randomized order, and were repaired automatically. The repairs were scheduled to result in slightly varying durations (for an average ~30s).

The auditory **communication task (Comm)** simulated pilot interaction with air traffic controller requests. Operators heard command messages to alter radio and frequency on one of four communications radios to a new five-digit frequency. The instructions were directed either to an ‘other’ ship or to the operator’s own ship by using a callsign designated during training, a distinction that is critical for later analyses. Requests could be ignored unless the own sign was called. Test trials contained equal numbers of own ship and other ship instructions (four each), and occurred once for each radio type with no overlap in the frequency entered.

In order to evaluate task switching preference to an alternative task (from tracking), *task event pair conflicts*, wherein an event in two different tasks occurred close together, were created. The arrival time of an event pair conflict varied randomly across trials, as did the order of the pair of tasks. Three types of paired events occurred, commensurate with a combination between monitoring, resource, and communications tasks (Table 1). The tracking task was not used because it was always the ongoing task, not defined by any discrete external initiating event. Events in the pair occurred within 500ms of each other, as simultaneous presentation was not possible in MATB.

	TRK	MON	RMAN	COMM
TRK	x	x	x	x
MON		x	2 conflicts	2 conflicts
RMAN			x	2 conflicts
COMM				x

Table 1. Graphical display of the types of conflicting event pairs that occurred during each of the test trials.

Two of each conflict pair occurred over the course of each test trial, resulting in 6 conflicts per trial. These represented times the operator could make a decision to switch from an OT to one of the two potential ATs, events represented in the event pair conflict. Determining which one of the two conflicting tasks was chosen when a task switch occurred consequently tested the attributes of relevance outlined in STOM. The conflict pairs were interleaved with numerous other single task switch opportunities.

The relative task difficulty and priority of the tracking task component of MATB was manipulated both within (difficulty) and between (priority) participants in a 2x2 mixed design. Based on the STOM model, three hypotheses were formulated and tested, two based on task difficulty and one based on task priority. First, because task switching was assumed to be effortful and resource limited, (*H1*) less switching should occur in difficult tracking than in easy tracking conditions. Secondly, (*H2*) a difficult task should garner proportionally fewer switches to it, whether difficulty is manipulated or measured. Finally, (*H3*) a higher priority task should lead to fewer switches away from it when it is the OT (a stay preference) and could also lead to more switches toward it (an AT attractiveness). Other task attribute ratings were assessed; however, we did not have specific hypotheses to offer here about relative or absolute outcomes.

Vitally, the manipulation of priority *concurrent* with manipulating difficulty represents the first integration of two of the task attributes in a task switching paradigm that may influence which AT is chosen when a switch occurs (but see Gopher, Brickner, & Navon, 1982, for effects of priority and difficulty on **concurrently** performed tasks). This helps populate the model and addresses gaps in the literature on task management.

METHODS

Participants

Eighty-one students at Colorado State University participated in return for optional, partial course credit.

Materials & Procedure

A computer with a standard mouse, stereo headphones and a Logitech joystick was used. The task screens were arranged in a square with approximately 1.16 degrees of visual angle separating two rows, and 0.19 degrees separating two columns.

Participants were introduced to the MATB II simulation through a series of instructional, self-paced slides adapted from Santiago-Espada et al. (2011). In one condition **(equal priority)** participants were told to perform all tasks as best as possible. In the other **(tracking priority)**, participants were told to prioritize tracking over all of the other tasks, while still performing them as best as possible (e.g., Gopher, Weil, & Siegel, 1989). Participants were instructed to only perform the tasks with a single, dominant hand, and they were not allowed to use two hands or the keyboard to respond. Thus, the participants were required to switch between mouse and joystick when necessary. This critical instruction allowed us to examine task switch behavior without the confounding influence of concurrent inputs.

