Journal of Experimental Psychology: Applied
2011, Vol. 17, No. 3, 233–246

1076-898X/11/$12.00 DOI: 10.1037/a0025171

Differences in Attentional Strategies by Novice and Experienced Operating Theatre Scrub Nurses

Ranieri Y. I. Koh and Taezoon Park
Nanyang Technological University

Christopher D. Wickens
University of Illinois

Lay Teng Ong and Soon Noi Chia
KK Women's and Children's Hospital, Singapore

This study investigated the effect of nursing experience on attention allocation and task performance during surgery. The prevention of cases of retained foreign bodies after surgery typically depends on scrub nurses, who are responsible for performing multiple tasks that impose heavy demands on the nurses' cognitive resources. However, the relationship between the level of experiences and attention allocation strategies has not been extensively studied. Eye movement data were collected from 10 novice and 10 experienced scrub nurses in the operating theater for caesarean section surgeries. Visual scanning data, analyzed by dividing the workstation into four main areas and the surgery into four stages, were compared to the optimum expected value estimated by SEEV (Salience, Effort, Expectancy, and Value) model. Both experienced and novice nurses showed significant correlations to the optimal percentage dwell time values, and significant differences were found in attention allocation optimality between experienced and novice nurses, with experienced nurses adhering significantly more to the optimal in the stages of high workload. Experienced nurses spent less time on the final count and encountered fewer interruptions during the count than novices indicating better performance in task management, whereas novice nurses switched attention between areas of interest more than experienced nurses. The results provide empirical evidence of a relationship between the application of optimal visual attention management strategies and performance, opening up possibilities to the development of visual attention and interruption training for better performance.

Keywords: attention allocation, interruption, experience, scrub nurse, SEEV model

Adverse events in health care due to human or system errors have been a subject of much scrutiny over the past 10 years. It is estimated that around 45% of all medical errors occur in the operating theater (Flin, Yule, McKenzie, Paterson-Brown, & Maran, 2006), and a half to two thirds of the adverse events were found to be attributed to surgical procedures (Brennan et al., 2004; Gawande, Thomas, Zinner, & Brennan, 1999; Thomas et al., 2000). Among these adverse events happening within the operating theater, more than half of them have been identified as preventable (Couch, Tilney, Rayner, & Moore, 1981; Thomas et al., 1999). Lincourt et al.(2007) claimed that one common, but poorly understood preventable medical errors in surgery is leaving a foreign object inside the patient after surgery, which causes additional treatment, prolonged hospital stays, provocation of litigation, as well as serious or fatal injuries involving the emotional trauma of the patient (Egorova et al., 2008).

The frequency of retained foreign objects (RFO) after surgery varies from 1:1000 to 1:18000 depending on the sources of the data, geographic location (Bani-Hani, Gharaibeh, & Yagha, 2005; Lincourt et al., 2007; Stark & Goldstein, 2004) and types of surgery (Hyslop & Maull, 1982). Considering the unreported or unfound cases, it is likely that the frequency can be much higher. The most commonly practiced method of preventing the case of RFOs after surgery is a manual count of surgical items typically performed by the scrub nurse and the circulating nurse on the surgical team (AORN, 1999). The procedure of counting with other simultaneous tasks is highly prone to errors as it relies on human consistency, accuracy, and efficiency which may be compromised under conditions of time pressure, distractions, workload, fatigue, and unexpected interruptions (Beyea, 2003). Scrub nurses mentally keep track of surgical items, such as "four gauzes down, one gauze in abdominal cavity", but their mental processes are vulnerable to working memory failures when the counts are not actively rehearsed.

Ranieri Y. I. Koh, Center of Human Factors and Ergonomics, Systems and Engineering Management Division, School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore; Taezoon Park, Systems and Engineering Management Division, School of Mechanical and Aerospace Engineering, Nanyang Technological University; Christopher D. Wickens, Human Factors Division, Institute of Aviation, University of Illinois; Lay Teng Ong and Soon Noi Chia, Nursing Division, KK Women's and Children's Hospital, Singapore.

Authors appreciate Dr. Bernard Chern, Dr. Yoel Donchin, and Prof. Martin Helander for their help and comments. This study is supported by Nanyang Technological University SUG08/08.

Correspondence concerning this article should be addressed to Taezoon Park, School of Mechanical and Aerospace Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798. E-mail: tzpark@ntu.edu.sg

Scrub nurses' responsibilities during the intraoperative phase of the surgery encompass cognitively demanding tasks (Roth et al., 2004). They traditionally entail providing skilled assistance to the surgeons and performing the swab/instrument counts during and at the end of the procedure (Taylor & Campbell, 2000). A scrub nurse works closely with the surgeons to anticipate the instrument needs of the surgeons, to be aware of the patient status and the progress of the surgery, and to aid the operation when extra help is needed (Koh et al., 2009). At the same time, the scrub nurse has to perform manual counts of swabs together with the circulating nurse throughout the operation, including a mandatory closure count before the closure of the cavity and a final count before the skin closure (AORN, 1999).

The role of scrub nurses during surgical operations involves high cognitive workload because the nurses must simultaneously handle multiple tasks with different priorities. The likelihood of error when performing these tasks increases with time pressure, interruptions requiring attention, and task switching. High cognitive workload when handling multiple competing tasks and demands for attention can contribute to errors in the mandatory counting protocol of scrub nurses (Roth & Patterson, 2005). Gawande, Studdert, Orave, Brennan, & Zinner (2003) found that the probability of retained foreign objects was eight times higher when an operation was performed on an emergency basis, and three times higher when the operation involved an unexpected change in procedure. These statistics reflect a stark increase in the possibility of retained foreign objects in emergencies when counting procedures may be partly or fully abandoned due to the disorganization and time pressure of the operation. Attention and task switching behaviors imposed by predicted or unpredicted interruptions require additional mental resources from the human operator (Trafton, 2007), here the scrub nurse. Although the strong causality of interruptions to adverse events is not empirically validated (Grundgeiger & Sanderson, 2009), a significant number of interruptions has been observed both from the surgical team members and external sources in the operating theater (Healey, Sevdalis, & Vincent, 2006), and a valid link between interruptions and adverse events (accidents and incidents) is well-established in aviation (Loukoupolis, Dismukes & Barshi, 2009).

Scrub nurses may develop a set of adaptive strategies to minimize the impact of multiple and simultaneous demands on attention during high workload stages of surgery (Dagi, Berguer, Moore, & Reines, 2007). In view of the primary goals of the scrub nurses during surgery, attention management and task management can therefore be ranked among the most important skills scrub nurses should acquire with experience. Task switching behaviors during interruptions or multi tasking are closely related to switches in the visual spatial attention (Fisher, 2011; Janczyk & Grabowski, 2011; Janssen & Brumby, 2010). A major objective of the current study is to understand and model experience-related differences in how scrub nurses control visual attention and manage the counting task. Visual scanning is a common measure of attention distribution, and several studies have documented expertise-related differences in attention allocation (primarily using visual scanning as a proxy for attention) in complex, high risk, real-world tasks, although not in health care surgery. The application of visual scanning in health care is limited to specific topics.

Seagull and colleagues (2004) used visual scanning measures to document attention allocation strategies by anesthesiology nurses during laparoscopic surgery, and found that a more distributed medical workplace imposed greater scanning. Another study on planning expertise in anesthesiology found that more experienced anesthesiologists search for visual cues that may indicate potential problems and provide information on obstacles to potential solutions (Xiao, Milgram, & Doyle, 1997). The performance and patterns of visual attention distribution of radiologists in medical image perception was extensively investigated (Krupinski, 2010), and causes of errors in searching, recognizing, and decision making have been discussed in relation to visual attention distribution (Kundel, 1989; Nodine, Mello-Thoms, Kundel, & Weinstein, 2002).

Charness and Tuffiash (2008) reviewed studies of expertise in attention in various domains including mental sports (Charness, Reingold, Pomplun, & Stampe, 2001; Ferrari, Didierjean, & Marmèche, 2008), physical sports (Hagemann, Schorer, Cañal-Bruland, Lotz, & Strauss, 2010; Savelsbergh, Van der Kamp, Williams, & Ward, 2005), radiology (Kundel & La Follette Jr, 1972), driving (Underwood, Chapman, Berger, & Crundall, 2003), industrial inspection (Megaw & Richardson, 1979), and in particular, aviation (Niessen, Eyferth, & Bierwagen, 1999). Findings from the eye-tracking studies are (a) experts capture information from a global scene as a chunked set within a short amount of time (Kundel, Nodine, Conant, & Weinstein, 2007), and (b) they focus on more relevant cues while ignoring nonrelevant information (Hagemann et al., 2010; Nodine, Kundel, Lauver, & Toto, 1996; Savelsbergh et al., 2005; Underwood, 2007).

Whereas the abovementioned studies have focused more on the channels of perceptual information for a single task (e.g., visual search), more directly related research about expertise in visual attention distribution with multiple tasks has been conducted primarily in the area of aviation. While they did not measure scanning, Wickens and Raby (1991) inferred expertise related attention allocation differences from activity recording, observing that more experienced pilots shifted attention between tasks more rapidly than did novices. Wickens et al. (2008) examined scanning in an advanced cockpit simulator. Although they did not group participants by experience per se, they did examine how the scan paths differed across levels of flight performance in this multitask environment (monitoring for traffic, while controlling the airplane). The authors employed an expected value model of attention allocation (described in more detail below) to assess the degree to which scan patterns matched optimal scanning (Sheridan, 1970; Wickens & McCarley, 2008), and discriminated different levels of multitask proficiency. Such discrimination is critical in the current study, and hence the expected value model is described in some detail as follows.

In high risk dynamic supervisory tasks, like vehicle control (flying and driving), air traffic control, and certainly operating room health care, it has been possible to model visual sampling and scanning on the basis of quantitative parameters, corresponding to objectively measured environmental and task variables, and hence specify models of optimal scan patterns (Moray, 1986). The roots of this work lay in the early studies by Senders and his collaborators (Carbonel, Ward, & Senders, 1968; Senders, 1964) who specified the importance of signal information content or bandwidth, in specifying the allocation of visual attention across channels of varying bandwidth. In a well calibrated supervisor, bandwidth should correspond to the expectancy with which events

occur in a channel; and optimal scan patterns, with fixation frequency proportional to bandwidth, can dictate the extent to which the fewest number of events will be missed across all channels. Here one can speak of an accurate mental model of the statistical properties of the environment, acquired through experience (Wickens & Hollands, 2000).

