Available online at www.sciencedirect.com

ScienceDirect

Procedia Manufacturing 00 (2015) 000–000

Procedia MANUFACTURING

www.elsevier.com/locate/procedia

6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015

Real Time Validation of Online Situation Awareness Questionnaires in Simulated Approach Air Traffic Control

Jan Kraemer[a], Heinz-Martin Süß[b]

[a]*German Aerospace Center (DLR), Lilienthalplatz 7, Braunschweig 38108, Germany*
[b]*Otto von Guericke University,Universitätsplatz 2 , Magdeburg 39106, Germany*

Abstract

Measuring Situation Awareness to evaluate an operator's ability to handle complex dynamic situations and the use of assistance systems have become a standard approach in Human Factors research. Ideally, the operator's work should be supported by enabling and disabling assistance systems depending on how well they are currently able to keep track of the situation. On the one hand, if the situation's complexity increases and therefore Situation Awareness is likely to be reduced, additional systems may help to overcome the operator's limitations of cognitive resources such as working memory and attention by taking control over the specific task or task components. On the other hand, there has been evidence showing that an overuse of assistance systems, which reduces the task of the operator to mere monitoring, reduces his ability to intervene in a timely manner if needed. Adjusting the usage of assistance systems based on the operator's Situation Awareness may help to overcome both limitations. However, existing measurement tools for Situation Awareness do not allow for an instant validation of the answers given by subjects. Instead, they require post-simulation analysis, e.g. replaying scenarios to compare the subject's answers with the actual state of the situation. This way it is not possible to make decisions based on the operator's Situation Awareness in real time, i.e., parallel with the situation at hand. Software capable of analyzing the current state of a given situation has been developed to overcome this limitation allowing real time assessment of Situation Awareness. This software was designed to measure Situation Awareness of approach air traffic controllers in the real time simulator NLR ATM Research Simulator. Situation Awareness was measured by online and offline probe questionnaires such as the Situation Present Assessment Method which were designed to be validated online as well. Before each item was presented, the simulation's log files were analyzed to provide the software with the correct answer before the item was presented to the operator. This way, the answers given could be validated immediately allowing for an instant evaluation of the operator's Situation Awareness. In a first study, a sample of 57 non-expert subjects was presented with online probe questionnaires in three different real time simulated approach air traffic control scenarios. It was found that the software was able to reliably analyze the answers given by the subject and to compare them with the actual situation. In the future, such systems could be used to make decisions about the need for further assistance while the situation is still happening. This way, operators would only get the necessary amount of assistance without reducing their work to passive monitoring.

Keywords: Situation Awareness, Online Validation, Air Traffic Control, Online Probe Techniques

1. Introduction

1.1 *Motivation*

Situation Awareness (SAw) has become one of the most important and widely used constructs in Human Factors and Ergonomics. The ability “to know what’s going on around you” [1] is considered crucial when operators have to make decisions while handling complex dynamic situations. The more complex and dynamic a situation, the harder is it to build and maintain sufficient SAw. In occupational areas, where decisions made by humans bear responsibility for peoples’ physical integrity or even lives, this is especially crucial.

Measuring of SAw has become one of the most widely used methods of evaluation for newly developed human machine interfaces. Thus, the quality of such systems is also measured by their capability to benefit or improve operator’s SAw. Unfortunately, existing objective measures of SAw found in the literature so far always require post-hoc validation. This makes it impossible to draw valid, objective conclusions about the person’s SAw during acting in the respective situation. Therefore, software was developed to overcome this limitation by measuring and validating SAw in real-time. In this paper, key features of the software are described and results of the first validation study in approach air traffic control (ATC) are reported.

1.2 Defining Situation Awareness

One of the first general definitions of SAw was provided by Endsley [2] as “the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future”. An overview of different SAw models was provided by Salmon et al. [3]. Although differing in the details, most of these theories agree on a three level structure of SAw as proposed by Endsley [1]. Gaining SAw starts by perceiving information in the environment (perception) and then drawing conclusions about the situation’s current state from them (understanding). On the third level, subjects make assumptions about the future state of the situation (projection). Endsley [1] supplied a theoretical framework using a Cartesian approach [4]. It is assumed that SAw is the product of acquired external information, activated long term memory aspects and processes in working memory. Combining these parts, subjects create detailed internal representations of the situation in their minds. The assumption of SAw requiring detailed information stored only in the mind is not undisputed. Change blindness, for instance, contradicts the assumption of detailed and stable internal presentations. Instead, it has been shown that subjects tend to miss even heavy changes to their environment in both pictures and reality [5]. Furthermore, complex situations consist of more than one object. Full attention, however, can only be given to one object at the same time [6].

