Available online at www.sciencedirect.com
SCIENCE DIRECT®
International Journal of Industrial Ergonomics 36 (2006) 463–474
International Journal of INDUSTRIAL ERGONOMICS
www.elsevier.com/locate/ergon

Situation awareness as judgment I: Statistical modeling and quantitative measurement

Alex Kirlik[a,*], Richard Strauss[b]

[a]*Human Factors Division, Beckman Institute, University of Illinois at Urbana-Champaign, Illinois USA*
[b]*FatWire Software, New York, NY, USA*

Available online 14 March 2006

Abstract

Situation awareness (SA) is a commonly used term to describe the challenges faced by human operators when interacting with controlled systems or work environments through technological interfaces. In this paper we conceptualize SA as judgment under uncertainty. As such, SA is conceived as the degree of correspondence between a set of human judgments and the distribution of true system or environmental states or events being judged. Statistical modeling and estimation techniques known as judgment analysis are used to decompose SA into seven individually measurable components for engineering analysis and design. We discuss how the model and measures complement existing qualitative models of SA (e.g., Endsley), human–automation interaction (e.g., Parasuraman, Sheridan & Wickens) and naturalistic decision making (NDM) (e.g., Klein). A companion article presents a demonstration of the judgment analysis approach to SA modeling and measurement.

Relevance to industry

Good situation awareness implies a high correlation between actual and judged system states. This paper provides a technique for decomposing this correlation into seven independent factors so that technology and training interventions target the specific barriers to situation awareness in any particular context or system.

Keywords: Situation awareness; Operator modeling; Judgment; Performance measurement

1. Introduction

Information technology and automation in industrial systems increasingly mediate the interaction between a human operator (or team) and a controlled system or work environment. The need for operators to maintain situation awareness (SA) in these contexts is frequently cited as a key to effective and efficient performance and the reduction of error (e.g., Adams et al., 1995; Durso and Gronlund, 1999; Endsley and Garland, 2001). Supporting human operators in intensively mediated situations through interface design and training promises to be a growing challenge for industrial ergonomics and human factors. Although there are many dimensions to SA, and its functional role depends on the specifics of the environment and task, few would disagree that continued advances into understanding and supporting SA are sure to depend on advances in measurement. Salas et al. (1995) summed up the situation well: "a central problem in understanding situation awareness is the lack of well-developed measurement tools" (p. 131).

In human factors, one of the most influential perspectives on SA has been put forth by Endsley, who has studied the phenomenon within the contexts of automation (Endsley, 1996), air traffic control (Endsley and Smolensky, 1998), and naturalistic decision making (NDM) (Endsley, 1997), among others. Informally, Endsley notes that SA concerns "knowing what is going on" (Endsley, 1995b, p. 36). More precisely, SA is defined as "the

*Corresponding author. Tel.: +1 217 244 8972.
E-mail address: kirlik@uiuc.edu (A. Kirlik).

0169-8141/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.ergon.2006.01.009

perception of the elements of the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future'' (Endsley, 1995b, p. 36).

Endsley's definition describes SA as a three-level construct involving elements of perception, comprehension, and projection. In this model, Endsley describes Level 1 SA as knowledge that results from the perception of disjointed ''elements'' in the environment, in which an element is equated with environmental objects and attributes. Level 2 SA, which is typified as a mental picture or model, is described as comprehension of the situation, or an understanding of its current significance. Level 3 SA is described as the ability to predict future states of Level 1 elements.

Other human factors researchers have also presented qualitative, cognitive models SA, defining it as a mental picture or an internal product. For example, in the context of land navigation, Wesler et al. (1998) equated SA to the contents of short-term memory. In aviation, Gibson and Garrett (1990) described SA as a Gestalt-like appreciation of a situation, and Taylor (1990) described SA as a ''veridical model of reality'' (p. 3-1).

1.1. Measuring situation awareness

Whatever cognitive processes account for SA, by necessity its measurement must have its basis in observable variables. One category of existing measurement techniques relies on largely verbal, qualitative data gathered through case-study examinations of critical incidents (e.g., Klein, 2000). These methods rely on subjective reports of retrospective memory for actual, past incidents and process-tracing (think out-loud) protocols for an examination of simulated or on-going incidents. Information is gathered on the environmental cues attended, active operator goals and expectations, and critical decision points. One primary focus is to identify any differences between novice and expert performers in how SA is achieved.

A second category of SA measurement techniques uses subjectively reported measures of SA. One well-known example is the Situation Awareness Rating Technique or SART (Taylor, 1990), although Pew (2000) has noted that the SART technique may confound SA measurement with workload measurement.

Perhaps the most popular and widely used SA measurement technique relies on probes or questions embedded within experimental or simulation studies in real-time, interactive contexts. Here, a task or dynamic situation is frozen at various points and operators are presented queries about the state of the controlled system or task environment. The best example of this approach is the Situation Awareness Global Assessment Technique or SAGAT (Endsley, 2000). SAGAT is used to design and administer queries pertaining to all three levels of SA described in Endsley's (1995b) model. The accuracy of these queries is then measured (percent correct) and aggregated to result in an overall or global metric of SA, and one that is considered to be direct: ''This type of measure is a direct measure of SA—it taps into the operator's perceptions rather than infers them from behaviors that may be influenced by many other factors besides SA'' (Endsley, 2000, p. 147). Endsley (2000) presents a comprehensive overview of the SAGAT technique including discussions of sensitivity, reliability, validity, and implementation.

2. SA measurement for systems engineering

Our goal in this article is to build upon previous SA research such as that described above to present a technique for modeling and measuring SA that is useful for engineering technological systems. The availability of techniques such as SAGAT, which can be used to measure the level of SA demonstrated by a performer given a particular system design are important; we too provide a technique grounded in the correspondence between an operator's perceptions or judgments and true system or environmental states. We suggest, however, that the realities of engineering operational human–machine systems require that an additional set of issues be addressed beyond those addressed by SAGAT, SART, or subjective report techniques. In the following, we present four such issues that influenced the development of our own SA modeling and measurement technique.

2.1. A systems approach

The levels of both actual and theoretically attainable SA in a technological system are only partially determined by the skill, ability, expertise, or knowledge of the human operator. Other contributions are made by the sensor, sensor integration, and display technologies responsible for providing the operator with a window to the controlled system or task environment. A more comprehensive approach to SA modeling and measurement requires techniques capable of representing and decomposing both the technological and psychological contributions to SA. Doing so requires a conceptualization of SA that spans the human-environment boundary, depicting how both technological and psychological processes may mediate the level of SA achieved. Modeling SA as distributed across this boundary is important in an engineering sense because only techniques capable of representing the external contributors to SA are capable of analyzing and predicting how technology design influences SA. These influences can either be positive (the availability of enhanced sensors—see Endsley, 2000) or negative (display design flaws associated with misleading or missing information—see Sarter and Woods, 1995).