Sheridan (1970) added the concept of value to a model of optimal scanning, regarding "how often a supervisor should sample", with value objectively characterizing the cost of failing to notice an event on a channel. Thus combining these two parameters—expectancy and value—it was possible to specify optimum expected value models of scanning, in the same tradition of optimum expected value models of decision making (Edwards, 1996). Moray (1976) has well summarized this work. Sheridan (1970) also added the concept of effort in his model of scanning. Integrating both expectancy and value in an operator's scan pattern refines the concept of the knowledge-based mental model in driving visual scanning, and leads to the assumption that more skilled operators will have a more calibrated mental model as reflected in their scan pattern.

Both expectancy and value components have been compared between pilots of different levels of experience in aviation. Bellenkes, Wickens, and Kramer (1997) established that both components drive pilot scanning within the cockpit, and further, that more experienced pilots (flight instructors) sampled more valuable instruments more often than did less experienced pilots, thereby documenting an expertise difference in the mental model driving scanning. Schriver, Morrow, Wickens, and Talleur (2008) also established the greater degree of visual attention given to more valuable cues by more experienced pilots (than novices) in an aviation failure management simulation.

While neither Bellenkes et al.(1997) nor Schriver et al.(2008) attempted to model sampling by optimal models in the tradition of Sheridan and Senders, Wickens et al. (2003) did so by developing and validating the SEEV (salience effort expectancy and value) model in three aviation experiments. They extended Sheridan's expected value model (the last two components of SEEV), by objectively quantifying the effort parameter as the distance between pairs of instruments or "areas of interest" (AOIs) given the assumption that greater separation, by increasing the effort of visual scanning and head movement, would inhibit that scanning, making transitions less likely. In this, and subsequent elaborations of the SEEV model (Steelman-Allen, McCarley, & Wickens, 2011), Wickens and colleagues have also borrowed the concept of salience (S) from psychological models of visual attention capture (Itti & Koch, 1999, 2000, 2001a, 2001b). It may be seen then that expectancy (bandwidth) and value are parameters that should, optimally, drive scanning, while salience and effort are parameters that may pull, or inhibit (respectively), scanning in a nonoptimal fashion.

Using only the expectancy and value components of SEEV, Wickens, Goh, Helleberg, Horrey, & Talleur (2003) found the model to predict a high amount of scanning variance in three simulations in the general aviation cockpit (over 90% variance accounted for in each). Using the full SEEV model (including salience and effort**)** in three more recent studies in the more automated cockpit (Sarter, Mumaw, & Wickens, 2007; Steelman-Allen et al., 2011; Wickens et al., 2008) observed slightly less variance in scanning predicted by the model (90% and 86% respectively).

Critically, while two of the aviation studies described above established expertise-related scanning differences in flight control and troubleshooting (Bellenkes et al., 1997; Schriver et al., 2008), only one study has apparently examined scan optimality across skill levels, according to optimal scanning models. Here, Wickens et al. (2008) reported two findings directly relevant to the current study. First, while all of their pilots were of the same general level of experience (all flight instructors), the investigators contrasted pilots who performed better on the multi task aspects of controlling the aircraft and monitoring for hazards (other air traffic) with those who performed more poorly. They found that the fit of the better-performing pilots' scan to the optimum expected value SEEV model was significantly greater than the fit of the more poorly performing pilots, hence establishing a direct link between scan optimality and pilot performance.

The second relevant finding to the current investigation is that Wickens et al. (2008) compared pilot model fits to the SEEV model including the Effort component, with fits to the simplified EV model (removing the inhibitory influence of effort, as well as that of salience), and noted that the removal of this Effort component had no influence on model fit. This finding suggested that, at least with these well-trained professionals (all pilots were flight instructors), their scan pattern was not inhibited by a reluctance to make longer scans (greater effort). In other words, if a visual area of interest (AOI) was important (valued) and more dynamic (higher bandwidth and expectancy), it would be sampled just as frequently if it was close to a neighbor, as if it was farther away. Using this finding in the current research, this study also employed the simpler two-parameter EV model. (Salience was not employed in either experiment, but see Steelman-Allen et al., 2011). In this way, we could also establish the simpler two parameter model as a "gold standard" for optimal scanning.

To our knowledge, optimum scanning models have not been applied in a dynamic supervisory health care setting such as the operating theater, let alone applied in a way to try to differentiate skill levels of trained professionals in real world (e.g., nonsimulated) conditions, and this was one major goal of the current study, to examine if skill levels of scrub nurses could be differentiated on the basis of their adherence to optimal scan strategies, where the latter were computed on the basis of the expected value component of the SEEV model. The current study therefore aimed to examine differences in experience of scrub nurses in their visual attention and task management skills, with a focus in the latter on managing interruptions of the count task. Extrapolating from other domains of attention-allocation research in high tempo/safety critical domains, together with supporting literature from the adaptive strategies of scrub nurses during periods of high workload and task density, we hypothesize the following:

> *H1:* All nurses will generally adhere to optimal attention allocation, as prescribed by the SEEV model, and as reflected in a significant correlation between model predictions and actual attention allocation. This hypothesis is also part of validation of SEEV in the health care field, ascertaining that its validity in aviation transfers to the current domain.

H2: More experienced nurses will adhere more closely to the optimum expected value model than less experienced ones, reflecting a more finely developed mental model calibrating expectancy and value on the basis of their greater experience.

H3: The benefits of experience on optimality will be enhanced in higher workload and higher risk phases, when attention is at more of a premium, and hence adhering to optimal scanning is more safety critical.

H4: Higher experience leads to better performance in task management. While we are restricted in our measures of task management performance measures, we were able to focus explicitly on interruption management of the critical task of counting.

H5: There is a linkage between experience-related performance differences in task interruption management and experience-related differences in visual attention allocation. More optimal scanners will be more effective in preventing interruptions by less important tasks.

Method

Participants

Institutional Review Board (IRB) and Human Use Committee and Patient Agreement approvals were obtained prior to the study.

Data collection was performed at a major hospital in Singapore, focusing on standard caesarean section surgery case. Participants involved 20 scrub nurses ($n = 20$), with varying experiences in Obstetrics and Gynecology in the hospital. According to the Association of PeriOperative Nurses (AORN), a registered nurse with two or more years of experience in a specialty is considered experienced, and a registered nurse with less than two years is considered a novice. Ten experienced nurses (M = 7.4 years, $SD = 6.3$ years) and 10 novice nurses (M = 11.5 months, $SD = 5.6$ months) were randomly selected by the staff nurses and they voluntarily participated in the data collection. All participants have certified registered nurses with either a Diploma in Nursing or a Bachelor of Nursing, and no other significant differences between experienced and novice groups were identified.

Variables & Measures

Scrub nurse experience served as the main independent variable. The dependent variables were percentage dwell time, attention switches, time taken for count, frequency of interruptions during counting, and correlation with predicted optimal attention allocation values. In order to capture the distribution of attention, eye tracking equipment was used. An Applied Science Laboratories (ASL) Mobile Eye Tetherless eye tracking system was used to track the eye movements, and both pupil and corneal reflections were sampled at 60Hz (see Figure 1).

The perceived workload of the participant in each surgery was measured through NASA-TLX (Hart & Staveland, 1988) for the whole operation for the purpose of checking whether there was a significant difference between experienced and novice nurses and detecting any extraordinary cases of the surgery, such as complications.

Procedure

Before the main data collection, a pilot study with one experienced nurse and one novice nurse was conducted to verify the feasibility of the experimental procedure. Twenty caesarean section surgery cases with 10 experienced and 10 novice scrub nurses each were then performed.

Prior to the start of the preoperative phase of each surgery, the participants were asked to sign an informed consent form before wearing the eye tracker. The eye tracker was then calibrated to each nurse's eye profile. Participants then proceeded to the operating theater to start scrubbing up and to prepare the surgical tray and instruments for the operation while wearing the eye tracker. Consent of the surgical team members was obtained before the start of the operation, and verbal consent from the patient was also obtained beforehand. A maximum of two experimenters were located in the operating theater at all times to provide technical aid if required. The recording of eye movement started from the point when the eye-tracker was calibrated to the point when it was removed after the surgery. After the removal of the eye tracker, the participants were asked to fill out a demographics questionnaire and the NASA-TLX form regarding the operation in which they had just participated.

Analysis

Areas of interest. A scrub nurse frequently uses four visual sources of information. The areas are defined by distinct boundaries within which the scrub nurses perform their tasks (see Figure 2). Shown relative to the patient's body, the important areas of interest (AOIs) are:

A. Incision area (operation site)—the abdominal cavity of the patient where the operation procedure takes place.

B. Patient's lower body—the area including the pelvis and thighs of the patient which are covered in drapes. This area is used as a place to temporarily place instruments that are currently used by the surgeon.

C. Mayo tray—the tray located above the lower legs of the patient. It is used to place the instruments more frequently requested by surgeons during the surgery.

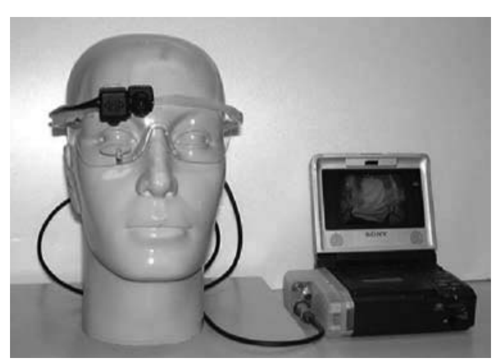

Figure 1. ASL Mobile Eye Tetherless Eye Tracking System.