Chiappe, Strybel and Vu [7] proposed situated SAw as a different approach to SAw. Here, operators do not rely solely on their own cognitive resources to maintain SAw. Instead, information can be stored externally to cope with the cognitive demands. Information-bearers may include a whole network of human and non-human agents. It is not necessary to have all the information stored in the mind to keep track of and perform well in complex situations. Instead, operators create partial representations which are constantly evaluated and adapted over time. Information stored externally is monitored constantly and can be reintegrated into the mind as soon as it becomes necessary to make decisions. Air traffic controllers, for example, use flight strips as external information sources. Information about incoming or outgoing aircraft within a sector are written down on pieces of paper or electronically. As soon as decisions have to made regarding the respective aircraft, the controllers can use the flight strips to reincorporate the information into their minds. The partial representations are assumed to be highly selective and created by interacting with the respective environment. Situated SAw combines the Construction-Integration model [8] with the Relevance Theory [9]. Construction-Integration is used to explain how SAw is derived from complex dynamic environments. Relevance Theory explains why some information is important to an operator and some is not. In short, SAw is built by continuously evaluating, filtering and aggregating external information. Important pieces of information are incorporated into a semantic network, unimportant ones are abandoned. In the end, a situational model is created by integrating the semantic representation with information stored in long term memory. This model can be used to make assumptions about the future development of a situation. Whether information is stored in memory or external sources depends on various aspects of the task (e.g., complexity, ease of access), characteristics of the operator (e.g., degree of expertise), and the work environment (e.g., system interfaces) [7].

1.3 Measuring Situation Awareness

Several measurement instruments for SAw have been provided by various researchers. In general, most instruments can be divided into subjective and objective methods. Subjective methods obtain information about the operator's SAw either by self-report [10] or third person ratings [11]. Self-reports may be useful to learn if humans feel to have a good SAw. They can be used to evaluate if operators are comfortable with the situation they are acting in. However, subjective measurements are unsuited to determine whether people actually have an adequate SAw or just believe to have it.

Probe techniques are the most often used instruments to acquire objective SAw data. Pieces of information are taken from the situation at hand and presented to the operator as questions about the respective information. SAw will be validated by comparing the actual information from the environment with the operator's answers. A higher accordance of actual information and answers given by the operator are considered to indicate a higher SAw. So far, this has to be done by replaying or reviewing the situation afterwards. The conditions under which the questions are presented separate offline from online probe techniques. In offline probe techniques such as the Situation Awareness Global Assessment Technique [2], the situation is paused while questions are given and answered. Additionally, all external information is blanked. This way, people have to answer questions only from internal memory. Item pools for SAw questionnaires are gathered using the Situation Awareness Requirement Analysis [12]. In cooperation with subject matter experts, the major goals of the respective task are identified and further divided into subgoals. For each subgoal the necessary decisions are worked out. In accordance to Endsley's model of decision making [1], the SAw requirements linked to each decision are found out. This way, an encompassing quantity of items is extracted. These items can now be used to create SAw questionnaires. Endsley and Jones [13] supplied an analysis of SAw requirements for Terminal Radar Approach Control tasks.

In online probe techniques, questions are presented parallel to the situation without freezing, and all agents bearing information are accessible as usual. As all information is accessible, solely relying on the correctness of answers would not exhaust all options. Therefore, the reaction time of correct answers is considered as indicator of SAw. It is assumed that questions are answered faster if information is stored internally. If information is stored externally but with consciousness about where it is, answering time will take a little bit longer. Finally, if no information is stored in either memory or external sources, questions wouldn't be answered correctly. So both latter options are considered as indicators of lacking or insufficient SAw.