As such, the technique we present has a systems, or ecological orientation, in that the intent is to capture *both* the cognitive *and* environmental (external to the performer) determinants of SA. While the majority of SA research has focused on the former, we agree with Pew (1995), who noted that "In order to adequately define SA we need to understand what we mean by a 'situation' and we need to know what it is about situations of which we must be aware" (p. 7). Other researchers have promoted the need for more systems-oriented, ecological definitions of SA. For example, Flach (1995) states that "SA defines the problem of human performance in terms of understanding the adaptive coupling between human and environment" (p. 153). He introduces the notion of *correspondence* to emphasize that "the human's awareness must correspond to the objective constraints of the situation" (p. 151). Smith and Hancock (1995) also promote an ecological focus, relying on Neisser's (1976) perceptual cycle to shape their definition. In their proposal, Smith and Hancock focus on elaborating the environmental components of SA. They do so using Gibson's (1979) "invariant" concept, to denote the meaningful environmental features or constraints to which a person must adapt in order to achieve SA. Sarter and Woods (1995) have provided an excellent case-in-point regarding the need to represent the technological contributors to SA, in their study of mode confusion and mode awareness in a modern cockpit.

As discussed previously, Pew (1995) also points toward a systems or ecological perspective on SA. Pew introduces two definitions, *ideal* and *obtainable awareness*. He defines *ideal awareness* as the awareness possible after all known information and knowledge requirements are satisfied. He then defines *obtainable awareness* as the level of *ideal awareness* possible after accounting for limitations in both knowledge and information. The possibility of such limitations, in either the external environment or in human cognition, suggests that SA modeling and techniques recognizing uncertainty would be beneficial.

2.2. Recognizing uncertainty

The existence of uncertainty is another reality of most operational human–machine systems. Just like we would like to be able to independently measure the technology design contributions to SA, we would also like to measure the contributions of any environmental uncertainty to SA. Say, for example, that a system evaluation yielded an average of 80% correct performance on probe questions used in SAGAT-type SA measurement. Could operators have achieved 100% had they been more experienced? Or might 80% be peak SA achievement given the existence of uncertainty in the relationship between displayed information and the actual environmental or system states or events represented by the display? One answer would direct remedial efforts toward the design of additional sensors or better displays. The other towards enhanced training or selection. As such, the SA modeling and measurement technique presented below makes explicit provisions for potential uncertainty mediating the relationships between the true environmental states or events of which the operator should be aware, the information displayed about those states or events, and the operator's resulting judgments of those states or events.

One important implication of the need to countenance the presence of uncertainty pertains to the unit of analysis used to model and measure SA achievement. In a less than perfectly predictable environment, the accuracy of individual human judgments often paints a misleading picture of human performance. As in weather forecasting, one can make the "correct" judgment (predicting rain) yet be found to be "in error" (due to chance), and one can make an "incorrect" judgment yet be found to be accurate (again, due to chance). As such, as many researchers studying judgment under uncertainty have realized (see Hammond and Stewart, 2001), evaluating human behavior in uncertain tasks requires measurement techniques that capture the degree of adaptation or calibration between an entire *set* of human judgments and a *distribution* describing the judgment criterion. This type of distributional, rather than instance- or case-based, analysis forms the basis of the SA modeling and measurement technique presented in the following.

2.3. A diagnostic approach

While measurement techniques capable of assigning a numerical value to SA in any given context are useful for system evaluation, remedial actions such as design or training typically require that one be able to determine why any particular level of SA was observed. While it is impossible to identify a complete set of factors that might lead to less than perfect SA, in the following we identify a restricted set of meaningful factors and a technique for statistical modeling and decomposition that allow these separate factors to be individually measured. For example, the technique described in the following is capable of distinguishing among factors such as the following as potential causes for less than perfect situation awareness: imperfect operator knowledge; imperfect operator perception of displayed cues; inconsistency in operator information processing, and inherent environmental uncertainty.

2.4. A functional approach

The SA modeling and measurement technique we present in the following has a pronounced lack of conceptual richness from a cognitive perspective. We do not rely on many constructs that are frequently used by other SA theorists: comprehension, attention, memory, projection, schemas, and mental models to name just a few. We focus instead on describing and then decomposing the functional, statistical relationship between directly observable variables. In particular, we view all instances of situation awareness to be one of a species of judgment under uncertainty, and use a combination of available judgment analysis (Cooksey, 1996) techniques as the basis for modeling and measurement.

For example, we have described Endsley's (1995b) three-level model of SA and a description of her SAGAT (Endsley, 2000) approach to SA measurement. The SAGAT approach is based on asking an operator to make a series of judgments, either about primitive perceptual elements (Level 1 SA), comprehension of meaning or relevance (Level 2 SA), or of future states (Level 3 SA). Despite the difference in *content*, we note that all three of these instances are judgments, and as such, are amenable to treatment from the perspective of judgment analysis (Cooksey, 1996), which forms the basis for the technique to be presented in the following. As such, one will see that our own modeling and measurement technique and Endsley's model and SAGAT measurement technique are complementary. More specifically, we suggest that SAGAT or a similar approach could be used to select and administer the set of operator queries (judgments) to be analyzed in SA measurement, and that the techniques presented here could be used to analyze the resulting data to potentially gain diagnostic insights into SA beyond what could be learned from the percent of queries answered correctly and incorrectly.

The SA-related judgments amenable to our technique can be of either past, present, or future situations or events, and are assumed to be based on the use of local, or proximal, information sources (such as those presented on an interface) to infer the existence of a situation, state or event present in a system or environment represented by that interface. This is exactly the role of the human operator or monitor in a large and growing class of systems of interest to industrial ergonomics.

2.5. Summary and overview

The modeling and measurement techniques we present in the following, while somewhat new to industrial ergonomics and human factors, and certainly new to the study of SA, are based on a long history of research and modeling in psychology, and more recent techniques borrowed from the weather forecasting literature. We construe SA as the ability to render accurate judgments about the external environment or controlled system, as a complement to existing approaches that characterize SA in other cognitive terms. Our approach results from the cumulative research of Brunswik (1956), who provided his lens model of judgment, the research of Hursch et al. (1964) and Tucker (1964), who provided a mathematical formulation for Brunswik's model, the research of Murphy (1988), who developed a diagnostic measure of weather forecasting skill, the research of Stewart (1990), who joined the lens model equation and Murphy's skill measure, and finally, the research of Stewart and Lusk (1994), who supplemented the previous research by providing resources for describing the contribution of sensing and information processing technology for modeling interface-mediated judgment.

In this article, we present these systems-oriented, ecological techniques in the hopes they may provide an advance in the measurement and support of situation awareness. Importantly, we also provide information on how these methods connect to, and complement, related theories of SA (e.g., Endsley, 1995a, b), NDM (e.g., Klein, 1999), and human–automation interaction, or HAI (Parasuraman et al., 2000).

3. SA modeling and measurement: a systems approach

Brunswik's (1956) ecological perspective on judgment under uncertainty is the initial basis for a model of the judgmental components of situation awareness (for a discussion of the relationship between Brunswikian theory and human factors, see Kirlik, 2000). Consider Fig. 1, which depicts the basic components of both Brunswik's theory of judgment and interface-mediated situation awareness.

