Buettner, Ricardo: Cognitive Workload of Humans Using Artificial Intelligence Systems: Towards Objective Measurement Applying Eye-Tracking Technology.
In KI 2013: 36th German Conference on Artificial Intelligence, September 16-20, 2013, Koblenz, Germany,
Vol. 8077 of Lecture Notes in Artificial Intelligence (LNAI), pp. 37-48, 2013.

Cognitive Workload of Humans using Artificial Intelligence systems: Towards Objective Measurement applying Eye-Tracking Technology

Ricardo Buettner

FOM University of Applied Sciences
Institute of Management & Information Systems
Hopfenstraße 4, 80335 Munich, Germany
ricardo.buettner@fom.de

Abstract. Replying to corresponding research calls I experimentally investigate whether a higher level of artificial intelligence support leads to a lower user cognitive workload. Applying eye-tracking technology I show how the user's cognitive workload can be measure more objectively by capturing eye movements and pupillary responses. Within a laboratory environment which adequately reflects a realistic working situation, the probands use two distinct systems with similar user interfaces but very different levels of artificial intelligence support. Recording and analyzing objective eye-tracking data (i.e. pupillary diameter mean, pupillary diameter deviation, number of gaze fixations and eye saccade speed of both left and right eyes) – all indicating cognitive workload – I found significant systematic cognitive workload differences between both test systems. My results indicated that a higher AI-support leads to lower user cognitive workload.

Keywords: artificial intelligence support, cognitive workload, pupillary diameter, eye movements, eye saccades, eye-tracking, argumentation-based negotiation, argumentation-generation

1 Introduction

Towards programming the "global brain" [1] and realizing real collective intelligence [2], the vision of flexibly connecting billions of computational agents and humans is constantly recurring (e.g. [3]). Behind this vision lies the assumption that artificial intelligence (AI) supports humans in solving tasks and distributing the human/cognitive workload across the "global brain" [1–7] (fig. 1). It is human nature to "off-load cognitive work onto the environment" [7, p. 628, claim 3].[1] However, information systems (IS) scholars have traditionally investigated a user's cognitive workload and its derivatives[2] primarily based on user-

[1] Because of the limits of human's information-processing abilities (e.g., limits to the attention and working memory of the human brain), we tend to exploit the environment in order to reduce cognitive workload [7].
[2] Such as concentration, mental strain, mental stress, e.g. [8].

Buettner, Ricardo: Cognitive Workload of Humans Using Artificial Intelligence Systems: Towards Objective Measurement Applying Eye-Tracking Technology.
In KI 2013: 36th German Conference on Artificial Intelligence, September 16-20, 2013, Koblenz, Germany,
Vol. 8077 of Lecture Notes in Artificial Intelligence (LNAI), pp. 37-48, 2013.

perceived/non-objective measures (e.g. MISQ: [9], JMIS: [10], DSS: [11], ISR [12]) or even discussed the need for user workload measurements without any measurement proposal (MISQ: [13]). Nevertheless, more and more IS scholars call for objective measurement techniques of user's cognitive workload and its derivatives (e.g. [14]) and a small group of IS researchers currently fosters the conducting of objective psychophysiological measures in IS research and recently formulated corresponding research calls (i.e. [15–17]).

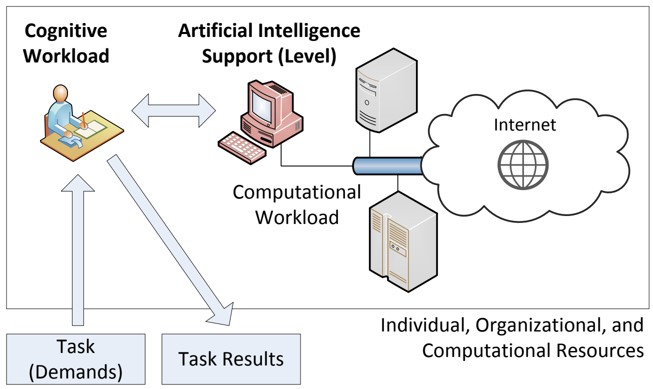

Fig. 1. Distributing human cognitive workload across the "global brain"

Replying to these research calls I show in this paper how the user's cognitive workload can be measured more objectively by capturing eye movements and pupillary responses via eye-tracking technology. Within a laboratory environment which adequately reflects a realistic working situation the probands had to use two distinct systems with similar user interfaces but very different levels of artificial intelligence support. In more detail I prototyped a test scenario derived from a typical real business environment in which extra-occupational MBA and bachelor students having professional working experience had to apply for jobs. The one system offers a chat function where the applicants had to situatively generate appropriate arguments on their own without any AI-support. The other system presented a set of AI-/system-generated arguments from which the users only had to select an appropriate argument. Recording and analyzing objective eye-tracking data (i.e. pupillary diameter mean, pupillary diameter deviation, number of gaze fixations and eye saccade speed of both left and right eyes) – all indicating cognitive workload – I found significant systematic cognitive workload differences between both systems.

Using this work I aim to contribute solutions to the current methodological problem of the objective measurement of user's cognitive workload when running AI systems (cf. [1, 3–6, 18–20]). In addition to these methodological contributions my results strongly emphasize the meaningfulness of the devel-

Buettner, Ricardo: Cognitive Workload of Humans Using Artificial Intelligence Systems: Towards Objective Measurement Applying Eye-Tracking Technology.
In KI 2013: 36th German Conference on Artificial Intelligence, September 16-20, 2013, Koblenz, Germany,
Vol. 8077 of Lecture Notes in Artificial Intelligence (LNAI), pp. 37-48, 2013.

opment of argumentation-based negotiation models using intelligent software agents (e.g. [21, 22]) from a human workload perspective.