Probe techniques are advantageous in terms of objectively measuring an operator's SAw. A disadvantage of both techniques, however, lies in their intrusiveness. Answering questionnaires while a situation is handled is a parallel task. In the past, it has been shown that offline SAw questionnaires do not affect performance in the major task significantly [14]. It was argued that this is due to SAw being highly activated throughout the questionnaires. After completing them, the simulation is continued from the exact same point. However, freezing a situation is not applicable to real-life situations. Following Durso et al. [15], insufficient SAw will result in longer response time and is therefore likely to cause more distraction from the task. This should apply to both online and offline probe techniques if the situation continues while answering SAw questionnaires. If real-time assessment of SAw should be used in real-life situations at some point in the future, only online probe methods appear suited. Safety reasons exclude offline probe questionnaires from this purpose.

To make real-time assessment of operator's SAw possible for real-life situations in the future, objective, non-intrusive methods of measurement are necessary. Furthermore, it should be possible to validate SAw immediately to react to insufficient SAw. Probe techniques are a good option to gather objective data about what an operator knows about the situation he is dealing with. Validating the answer given by the operator immediately after a question is presented, would make it possible to assess SAw in real-time. If such an instrument proved to be non-intrusive on top of that, it could very likely qualify as a real-life mean of measuring SAw. In a first approach to create such an instrument, software capable of validating SAw in simulated approach ATC was developed. In this study, it was examined if it could be used to measure and validate SAw of subjects. As good SAw is considered essential to handle complex dynamic situations, higher SAw should result in better performance. Higher levels of SAw are depending on the lower ones. Thus, all three levels of SAw collected by the questionnaires should be correlated positively. Lastly, if the additional measurement of SAw during a scenario is non-intrusive, there should be no performance differences in scenarios where questionnaires are presented compared to scenarios without questionnaires.

2. Methods

2.1. Subjects

A total of 57 student subjects (35 male, 22 female) from the Technical University of Braunschweig (Germany) were invited to take part in the experiment, aged between 18 and 31 years (M = 24.41, SD = 1.38). Participation in the experiment was voluntary. Subject were paid 40 Euros if they completed the study.

2.2. Experimental setup

The experiment was conducted in the Air Traffic Management and Operations Simulator of the Institute of Flight Guidance at German Aerospace Center (DLR) in Braunschweig, Germany. The simulator room was designed to provide silent and undisturbed working stations for ATC and management research. Subjects were tested in groups of up to six people working at separate computer working places. Two 20” monitors each with a resolution of 1440x900 pixels were provided for every PC. Data were input through ordinary computer keyboards and mice. All tasks were run from the experimenter’s computer located next door. A sound-proof glass wall allowed for view on the subject at all times. During experimental runs, the experimenter was present in the subjects’ room.

2.3. Instruments

The main task in this experiment was to control approach traffic in a total of eight real-time simulated scenarios. Scenarios were presented using the NLR Research Simulator (NARSIM) [16], a complex real-time air traffic control simulator which is also used in expert controller studies. Scenarios consisted of aircraft approaching Düsseldorf (EDDL) airport in Germany. Aircraft were coming in from five different entry points which marked the beginning of the standard arrival routes. They automatically followed the routes until they reached the downwind. At that point, subjects had to give the order to intercept on the centerline so the aircraft could reach the airport. Directions were given using a control interface additionally containing information about all aircraft in EDDL control zone. Each aircraft was listed in a separate line with its own command input field. Changes to speed and altitude were made by entering the respective commands and the desired value into the command field of the respective aircraft. All through the approach process, subjects had to control speed and altitude of every aircraft in order to prevent violations of limitations during the process. Additionally, a minimum separation of three nautical miles between aircraft had to be maintained as long as vertical separation was less than 1,000 feet or at all times after passing the final approach fix. A separation tool for measuring distances of two or more aircraft was included in the simulation. Out of the five scenarios, two were used for preparation and training purposes, each lasting ten minutes. Three 30 minutes scenarios were used for the actual performance runs.