The left side of Fig. 1 depicts the situation, or what Brunswik referred to as the *environment*, and the right side of the figure depicts awareness, or what Brunswik referred to as the *organism* (we shall use "operator"). Mediating the situation–awareness relationship, or generally the environment–organism relationship, are both information and the uncertainty that accompanies it (i.e., the degree to which the information is capable of adequately specifying the state of the remote, or distal, situation or environment). Brunswik (1956) originally proposed measuring the quality of judgment in terms of a correspondence between the judged situation and the actual, environmental situation, measured by linear association, or bivariate correlation. Thus, the measure of achievement is obtained by correlating the operator's judgments with the true states of the situation being judged. It is important to note that achievement is measured as the correlation between two observables; the concept of an "internal model" or "internal representation" plays no role in this conceptualization. Denoting this correlation coefficient r_{YO}, where the subscripts represent the contribution of the operator's

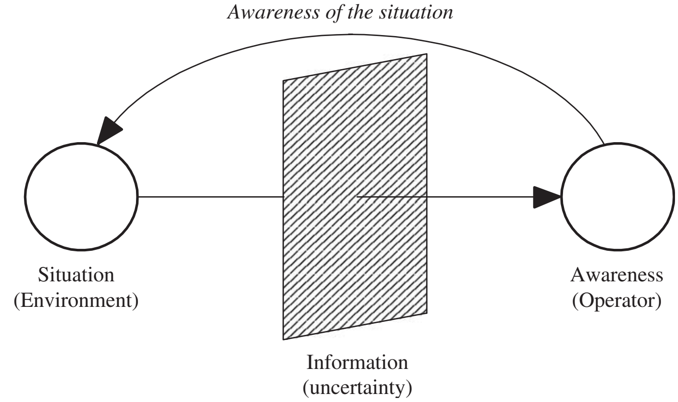

Fig. 1. A systems or ecological framework for modeling judgment or mediated SA.

judgment (Y) and the situation or "Object" of judgment (O), the higher the correlation ($0 \leqslant r_{YO} \leqslant 1$), the better the correspondence, the better the operator's achievement, and the better the SA.

3.1. Limitations of the correlation coefficient

Although the correlation coefficient provides a useful measure of correspondence, it has properties that limit its sensitivity, and thus utility, for measuring judgment quality (Cronbach and Gleser, 1953). Correlation captures only *shape* differences between two sets of variables, i.e., their shared pattern of ups and downs, without distinguishing the differences in either their *magnitude* or *scale*. A depiction of the correlation coefficient's insensitivity to both magnitude and scale is shown in the graphs in Fig. 2. Note that the correlation is 1.0 in all four graphs.

Graphs (a) and (b) show a difference in magnitude not captured by correlation. Graph (a) depicts a relationship between judgments (e.g., predicted distance of an approaching aircraft) and the true state of the situation (e.g., true distance of an approaching aircraft). In this case, average judgment is 11 (miles). Graph (b) shows another set of judgments for the same situation, but with an average judgment of 80 (miles). This inability of correlation to distinguish between these cases is evidence of the insensitivity of correlation to differences in magnitude. Graphs (c) and (d) in Fig. 2 show a difference in scale not captured by linear correlation. In graph (c), judgments of a situation have a standard deviation of 5.9 (miles), while in graph (d), the standard deviation is 17.7 (miles), indicating the insensitivity of correlation to differences in scale.

3.2. Absolute distance measures of judgment quality

These deficiencies of the correlation coefficient have motivated several researchers to look for more sensitive measurements of judgment correspondence. One alternative has been to look at the distance between data sets rather than their shared shape, a strategy often found in studies of meteorological forecasting (e.g., Murphy, 1988). Mean Square Error (MSE), a measure of the squared Euclidean distance between two data sets (Cooksey, 1996), has been regularly adopted for this purpose (e.g., Lee and Yates, 1992; Stewart and Lusk, 1994). MSE defines distance using the following equation:

$$\mathrm{MSE}_Y = \left(\frac{1}{n}\right) \sum (Y_i - O_i)^2. \quad (1)$$

Here, the two data sets are the judgments of the operator and the corresponding true states of the situation being judged. These two sets are used to form n pairs, where one

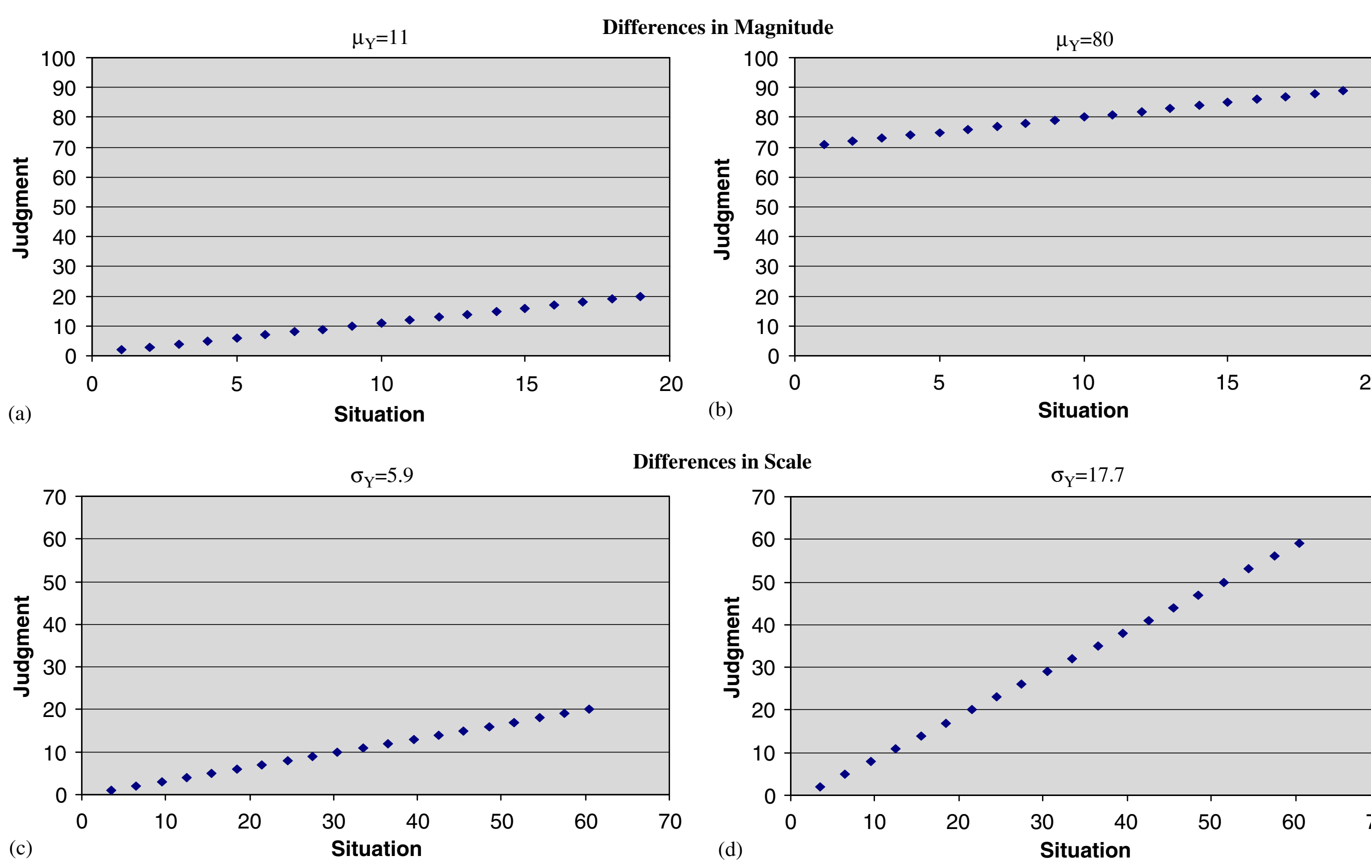

Fig. 2. Insensitivity of the correlation coefficient to magnitude and scale.

element of the pair comes from each set (Y_i and O_i denote the ith judgment and ith true state, respectively). When the judgments are perfect, MSE is equal to zero. As a replacement for the correlation coefficient, MSE would be unremarkable except that it can be partitioned into three distinct components representing shape, scale, and magnitude. Here, we present the decomposition proposed by Murphy (1988), introducing first, however, his skill score (SS) measure of judgment quality.