Participants then completed a training trial containing all of the elements of the MATB simulation used during later experimental test trials. Before beginning the test trials, the experimenter reminded participants of their group instruction (e.g., equal or tracking priority). Subsequently, participants performed three test trials of varying tracking task difficulty (easy, difficult, and transition). The difficulty of one task in MATB (tracking) was manipulated within subjects by altering the update rate of the tracking task (i.e., changing bandwidth, Wickens & Hollands, 2000) while controlling task input rate. In other words the amount of change able to be affected by stick movements was "low" and MATB was used to vary the frequency and deviation of the correction needed to track successfully. Easy and difficult trials were counterbalanced.

Multiple events in all four tasks were presented, randomly interspersed with single events and participants attempted to respond to all task events. After the final test trial (transition), in order to help assess attribute weights, a survey was administered which asked participants to make paired comparison ratings to determine which tasks were more difficult to perform, more interesting, and of higher priority. Comparison ordering was mixed between rating variables.

RESULTS

Performance data for each task were recorded, but as the focus is on switching choice behavior, for brevity, those results are not described here. Furthermore, again because of space constraints, our reporting of switching data are confined to the most informative effects to assist in populating the STOM model (see Gutzwiller, 2014). Two participants' data were removed due to outliers greater than 3SD, and five participants' data were missing for the survey and were not included in those results.

Task Switching

Task switching was measured by examining actions taken in each of the four tasks in MATB over the course of each test trial. As tracking was the manipulated task, all comparisons reported are focused on switches to and from tracking. Two hypotheses were related to tracking task difficulty, and were addressed in the following ANOVA. First, there should be fewer switches in general under difficult tracking conditions. Second, there should be more switches *to* tracking under easy, than difficult tracking. Third, higher task priority should result in more switches toward (and fewer switches away from) a task compared to lower priority.

A 2 (tracking difficulty) x 2 (priority group) x 2 (counterbalance condition) repeated measures ANOVA was conducted. A main effect of tracking difficulty was found, confirming the first hypothesis – there were fewer task switches when tracking was difficult (*M*=43.26) than when it was easy (*M*=48.75; *F*(1,75)= 17.22, *p*<.01, η_p^2=.19). No main effect of priority or counterbalancing group (*F*<1) on number of switches related to tracking was found.

A comparison of only switches "to tracking" was undertaken to specifically address *H2* (fewer switches to difficult tasks), Participants switched to the tracking task less when it was difficult (*M*=43.54), than easy (*M*=48.28; (*t*(78)=3.55, *p*<.01) confirming the second hypothesis.

Although hypothesized (*H3*) no effect of priority was found when analyzing switches. However a visible trend could be seen between difficulties for the task pair of tracking and resource management, especially when tracking was prioritized (Fig. 2) rather than equal priority (Fig. 1).

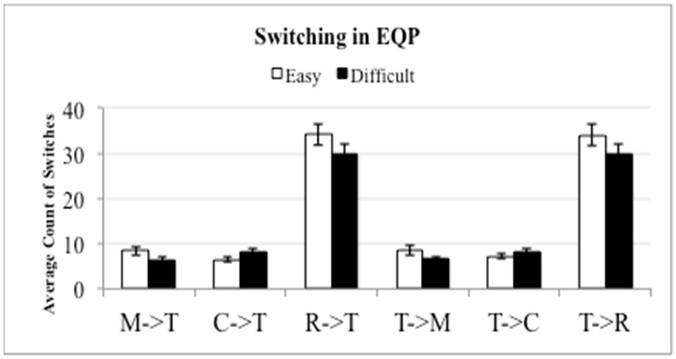

Figure 1. **Equal priority tracking condition.** Switches **to** tracking (left triad of bar pairs) and **from** tracking (right triad of bar pairs), under easy (white) and difficult (black) tracking conditions. Error bars are standard error of the mean. C=Comm, M=Mon, R=Rman, T=Trk.

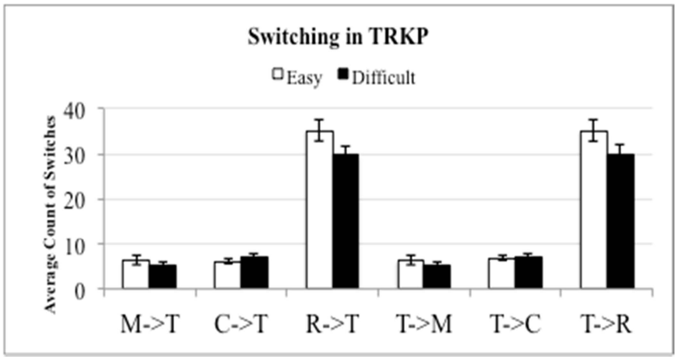

Figure 2. Tracking priority condition. Switching to and from tracking, as in Figure 1.

