Extending the Reach of Augmented Cognition To Real-World Decision Making Tasks

Frank L. Greitzer

Pacific Northwest National Laboratory
Richland, WA 99352
Frank.Greitzer@pnl.gov

Abstract

The focus of this paper is on the critical challenge of bridging the gap between psychophysiological sensor data and the inferred cognitive states of users. It is argued that a more robust behavioral data collection foundation will facilitate accurate inferences about the state of the user so that an appropriate mitigation strategy, if needed, can be applied. The argument for such a foundation is based on two premises: (1) To realize the envisioned impact of augmented cognition systems, the technology should be applied to a broad, and more cognitively complex, range of real-world problems. (2) To support identifying cognitive states for more complex, real-world tasks, more sophisticated instrumentation will be needed for behavioral data collection. It is argued that such instrumentation would enable inferences to be made about higher-level semantic aspects of performance. The paper describes how instrumentation software developed to support information analysis R&D may serve as an integration environment that can provide additional behavioral data, in context, to facilitate inferences of cognitive state that will enable the successful augmenting of cognitive performance.

Keywords: *augmented cognition; workstation monitoring; complex task performance; behavioral instrumentation; cognitive states.*

1 Introduction

A grand challenge for augmented cognition research is the ability—in real time—to (a) monitor and assess the user's cognitive state through behavioral and psychophysiologically derived measures ("cognitive state gauges") acquired from the user while interacting with the computer system; and (b) adapt or augment the computational interface to improve performance of the user-computer system. The first objective is a major focus of the first phase of the augmented cognition program that enables research to progress toward the second, and ultimate, objective.

To meet this challenge, it is not only necessary to devise and validate sensor systems to acquire the cognitive state gauge data, but also to develop methods for determining the cognitive state based on these data. This requires behavioral research and analyses to identify trends or patterns, within the appropriate context, that enable correct inferences about the user's cognitive state. Coupled with appropriate intelligent support and a computational architecture, this integrated system comprises the minimum requirements for an augmented cognition system capable of mitigating human information processing/decision making limitations, thereby "closing the loop" and successfully augmenting the user's cognition (Schmorrow and Kruse, 2004). This paper describes an approach to behavioral data collection that may be applied to benefit the Augmented Cognition program by helping to bridge the gap between the sensor data and the inferred cognitive states of users, and by extending the domain of instrumented augmented cognition research to include performance on more real-world tasks.

2 Motivation and Objectives

In the first phase of DARPA's augmented cognition program, a major Technical Integration Experiment was conducted to measure the cognitive state using data from more than twenty cognitive state gauges during the conduct of a "relatively complex cognitive task that was derived from the real world decision making requirements seen with tactical decision makers" (St. John, Kobus, and Morrison, 2003). As reported, many of the sensors could be reliably linked with behavioral measures during the performance on the test task (Warship Commander, a computerized command and control task described as fast-paced and requiring cognitive activities of detection, identification, and memory recall). The results of this study were important because, perhaps for the first time, a set of bio-sensors could be shown to correlate with observations that could be linked with mental state constructs[1]. Nevertheless, as is always the case with such empirical studies, and as pointed out by the authors of this study, "… generalization to other tasks and other situations must be drawn with care."

The high-tempo, command-control monitoring nature of the Warship Commander Task (WCT) makes it well-suited to generalize to tactical decision making or air traffic control—high risk/high stress activities with a strong focus on workload as embodied in tempo of operations, attention/divided attention, and multiple task performance. Nevertheless, as the authors point out, "… not all aspects of user cognition and workload were specifically manipulated by the WCT. For instance, the amount and complexity of relatively high-level decision-making was limited. Therefore, the assessment of any gauges claiming to measure higher order cognitive processing was limited." (St. John *et al.*, 2003; p. 121). To stimulate and provide context for refinement of cognitive state gauges for higher-level cognitive processes such as memory, learning/comprehension, and problem solving, augmented cognition technology should be applied to more complex tasks such as information analysis and decision making. This has the added benefit of expanding the application of the technology to a broader range of real-world tasks. The conceptual challenge of specifying more precisely the psychological construct of cognitive state gauges requires data that enable a more direct association between the sensor data and behavioral data. Therefore, a more robust behavioral data collection foundation is needed to facilitate accurate inferences about the state of the user.

In short, the motivation for the present paper is based on the following premises (and objectives):

- To realize the envisioned impact of augmented cognition systems, the technology should be applied to a broad, and more cognitively complex, range of real-world problems (broadening the task set).
- To support identifying cognitive states for more complex, real-world tasks, more sophisticated instrumentation will be needed for behavioral data collection (enhancing the behavioral instrumentation).