3.3. The skill score as a measurement of judgment quality

To develop his decomposition of MSE, Murphy (1988) used the concept of *skill*, which he defined as judgment performance above chance. Chance performance is defined to be the degree of correspondence that would be obtained had a person always provided the same (constant) judgment based on the average, base-rate value of the situations being judged. In Eq. (2), the quality of this standard is the MSE_R that would be expected if the standard were always used:

$$\mathrm{MSE}_R = \left(\frac{1}{n}\right)\sum(\mathrm{O}_i - \bar{\mathrm{O}})^2. \tag{2}$$

Here $\bar{\mathrm{O}}$ is the mean, or base rate, of the observed event being judged. Deriving the skill score requires measuring the ratio between the MSE of the operator's judgment (Eq. (1)) and the MSE of the standard (Eq. (2)). This ratio is then subtracted from unity to create the skill score (SS). This relationship is shown in Eq. (3). In this basic form, the skill score provides overall evaluation of the quality of an operator's judgments as compared to chance. When SS is

$$\mathrm{SS} = 1 - \left[\mathrm{MSE}_Y/\mathrm{MSE}_R\right] \tag{3}$$

positive, the operator's judgments are better than chance ($\mathrm{MSE}_Y < \mathrm{MSE}_R$); when it is zero, the judgments of the operator are equal to chance performance ($\mathrm{MSE}_R = \mathrm{MSE}_Y$); and when SS is negative, the operator's judgments are worse than chance ($\mathrm{MSE}_Y > \mathrm{MSE}_R$).

3.4. Murphy's decomposition of the skill score

Murphy (1988) developed the SS to enable the MSE to be decomposed. By substituting the equations for MSE_Y (Eq. (1)) and MSE_R (Eq. (2)) into the form of the skill score (Eq. (3)), Murphy (1988) showed how to derive the desired decomposition. A conceptual representation of his decomposition is presented below:

$$\text{Judgment Quality (SS)} = [\text{Shape (correlation)} - \text{Scale Error} - \text{Magnitude Error}]. \tag{4}$$

Here, the SS is partitioned into three components, and thus shape (correlation) is separated from errors associated with differences in magnitude and scale. The result is a more sensitive and diagnostic measurement than correlation alone. The scale error component has been called *Regression Bias*, as it measures whether the operator has appropriately scaled judgmental variability to situational variability given the level of achievement (r_{YO}) attained. It is zero when the slope of the regression line predicting the observed events from the operator's judgments is 1.0 (Stewart and Lusk, 1994). For example, a submarine sonar technician with a regression bias might on average judge the range of an approaching enemy to be between 20 and 100 nm, when the actual interval of ranges is between 50 and 80 nm. A regression bias is a tendency to produce judgments in either a smaller or larger range than in the actual situation than is warranted by the attained level of achievement (r_{YO}).

Finally, the magnitude error component of Murphy's Eq. (4) has been called *Base Rate Bias* (Stewart, 1990). It measures the overall (unconditional) bias in the operator's judgments, thus diagnosing a tendency to over- or underestimate the judged situation. This bias equals zero only when the mean of the operator's judgments equals the mean of the judged states (i.e., the objective base rate, and is non-zero when the mean operator's judgment is too high or low).

Murphy's decomposition of the skill score provides a sensitive measure of the judgmental components of situation awareness, as it disentangles the joint contributions of shape, scale, and magnitude in the measurement of judgment quality. The mathematical decomposition is presented below in Eq. (5), and its components are summarized in Table 1. It is important to note that the decomposition shown in Table 1 is meant to capture regularities in the relationship between judgments and a

Table 1
Components of Murphy's decomposition of the skill score

Component	Name	Description
SS	Skill score	A relative measure of "actual" judgment quality
r_{YO}	Correlation coefficient	*Shape*—degree of linear association between judgments and situation. "Potential" skill in judgment
$\left[r_{YO} - (S_Y/S_O)\right]^2$	Conditional/regression bias	*Scale*—degree that standard deviation of judgments accounts for imperfect correlation; for the bias to vanish, S_Y must be adjusted to equal $r_{YO}(S_O)$
$\left[((\bar{Y} - \bar{O})/S_O)\right]^2$	Unconditional/base rate bias	*Magnitude*—degree that average judgment equals the base rate of occurrence in the situation

criterion, and thus describes the *statistical* criteria that must be achieved for a high level of situation awareness (SA). These criteria do not map into Endsley's (1995b) "level-based" decomposition of SA, because the statistical decomposition applies to *each* of Endsley's levels (depending on whether the judgments being analyzed are level 1 perceptual judgments, level 2 "meaning comprehension" judgments, or level 3 projection judgments). Our modeling and measurement technique is amenable to analyzing SA achievement at any of these levels:

$$\mathrm{SS} = (r_{YO})^2 - \left[r_{YO} - \left(S_Y/S_O\right)\right]^2 - \left[\left((\bar{Y} - \bar{O})/\mathrm{S}_O\right)\right]^2. \quad (5)$$

3.5. Augmenting the skill score with the lens model equation

The decomposition of the skill score, as shown in Eq. (5) and Table 1, can improve the diagnosticity of the judgmental components of SA. Additional diagnosticity can be gained by taking this decomposition one step further by decomposing the correlation coefficient (r_{YO}). To do so, consider a more detailed depiction of Brunswik's lens model as shown in Fig. 3.

The lens model shares the same configuration as Fig. 1, yet depicts information as an enumerable set of cues, or items of information (labeled X_i's). Recall that Brunswik measured achievement (the top arc in Fig. 3) as a linear correlation, in our case, r_{YO}. Each relationship between a cue and a situation can be assigned an *ecological validity*, or the degree to which a cue informative about the situation. In addition, each relationship between a cue and operator judgment can be assigned a *cue utilization*. The lens model can be used to further decompose Murphy's Skill Score by examining the relations among these variables.

3.6. The lens model equation

One of the most important extensions to Brunswik's lens model was the development of the lens model equation (Hursch et al., 1964; Tucker, 1964). The lens model equation (LME) provides a mathematical representation of the lens model and partitions the overall correlation represented by the level of achievement or r_{YO} into correlations related to ecological validities of cues, cue utilizations, the predictability of the environment, and the consistency with which an operator implements his or her judgment (cue-weighting) strategy.