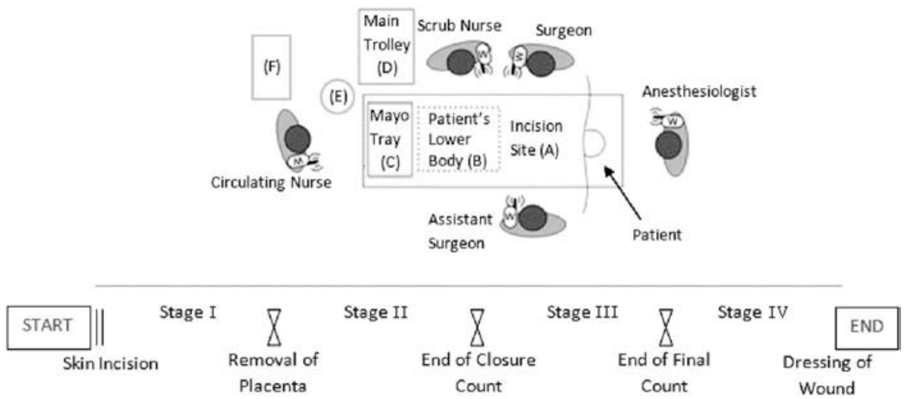

Figure 2. Time Line and Spatial Layout of the Key Phases and Areas During Surgery.

D. Main instrument trolley—the main trolley where the scrub nurse places all other instruments, gauzes and swabs. It is also the area for preparing the instruments, such as sutures, injections, and so forth.

Area E (kick about and sterile basin) and F (swab count rack) are visual sources of information supporting task of housekeeping, while G (others) include the scrub nurse's environment outside of her workstation. Area E, F, and G were rarely scanned, particularly area E, and so are excluded from the scanning analysis.

Stages of operation. Shown across the bottom of Figure 2 is a sequential time line of the four major stages of a typical caesarean section operation. The four stages were defined based on the critical events of a caesarean section elicited during the initial interviews with 33 experienced scrub nurses and nurse managers, which will be described later. Stage I spans from the initial skin incision to the removal of the placenta after the delivery of the baby. Stage II commences from the removal of the placenta until the end of the closure count. The closure count is a mandatory count to be performed in every caesarean section surgery, which includes accounting of all the instruments, gauzes, and swabs introduced into the sterile field throughout the operation before the closure of a cavity. A final count is a similar mandatory count of all the instruments, gauzes, and swabs introduced into the sterile field throughout the operation, but just before skin closure. Other counts that are performed throughout the operation include counts of gauzes, swabs, or instruments. These counts are constantly performed to check that all the items are accounted for throughout the surgery. Stage III spans from the end of the closure count to the end of the final count, and the final stage (Stage IV) starts at the end of the final count and ends at the dressing of the wound.

Attention allocation. Each video recording from the eye tracker was manually transcribed onto a computer using Eyevision™, the software accompanying the ASL Mobile Eye Tetherless Eye Tracking System. Each video was then analyzed manually to obtain dwell time and attention switches among different areas of operating theater during the surgery.

The attention distribution in each area of interest was captured in seconds and categorized throughout the intraoperative phase of the operation procedure, that is, from initial incision to dressing. When the nurse's gaze entered an area, the video would be paused and the fixation time recorded; the video would then be played, paused, and time recorded again when the nurse's eye was leaving that area. Throughout the analysis, glances away from the intended area of interest totaling less than 1 second were considered to be of no cognitive significance, and the nurse would be considered as still looking within the area of interest. The total dwell time (DT) of the scrub nurses in each of the areas were totaled, and calculated as a percentage of the total operation time within each of the stages (percentage dwell time, PDT). Their attention distribution was also plotted to facilitate a qualitative analysis of the patterns between and within stages. Along with that, the frequencies of attention switches between AOIs were also tallied.

SEEV model. The analysis involving the expected value version of the SEEV model predicted percentage dwell times for each area and stage and correlated the predicted PDTs with the actual PDTs of the novice and experienced nurses.

Table 1 shows the SEEV coefficients of expectancy and value that were used to generate the predicted percentage dwell times for each AOI and stage. These coefficients were assigned to each AOI and stage through a systematic analysis of scrub nurses' tasks and their relevance to locations across the operation stages.

The general heuristic for assigning expectancy/ bandwidth was the reported activity level during each stage. This was particularly heightened during Stage I in AOI A because here the surgeon's hands generated the activity, in a manner that was less predictable than observing motion caused by the nurses' own hands (picking and arranging instruments and swabs in the other areas).

The heuristic for assigning value parameters was based primarily on the relevance of the different areas, during different stages,

Table 1
Parameter Values for SEEV Optimal Model Predictions

Parameter	Stage I	Stage II	Stage III	Stage IV
Bandwidth/Expectancy (EX)				
Area A	3	3	2	2
Area B	2	2	1	1
Area C	1	2	2	1
Area D	2	3	3	2
Values (V)				
Area A	4	4	3	2
Area B	2	2	2	1
Area C	2	2	2	1
Area D	1	2	2	1

Table 2
Percentage Dwell Times in the 4 Main Areas of Interest Across the 4 Stages of a Standard Caesarean Section Surgery

Group	Area A *M* (*SD*)	Area B *M* (*SD*)	Area C *M* (*SD*)	Area D *M* (*SD*)	Area A *M* (*SD*)	Area B *M* (*SD*)	Area C *M* (*SD*)	Area D *M* (*SD*)
	Percentage Dwell Time in Stage I				Percentage Dwell Time in Stage II			
Experienced	0.522 (0.149)	0.244 (0.112)	0.064 (0.033)	0.044 (0.022)	0.371 (0.095)	0.238 (0.068)	0.054 (0.024)	0.173 (0.031)
Novice	0.443 (0.169)	0.313 (0.080)	0.102 (0.056)	0.047 (0.045)	0.280 (0.091)	0.301 (0.066)	0.082 (0.030)	0.225 (0.048)
Overall	0.483 (0.160)	0.278 (0.101)	0.083 (0.049)	0.046 (0.035)	0.326 (0.102)	0.270 (0.073)	0.068 (0.030)	0.199 (0.048)
	Percentage Dwell Time in Stage III				Percentage Dwell Time in Stage IV			
Experienced	0.217 (0.146)	0.183 (0.078)	0.086 (0.072)	0.329 (0.088)	0.353 (0.164)	0.118 (0.101)	0.012 (0.014)	0.362 (0.152)
Novice	0.113 (0.104)	0.215 (0.042)	0.116 (0.072)	0.455 (0.130)	0.255 (0.070)	0.171 (0.065)	0.008 (0.010)	0.427 (0.084)
Overall	0.165 (0.135)	0.199 (0.063)	0.101 (0.072)	0.392 (0.126)	0.300 (0.127)	0.147 (0.085)	0.010 (0.012)	0.398 (0.119)

Note. Stage I = Initial incision to removal of placenta; Stage II = Removal of placenta to end of closure count; Stage III = End of closure count to end of final count; Stage IV = End of final count to plastering of wound; Area A = Incision area; Area B = Patient's legs; Area C = Mayo tray; Area D = Main instrument trolley.

for the most important tasks. Through the consultation with subject matter experts, experienced scrub nurses, four major tasks were identified and prioritized. The first of two most important tasks was monitoring the patient's status, in other words, keeping watch on the incision area particularly when the body cavity was opened. Of equal importance was responding to the surgeon's requests for items or help. The third task of lower importance was performing and maintaining an accurate count of surgical items. Lastly, housekeeping was considered the fourth and lowest priority main task that the scrub nurses had to perform. Housekeeping included subtasks that the scrub nurses had to accomplish, namely checking the instrument count checklists, preparing injections, sutures, and gauzes, ordering new items, washing instruments, and arranging instruments. However, when in conflict with one of the three higher priority tasks, the latter should take precedence (Fletcher, McGeorge, Flin, Glavin, & Maran, 2002). In order to monitor the patient, the scrub nurse was expected to pay most of her attention to the patient, especially the incision area (Area A).To assist the surgeon, attention needed to be distributed across all four areas, but with greater focus on the incision area. To perform the counting task, the scrub nurse also had to process information from the main instrument trolley (Area D) and Mayo tray (Area C), and sometimes the incision area (Area A) or the area in between the incision area and Mayo tray (Area B) to check for instruments or gauzes in use. Housekeeping tasks were mostly related with the Mayo tray (Area C) and main trolley (Area D).

Two human factors engineers, with knowledge of the surgical procedure (one with extensive knowledge) independently evaluated the parameters for the four tasks of scrub nurses, by establishing relevance of each AOI to each task, and multiplying task relevance by task priority in the manner prescribed by Wickens et al. (2003). Then the overall value parameter (V) of each AOI in each stage was calculated and approximated to integer values (Wickens et al., 2003), which, in additive combination with bandwidth, was used as the predicted optimal distribution of attention. The interrater reliability between two evaluators, as measured by Cohen's kappa ($k = 0.685$), showed fairly high reliability, which was judged as sufficient.

Counting task. A detailed analysis on the task "surgical counts" was performed to identify all instances of instrument, gauze and swab counts that took place during the intraoperative phase (i.e., from initial incision to dressing of the wound). Subtasks that were coded for this study included: gauze count, swab count, instrument count, closure count, and final count. The frequency of counts, duration of counts, number of interruptions during counts, and the stages within which the counts happened were documented. Among these subtasks, the closure counts and final counts were further analyzed. The contents of the counts were similar for both the closure counts and final counts, encompassing all the instruments, gauzes and swabs that had been used throughout the surgery. The duration of each final count and closure count was recorded, together with the number and qualitative descriptions of the interruptions that happened during the counts and incorrect counts (if any). The sources of interruptions, such as requests by other surgical team members, procedural violation, and other housekeeping duties, were identified and tallied to capture the nature of interruptions.

Workload and risk distribution. Independent from the data collection in the operating theater, a questionnaire survey with scrub nurses was conducted to solicit information on the workload distribution throughout a c-section surgery. Thirty-three c-section scrub nurses and nurse managers with a mean of six years of experience ($SD = 5$ years) participated in the survey which evaluated the workload of a c-section surgery divided into four stages. Each of the ratings was compiled into a matrix, tallying the total ranks of '4', '3', '2', and '1' for each of the stages, where 4 was the highest workload, and 1 the lowest. Each count was then multiplied by the rank number, and added up to form the total score for the stage. The values did not serve as absolute indicators of workload for the stages but as relative scores that could be used to compare between the four stages.