To assess SAw of subjects in real-time during simulation runs, software was developed to present SAw questionnaires. The software was designed to read and analyze the NARSIM’s log files. Doing so, it was possible to know the right answer to every question in advance. Immediately following a subject’s answer, the responses could be validated. As reading and analyzing the log files did not interfere with the simulator, i.e., technical delays were not caused by evaluating the completed questionnaires. In accordance to the Situation Present Assessment Method

[15], every questionnaire was announced by a separate warning message containing an accept button. This should prevent questionnaires from being prompted during crucial or stressful situations. Immediately after accepting the incoming request, the software randomly chose an item to be presented. After gathering the correct answer, the item was presented on the screen to be answered by the operator. All questionnaires could be presented as online or offline probes. In online mode questionnaires, access to all sources of information was given at any time while responding to an item. In offline probes, blank panels blocked the view on all interfaces on all displays. Thus, it was impossible for subjects to retrieve information from external sources during questionnaires. The questionnaires were presented in separate windows in order not to block the view on other interfaces during presentation. As recommended, questionnaires began six minutes into the scenario with items separated by one minute [12]. Only one item was presented per questionnaire.

Items used in the SAw questionnaires were obtained from a SAw requirement analysis for Terminal Radar Approach Control procedures [13] and reduced to items which appeared suitable for non-expert subjects and the simulation's capabilities, resulting in a total of 14 items.

This processing offered two big advantages. First, rerunning simulations in order to compare answers from SAw questionnaires to the actual situations became obsolete because the answer was already known to the software. Second, right after answering an item, SAw of the operator could be validated immediately. In the future, this might allow reacting to insufficient operator's SAw while the situation is still present. This might prove especially useful for validating assistance systems or automation research in general.

2.4 Procedure

One week before the actual experiment began, subjects were sent a link to an online presentation. It served to provide basic information on the process of approach ATC and the simulation. It included information on the traffic management area of the airport Düsseldorf (EDDL), the simulation environment and command options. Subjects were allowed to repeatedly complete the presentation to prepare for the experiment. Completing the presentation was not mandatory to participate in the experiment.

The actual experiment was completed in one session per group, lasting about two hours with several breaks between the five scenarios. First, an introduction was read to the participants. It contained information about the experiment in general. After that participants completed biographical questionnaires.

Following the biographical questionnaires, subjects were given written instructions of the simulator task. Instructions included a schematic map of EDDL airspace, a screenshot of the control interface and a summary of commands used during the simulation. Prior to the actual scenarios, subjects completed two training scenarios. In the first training, the examiner introduced the participants to the simulation environment, explaining the interfaces and controls. A second training scenario was completed afterwards by subjects to further accustom them to the task. After training, the first of the 30 minute scenarios was started. The first scenario always served as the baseline scenario. No SAw questionnaires were presented during this run. During the second and third scenario, SAw questionnaires were presented beneath the control interface. Online and offline probes were used for one scenario each. Scenarios and order of presentation modes were randomized between groups of subjects.

3. Results

ATC performance was calculated as the total duration of all conflicts and speed/height violations occurring throughout a scenario in relation to total scenario time. Performance was calculated for each scenario, separated by experimental condition (baseline, online probes and offline probes). Descriptive statistics of ATC performance are presented in Table 1.

Table 1. Descriptive statistics of subjects' Air Traffic Control performance in each scenario. In baseline condition, no Situation Awareness questionnaires were presented. Performance was calculated as proportion of total duration of violations of all limitations in relation to total scenario time.

Scenario	**M**	**SD**	**Min.**	**Max.**
Baseline	.22	.04	.17	.39
Online Probes	.20	.04	.17	.39
Offline Probes	.21	.03	.17	.35

SAw questionnaires were analyzed by response time of correctly answered items in online probe condition. For offline probes, accuracy was used as index of performance. Total SAw for each of the three levels was calculated as mean response time in online probes. Offline probe scores were calculated as sum of correct answers for each level. No distinction was made between different ATC scenarios performed during presentation. Descriptive statistics are presented in Table 2. Spearman Rho correlations of mean response time in online scores showed significant positive correlations between all three levels of SAw (see Table 3). For offline probes, correlations between any levels were not significant at all.

Table 2. Descriptive statistics of subjects' Situation Awareness. Offline probe statistics are given as mean numbers of correct answers. Mean response time of correct answers (ms) was used in online probe questionnaires.