Additionally, an unreported performance analysis (see Gutzwiller, 2014) showed the effect of tracking difficulty was only evident for the tracking and resource management tasks, conceptually linking them together. Therefore, two separate ANOVAs were run on combined tracking and resource management switches, one for each group, with the goal of determining whether priority influenced switches.

In the analysis of the equal priority group, a marginally significant effect of tracking task difficulty was revealed, with fewer switches on the difficult trial (M=74.8) than on the easy tracking trial (M= 83.38; $F(1,36)=3.88$, $p=.06$, $\eta_p^2=.10$). The same effect, now significant, was found for the tracking prioritized group with fewer switches on difficult (M=72.5) compared to the easy tracking trial (M= 83.05; $F(1,41)=13.27$, $p<.01$, $\eta_p^2=.24$). The analyses did not reveal an influence of task priority in terms of switching.

Despite the lack of an effect of task priority for switching frequencies, it could have been the case that participants simply stayed longer with tasks under higher priority or higher difficulty tracking conditions (in other words, spent less time on other lower priority, or easier tasks). Time spent tracking was assessed by a 2 (priority condition) x 2 (difficulty condition) x 2 (counterbalance condition) repeated measures ANOVA. Significantly more time was spent in the tracking task under difficult (M=526s) compared to easy (M=518s) conditions, revealing a main effect of difficulty ($F(1,75)$= 11.53, $p<.01$, $\eta_p^2=.13$). No main effect of priority or counterbalancing was found ($F<1$). Critically, a marginally significant interaction ($F(1,75)$= 3.78, $p=.06$, $\eta_p^2=.05$) suggested the difference between time spent in the easy and difficult tracking conditions was larger when tracking was prioritized (M_{easy}=515s; $M_{difficult}$=528s), than when it was not. This difference was significant in the prioritized tracking condition ($t(41)=-4.51$, $p<.01$) but failed to reach significance for the equal priority condition ($t(36)=-.43$, $p>.05$). However priority never directly influenced the time on the tracking task.

Subjective Ratings

Participants provided paired task comparison ratings for three main categories of relevance to the STOM model: priority, difficulty and interest. No known work with MATB has examined participant ratings of these three attributes of component tasks. Based on Table 2, in the tracking priority condition, both the tracking and resource management tasks should be the highest "attractiveness" (being switched to), depite their high difficulty (compared to monitoring and communications), a finding backed up by Figs. 1 and 2.

	Priority		**Interest**		**Difficulty**	
Task	**EQP**	**TRK P**	**EQP**	**TRK P**	**EQP**	**TRK P**
Mon	-3.13	-3.98	-3.91	-4.29	-2.91	-3.56
Comm	0.06	-2.20	-0.09	-2.10	-2.03	-2.24
RMan	2.65	-0.24	3.34	1.51	4.69	3.10
Trk	0.53	6.41	0.66	4.88	0.25	2.71

Table 2. Subjective ratings summed for each of the four tasks in MATB for each condition. EQP=equal priority; TRKP=tracking priority. Higher ratings=a more positive global attractiveness score for priority and interest. Lower ratings are more attrative for task difficulty. Priority was rated higher for the tracking task in the TrkP condition than in the EqP condition.

We suggest that this is in part because of high interest values of both tasks (*Trk*=+4.88; *Rman*=+1.51) in the tracking priority condition, and their high perceived priority (*Trk*=+6.41; *Rman*=-0.24), both well above the other two tasks, whose values were far into the negative range. The subjective ratings were useful in interpreting the results of the paired conflict events.