2.1 Broadening the Task Set

An important and interesting class of real-world cognitive tasks associated with high risk and high stress is tactical or strategic decision making associated with information analysis (IA). Heuer (1999) describes this as a fluid, dynamic, fast-paced, and multi-process activity that is affected by human information processing limitations and biases. Analysts must make difficult judgments to assess the relevance, reliability, and significance of disparate pieces of information that requires complex reasoning processes to determine "the best explanation for uncertain, contradictory and incomplete data" (Patterson, Roth, & Woods, 2001, p. 225). The combination of data uncertainties and volume, complex judgments/reasoning processes, time pressure and high stakes decision making creates a challenging and stressful environment. Woods, Patterson and Roth (2002) conducted empirical studies of information overload to identify its impacts and to suggest mitigation approaches. Fundamental to this research is the assertion that data can best be understood in terms of the user's context. By focusing on how people deal with large quantities of data, this research seeks to provide guidance for using the context sensitive nature of data to assist the analysis process. Some of the key findings of this area of research are that analysts use "search narrowing" strategies that contribute to a tendency to miss "high profit" information; and that some analysts are less persistent in their search/evaluation strategies, as reflected in the amount of time spent searching and reading. The search narrowing effect was also documented by Nordlie (1999), whose work suggests that attempting to fit information into an incomplete mental model is one reason for incomplete, inadequate or digressive searches.

There are several large, ongoing programs of research on IA aimed at developing information technology tools to augment cognitive processes and overcome human information processing limitations and biases. There are sharp differences in the type of cognitive load imposed by the WCT studied to date in augmented cognition research and that associated with IA tasks. For example, in the sense of NASA's Task Load Index (Hart & Staveland, 1988), the WCT has a high *temporal* demand but relatively short in duration; in contrast, IA decision making has a relatively large *mental* demand component but a longer duration. To broaden the reach of augmented cognition in the real world, it would be desirable to conduct a technical integration experiment on the more cognitively complex sorts of information discovery and decision making tasks that comprise IA. In turn, it is conceivable that the data provided

[1] Computer-based simulation tasks or "games" have been used profitably for decades in research investigating impacts of cognitive load on information processing performance (e.g., Greitzer, Hershman, and Kelly, 1981), but technologies have only recently become available to integrate portable sensing equipment with computer-based tasks in research environments.

by cognitive state gauges may provide additional context that will ultimately benefit augmented cognition enabled systems to identify when and what type of mitigation intervention or support is needed. The role of context (derived from time-stamped behavioral and bio-sensor data) will become critical in reliably inferring the user's cognitive state and identifying limitations or biases that require intervention. This is discussed in the next section.

2.2 Enhancing the Behavioral Instrumentation

As we focus on more complex cognitive tasks, we will find that it is increasingly difficult to make inferences and determine correlations that connect cognitive state gauge data to hypothesized or observed limitations/bottlenecks in human performance. The ability to infer the cognitive state of the user is extremely valuable under circumstances of high stress, information overload, or when performance limitations or biases are likely to occur. On the basis of behavioral measures alone, it is extremely difficult to infer cognitive states because higher level cognitive processes are not manifested in typical data that may be recorded by monitoring instrumentation software at the analyst's workstation (such as keystrokes, mouse clicks, and the like). Similarly, psychophysiological sensor data collection alone—a fundamental component of research on augmented cognition—suffers from limitations inherent in the lack of correspondence with specific performance measures associated with tasks. By combining behavioral and psychophysiological sensor data collection methods, correlations between these data may facilitate inferences about higher-level semantic processes and cognitive states associated with performance limitations. With such detailed, time-stamped behavioral and sensor data, it may be possible to define the desired cognitive constructs associated with cognitive gauge data. The notion is illustrated in Figure 1. Time-stamped data captured by monitoring/instrumentation software on the user's workstation may be aggregated with cognitive gauge sensor data and then analyzed to assess behavioral correlates of the cognitive state constructs. Convergence of measures derived from such integrated and coordinated data may afford an opportunity to improve our understanding of and ability to act upon feedback on the user's cognitive state.