Mode	**Level of SAw**	**M**	**SD**	**Min.**	**Max.**
Offline	Perception	1.62	1.92	0.00	3.54
	Understanding	2.62	1.00	0.00	4.00
	Projection	2.40	1.09	1.00	5.00
Online	Perception	9,674.98	3,477.31	3,998.00	22,090.00
	Understanding	6,647.85	2,380.58	2,523.00	17,342.00
	Projection	7,737.41	3,211.91	2,790.33	26,053.00

Several relationships between ATC performance and SAw were extracted from the data. ATC performance was significantly correlated with mean response time of online probes in level one (ρ = -.40, p = .001) and two (ρ = -.33, p = .008) but not in level three (ρ = -.08, p = .277). No significant correlations between ATC performance and offline probe measures were found on any level.

Intrusiveness of SAw questionnaires was examined by comparing performance between experimental conditions (see Figure 1). A significant main effect of presentation mode (baseline, online, offline) on ATC performance was found ($F(2, 110) = 8.20$, $p < .001$, $\eta^2 = .13$). Bonferroni-corrected post-hoc t-tests showed significantly better performance for scenarios in which offline probes were presented. This was found compared to both baseline condition ($p < .001$) and online probe condition ($p < .001$). No difference in performance was found between baseline and online probe condition ($p = .929$).

Differences between subjects who completed the preparation survey before the experiment and those who did not were examined. Neither ATC performance, accuracy of offline probes or response times of online probes showed any significant differences between both groups.

Table 3. Spearman Rho correlations between levels (Lv) of Situation Awareness for online (On) and offline (Off) probes. Correlations marked with '***' indicate levels of significance below .001.

		1	2	3	4	5
1	On Lv 1					
2	On Lv 2	.48***				
3	On Lv 3	.50***	.57***			
4	Off Lv 1	-.07	-.03	-.08		
5	Off Lv 2	-.06	-.14	-.03	.14	
6	Off Lv 3	-.16	.10	-.10	.06	.22

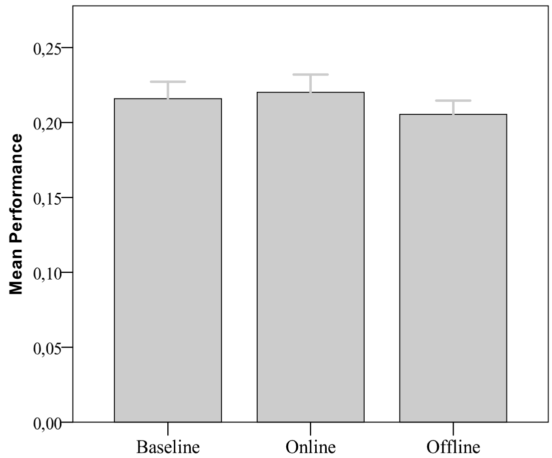

Figure 1. Mean ATC performance over experimental conditions (Baseline, online probes, offline probes). Whiskers equal 1.96 standard errors of mean.

4. Discussion

The presented results indicate that the introduced software is capable to measure SAw during simulation runs in the NARSIM. It was shown that all levels of SAw measured by the instrument are intercorrelated significantly in online probe conditions. Furthermore, significant relationships between ATC performance and SAw questionnaires were found. It can be concluded that using the software it is possible to validate SAw in real time during a simulation run. This way, it is possible in future studies to assess and evaluate operators' SAw while they are completing different scenarios. No replays of the scenarios are necessary to do so.

Data has shown that the presentation of SAw questionnaires during simulation runs do not affect ATC performance negatively. Although this speaks for the usability of the instrument, it should be noted that operators still have to interrupt their major task to respond. Therefore, cognitive resources have to split up between both tasks. This problem was inevitable because questions were presented visually. To reduce this effect, it is possible to present questionnaires acoustically [15]. In this study, this was impracticable as subjects were tested in groups. Even if questionnaires were to be presented via headphones, answering items would likely have caused distraction among subjects. However, it is not possible to solve this problem completely in probe techniques. Hence, SAw questionnaires still cause distraction. However, the goal of this study was to validate the introduced software in terms of measuring SAw. Adapting the principle to make it more suitable for actual application in real-life situations is a topic for future studies.

One might wonder why no significant correlations of level three SAw questionnaires and ATC performance were found in the data. This is likely because the software is still imperfect to analyze future states of the simulation from the present state. Simplified algorithms were used to calculate future positions of aircraft while the actual physical models used in the NARSIM are much more complex. In the future, this shall be fixed by making use of trajectory predictors. These tools are capable of predicting the future positions of aircraft over a period of time. They are much more accurate because they make use of the actual NARSIM models. Unfortunately, they were not available by the time this experiment was done but will be so in upcoming studies.