Paired Conflict Events

The most informative analysis of comparative task switching was the switch from tracking as an ongoing task, to either Rman or Comm (the Mon task was excluded from this analysis because it was consistently an extremely "unfavored" task across all attributes in both groups). Our interest was in the trade-off of attributes as collapsed across groups. Rman and Comm allowed a tradeoff to be examined because Comm was rated easier (M= -2.13), but less interesting (M= -1.10), while Rman was considerably more difficult (M= +3.89) but more interesting (M= +2.42; see Table 2). We assessed task switch preference only for ownship Comm events (i.e., when the Comm task was relevant for the participant). These data are shown in Figure 3.

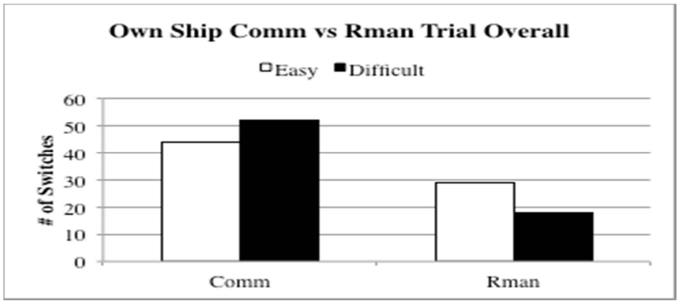

Figure 3. Number of switches summed across both priority groups in the own ship event.

An overall Comm over Rman preference ratio of 48:24 (2:1) [67% vs 33%] was amplified to a ratio of 52:18 (3:1) [74% vs 26%] when tracking was difficult, and reduced (but still present) to a 44:29 [60% vs 40%] ratio when tracking was easy. Tests of proportions indicated that the overall preference for Comm was significantly greater than 50% (by 95% confidence intervals [.55- .76]), which was also the case when split into difficult tracking [CI: .63- .83], but only marginally significant when tracking was easy [95% CI: .49- .71].

Both of these effects: the overall switch choice preference for Comm, and its amplification when resources are scarce can be explained via the task attributes in Table 2. While Rman was perceived as more interesting and higher priority (even when equal priority instructions were given), it was also perceived as much more difficult. Furthermore, although not rated as an attribute, Comm has a higher *salience* attribute within the STOM model, being a task that is "announced" by an auditory rather than visual event (Lu et al., 2013).

DISCUSSION

Returning to the three hypotheses, **first**, we predicted fewer switches in general during difficult tracking conditions.

Indeed this was the case, as participants switched approximately 13% more often in easy, than in difficult tracking conditions, a finding that fits nicely with a general switch cost avoidance behavior under cognitive load (Kool et al., 2010). **Secondly**, we expected there would be fewer "to" tracking switches when tracking was difficult compared to easy, and this was also found (12% less often). Critically this upholds a general propensity to switch to an easier task as found in a recent meta-analysis of basic and applied task switching literature (Wickens et al., 2014).

Third, we predicted a main effect of priority group, such that we should (but did not) find more switching to tracking in the tracking prioritized condition compared to the equal priority condition. However, priority did exert an effect by changing operators' delegation of time to an easy (less time) compared to a difficult (more time) tracking task. This suggests a relatively minor role of priority as a factor in isolation with regard to task switching (versus what has been found previously with concurrent multitasking; Gopher et al., 1982, 1989). It may be that most events with high priority that are switched to in the real world (e.g., avoiding a hazard during driving) also are co-occurring with other attributes such as task interest and value. In the general examination of all of the paired conflicts (see Gutzwiller, 2014) a high priority rating was never the only explanation for a choice. Although it was a consistent factor, it generally occurred along with greater salience, ease of the task, or higher interest.

In the specific conflict pair examined, the subjective ratings played a useful role, allowing for an examination of switch choice (between Comm and Rman tasks) and what attributes (interest, priority and task difficulty) corresponded to the observed choice. In this case it pointed to the high influence (potential weightings), of task difficulty and task salience.

The difficulty of an ongoing task exerted several of the predicted effects (and one unpredicted interaction), and represents the most validated and successful parameter of the model, reflecting the overall cost of limited cognitive resources, and the effect of this cost on decision making (Kahneman, 2011). Future directions also aim to incorporate the role of time on task as related to the stay/switch preference (see Gutzwiller, 2014; Wickens et al., 2014), and operator fatigue effects in order to further inform the STOM model.