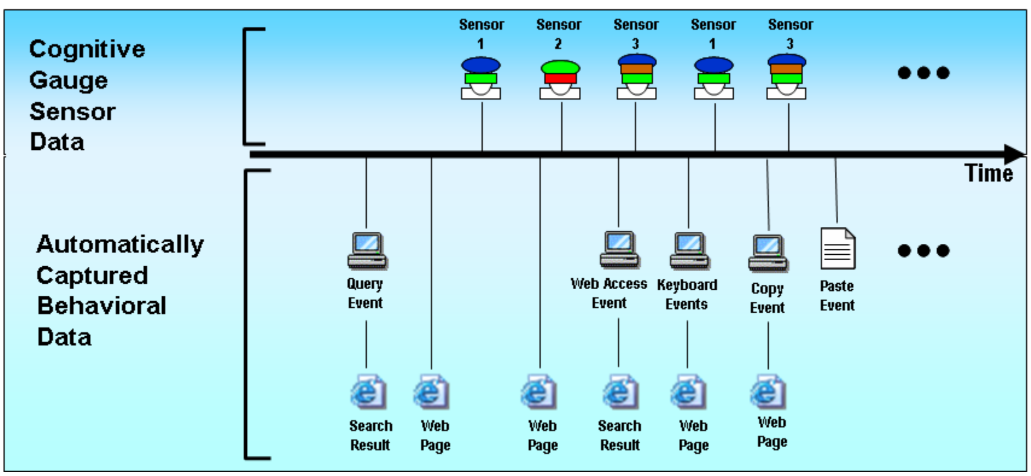

Figure 1. Notional illustration of cognitive gauge data collected in coordination with extensive behavioral data in an information analysis task.

3 A Behavioral Monitoring Instrumentation Environment for Monitoring Information Analysis

The Advanced Research and Development Activity initiated the Novel Intelligence from Massive Data (NIMD) program to address the very complex and difficult problem of anticipating strategic surprise by augmenting the intelligence analysis (IA) process. As part of this program, a Glass Box Analysis Environment (GBAE) has been developed (Cowley, Nowell, and Scholtz, 2005) to provide data and visualization tools to increase researchers' understanding of cognitive foundations of IA and to provide an integration/test environment for NIMD-developed IA tools. The remainder of this paper discusses ways in which the GBAE instrumentation may be integrated into a test environment that can provide additional behavioral data, in context, to facilitate inferences of cognitive state that will enable the successful augmenting of cognitive performance.

3.1 Overview of Glass Box Instrumentation Software Architecture

The Glass Box provides the infrastructure through which data on the analytic process is collected, research products are evaluated, and disparate analytic research applications communicate, integrate and coordinate with each other. A web-services architecture enables distribution across wide area networks and firewall boundaries while simultaneously reducing the resource footprint on the user's workstation. Shared Analytic Event Servers maintain the current state, notify clients (applications) of current events, provide publish/subscribe communications between clients (some of which may reside on the user's workstation), and accumulate a long term historical record of the analytic events and communications between clients. While the Glass Box instrumentation has been used to support integration of software products designed to enhance human information processing, it can also be used to support integration of cognitive state sensors being studied in the Augmented Cognition research program.

3.2 Data Collected by the Glass Box Instrumentation Software

The Glass Box instrumentation captures analyst workstation activities including keyboard/mouse data, window events, file open and save events, copy/paste events, instant messaging, and web browser activity. The Glass Box makes extensive use of a relational database to store time-stamped events and a hierarchical file store where files and the content of web pages are stored. The Glass Box "snatches" a copy of every file the analyst opens and saves so there is a complete record of the evolution of documents. The material on every web page the analyst visits is explicitly stored so that each web page can be later recreated (by researchers) as it existed at the time it was accessed by the analyst; screen images are also captured. Through its Application Programming Interface (API), the Glass Box serves as a "sensory mechanism" for applications on the analyst's workstation. The API enables applications to access user interactions and system activity, to log activities of interest, and to communicate through a publish/subscribe mechanism. Thus, the API allows applications to find out what the analyst is doing in near real time, determine a course of action, and to coordinate their response. Researchers and those evaluating the applications can use the resulting data to determine how the application was used, where the analysts had problems, where they took unexpected paths, and what conditions were present when problems developed.