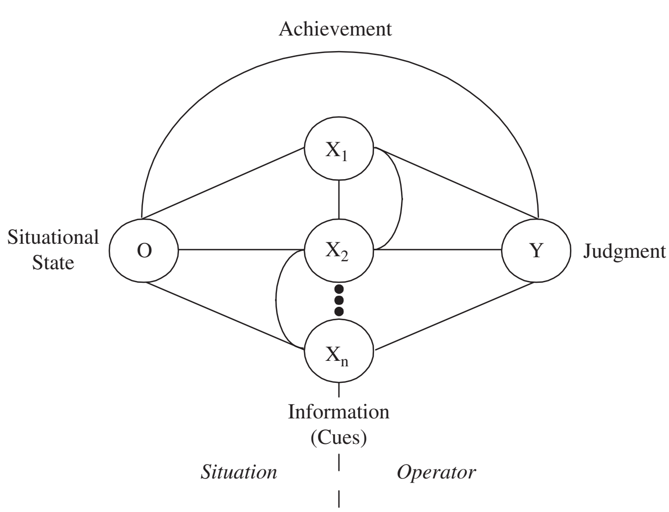

Fig. 3. Brunswik's lens model.

At the basis of the LME are two parallel models, which represent the *Situation* side and the *Operator* side of the lens model shown in Fig. 3. Both models are typically implemented with multiple linear regression models, but this need not be the case (e.g., see Rothrock and Kirlik, 2003, for a discussion of this issue and an alternative formulation in terms of rule-based modeling). The situation model describes the overall correspondence between the cues (X_i's) and the situation (O), and the operator model describes the overall correspondence between the cues (X_i's) and the operator's judgment (Y). Based on these two models, the resulting decomposition of achievement is depicted conceptually as

$$(r_{YO}) = \text{Environmental predictability} \times \text{Consistency} \times \text{Knowledge} + \text{Error}. \quad (6)$$

Environmental Predictability (or equivalently situational predictability) is the correspondence between the cues and the situation. *Consistency* is the correspondence between the cues and the operator's judgments as reflected in the operator model. Thus, a lower correlation between the cues and judgment behavior is less than fully predictable based on knowledge of the cues.

Knowledge is the degree of correspondence between the outputs of the situation and the operator models. Outputs from these models represent the predictable aspects of the situation and operator based on their respective degrees of correspondence with the cues. This component is called *Knowledge* because it indicates the degree to which the operator correctly models the regularities of the situation, or weights the cues adaptively. *Model Error* in Eq. (6) is the degree of correspondence between the unpredictable portions of the situation and operator models. Typically, this value is found to be marginal in human judgment (Brehmer, 1994).

The decomposition shown in Eq. (7) is accomplished with multiple linear regression. Thus, *Environmental Predictability*, *Consistency*, *Knowledge*, and *Model Error* are measured using multiple correlation statistics. The mathematical form of the LME is shown in Eq. (7), and its components are summarized below in Table 2.

$$r_{YO} = R_{OX}\mathrm{G}R_{YX} + C\sqrt{1 - R_{YX}^2}\sqrt{1 - R_{OX}^2} \quad (7)$$

4. Expanding lens model: sensing and information processing

To integrate the benefits of both the LME's decomposition of achievement and Murphy's (1988) decomposition of the SS, Stewart (1990) presented a single equation showing

Table 2
Components of the lens model equation

Component	Name	Description
r_{YO}	Achievement	Correlation between judgments and situation
R_{OX}	Environmental predictability	Correlation between situation and cues
G	Knowledge	Correlation between the model of the situation and model of the operator
R_{YX}	Consistency	Correlation between cues and judgments
C	Model error	Correlation between the residuals (or errors) in both models

how these decompositions could be combined. In Stewart's model, r_{YO} from the SS in Murphy's decomposition (see Eq. (7)) is expanded using a partial form of the LME (one excluding the typically negligible C or error component). A conceptual form of Stewart's decomposition is presented below in Eq. (8) and its mathematical form is presented in Eq. (9).

The first three terms in Eq. (8) are from the LME decomposition of r_{YO}. These three terms are, respectively, translated into the first three measures shown in Eq. (9). The last two terms in Eq. (8) (*Scale Error* and *Magnitude Error*) remain unchanged from Murphy's original decomposition of the SS. They are translated into the last two measures of Eq. (9):

$$\text{SS} = \text{Env. predictability} \times \text{knowledge} \times \text{consistency} - \text{regression bias} - \text{base rate bias} \tag{8}$$

$$\text{SS} = (R_{OX}\text{G}R_{YX})^2 - \left[r_{YO} - \left(S_Y/S_O\right)\right] - \left[\left((\bar{Y} - \bar{O})/S_O\right)\right]^2. \tag{9}$$

4.1. The expanded lens model

When considering the structure of the model presented in Eqs. (8) and (9), Stewart and Lusk (1994) recognized that *Environmental Predictability* (the correlation between the cues and a situation) could be further decomposed into two aspects: (1) the process that extracts data from the situation and (2) the process that transforms extracted data into cues available to the operator. These two processes represent a sequence typical in many systems with interface technology. For example, in a submarine the first process is represented by the sensing technology that gathers raw data from the underwater environment. The second process is represented by the fusing and display technology that transforms this data into the cues presented to the operator. Note that in this example, noise in the sensors or underwater environment can be passed on to the raw data, which can ultimately be passed on to the cues.

Decomposing *Environmental Predictability* into a two-stage sequence provides a more diagnostic description of the situation–information relationship. Moreover, it can help to identify constraints, such as malfunctioning sensors, that attenuate higher degrees of correspondence.

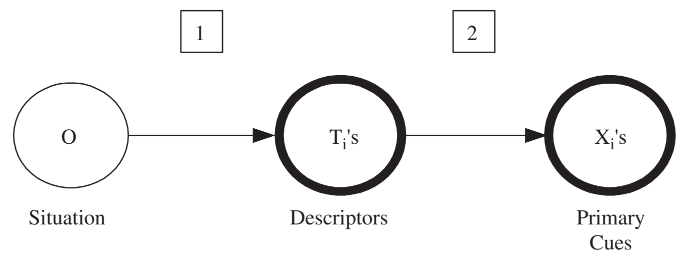

Fig. 4. Two-stage sequence of technological information processing.

This sequence of "technological information processing" is presented in Fig. 4.

As shown in Fig. 4, the two stages are labeled [1] and [2]. Stage [1] depicts the extraction of data or *descriptors* (labeled T_i's) from the situation. In Stage [2], these descriptors are transformed into the primary cues provided to the operator via interface displays. The thickened outlines around both the descriptors and the primary cues indicate their status as sets. Note that the situation and primary cues (O and X_i's, respectively) correspond to the situation and cues depicted in the traditional lens model.

In a similar manner, Stewart and Lusk also recognized that *Consistency*, a measurement of the correspondence between information (cues) and operator judgments, could be further decomposed into two processes: (1) the process that an operator uses to acquire information, and (2) the process that the operator uses to transform that information into a judgment.