ACKNOWLEDGMENTS

This work was supported by NASA under Grant NNX12AE69G (PI: Angelia Sebok), technical monitor Dr. Jessica Marquez and technical sponsor Dr. Brian Gore. RSG's contribution was supported by a DoD SMART scholarship through Space and Naval Warfare Systems Central – Pacific. We also thank Tyler Scott for extensive aid in extracting data from MATB II.

REFERENCES

Altmann, E. M., & Trafton, J. G. (2002). Memory for goals: An activation-based model. *Cognitive Science, 26*(1), 39–83.

Comstock, J. R., & Arnegard, R. J. (1992). The Multi-Attribute Task Battery for human operator workload and strategic behavior research (Tech Memorandum 104174). Hampton, VA: NASA Langley Research Center.

Dismukes, R. (2010). Remembrance of things future: prospective memory in laboratory, workplace, and everyday settings. *Reviews of Human Factors and Ergonomics,* 1–86.

Freed, M. (2000). Reactive prioritization. *In proceedings of the 2nd NASA International Workshop on Planning and Scheduling in Space.* Washington, DC: National Aeronautics and Space Administration.

Gopher, D., Brickner, M., & Navon, D. (1982). Different difficulty manipulations interact differently with task emphasis: Evidence of multiple resources. *Journal of Experimental Psychology: Human Perception & Performance*, *8*(1), 146-157.

Gopher, D., Weil, M., & Siegel, D. (1989). Practice under changing priorities: An approach to the training of complex skills. *Acta Psychologica, 71*, 147–177.

Gutzwiller, R. S. (2014). Switch choice in applied multi-task management. Unpublished doctoral dissertation. Colorado State University, Fort Collins.

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General, 139*(4), 665–682.

Kahneman, D. (2011). *Thinking fast and slow.* New York: Farrar, Strauss, & Giroux.

Lu, S. A., Wickens, C. D., Prinet, J. C., Hutchins, S. D., Sarter, N., & Sebok, A. (2013). Supporting interruption management and multimodal interface design: Three meta-analyses of task performance as a function of interrupting task modality. *Human Factors.*

Salvucci, D. D., & Taatgen, N. A. (2011). *The multi-tasking mind.* Oxford, U.K.: Oxford University Press.

Santiago-Espada, Y., Myer, R. R., Latorella, K. A., & Comstock, J. R. (2011). The Multi-Attribute Task Battery II (MATBII): Software for human performance and workload research: A user's guide. *NASA Tech Memorandum 217164*.

Sheridan, T. B. (1970). On how often the supervisor should sample. *IEEE Transactions on Systems Science & Cybernetics, SSC-6*(2), 140-145.

Trafton, J., & Monk, C. (2007). Task interruptions. *Reviews of Human Factors and Ergonomics*, 111–126.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors*, *50*(3), 449–455.

Wickens, C. D. (2013). Attention. In J. Lee & A. Kirlik (Eds.), *Handbook of Cognitive Engineering*. Oxford, U.K.: Oxford University Press.

Wickens, C. D. (2014). Noticing events in the visual workplace: The SEEV and NSEEV models. In R. Hoffman & R. Parasuraman (Eds.), *Handbook of Applied Perception.* Cambridge, U.K.: Cambridge University Press.

Wickens, C. D., Gutzwiller, R. S., & Santamaria, A. (2014, under review). Discrete task switching in overload: A meta-analysis and a model.

Wickens, C. D., Laux, L., Hutchins, S., & Sebok, A. (2014, HFES conference). Effects of sleep restriction, sleep inertia, and overload on complex cognitive performance before and after workload transition: A meta-analysis and two models.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance.* Upper Saddle River, MI: Prentice-Hall.

Wickens, C. D., Santamaria, A., & Sebok, A. (2013). A computational model of task overload management and task switching. *Proceedings of the Human Factors and Ergonomic Society, 57*(1), 763–767.