Approximately 88% of the scrub nurses indicated Stage I as the highest or the second highest workload. Stage II was also ranked very high in terms of the workload. Ninety-four percent of the scrub nurses rated it either the highest or the second highest. There was no clear distinction between the two stages in their workload. However, 67% of scrub nurses clearly ranked Stage III as relatively lower workload as compared to Stages I and II. Seventy-nine percent of scrub nurses ranked Stage IV as the stage of lowest workload. The evaluated ranks were converted into a score by

giving 4 for the highest ranking and 1 for the lowest ranking. The frequency and score were multiplied and summed up to get a single score per each stage. Stage I (109) and II (114) were interpreted to be the stages of highest workload, followed by Stage III (71) then Stage IV (44).

Corresponding to the distribution of workload and the qualitative description of the stages by the 33 SMEs, the four stages were also discriminated in terms of risk to the patient. The risk of the patient is usually highest in Stage I and Stage II, followed by Stage III, and then Stage IV. Solicited from the questionnaires, the surgery approaches its risk climax toward the end of Stage I and at the start of Stage II, where the chance of bleeding (which is detrimental to patient status) is highest. The risk level gradually decreases in Stage III as the surgeon completes the crucial steps of the operation, and reaches its lowest in Stage IV after the cavity closure.

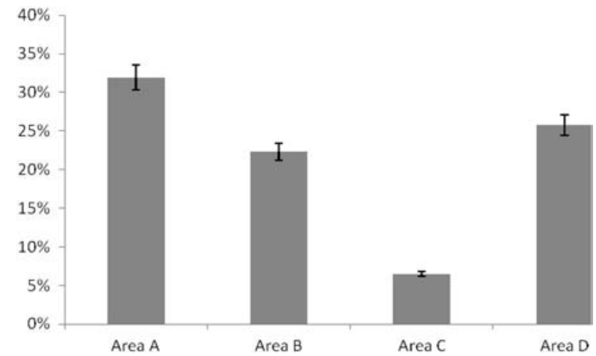

Figure 3. Percentage Dwell Time Across All 4 Areas. *Area A = Incision site; Area B = Patient's legs; Area C = Mayo tray; Area D = Main instrument trolley.*

Results

Scan Analysis

The normality of the data was checked by normal quantile plots and Shapiro-Wilk test of normality. The normal quantile plot reflected that the results generally follow a normal distribution, and Shapiro-Wilks test of normality confirmed that the normality of the data was not violated.

The analysis of visual scanning helped to confirm the relative importance of different areas, and also, as we describe below, how this importance was modulated by both surgical phase and level of experience. A 4 (stages) by 4 (AOI) by 2 (level of experience) ANOVA was carried out, with percentage dwell time (PDT) as the dependent variable, and the results are discussed in the following section focusing first on the overall pattern of scanning, and then how this pattern was moderated by experience.

The main effect of stages was not significant, $F(3, 292) = 0.19$, $p = .91$, $\eta^2 = 0.002$, an unsurprising result because the measure PDT was constrained to sum to 1.0 within each stage. However, more importantly the main effect of areas was highly significant, $F(3, 292) = 117.80$, $p < .0001$, $\eta^2 = 0.552$, indicating that the incision area (Area A) received highest attention (31.9%) compared to the patient's lower body (Area B; 22.2%, $p < .0001$), Mayo tray (Area C; 6.5%, $p < .0001$), and main trolley (Area D; 25.7%, $p = .0001$). There was no significant difference ($p = .07$) between the patient's lower body (Area B; 22.2%) and main trolley (25.7%), but Mayo tray (6.5%) received lowest attention compared to other areas ($PDT_C < PDT_A$, $p < .0001$; $PDT_C < PDT_B$, $p < .0001$; $PDT_C < PDT_D$, $p < .0001$). This general pattern was consistent with the critical importance of patient direct monitoring (A), and the diminished attention to the Mayo tray. (see Figure 3)

Participants were found to allocate attention differently across surgical stages, $F(9, 292) = 42.97$, $p < .0001$, $\eta^2 = 0.574$, as indicated in Figure 4 which depicts the interaction between stages and areas. The most pronounced contribution to this interaction appeared to be in Stage I, within which the effect of AOI was highly significant, $F(3, 292) = 104.53$, $p < .0001$; PDT_A 48.3% > PDT_B 27.8%, $p < .0001$; PDT_B 27.8% > PDT_C 8.3%, $p < .0001$, reflecting the heavy attention allocated to the incision area during the opening of body cavity, delivery of baby and subsequently the removal of the placenta. Similarly, in Stage II, the attention on Area A was significantly higher than Area C and Area D, $F(3, 292) = 31.36$, $p < .001$; PDT_A 32.5% > PDT_C 6.7%, $p < .0001$; PDT_A 32.5% > PDT_D 19.8%, $p = .0008$, although there was no significant difference between Area A and Area B (PDT_A 32.5% = PDT_B 26.9%, $p = .8139$). However, the attentional interest in Area A declined substantially after the placenta had been removed, and greatest interest was placed on Area D in Stage III, $F(3, 292) = 39.71$, $p < .0001$; PDT_D 39.2% > PDT_B 19.9%, $p < .0001$; PDT_B 19.9% > PDT_C 10.1%, $p < .0001$, and Stage IV, $F(3, 292) = 65.37$, $p < .0001$; PDT_D 39.5% > PDB_B 14.4%, $p < .0001$; PDT_B 14.4% > PDT_C 0.8%, $p < .0001$, due to the need to perform constant counts of instruments, gauzes and swabs to account for all the instruments within the sterile field.

Figure 5 shows the highly significant interaction between experience and area, $F(3, 292) = 12.85$, $p < .0001$, $\eta^2 = 0.119$, supporting hypothesis H2. Multiple comparison of least square means using Tukey-Kramer adjustment showed that there were significant differences between experienced and novice nurses on Area A (incision area; PDT_A(experienced) 36.5% > PDT_A(novice) 27.3%; $p < .0001$) and Area D (main trolley; PDT_D(experienced) 22.7% < PDT_D(novice) 28.8%; $p = .046$), but nonsignificant differences on Area B (lower body of the patient; $p = .12$) and Area C (Mayo tray; $p = .92$). Thus we see that more skilled nurses looked considerably more at the incision area, indicating about 30% more scanning.

To better understand the pattern of scanning, and how it differed between experienced and novice nurses, the percentage dwell times were compared with the predicted optimal attention allocation generated using the EV components of the SEEV model.

SEEV Predicted PDT Versus Actual PDT

The details of the highly significant level of experience × AOI interaction (see Figure 5) were examined first through the SEEV model fit. Figure 6 shows the scatter plot of model predictions versus obtained PDT for the experienced group averaged over all experienced nurses (Figure 6a) and the novice group averaged over all novice nurses (Figure 6b). It revealed that there was a higher prediction by the expected value modeling for those of higher ($R^2 = 0.63$) than those of lower ($R^2 = 0.36$) level of experience, even as sampling of both groups revealed a substantial adherence to the model, with both correlations being highly significant ($p < .0001$). Therefore H1 was supported.

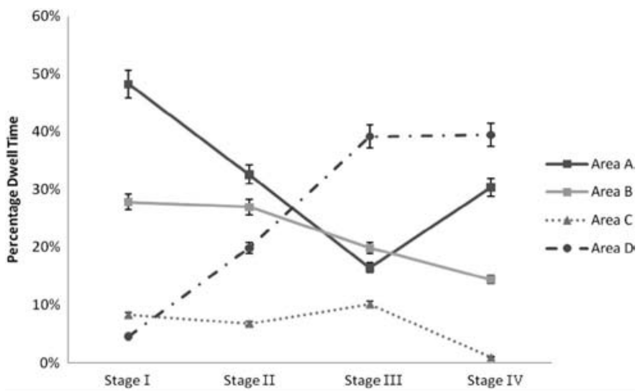

Figure 4. Plot of Percentage Dwell Time Across the 4 Areas and 4 Stages. *Stage I = Initial incision to removal of placenta; Stage II = Removal of placenta to end of closure count; Stage III = End of closure count to end of final count; Stage IV = End of final count to plastering of wound; Area A = incision area; Area B = patient's legs; Area C = Mayo tray; Area D = main instrument trolley.*

In order to examine the experience difference statistically, separate model fits were computed for each nurse, with each correlation based on an *N* of 16 points (the four stages and four AOIs). The model fit was then calculated within each stage, so that each correlation was based on an *N* of 4 points (the four AOIs). These correlations then became the raw data for statistical analysis, a set of separate *t*-tests comparing the two experience groups on an overall basis and then within each stage of surgery. Values that were larger than 2.5 standard deviations were considered outliers and removed from the data. Only one observation point from the experienced nurses in Stage I was removed. The *t*-test comparing the experienced and novice group revealed significantly higher correlations for the experienced group than the novice group, with $t(18) = 4.02$, $p = .0008$, supporting H2. The results of the *t*-tests comparing the two experience groups for the individual stages are shown in the top part of Table 3, reflecting significantly higher correlations for the more experienced group in stage I and II, but not 3 and 4. These results served to support H3.