Decomposing *Consistency* into a two-stage sequence increases the diagnosticity of the description of the information-operator relationship. Moreover, it can help to identify human information processing constraints that attenuate higher degrees of correspondence. The two-stage sequence of "operator information processing" (IP) is presented above in Fig. 5. As shown in Fig. 5, the two stages of operator IP are labeled [1] and [2]. Stage [1] depicts the operator's acquisition of *secondary cues* (labeled U_i's) from primary cues. In Stage [2], these secondary cues are transformed into a judgment. The primary cues in Figs. 4 and 5 are the same.

By inserting the sequences of technological IP and operator IP into the lens model, Stewart and Lusk (1994) developed the graphical form of the expanded lens model (ELM). The result is shown in Fig. 6. Fig. 6 maintains the basic configuration of the lens model, but shows the

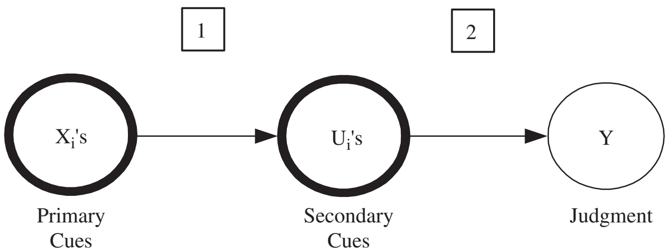

Fig. 5. Two-stage sequence of operator information processing.

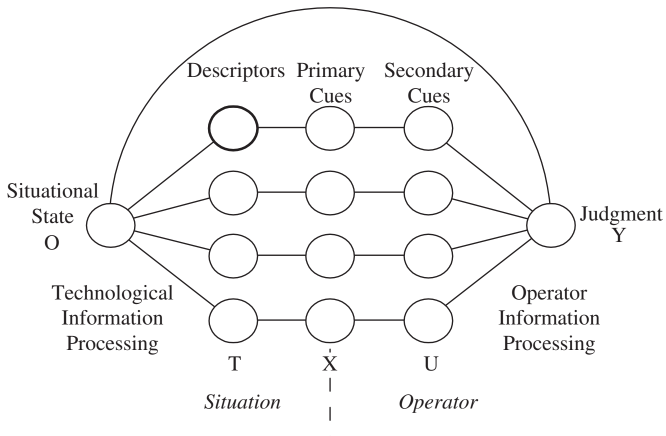

Fig. 6. The expanded lens model (ELM).

sequence of technological IP inserted into the *situation* side, and the sequence of operator IP inserted into the *operator* side. Expanding the LM in this manner introduces a new set of relationships between the model's components.

For technological IP, Stewart and Lusk (1994) used correlations to measure the two relevant correspondences: (1) between the situation and the descriptors and (2) between the situation and the primary cues. Calling the correlation between the situation and primary cues R_{OX} and the correlation between the situation and the descriptors R_{OT}, Stewart and Lusk derived Eq. (10). This equation shows the decomposition of R_{OX} into the product of R_{OT} (sensing accuracy) and the ratio of R_{OX} to R_{OT} (labeled V_{TX}). V_{TX} is called the *Fidelity of the Information System* (Stewart and Lusk, 1994):

$$R_{OX} = R_{OT}\left(R_{OX}/R_{OT}\right) = R_{OT}V_{TX}. \tag{10}$$

It captures the proportion of the variance within the descriptors, relative to the situation, that is maintained in the primary cues. For example, a submarine with display technology that perfectly transformed sensor information into the cues displayed to an operator would have high fidelity (i.e., V_{TX} would equal one and thus R_{OX} would equal R_{OT}). Note that these circumstances do not guarantee that the task environment is completely predictable; i.e., R_{OT} could still be less than one (e.g., the underwater environment may be inherently noisy):

$$R_{YX} = R_{YU}\left(R_{YX}/R_{YU}\right) = R_{YU}V_{UX}. \tag{11}$$

Table 3
The components of the expanded lens model

ELM component		Name
	SS	Skill score
(1)	R_{OT}	Environmental predictability
(2)	V_{TX}	Fidelity of the information system
(3)	G	Knowledge
(4)	V_{UX}	Consistency of information acquisition
(5)	R_{YU}	Consistency of information processing
(6)	$\left[r_{YO} - \left(S_Y/S_O\right)\right]^2$	Regression bias
(7)	$\left[\left((\bar{Y} - \bar{O})/S_O\right)\right]^2$	Base rate bias

Treating operator information processing analogously (see Eq. (11)), the complete mathematical form of Stewart and Lusk's expanded lens model is shown below:

$$\begin{aligned}\mathrm{SS} = {} & [(\overset{\boxed{1}}{R_{OT}})(\overset{\boxed{2}}{V_{TX}})(\overset{\boxed{3}}{G})(\overset{\boxed{4}}{V_{UX}})(\overset{\boxed{5}}{R_{YU}})]^2 \\ & - \overset{\boxed{6}}{\left[r_{YO} - \left(S_Y/S_O\right)\right]^2} - \overset{\boxed{7}}{\left[\left((\bar{Y} - \bar{O})/S_O\right)\right]^2}.\end{aligned} \tag{12}$$

Seven components comprise the final ELM decomposition. Components [6, 7] are the regression and base-rate biases discussed previously. Components [1–5] result from decompositions of *Achievement*, *Environmental Predictability*, and *Consistency* from the LME. Component [3] survives from the original LME and represents the operator's adaptation to cue-criterion correlations. Components [1, 2] capture the quality of the technological sensors that extract data from the situation and the technological IP transforming it into the cues presented to the operator. Components [4, 5] capture the quality of the operator IP that acquires the primary cues to form secondary cues, and transforms these secondary cues into a situational judgment. These components are summarized in Table 3.

4.2. Practical considerations in SA measurement

Modeling and measuring interface-mediated SA using this approach requires that the judgment task be analyzed in accordance with the structure depicted in Fig. 6: the situation must be defined as a criterion to be judged, the operator's task must be defined in terms of the judgments to be made, and information must be defined as a set of cues. Moreover, to implement the (typical) regression modeling procedure, both situation and judgment must be assigned quantitative values, and the numerical characteristics of the cues, such as their values and ranges, must be defined. This may not always be possible, or natural, and of course the model can make high demands for data collection due to the need to individually estimate its many diagnostic parameters (e.g., see Bisantz et al., 2001).

In addition to these limitations, the approach is not equipped to investigate all aspects of SA. For example, cognitive constructs such as memory, mental pictures, attention, and schemata (Endsley, 1995b; Wilson, 1995), are outside the purview of the ELM. Furthermore, the approach is not equipped to capture the relationships between SA and communication (Schreiber et al., 1998), or the emotional and phenomenological aspects of SA (Gerson, 1997).