The paper is organized as follows: After this introduction I firstly present in section 2 the state of the art concerning pupillary responses and eye movements as cognitive workload indicators in human psychophysiology and IS research. In section 3 I define the objectively measurable cognitive workload indicators, determine the hypotheses, and describe the test systems, the laboratory setting, as well as the sampling strategy. Next, in section 4 I present the objectively measured cognitive workload indicators and the results on the hypotheses evaluation. Results are then discussed in section 5. Finally, I discuss the contributions and limitations of my results and indicate future work in section 6.

2 Related work

Despite the high level of interest in cognitive workload, there is still no universally accepted definition of this mental construct [8]. However, it is clear that cognitive workload results from mental processes when performing tasks – depending on the users's capabilities and the task demands, e.g. [23–25]. The corresponding user's cognitive workload measurement techniques can be roughly separated into two categories [8]: subjective self-assessment and rating scales (e.g. NASA TLX), and objective psychophysiological measures (e.g. pupillary responses). In the following two sections I concentrate on eye-related psychophysiological measures indicating cognitive workload and measurable by eye-tracking technology.

2.1 Pupillary responses and eye movements as cognitive workload indicators in human psychophysiology research

The initial work on the relationship between cognitive workload and pupillary responses stems from Hess and Polt [24] and was published in 1964 in the `Science` journal. Hess and Polt [24] measured the cognitive workload of 5 probands by capturing the task-evoked pupillary diameter, but only based on simple multiplication tasks. Kahneman and Beatty [25] showed that the rate of the task-evoked pupillary diameter changes strongly in relation to task difficulty. Bradshaw [26] found the post-solution drop of the task-evoked pupillary diameter after finishing the task. Simpson [27] found that a subsequent indication of task completion causes a higher pupillary dilation during the preceding cognitive task. Based on testing 17 students, Kahneman, Beatty and Pollack [28] showed the stability of the correlation between cognitive workload and pupillary diameter for much more complex tasks (listening, calculating, talking) under different conditions. Following the fundamental investigations of Kahneman and colleagues [25, 28], the amount of user cognitive workload clearly corresponds with the pupillary dilation, e.g. [23, 26, 29].

Besides the diameter of the pupillary, some data from eye movements also indicate the user's cognitive workload level: Eye saccades are the rapid eye movements between points of fixation and often used for cognitive analysis [30, 31].

Buettner, Ricardo: Cognitive Workload of Humans Using Artificial Intelligence Systems: Towards Objective Measurement Applying Eye-Tracking Technology.
In KI 2013: 36th German Conference on Artificial Intelligence, September 16-20, 2013, Koblenz, Germany,
Vol. 8077 of Lecture Notes in Artificial Intelligence (LNAI), pp. 37-48, 2013.

Information can only be perceived during fixations and not 75msec before saccades starting, during saccades, and 50msec after saccades. Since long fixations (>500 msec) indicate a deeper cognitive processing, cognitive workload is clearly positively correlated with the frequency of long fixations but negatively correlated with the saccade speed, e.g. [30, 32–34].

In general most of the psychophysiological work was based on pupillary-photographing technology, thus limited to measurements of well-seperated rudimental/basic tasks. But, IS usage is regularly very dynamic. However, very recently Wierda et al. [35] showed how eye-tracking technology can be used for high-temporal-resolution tracking of cognitive workload.

2.2 Pupillary responses and eye movements as cognitive workload indicators in IS research

IS scholars have traditionally investigated user's cognitive workload and its derivatives [8] primarily based on user-perceived/non-objective measures (e.g. [9–12]) or even discussed the need for user workload measurements without any measurement proposal (e.g. [13]).[3] In the seldom case of using objective psychophysiological measures, IS research has mainly applied pupillary-based techniques indicating cognitive workload within the human-computer interaction domain, especially for adaption and personalization purposes (essential publications: [15, 37–43]).

When focusing on "AI-specific" work (in a very broad sense) in more detail, it can be summarized that determining the user's cognitive workload is often mentioned as a fundamental problem in human-machine systems (e.g. [18, 20]). The discourse on measuring the machine intelligence of human-machine cooperative systems (e.g. [6]) showed the need to quantify the cognitive workload of machine users and postulated the need for research on workload measures based on objective parameters such as behavioral signals, eye scanning movements, or physiological variables. Also the discussions about metrics for human-robot interaction emphasized the need for research into a more objective cognitive workload measurement technique (e.g. "At this point in time, there is a need to identify non-intrusive measures of workload..." [19, p.,38]). Accordingly a lot of trials and rudimentary/simple approaches on measuring the user's cognitive workload when using AI systems exist. For example, Pomplun and Sunkara [44] used the pupillary dilation as a cognitive workload indicator within a simple visual experiment asking users to find numbers in ascending order and read them out loud. Longo [45] sketched a very rudimentary framework for cognitive workload assessment using information technology. Cegarra and Chevalier [46] experimentally evaluated the cognitive workload of users solving a Sudoku puzzle

[3] Other IS-relevant disciplines show the same situation concerning user-perceived/non-objective cognitive workload measures. For example, Loft et al. [36] summarizes the state of the art concerning 22 existing models which predict cognitive workload in air traffic control. It is remarkable that all of 22 developed models were based on subjective workload ratings.