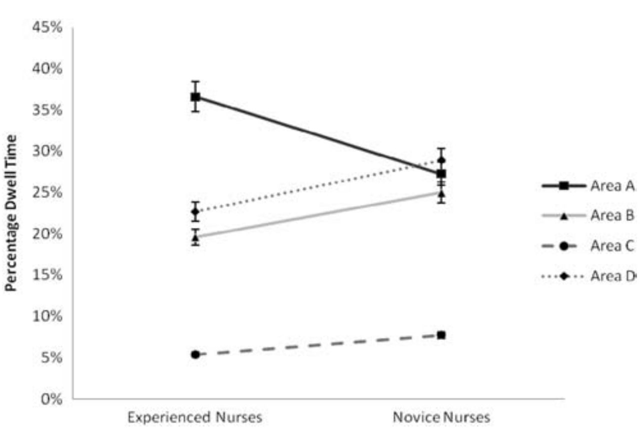

Figure 5. Plot of Percentage Dwell time Across Areas for Experienced nurses and Novices. *Area A = incision area; Area B = patient's legs; Area C = Mayo tray; Area D = main instrument trolley.*

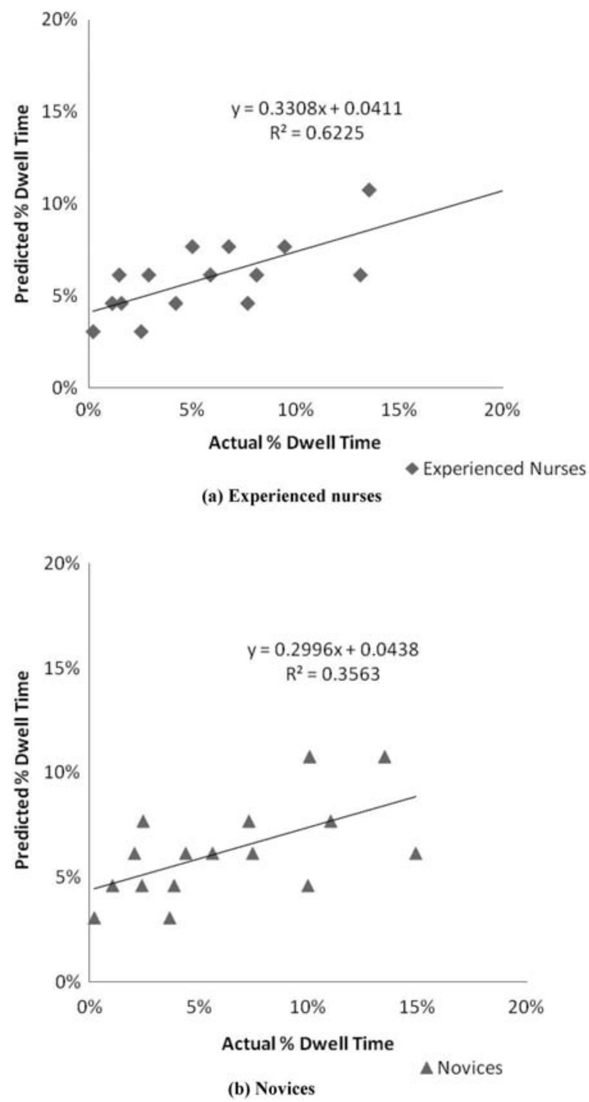

Figure 6. Model Prediction versus Scan (Percentage Dwell Time) Performance.

Having then established differences in the levels of experience between two of the four stages, we proceeded to examine the primary source of these differences by performing post-hoc tests for each AOI within all of the stages where differences in model fit were obtained. The focus here was on the AOI that showed the largest difference within each stage. Both stages I and II showed that the experienced group spent significantly more time scanning Area A (the body cavity) than the novice group. This was confirmed by a 2 (level of experience) by 4 (stages) ANOVA on PDT within AOI A, which revealed a significant effect of level of experience, $F(1, 70) = 10.36$, $p = .002$, $\eta^2 = 0.129$. Indeed contrasts revealed that experienced nurses focused on Area A significantly more than novices during all stages ($p < .05$) except Stage III ($p = .08$).

Table 3
Model Fits and Discriminating Areas of Interest Between Experienced and Novice Scrub Nurses Within Each Stage

Model Fit	Stage I	Stage II	Stage III	Stage IV
Experienced	R = .98	R = .71	R = .33	R = .69
Novice	R = .79	R = .30	R = .22	R = .56
Difference	T = 3.17, p = .0055	T = 2.83, p = .011	T = 0.72, p = .48	T = 1.17, p = .26

Most Discriminating AOIs	A	B	D	A	NA	NA
Experienced	13.6%	5.88%	1.14%	20%		
Novice	10.1%	7.46%	1.04%	13.5%		
Difference	p = .044	p = .047	p = .012	p = .013		

Surgical Counts and Interruptions

The process of counting is a critical aspect of the scrub nurse's responsibility. Of the 140 counts recorded in the 20 surgeries, 136 counts occurred in Stage II and Stage III (97.14%); two counts occurred in Stage I (1.42%), and two counts in Stage IV (1.42%). The results illustrated in Table 4 show that scrub nurses performed an average of seven counts per surgery, spending a mean of 8.57% of their time counting, with the experienced nurses spending 7.41% of their time counting and novice nurses spending 9.15%. They averaged 27.1 seconds per count ($M_{experienced}$ = 23.8s, M_{novice} = 28.7s), and encountered a mean of 2.2 interruptions during their counts.

Among the different types of counts, time taken for the closure and final count were compared to find any differences according to experience. Since other instrument and gauze counts optionally performed by scrub nurses entail different number of items each time, only the mandatory counts were compared. A 2 (level of experience) by 2 (count type) ANOVA analysis for the time taken for counts reflected significant main effect of count type, $F(1, 36) = 58.73$, $p < .0001$, $\eta^2 = 0.620$, and interaction effect of level of experience and counts, $F(1, 36) = 4.56$, $p = .0396$, $\eta^2 = 0.112$. Contrast result showed that experienced nurses spent significantly less time on the final count as compared to novices, $F(1, 36) = 6.24$, $p = .0172$, $\eta^2 = 0.1477$, supporting hypothesis H4, but no significant difference was found on the closure count, $F(1, 36) = 0.27$, $p = .604$, $\eta^2 = 0.0075$.

To test for hypothesis H5, further analysis based on the frequencies and nature of interruptions during counts was conducted. Out of the 140 counts (experienced: 77; novice: 63) performed, experienced nurses encountered fewer interruptions (17 times; 38%) than novice nurses (28 times, 68%), a difference that approached conventional levels of statistical significance, $c^2(1, N = 45) = 2.69$, $p = .051$. An analysis of the causes of interruptions during the surgical counts reflected various reasons, including assisting the surgeon, unable to find instruments, incorrect counts, interruption by circulating nurse, and housekeeping duties. Ideally, the process of counting should only be interrupted by an interrupting task of higher priority (Wickens & McCarley, 2008), which should only be assisting the surgeon, according to the task importance ratings used for the SEEV parameter table. The correlation between the number of interruptions experienced by each nurse and the optimality score of the SEEV model fit was not significant for either the experienced ($r = .445$, $p = .197$) or the novice ($r = .343$, $p = .332$) nurses. Interruptions to assist the surgeon were equally distributed between novices (7) and experienced (9) nurses $c^2(1,16) = 0.25$, $p = .308$). However non-surgeon-triggered interruptions were more frequently observed with novices (19) than experienced nurses (10) $c^2(1,29) = 2.79$, $p = .047$.

Attention Switching

The frequency of total attention switches per minute between areas within each stage was tabulated and plotted (See Figure 7).

The results suggested that novices switched attention from AOI to AOI more often than experienced nurses. A two-way ANOVA of experience and stages for the attention switches showed significant differences in the main effect of experience, $F(1, 70) = 7.37$, $p = .008$, $\eta^2 = 0.095$. Novices were found to switch between AOIs about 40% more often than the experienced nurses in Stage I (M_{novice} = 18.04, $M_{experienced}$ = 12.90; $F(1, 70) = 4.41$, $p = .0395$). In stages II, III, and IV, no significant differences on attention switching were found by level of experience.

Table 4
Analysis of the Counting Tasks Performed by the Scrub Nurses Throughout the Intra-Operative Phase of Surgeries

Group	Number of Counts: Stage I	Stage II	Stage III	Stage IV	Time for Counts (secs)	Average Time per Count (secs)	Time for Surgery (secs)	Percentage Time for Counts
Experienced								
TOTAL	1	57	17	2	1759		27499	
AVG	0.1	5.7	1.7	0.2	175.9	23.8	2749.9	7.41%
Novice								
TOTAL	1	46	16	0	1685		20568	
AVG	0.1	4.6	1.6	0	168.5	28.7	2056.8	9.15%

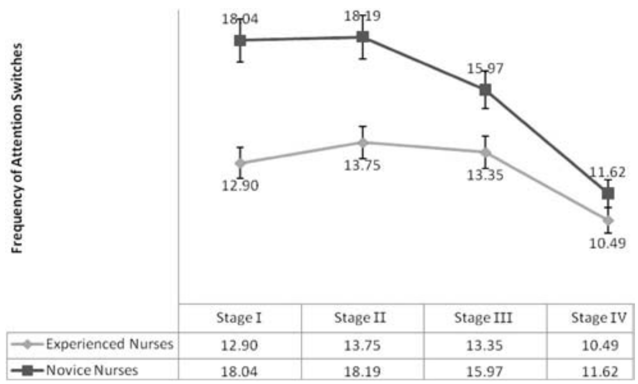

Figure 7. Graph Illustrating the Frequency of Attention Switches per Minute Between All Areas of Interest. *The analysis of the attention switches are also broken down to individual stages, to analyze the attention switches with respect to differing periods of workload.*

The main effect of stages on the attention switches was also found to be significant, $F(3, 70) = 3.103$, $p = .032$, $\eta^2 = 0.117$, indicating less switching during stage IV than during the prior stages. The interaction effect of level of experience and stages was not significant, $F(3, 70) = 0.529$, $p = .664$, $\eta^2 = 0.022$.

NASA-TLX Subjective Mental Workload Ratings

Out of the 20 subjective mental workload forms that were collected, two forms were incomplete and could not be included in the analysis. There was no significant difference between the subjective mental workload of the experienced and novice scrub nurses, $F(1, 17) = 1.980$, $p = .179$, although novice nurses showed higher subjective workload than experienced nurses ($M_{experienced} = 7.14$, $SD = 2.95$; $M_{novice} = 9.21$, $SD = 3.30$).

Discussion

The purpose of the current experiment was to establish the linkages between experience and visual attention (scanning) in medicine, and the possible link between these and measures of task/interruption management. These linkages were collectively expressed in a set of hypotheses formulated partially on the basis of the intersection of prior research in these areas. We were able to study the role of attention in actual clinical practice, rather than needing to extrapolate to these conditions from simulated conditions.