5. Related approaches

We began this article with a discussion of Endsley's (1995b) three-level model of SA and a description of her SAGAT (Endsley, 2000) approach to SA measurement. The SAGAT approach is based on asking an operator to make a series of judgments, either about primitive perceptual elements (Level 1 SA), comprehension of meaning or relevance (Level 2 SA), or of future states (Level 3 SA). Despite the difference in content, we note that all three of these instances are judgments, and as such, are amenable to treatment from the perspective of judgment analysis (Cooksey, 1996). The research of Murphy (1988) and Stewart and Lusk (1994) on which we have drawn, and our own integration and application to the context of SA and systems engineering, all fall within the realm of judgment analysis. As such, one can see that our own modeling and measurement technique and Endsley's SAGAT measurement technique are complementary. More specifically, we suggest that SAGAT or a similar approach could be used to select and administer the set of operator queries (judgments) to be analyzed in SA measurement, and that the techniques presented here be used to analyze the resulting data to potentially gain diagnostic insights into SA beyond what can be learned from the percent of queries answered correctly and incorrectly. We hope this article has illustrated some of the benefits that might be gained from such a process.

5.1. Human–automation interaction

Parasuraman et al. (2000) provided a framework for human–automation interaction (HAI) by defining four "stages" of automation. Stage 1 automation concerns the acquisition of information from the task environment. Stage 2 represents a processing, fusing, or filtering of this information prior to information display. Stage 3 concerns decision support in selecting a course of action. Stage 4 concerns action implementation. For a review including a variety of applications of this framework to studies of HAI, see Wickens and Xu (2002).

Various components of our SA modeling and measuring technique map directly onto Parasuraman et al.'s framework. For example, the model decomposes overall task predictability (from the operator's perspective) into its environmental predictability and the predictability inherent in the fidelity of sensing and (automated) IP system. These measures could be used to assess, and perhaps predict, the impact of additional sensors (Stage 1 automation) or data fusion, filtering, etc. technology (Stage 2 automation). More generally, our approach to SA modeling allows one to diagnose and localize any positive or negative effects of automation on various measures (e.g., to consistency of information processing, to task knowledge, etc.). We believe that used collectively, the Parasuraman et al. framework and systems-oriented, ecological approach may advance understanding of the impact of various types of automation on SA.

5.2. Naturalistic decision making

The NDM paradigm (Klein, 1999), has come to represent a broadened view of judgment and decision making, with a focus on studying "how people use their experience to make decisions in field settings" (p. 97). It is crucial to note that, like NDM, the present systems approach to SA measurement and modeling does not have its basis in either classical decision theory (CDT) or behavioral decision making (BDM) approaches to the study of cognition. Both of those approaches have their foundation in models based on the internal *coherence* of cognition (e.g., expected utility theory, Bayes theorem, etc.). Like NDM, we instead use a *correspondence-based* approach, where cognition and behavior are evaluated in terms of adaptive achievement (for a discussion of the contrast between coherence- and correspondence-based approaches, see Hammond, 1999).

There are substantive ties between the two approaches as well. First, NDM acknowledges that environmental uncertainty places a ceiling on achievement. As Lipshitz et al. (2001a) have (qualitatively) put it, "Uncertainty is intimately linked with error: the greater the uncertainty, the greater the probability of making an error" (p. 339). Second, NDM, due to its historical roots in Klein's Recognition-Primed Decision (RPD) model, places a heavy emphasis on identification of diagnostic cues (Klein, 1999) supporting inference (e.g., Crandall and Getchell-Reiter, 1993). Third, owing to the methodological roots of the present approach in the work of Brunswik (1956), our approach to the study of SA shares with NDM the goal of "conducting one's study with representative samples of subjects, tasks and contexts to which one wishes to generalize" (Lipshitz et al., 2001b, p. 386).

Where our approach may differ from NDM, however, is that we do not necessarily agree that scientific research should include, as an "essential characteristic" a commitment to "informal modeling" (Lipshitz et al., 2001a, pp. 334–335). NDM's commitment to informality seems to us to arise out of its cognitive rather than ecological focus: "NDM places the human … at the center of interest and as its basis for prescription" (Lipshitz et al., 2001a, p. 333). As such, like the work of Endsley discussed previously, NDM places a premium on the richness of its psychological constructs (e.g., "mental pictures," "metacognition," etc.)

as these are seen to be required to paint a faithful phenomenological account of naturalistic cognition.

We recognize the value of naturalistic observation, and have made it a point to admit that our techniques do not capture all of the possible dimensions of SA. However, our view is that, while *premature* formalization is not appropriate, aiming toward increased formality should remain a guiding light in human factors, as with formality comes abstraction, the central means by which generalization can be assured, or at a minimum, can be empirically tested and evaluated.

6. Conclusion

We have presented a systems-oriented, ecological perspective on modeling and measuring interface-mediated SA. While the approach does not speak to every theoretical construct, or deal with every phenomenon previously claimed to contribute to SA, we do believe it should be considered as an addition to the human factors toolbox, as advances in theory and application depend on advances in measurement. We have also discussed how our methods complement related approaches, not only to SA, but also to HAI and NDM. Viewed abstractly, our techniques support separating signal and noise in the performance of uncertain tasks. We appreciate that the data needed to implement these techniques may be inconsistent with the realities of field studies. Thus, we naturally do not recommend these methods for every study touching on SA issues.

We note, however, that these techniques, seen as tools for filtering signal and noise in behavior in uncertain situations, are also useful for measuring the degree to which any claims about internal psychological processes can be sustained based on sparse data sets. While the need for extensive sampling of both situations and behavior in uncertain tasks can be viewed as a legitimate limitation of our approach, this same issue should caution us about the reliability of *any* inferential technique used to identify the factors contributing to SA based on sparse data sets.

A companion article (Strauss and Kirlik, 2006) presents a demonstration of the judgment-based SA modeling and measurement approach. This demonstration was useful in diagnosing and isolating the effects of display design differences and individual differences in SA achievement in a laboratory context. This demonstration is meant to serve as an illustrative example of the class of operational and industrial contexts in which our judgment-based approach to SA modeling and measurement is applicable. This class includes any situation in which human operators required to use judgment to coordinate their interactions with work environments or controlled systems through technological interfaces.

Acknowledgements

We thank Johns Hopkins Applied Physics Lab and the Naval Air Warfare Center for support.

References

Adams, M.J., Tenney, Y.J., Pew, R.W., 1995. Situation awareness and cognitive management of complex systems. Human Factors 37 (1), 85–104.

Bisantz, A., Kirlik, A., Gay, P., Phipps, D., Walker, N., Fisk, A.D., 2001. Modeling and analysis of a judgment task using a lens model approach. IEEE Transactions on Systems, Man and Cybernetics – Part A: Systems and Humans 30 (6).

Brehmer, B., 1994. The psychology of linear judgment models. Acta Psychologica 87, 137–154.

Brunswik, E., 1956. Perception and the Representative Design of Psychological Experiments. University of California Press, Berkeley, CA.

Cooksey, R.W., 1996. Judgment Analysis: Theory, Methods, and Applications. Academic Press, Inc, San Diego, CA.

Crandall, B., Getchell-Reiter, K., 1993. Critical decision method: a technique for eliciting concrete assessment indicators from the intuition of NICU nurses. Advances in Nursing Science 16 (1), 42–51.

Cronbach, L.J., Gleser, G.C., 1953. Assessing similarity between profiles. Psychological Bulletin 50 (6), 456–473.