3.3 Using the Glass Box Data

By itself, the low level data collected by the Glass Box has limited value in studying the higher cognitive aspects of the information analysis process. These data become interesting, however, when they aggregated, distilled or combined with other data to indicate what we call "analytic events." We define an analytic event as a primitive or composite event type that is significant to the analytic process. The NIMD community is still wrestling with what data may be defined as an analytic event. Of immediate and obvious significance are events such as issuance of a search engine query. On the other hand, keystroke data may not be useful unless combined with other data or processed further (e.g., to derive words, sentences, or paragraphs; or to derive temporal measures related to workload or higher-level cognitive processes). Thus, when aggregated or synthesized, the Glass Box data may shed light on higher-level cognitive processes. Sanquist, Greitzer, Slavich, Littlefield, Littlefield, and Cowley (2004) used temporal measures derived from the Glass Box data to describe the task-level activities of intelligence analysts working in a Glass Box environment. While most research employing temporal measures to evaluate links between external behavior and cognitive processes has focused on the millisecond time domain (e.g., Meyer, *et al.*, 1988), Sanquist *et al.* apply the same principles to the temporal aspects of cognitive tasks. Such temporal relationships may illuminate cognitive activities that are amenable to augmentation. For example, Sanquist *et al.* found that time measures facilitate screening data to characterize the nature of the analyst session, identifying specific areas where the analyst shows interest, identifying the material that the analyst filtered out with quick views, and (perhaps to a lesser extent) portraying the flow of information across the task structure. Similarly, researchers at the National Institute of Standards and Technology used Glass Box data to study the growth of a report over time by combining individual time stamped characters and cut/paste events (Scholtz *et al.* 2004).

4 Conclusions

The Glass Box generates a very large amount of data that may be used for several purposes. To date, the focus of research involving these data has been on evaluation of information technology tools, but the present discussion has sought to articulate its potential utility as an environment in which to integrate cognitive gauge sensors to advance augmented cognition research. By integrating cognitive state sensors into a Glass Box environment, it is possible to closely associate psychophysiological measures collected concurrently with time-stamped behavioral data that are automatically logged for users engaged in real-world, complex cognitive tasks such as information analysis. Such an integration and exploitation of this substantial data collection capability can provide a source of behavioral data that will complement the augmented cognition sensor data and help to bridge the gap between the sensor data and the inferred cognitive states of users. Identification and utilization of the human decision maker's cognitive state is a major thrust of the NIMD program aimed at applying more effective, intelligent support to analysts. Indeed, it is possible that the convergence of behavioral and psychophysiological measures that would be obtained by integrating cognitive state sensors into a Glass Box analytic environment may yield a better recognition and identification of cognitive states that have been heretofore unobservable.

References

Cowley, P. J., Nowell, L. T., Scholtz, J. (2005). Glass Box: An Instrumented Infrastructure for Supporting Human Interaction with Information. *Hawaii International Conference on System Sciences.* January 2005.

Greitzer, F. L., Hershman, R. L., and Kelly, R. T. (1981) The air defense game: A microcomputer program for research in human performance. *Behavior Research Methods and Instrumentation, 13*, 57-59.

Hart, S. G., and Staveland, L. E. (1988). Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload*. Amsterdam, The Netherlands: Elsevier.

Heuer, R. J. (1999) *Psychology of intelligence analysis.* Washington, D.C.: Central Intelligence Agency, Center for the Study of Intelligence.

Meyer, D.E., Osman, A.M., Irwin, D.E., and Yantis, S. (1988). Modern mental chronometry. *Biological Psychology*. *26*: 3 – 67.

Nordlie, R. (1999). "User revealment"—a comparison of initial queries and ensuing question development in online searching and in human reference interactions. *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Berkeley, CA, ACM Press.

Patterson, E. S., Roth, E. M. & Woods, D. D. (2001). Predicting Vulnerabilities in Computer-Supported Inferential Analysis Under Data Overload. *Cognition, Technology & Work*, *3*, 224-237.

Woods, D. D., Patterson, E. S., & Roth, E. M. (2002) Can We Ever Escape from Data Overload? A Cognitive Systems Diagnosis. *Cognition, Technology, and Work, 4,* 22-36.

Sanquist, T. F., Greitzer, F. L., Slavich, A., Littlefield, R., Littlefield, J., and Cowley, P. (2004) Cognitive tasks in intelligence analysis: Use of event dwell time to characterize component activities. *Proceedings, Human Factors and Ergonomics Society 48th Annual Meeting*. New Orleans, LA: September 2004.

Schmorrow, D. D., and Kruse, A. A. (2004). Augmented Cognition. In W.S. Bainbridge (Ed.), *Berkshire Encyclopedia of Human Computer Interaction.* Great Barrington, MA: Berkshire Publishing Group. (54-59)

Scholtz, J., Morse, E., and Hewett, T. (2004). In depth observational studies of professional intelligence analysts. *Human Performance, Situation Awareness and Automation Technology Conference*. Daytona Beach, FL: March 2004.

St. John, M., Kobus, D. A., and Morrison, J. G. (2003) *DARPA Augmented Cognition Technical Integration Experiment (TIE).* Technical Report 1905. San Diego: Pacific Science and Engineering Group, Inc.