Offline probe questionnaires showed poor results compared to the online probes. Very likely, this is to blame on the lack of ATC experience of subjects. In offline probes subjects could not make use of any external information and had to rely solely on internal memory. While experts make use of mental models and cognitive schemata to store larger amounts of information, subjects are put under high cognitive demands here. If expert air traffic controllers complete the offline probes, significant correlations between levels of SAw and ATC performance are to be expected.

Endsley [12] described how SAw questionnaires can be adapted to the current situation. For example, questions about separation loss between aircraft could be left out in questionnaires if none are present. This indicates another approach to SAw questionnaires. Using the software presented in this paper, analyzing the situation could be used to adapt the point in time at which questionnaires are presented. Instead of solely presenting them at specified points in time, present or upcoming conflicts could trigger questionnaires. This way, it would be possible to find out if subjects already registered an emergency. Depending on the answer, a warning message could be implemented to point out unrecognized emergencies to the operator.

Being able to objectively evaluate an operator's SAw while he's handling a situation might be crucial for future progress in human-in-the-loop automation research. Different effects of automation on performance have been reported in the literature [17]. However, mental overload is likely to occur if a situation's complexity exceeds the operator's capabilities. To prevent this, assistance systems were designed to support the operator. Equally, mental underload may occur if operators do nothing but monitoring the situation without intervening actively. This can happen if a system is completely automated and operators are only required to act if an error or emergency causes the system to fail. Evaluating the operator's SAw during the task could help to find a balance between both extremes. Depending on the current SAw of the operator, it could be decided whether assistance systems should be turned on or off. Previously, without an instrument capable of analyzing SAw on the run, this was not possible.

References

[1] M.R. Endsley, in: M.R. Endsley, D.J. Garland (Eds.), Situation Awareness Analysis and Measurement, Lawrence Erlbaum Associates, Mahwah, NJ, 2000a, pp. 3 – 32.
[2] M.R. Endsley, Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 1988, pp. 97–101.
[3] P.M. Salmon, N.A. Stanton, G.H. Walker, C. Baber, D.P. Jenkins, R. McMaster, M.S. Young, Theoretical Issues in Ergonomics Science 9 (2008) 297–323.
[4] S.W.A. Dekker, D.H. Hummerdal, K. Smith, Theoretical Issues in Ergonomics Science 11 (2010) 131–135.
[5] D. Simons, D.T. Levin, Psychonomic Bulletin and Review 5 (1998) 644- 649.
[6] R. Rensink, Visual Cognition 7 (2000) 17-42.
[7] D.L. Chiappe, T.Z. Strybel, K.-P.L. Vu, Theoretical Issues in Ergonomics Science 13 (2012) 625-647.
[8] F.T. Durso, K.A. Rawson, S. Girotto, in: Handbook of Applied Cognition, Wiley, Hoboken, NJ, pp. 163-193.
[9] D. Sperber, D. Wilson, Relevance: Communication and Cognition, second ed., Blackwell, Oxford, 1995.
[10] R.M. Taylor, Advisory Group for Aerospace Research & Development Conference Proceedings, 1990, pp. 3.1-3.17.
[11] W.L. Waag, M.R. Houck, Aviation, Space, and Environmental Medicine 65 (1994) A13-A19.
[12] M.R. Endsley, in: M.R. Endsley, D.J. Garland (Eds.), Situation Awareness Analysis and Measurement, Lawrence Erlbaum Associates, Mahwah, NJ, 2000b, pp. 147 – 173.
[13] M.R. Endsley, D.G. Jones, Situation Awareness Requirement Analysis for TRACON Air Traffic Control (TTU-IE-95-01), Texas Tech University, Lubbock, TX, 1995.
[14] M.R. Endsley, Human Factors 37 (1995) 65-84.
[15] F.T. Durso, M.K. Bleckley, A.R. Dattel, Human Factors 48 (2006) 721-733.
[16] J.M. ten Have, Simul. Pr. Theory 1 (1993) 31-39.
[17] D.B. Kaber, M.R. Endsley, Theoretical Issues in Ergonomics Science 5 (2003) 1-40.