Durso, F.T., Gronlund, S., 1999. Situation awareness. In: Durso, F.T., Nickerson, R., Schvaneveldt, R.W., Dumais, S.T., Lindsay, D.S., Chi, M.T.H. (Eds.), The Handbook of Applied Cognition. Wiley, New York, pp. 284–314.

Endsley, M.R., 1995a. Measurement of situation awareness in dynamic systems. Human Factors 37 (1), 65–84.

Endsley, M.R., 1995b. Toward a theory of situation awareness in dynamic systems. Human Factors 37 (1), 32–64.

Endsley, M.R., 1996. Automation and situation awareness. In: Parasuraman, R., Mouloua, M. (Eds.), Automation and Human Performance: Theory an Applications. Lawrence Erlbaum Associates, Mahwah, NJ, pp. 163–181.

Endsley, M.R., 1997. The role of situation awareness in naturalistic decision making. In: Zsambok, C.E., Klein, G. (Eds.), Naturalistic Decision Making. LEA, Mahwah, NJ, pp. 269–283.

Endsley, M.R., 2000. Direct measurement of situation awareness: Validity and use of SAGAT. In: Endsley, M.R., Garland, D.J. (Eds.), Situation Awareness Analysis and Measurement. Erlbaum, Mahwah, NJ, pp. 147–174.

Endsley, M.R., Garland, D.J. (Eds.), 2001. Situation Awareness: Analysis and Measurement. Erlbaum, Mahwah, NJ.

Endsley, M.R., Smolensky, M.W., 1998. Situation awareness is air traffic control: The picture. In: Smolensky, M.W., Stein, E.S. (Eds.), Human Factors in Air Traffic Control. Academic Press, San Diego, CA.

Flach, J.M., 1995. Situation awareness: proceed with caution. Human Factors 37 (1), 149–157.

Gerson, C.W., 1997. Situation awareness and dynamic performance training systems: Some reflections on the literature. Journal of Educational Technology Systems 25 (4), 373–407.

Gibson, C. P., Garrett, A. J., 1990. Towards a future cockpit—the prototyping and pilot integration of the mission management aid (MMA). Paper presented at the Situational Awareness in Aerospace Operations, Copenhagen, Denmark.

Gibson, J.J., 1979. The Ecological Approach to Visual Perception. Lawrence Erlbaum Associates, Inc, Hillsdale, NJ.

Hammond, K.R., 1999. Coherence and correspondence theories of judgment and decision making. In: Connolly, T., Arkes, H.R., Hammond, K.R. (Eds.), Judgment and Decision Making: An Interdisciplinary Reader. Cambridge University Press, New York, pp. 53–65.

Hammond, K.R., Stewart, T.R., 2001. The Essential Brunswik. Oxford University Press, New York.

Hursch, C.J., Hammond, K.R., Hursch, J.L., 1964. Some methodological considerations in multiple-cue probability studies. Psychological Bulletin 71, 42–60.

Kirlik, A., 2000. Human factors. In: Hammond, K.R., Stewart, T. (Eds.), The Essential Brunswik. Oxford University Press, New York.

Klein, G.A., 1999. Applied decision making. In: Hancock, P.A. (Ed.), Handbook of Perception and Cognition: Human Performance and Ergonomics. Academic Press, New York, pp. 87–108.

Lee, J.W., Yates, J.F., 1992. How quantity judgment changes as the number of cues increases: An analytical framework and review. Psychological Bulletin 112 (2), 363–377.

Lipshitz, R., Klein, G., Orasanu, J., Salas, E., 2001a. Focus article: Taking stock of naturalistic decision making. Journal of Behavioral Decision Making 14, 331–352.

Lipshitz, R., Klein, G., Orasanu, J., Salas, E., 2001b. Rejoinder: a welcome dialogue—and the need to continue. Journal of Behavioral Decision Making 14, 385–389.

Murphy, A.H., 1988. Skill scores based on the mean square error and their relationships to the correlation coefficient. Monthly Weather Review 116, 2417–2424.

Neisser, U., 1976. Cognition and Reality: Principles and Implications of Cognitive Psychology. W. H. Freeman, San Francisco, CA.

Parasuraman, R., Sheridan, T.B., Wickens, C.D., 2000. A model for types and levels of human interaction with automation. IEEE Transactions on Systems, Man, & Cybernetics—Part A: Systems and Humans 30 (3), 286–297.

Pew, R.W., 1995. The state of situation measurement: Circa 1995. Paper presented at the International Conference on Experimental Analysis and Measurement of Situation Awareness, Daytona Beach, FL.

Pew, R.W., 2000. The state of situation awareness measurement: Heading toward the next century. In: Endsley, M.R., Garland, D.J. (Eds.), Situation Awareness Analysis and Measurement. Erlbaum, Mahwah, NJ, pp. 33–50.

Rothrock, L., Kirlik, A., 2003. Inferring rule-based strategies in dynamic judgment tasks. IEEE Transactions on Systems, Man, and Cybernetics: Part A: Systems and Humans 33 (1), 58–72.

Salas, E., Prince, C., Baker, D.P., Shrestha, L., 1995. Situation awareness in team performance: implications for measurement and training. Human Factors 37 (1), 123–136.

Sarter, N.B., Woods, D.D., 1995. How in the world did we ever get into that mode? Mode error and awareness in supervisory control. Human Factors 37 (1), 5–19.

Schreiber, B.T., Bell, H.H., Raspotnik, W.B., 1998. Investigating communication and situation awareness in air combat. Paper presented at the Human Factors and Ergonomics Society 42nd Annual Meeting.

Smith, K., Hancock, P.A., 1995. Situation awareness is adaptive, externally directed consciousness. Human Factors 37 (1), 137–148.

Stewart, T.R., 1990. A decomposition of the correlation coefficient and its use in analyzing forecasting skill. Weather and Forecasting 5, 661–666.

Stewart, T.R., Lusk, C.M., 1994. Seven components of judgmental forecasting skill: Implications for research and improvement of forecasts. Journal of Forecasting 13, 579–599.

Strauss, R., Kirlik, A., 2006. A systems perspective on situation awareness II: Experimental evaluation of a modeling & measurement technique, International Journal of Industrial Ergonomics, this issue.

Taylor, R.M., 1990. Situational Awareness Rating Technique (SART): the development of a tool for aircrew systems design. Paper presented at the Situational Awareness in Aerospace Operations, Copenhagen, Denmark.

Tucker, L.R., 1964. A suggested alternative formulation in the developments of Hursch, Hammond, and Hursch, and Hursch, Hammond & Todd. Psychological Review 71 (6), 528–530.

Wesler, M.M., Marshak, W.P., Glumm, M.M., 1998. Innovative measures of accuracy and situational awareness during landing navigation. Paper presented at the Human Factors and Ergonomics Society 42nd Annual Meeting.

Wickens, C. D., Xu, X., 2002. Automation trust, reliability and attention. University of Illinois, Human Factors Division Technical Report AHFD-02-14/MAAD-02-02.

Wilson, G.F., 1995. Psychophysical assessment of SA? Paper presented at the International Conference on Experimental Analysis and Measurement of Situation Awareness, Daytona Beach, FL.