The first hypothesis was that nurses' attentional strategies would be predicted by the expected value model of visual attention. The participants were all well-trained with, presumably, a good mental model of environmental events and their job responsibilities, and the scanning required to carry out these responsibilities. Thus, while confirming that expected value drives scanning is itself somewhat expected, such confirmation has never before been examined with health care professionals. The obtained correlations between scanning and the EV predictions for both groups, shown in both Figure 6 and Table 3 were reasonably strong and positive. Here, we note that the three parameters of the SEEV model (including both relevance and task priority, used to compute value) were established independently of the data (i.e., were not adjusted post hoc to optimize the fit with the data); and these were based on plausible indicators of optimal sampling, independently validated in prior research: expectancy (bandwidth) by Senders and his colleagues (Carbonel et al., 1968; Senders, 1964, 1966), and Value or AOI importance (Moray, 1976; Sheridan, 1970; Wickens et al., 2003).

At the same time, the present experiment provided additional validation of the SEEV model to a domain (health care) beyond that of driving (Horrey et al., 2006) and aviation (Steelman-Allen et al., 2011; Wickens et al., 2003; Wickens et al., 2008) to which it had already been applied, hence expanding the generality of its predictions to real world (e.g., extralaboratory) dynamic environments. This was the first study to have been applied to actual operators performing a "live" task.

The second hypothesis was that more experienced nurses would use more optimal attention allocation strategies, as reflected in a better fit of their data to the SEEV model. Such a hypothesis was based on ample findings of novice-expert differences, or experience-related changes in scan strategies in other dynamic environments (Bellenkes et al., 1997; Huestegge, Skottke, Anders, Musseler, & Debus, 2010; Mourant & Rockwell, 1972; Nodine et al., 1996; Page, Bates, Long, Dawes, & Tipton, 2011; Pollatsek, Fisher, & Pradhan, 2006; Pradhan, Pollatsek, Knodler, & Fisher, 2009; Schriver et al., 2008), and investigators have interpreted such differences qualitatively by observing that those with more experience use more relevant information (e.g., Bellenkes et al., 1997; Schriver et al., 2008). But such differences have not before been examined in health care nor in any prior studies interpreted against a "gold standard" of optimality to determine the quantitative degree of adherence to or departure from this gold standard by workers at different experience levels.

The evidence in support of our second hypothesis was compelling. The ANOVA established the statistical differences in scanning between the two experience groups, and statistical differences in the more refined fit to the EV model established the clearly more optimal scanning of the experienced group than of the novice group. Furthermore, as reflected in Hypothesis 3 and shown in Table 3, this difference was observed exclusively during the two highest workload and highest risk stages (Stage I and II), when the body cavity was open, as this greater workload and risk was independently established by the SME interviews. Such enhancement of experience-related optimality of resource allocation at times of high workload and risk is both understandable and an important and valued manifestation of expertise. Stated simply, when patient risk is highest, the costs of errors in attention allocation are most severe.

At this point, it is worth spending some time reviewing the data for **how** the two groups differed, and here we return to the ANOVA and the data in Table 3 for details. Such data indicate that the major difference lies in the greater attention allocation by the experienced nurses to the body cavity, nearly 1.5 times as much. This difference presumably reflects in part, a greater vigilance toward the surgeon's needs, in particular, anticipating the surgeon's instrument needs from his or her observed manual actions. In this regard, one of the important cognitive nursing skills highlighted by Flin, O'Connor, and Crichton (2007) includes situation awareness, where a higher level of experience is usually associated with more adept anticipation, or level 3 situation awareness (Fletcher et al., 2002; Mishra, Catchpole, & McCulloch, 2009; Yule, Flin, Paterson-Brown, & Maran, 2005). In an attempt to define some properties of expertise in the control of complex and

dynamic environments, Cellier, Eyrolle, and Marine (1997) also characterized experienced nurses to have greater skill in anticipating and processing cues proactively. Bellenkes et al. also observed that more experienced pilot scanning reflected greater attention to areas associated with anticipating aircraft behavior.

Because this favoring of the surgical site by the experienced nurses persists after Stage I, it may also reflect their greater concern for overall patient health, as we defined formally to be a first task in the SEEV analysis, visual information that can only be extracted from this area (Schulz-Stubner, Jungk, Kunitz, & Rossaint, 2002). However as a third explanation, it also may have been the case that the novices spent less time attending to this vital area because their limited visual resources were more preoccupied with managing the main trolley, a skill that presumably develops with experience. Here again an analogy with the flight expertise/scanning study of Bellenkes et al. (1997) is direct. The novice pilots in that study required nearly twice as much time (e.g., more visual resources) to extract the flight stability information from the moving horizon display, hence availing less time to monitor the predictive flight instruments. Hence the absence of skill development on component tasks requires added resources that might otherwise be deployed to more important tasks.

The link to performance, (different skill levels in instrument management) in the above analysis was of course inferred rather than directly observed, and we sought more direct evidence for performance differences to support Hypothesis 4. While more experienced workers may not always be better performers, emphasizing the distinction between experience and expertise, we did anticipate finding some differences. Although we could not observe or code many aspects of performance quality, we were able to identify differences (in favor of experience) more directly related to the counting task that underlay the retained foreign object problem. This problem of course initiated the current investigation, with an overall assumption that experience would make counting more fluent and less interrupted, thereby better supporting patient safety.

Fortunately, from a patient safety standpoint, there were no RFO incidents (nor near incidents) in our data sample, so our investigation then focused on counting. Here we found that experienced nurses performed the final counting significantly faster (with no loss of accuracy). We also found that experienced nurses allowed fewer interruptions in general ($p = .51$), and significantly fewer interruptions in particular from all sources other than surgeon requests, the latter being the highest priority task (10 vs. 19, $p = .047$). While the sample size (29) was small, limiting statistical power, it was still in a direction to suggest that experienced nurses were better at staying "on task" (here, counting) in the face of lower priority interruptions, than were novices.

Hypothesis 5 concerned the extent to which we could draw an explicit link between scan optimality differences and performance differences. Of course this link is implicit because we observed group differences in measures of both scan optimality and performance, in a direction that suggested novices were lower on both measures. However, such between-groups covariance is only weak evidence for causality. Thus, here we turned to two sources of data. First, we correlated the number of count interruptions experienced by each nurse, with his or her optimality score. While such correlations were nonsignificant, they were also relatively uninformative because the small number of interruptions was fairly evenly spread across all of the nurses. Correlation is a powerful measure only to the extent that both X and Y variables show considerable variance. Here, while optimality scores did, count interruptions did not.

Second, we looked at the behavior of one outlier who did show a high number of interruptions. Participant #8 exhibited seven of the 17 interruptions shown by the high experience cohort. Only one other nurse in that group showed more than two. At the same time, the same participant was the only nurse in either group classified as an outlier in the EV model fit analysis, with an optimality correlation of more than 2.5 *SD* away from (and here, less) than the mean. Such a finding is certainly consistent with an association between less optimal scanning, and greater interruptability, but is far from conclusive, and Hypothesis 5 must await more controlled simulation experiments to precisely establish the link between scanning optimality and performance in health care (but see Wickens et al., 2008 for such link in aviation).

Returning to the finding that experienced nurses stayed on task more for the counting task, this finding is also consistent with our observation that this group showed about 30% fewer attention switches, particularly during those high workload stages I and II (see Figure 7). While such an observation contradicts the finding by Raby and Wickens (1994), it is noteworthy that those authors based their conclusions directly on task switching, extracted from a flight deck activity recording log, whereas the current conclusions were extracted from visual attention switching.

There is also a second important difference between the two studies that can account for the opposing findings. In the flight deck environment, some of the tasks switched between are of near coequal importance. In contrast, in the surgical environment during stages I and II, the body cavity possesses twice the importance (Value) and 50% greater Expectancy as any other AOI (see Table 1). Hence more optimal scanning will require fewer transitions away from this most important area, and hence longer dwells (duration of stay) once the eye enters it. More experienced nurses exhibited this sustained scan on the cavity more than the novices.

A third factor differentiating the flight and surgical environments (and hence explaining the different findings of the two studies) is that in aviation, at least two of the tasks—flight stability and navigation—are affected by environmental sources of uncertainty (wind turbulence). In surgery, however, nearly all of the environmental uncertainty is associated with the physicians' hands and the patient's response, both within the body cavity. The sources of activity (and hence Expectancy in SEEV) associated with other AOI's are primarily related to the nurses own activity which is, by definition, less "uncertain" from her point of view because she generated it, and hence does not need visual attention to monitor for it.

Limitations and Future Directions

A major limitation of the current research was of course that it was not a true experiment, with control exerted over the independent variables (e.g., stages of the surgical process), nor did we have the ability to manipulate others, such as the frequency of interruptions or the timing of the count tasks. Such constraints are, unfortunately inevitable in the naturalistic (nonsimulated) environment in which our data were collected, but we are convinced that the benefit of having participants with fully realistic motivations can offset the loss of control. That is, the "applied" aspect here

balances the "experimental" aspect articulated in the title of this journal.

A second limitation, also resulting from the naturalistic environment is lower than desired statistical power for some of the elements we analyzed, in particular the count interruptions. Similar to a recent naturalistic study of air traffic controller response to false alarms in a conflict alerting system (Wickens et al., 2009), we were at the mercy of our conditions to generate a sufficient number of events of interest, to assess differences within. Here, as in that prior study, we were somewhat fortunate. Although the number of count interruptions was relatively small, it still remained large enough so that between group differences could be identified with some confidence when they occurred (e.g., in non physician interruptions). A third limitation could be seen in our establishment of the Expectancy and Value parameters of the SEEV model. We did not evaluate the effort parameter of the model as has been done in some other simulations. To do so requires a dynamic simulation version of SEEV (Wickens & McCarley, 2008) which was beyond the scope of this effort; and merit can be placed on parsimony: the ability of the simplest two parameter linear additive equal weighting model chosen here to account for the data. Since the EV version of the model is the only one that is truly optimal, and because skilled pilots conform to this version as well as to the full SEEV model (Wickens et al., 2008), it can be argued that the simpler model used here is the superior version for expertise evaluation.

A fourth limitation and one that applies to all SEEV (or EV) model applications is that the Value parameter can only be established by subjective estimates of model users, intimately familiar with the work domain, who must assess the importance of tasks and the relevance of each task to each area of interest. This requires a careful cognitive task analysis. We believe that the first and second authors, the first having observed all of the surgeries and collected interviews from many of the scrub nurses, possessed the capability to make these assignments reliably.

In spite of these limitations, we believe the data make a strong case for the importance of visual scanning as a documented feature differentiating experience levels in high risk, high tempo nursing environments. We believe the association between scanning optimality and interruption management, in this and a variety of other environments as suggested by these data, clearly warrants future experimentation in a controlled simulation environment, in a manner paralleling the intriguing work on nursing interruptions carried on by Grundgeiger, Sanderson, MacDougall, and Venkatesh (2010). Our intention is also to pursue the development of technology aids to assist with the counting task, given that we found a substantial number of interruptions with this task. Although new technologies, such as radio-frequency identification (RFID) scanning, are being introduced in surgical counts (Macario, Morris, & Morris, 2006), the development of automation alongside the assistance of interruption management has yet to be explored.

Practical Implications

We see three major practical implications of the current work. First, in identifying the major scanning differences between the two groups, we believe that relatively simple instructional training material can be provided, focusing on the need for ample attention to the incision area. Such training may be coupled by intensive part task training of instrument management in the other areas, in order to develop high proficiency ("automation") with this task, availing, for less experienced nurses, ample resources to the incision area. The potential importance of some training in interruption management (Dodhia & Dismukes, 2009; Trafton, 2007) can also be viewed as an important implication. Finally, we feel these results suggest that the EV component of the SEEV model can itself serve as an important benchmark or standard for evaluating training in complex environments.

Conclusion

The study of the impact of differing levels of experience, and how this is manifested in cognition continues to be a hallmark of research in applied psychology (e.g., Cellier et al., 1997; Charness & Tuffiash, 2008; Ferrari et al., 2008). Overlapping this issue, discussions of the degree of optimality of human performance in general, and decision performance in particular (Kahneman & Klein, 2009), offers an equally intriguing link between more basic and applied psychology. The current data provide a useful link between these two approaches to applied cognitive psychology, considering the direction of visual scanning as a decision of where to deploy visual attention in the service of the multitask demands of surgery.

References

AORN. (1999). Recommended practices for sponge, sharp, and instrument counts. *AORN Journal,* 459–468.

Bani-Hani, K. E., Gharaibeh, K. A., & Yagha, R. J. (2005). Retained surgical sponges (Gossypiboma). *Asian Journal of Surgery, 28,* 109–115. doi:10.1016/S1015-9584(09)60273-6

Bellenkes, A. H., Wickens, C. D., & Kramer, A. F. (1997). Visual scanning and pilot expertise: The role of attentional flexibility and mental model development. *Aviation Space and Environmental Medicine, 68,* 569–579.

Beyea, S. C. (2003). Counting instruments and sponges. *AORN, 78,* 290–294. doi:10.1016/S0001-2092(06)60778-3

Brennan, T. A., Leape, L. L., Laird, N. M., Herbert, L., Localio, A. R., Lawthers, A. G., . . . Hiatt, H. (2004). Incidence of adverse events and negligence in hospitalized patients: Results of the Harvard Medical Practice Study I (Reprinted from New England Journal of Medicine, Vol. 324, pg 370–7, 1991). *Quality & Safety in Health Care, 13,* 145–151. doi:10.1136/qshc.2002.003822

Carbonell, J. R., Ward, J. L., & Senders, J. W. (1968). A queuing model of visual sampling experimental validation. *IEEE Transactions on Man-Machine Systems, MMS93,* 82–87.

Cellier, J. M., Eyrolle, H., & Marine, C. (1997). Expertise in dynamic environments. *Ergonomics, 40,* 28–50. doi:10.1080/001401397188350

Charness, N., Reingold, E., Pomplun, M., & Stampe, D. (2001). The perceptual aspect of skilled performance in chess: Evidence from eye movements. *Memory & Cognition, 29,* 1146–1152.

Charness, N., & Tuffiash, M. (2008). The role of expertise research and human factors in capturing, explaining, and producing superior performance. *Human Factors, 50,* 427–432. doi:10.1518/001872008X312206

Couch, N. P., Tilney, N. L., Rayner, A. A., & Moore, F. D. (1981). The high cost of low-frequency events: The anatomy and economics of surgical mishaps. *New England Journal of Medicine, 304,* 634–637. doi:10.1056/NEJM198103123041103

Dagi, T. F., Berguer, R., Moore, S., & Reines, H. D. (2007). Preventable errors in the operating room: Retained foreign bodies after surgery - Part

II. *Current Problems in Surgery, 44,* 347–381. doi:10.1067/j.cpsurg.2007.04.001

Dodhia, R. M., & Dismukes, R. K. (2009). Interruptions Create Prospective Memory Tasks. *Applied Cognitive Psychology, 23,* 73–89. doi:10.1002/acp.1441

Edwards, W. (1996). European Behavioral Decision Theory, as SPUDM-14 displayed It. In Contributions to Decision Making: I. J. P. Caverni, M. Bar-Hillel, F. H. Barron, & H. Jungermann (Eds.), Amsterdam: Elsevier Science, 1995, ISBN 0-444-82181-3, 361pp. *Journal of Behavioral Decision Making, 9,* 301–303. doi: 10.1002/(SICI)1099-0771(199612)9:4<301::AID-BDM234>3.0.CO;2-A

Egorova, N. N., Moskowitz, A., Gelijns, A., Weinberg, A., Curty, J., Rabin-Fastman, B., . . . Greco, G. (2008). Managing the prevention of retained surgical instruments - What is the value of counting? *Annals of Surgery, 247,* 13–18. doi:10.1097/SLA.0b013e3180f633be

Ferrari, V., Didierjean, A., & Marmèche, E. (2008). Effect of expertise acquisition on strategic perception: The example of chess. *The Quarterly Journal of Experimental Psychology, 61,* 1265–1280. doi:10.1080/17470210701503344

Fisher, A. V. (2011). Automatic shifts of attention in the Dimensional Change Card Sort task: Subtle changes in task materials lead to flexible switching. *Journal of Experimental Child Psychology, 108,* 211–219. doi:10.1016/j.jecp.2010.07.001

Fletcher, G. C. L., McGeorge, P., Flin, R. H., Glavin, R. J., & Maran, N. J. (2002). The role of non-technical skills in anaesthesia: A review of current literature. *British Journal of Anaesthesia, 88,* 418–429. doi:10.1093/bja/88.3.418

Flin, R., O'Connor, J. P. B., & Crichton, M. T. (2007). *Safety at the Sharp End: A Guide to Non-Technical Skills.* Burlington, VT: Ashgate.

Flin, R., Yule, S., McKenzie, L., Paterson-Brown, S., & Maran, N. (2006). Attitudes to teamwork and safety in the operating theatre. *Surgeon-Journal of the Royal Colleges of Surgeons of Edinburgh and Ireland, 4,* 145–151.

Gawande, A. A., Studdert, D. M., Orav, E. J., Brennan, T. A., & Zinner, M. J. (2003). Risk factors for retained instruments and sponges after surgery. *New England Journal of Medicine, 348,* 229–235. doi:10.1056/NEJMsa021721

Gawande, A. A., Thomas, E. J., Zinner, M. J., & Brennan, T. A. (1999). The incidence and nature of surgical adverse events in Colorado and Utah in 1992. *Surgery, 126,* 66–75. doi:10.1067/msy.1999.98664

Grundgeiger, T., & Sanderson, P. (2009). Interruptions in healthcare: Theoretical views. *International Journal of Medical Informatics, 78,* 293–307. doi:10.1016/j.ijmedinf.2008.10.001

Grundgeiger, T., Sanderson, P., MacDougall, H. G., & Venkatesh, B. (2010). Interruption management in the intensive care unit: Predicting resumption times and assessing distributed support. *Journal of Experimental Psychology-Applied, 16,* 317–334. doi:10.1037/a0021912

Hagemann, N., Schorer, J., Cañal-Bruland, R., Lotz, S., & Strauss, B. (2010). Visual perception in fencing: Do the eye movements of fencers represent their information pickup? *Attention, Perception, and Psychophysics, 72,* 2204–2214.

Hart, S. G., & Steveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload*. Amsterdam, Holland: North-Holland. doi:10.1016/S0166-4115(08)62386-9

Healey, A. N., Sevdalis, N., & Vincent, C. A. (2006). Measuring intra-operative interference from distraction and interruption observed in the operating theatre. *Ergonomics, 49,* 589–604. doi:10.1080/00140130600568899

Horrey, W. J., & Wickens, C. D. (2006). Examining the Impact of Cell Phone Conversations on Driving Using Meta-Analytic Techniques. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 48,* 196–205.

Huestegge, L., Skottke, E. M., Anders, S., Musseler, J., & Debus, G. (2010). The development of hazard perception: Dissociation of visual orientation and hazard processing. *Transportation Research Part F-Traffic Psychology and Behaviour, 13,* 1–8. doi:10.1016/j.trf.2009.09.005

Hyslop, J. W., & Maull, K. I. (1982). Natural history of the retained surgical sponge. *Southern Medical Journal, 75,* 657–660. doi:10.1097/00007611-198206000-00006

Itti, L., & Koch, C. (1999). A comparison of feature combination strategies for saliency-based visual attention systems. In B. E. Rogowitz & T. N. Pappas (Eds.), *Human vision and electronic imaging IV* (Vol. 3644, pp. 473–482).

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40,* 1489–1506. doi:10.1016/S0042-6989(99)00163-7

Itti, L., & Koch, C. (2001a). Computational modelling of visual attention. *Nature Reviews Neuroscience, 2,* 194–203. doi:10.1038/35058500

Itti, L., & Koch, C. (2001b). Feature combination strategies for saliency-based visual attention systems. *Journal of Electronic Imaging, 10,* 161–169. doi:10.1117/1.1333677

Janczyk, M., & Grabowski, J. (2011). The focus of attention in working memory: Evidence from a word updating task. *Memory, 19,* 211–225. doi:10.1080/09658211.2010.546803

Janssen, C. P., & Brumby, D. P. (2010). Strategic adaptation to performance objectives in a dual-task setting. *Cognitive Science, 34,* 1548–1560. doi:10.1111/j.1551-6709.2010.01124.x

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist, 64,* 515–526. doi:10.1037/a0016755

Koh, R. Y. I., Yang, X., Shanqing, Y., Ong, L. T., Donchin, Y., & Park, T. (2009). *Measuring Attention Patterns and Expertise of Scrub Nurses in the Operating Theatre in Relation to Reducing Errors in Surgical Counts.* Paper presented at the Human Factors and Ergonomics Society 53rd Annual Meeting, San Antonio, TX.

Krupinski, E. A. (2010). Current perspectives in medical image perception. *Attention, Perception, and Psychophysics, 72,* 1205–1217. doi:10.3758/APP.72.5.1205

Kundel, H. L. (1989). Perception errors in chest radiography. *Seminars in Respiratory Medicine, 10,* 203–210. doi:10.1055/s-2007-1006173

Kundel, H. L., & La Follette Jr., P. S. (1972). Visual search patterns and experience with radiological images. *Radiology, 103,* 523–528.

Kundel, H. L., Nodine, C. F., Conant, E. F., & Weinstein, S. P. (2007). Holistic component of image perception in mammogram interpretation: Gaze-tracking study. *Radiology, 242,* 396–402. doi:10.1148/radiol.2422051997

Lincourt, A. E., Harrell, A., Cristiano, J., Sechrist, C., Kercher, K., & Heniford, B. T. (2007). Retained foreign bodies after surgery. *Journal of Surgical Research, 138,* 170–174. doi:10.1016/j.jss.2006.08.001

Loukopoulos, L. D., Dismukes, K., & Barshi, I. (2009). *The Multitasking Myth: Handling Complexity in Real-World Operations,* Ashgate.

Macario, A., Morris, D., & Morris, S. (2006). Initial clinical evaluation of a handheld device for detecting retained surgical gauze sponges using radiofrequency identification technology. *Archives of Surgery, 141,* 659–662. doi:10.1001/archsurg.141.7.659

Megaw, E. D., & Richardson, J. (1979). Eye movements and industrial inspection. *Applied Ergonomics, 10,* 145–154. doi:10.1016/0003-6870(79)90138-8

Mishra, A., Catchpole, K., & McCulloch, P. (2009). The Oxford NOTECHS System: Reliability and validity of a tool for measuring teamwork behaviour in the operating theatre. *Quality & Safety in Health Care, 18,* 104–108. doi:10.1136/qshc.2007.024760

Moray, N. (1976). Attention control, control, and sampling behaviour. *Monitoring Behaviour and Supervisory Control,* 221–244.

Mourant, R. R., & Rockwell, T. H. (1972). Strategies of visual search by novice and experienced drivers. *Human Factors, 14,* 325–335.

Niessen, C., Eyferth, K., & Bierwagen, T. (1999). Modelling cognitive processes of experienced air traffic controllers. *Ergonomics, 42,* 1507–1520. doi:10.1080/001401399184857

Nodine, C. F., Kundel, H. L., Lauver, S. C., & Toto, L. C. (1996). Nature of expertise in searching mammograms for breast masses. *Academic radiology, 3,* 1000–1006. doi:10.1016/S1076-6332(96)80032-8

Nodine, C. F., Mello-Thoms, C., Kundel, H. L., & Weinstein, S. P. (2002). Time course of perception and decision making during mammographic interpretation. *American Journal of Roentgenology, 179,* 917–923.

Page, J., Bates, V., Long, G., Dawes, P., & Tipton, M. (2011). Beach lifeguards: Visual search patterns, detection rates and the influence of experience. *Ophthalmic and Physiological Optics, 31,* 216–224. doi: 10.1111/j.1475-1313.2011.00824.x

Pollatsek, A., Fisher, D. L., & Pradhan, A. (2006). Identifying and remedying failures of selective attention in younger drivers. *Current Directions in Psychological Science, 15,* 255–259. doi:10.1111/j.1467-8721.2006.00447.x

Pradhan, A. K., Pollatsek, A., Knodler, M., & Fisher, D. L. (2009). Can younger drivers be trained to scan for information that will reduce their risk in roadway traffic scenarios that are hard to identify as hazardous? *Ergonomics, 52,* 657–673. doi:10.1080/00140130802550232

Raby, M., & Wickens, C. D. (1994). Strategic workload management and decision biases in aviation. *The International Journal of Aviation Psychology, 4,* 211–240. doi:10.1207/s15327108ijap0403_2

Roth, E. M., Christian, C. K., Gustafson, M. L., Sheridan, T. B., Dwyer, K., Gandhi, T. K., . . . Dierks, M. M. (2004). Using field observations as a tool for discovery: Analyzing cognitive and collaborative demands in the operating room. *Cogn Tech Work, 6,* 148–157. doi:10.1007/s10111-004-0156-0

Roth, E. M., & Patterson, E. S. (2005). Using observational study as a tool for discovery: Uncovering cognitive and collaborative demands and adaptive strategies. In H. Montgomery, R. Lipshitz, & B. Brehmer (Eds.), *How professionals make decisions* (pp. 379–393).

Sarter, N. B., Mumaw, R. J., & Wickens, C. D. (2007). Pilots' monitoring strategies and performance on automated flight decks: An empirical study combining behavioral and eye-tracking data. *Human Factors, 49,* 347–357. doi:10.1518/001872007X196685

Savelsbergh, G. J. P., Van der Kamp, J., Williams, A. M., & Ward, P. (2005). Anticipation and visual search behaviour in expert soccer goalkeepers. *Ergonomics, 48,* 1686–1697. doi:10.1080/00140130500101346

Schriver, A. T., Morrow, D. G., Wickens, C. D., & Talleur, D. A. (2008). Expertise differences in attentional strategies related to pilot decision making. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 50,* 864–878. doi:10.1518/001872008X374974

Schulz-Stubner, S., Jungk, A., Kunitz, O., & Rossaint, R. (2002). Analysis of the anesthesiologist's vigilance with an eye-tracking-device. A pilot study for evaluation of the method under the conditions of a modern operating theatre. *Anaesthesist, 51,* 180–186.

Seagull, F. J., Ward, R., Mills, J., Goodrich, C., & Xiao, Y. (2004). Measuring awkwardness of workplace layout: Dispersion of attentional and psychomotor resources within the anesthesia workspace. *Human Factors and Ergonomics Society Annual Meeting Proceedings, 48,* 1755–1758.

Senders, J. (1964). The human operator as a monitor and controller of multidegree of freedom systems. *IEEE Transactions on Human Factors in Electronics, HFE-5*. doi:10.1109/THFE.1964.231647

Senders, J. W. (1966). A re-analysis of pilot eye-movement data. *IEEE Transactions on Human Factors in Electronics, HFE7,* 103–106. doi: 10.1109/THFE.1966.232330

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

Stark, K., & Goldstein, J. (2004). When surgical instruments are left behind - in patients. *Philadelphia Inquiry* (February).

Steelman-Allen, K., McCarley, J. S., & Wickens, C. D. (2011). Modeling the control of attention in visual workspaces. *Human Factors, 53,* 142–153. doi:10.1177/0018720811404026

Taylor, M., & Campbell, C. (2000). The multi-disciplinary team in the operating department. In D. Plowes (Eds.), *Back to basics: Perioperative practice principles,* (pp. 44–48). Harrogate: National Association of Theatre Nurses.

Thomas, E. J., Studdert, D. M., Burstin, H. R., Orav, E. J., Zeena, T., Williams, E. J., . . . Brennan, T. A. (2000). Incidence and types of adverse events and negligent care in Utah and Colorado. *Medical Care, 38,* 261–271. doi:10.1097/00005650-200003000-00003

Thomas, E. J., Studdert, D. M., Newhouse, J. P., Zbar, B. I. W., Howard, K. M., Williams, E. J., & Brennan, T. A. (1999). Costs of medical injuries in Utah and Colorado. *Inquiry-the Journal of Health Care Organization Provision and Financing, 36,* 255–264.

Trafton, J. G. (2007). *Dealing with interruptions.* Santa Monica, CA: Human Factors.

Underwood, G. (2007). Visual attention and the transition from novice to advanced driver. *Ergonomics, 50,* 1235–1249. doi:10.1080/00140130701318707

Underwood, G., Chapman, P., Berger, Z., & Crundall, D. (2003). Driving experience, attentional focusing, and the recall of recently inspected events. *Transportation Research Part F: Traffic Psychology and Behaviour, 6,* 289–304. doi:10.1016/j.trf.2003.09.002

Wickens, C. D., Goh, J., Helleberg, J., Horrey, W. J., & Talleur, D. A. (2003). attentional models of multitask pilot performance using advanced display technology. *Human Factors, 45,* 360–380. doi:10.1518/hfes.45.3.360.27250

Wickens, C. D., & Holland, J. G. (2000). *Engineering Psychology and Human Performance.* Upper Saddle River, NJ: Prentice Hall.

Wickens, C. D., & McCarley, J. S. (2008). *Applied Attention Theory.* Boca Raton, FL: CRC Press.

Wickens, C. D., McCarley, J. S., Alexander, A. L., Thomas, L. C., Ambinder, M., & Zheng, S. (2008). Attention-situation awareness (A-SA) model of pilot error. In D. Foyle & B. Hooey (Eds.), *Human performance models in aviation*. Boca Raton, FL: Taylor & Francis.

Wickens, C. D., & Raby, M. (1991). *Individual differences in strategic flight management and scheduling.* Paper presented at the Proceedings of the Sixth International Symposium on Aviation Psychology, Ohio State University, Columbus, OH.

Wickens, C. D., Rice, S., Keller, D., Hutchins, S., Hughes, J., & Clayton, K. (2009). False alerts in air traffic control conflict alerting system: Is there a "cry wolf" effect? *Human Factors, 51,* 446–462. doi:10.1177/0018720809344720

Xiao, Y., Milgram, P., & Doyle, D. J. (1997). Capturing and modeling planning expertise in anesthesiology: Results of a field study. In C. E. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 197–205).

Yule, S., Flin, R., Paterson-Brown, S., & Maran, N. (2005). NOTSS: A non-technical skills taxonomy for surgeons. *British Journal of Surgery, 92,* 103–103.

Received September 9, 2010
Revision received June 16, 2011
Accepted June 29, 2011 ■