Understanding systems that are designed to support human cognition

Lewis L. Chuang
Department for Perception, Cognition, and Action,
Max Planck Institute for Biological Cybernetics
Max Planck Ring 8, 72076, Tübingen, Germany
lewis.chuang@tuebingen.mpg.de

Albrecht Schmidt
Human Centered Ubiquitous Media Group, Department of Computer Science
Ludwig-Maximilians-UniversitÃď't München, Frauenlobstr. 7a, 80337, München, Germany
albrecht.schmidt@um.ifi.lmu.de

 Please address correspondence to lewis@humanmachinesystems.org
Rethinking Interaction: From Instrumental Interactions to Human-Computer Partnerships at CHI'18, April 21–26, 2018, Montreal, Canada.

Abstract
The prevalence of automation and user-adaptive systems has created a demand for human-machine interfaces that are designed to be aware and sensitive to the cognitive states of the user. We argue that the unitary concept of *mental workload* is insufficient in capturing the benefits that computing systems can deliver to their users. More specifically, we demonstrate with examples from our own research how to target more specific and robustly defined cognitive processes, with the use of non-obtrusive methods such as gaze-tracking, electroencephalography, and physiological measurements. Our examples, particularly in the evaluation of auditory notification design and *in situ* displays, will demonstrate that it is advantageous to target specific cognitive processes and mechanisms in accordance to the design purpose of a given interface, within the constraints of contemporary models of psychology and neuroscience.

Author Keywords
cognition, context-aware, mental workload, automation, adaptive systems, EEG

Towards machines that support cognition
Two recent trends in computing have created an impetus in redefining the role of human cognition in human-machine interactions. More specifically, we refer to the rapid adoption of automation that is able to perform work without our

Figure 1: The exponential concern of mental workload in the use of computing systems. Search performed on 2018-02-08.

Figure 2: Computing machines that allow for abstract calculations have developed over time in terms of their size and their programmable functional flexibility. *Top-left (clockwise): Chinese Abacus (ca. CE 190), Babbages Analytical Engine (1834-1871), Programmable calculator (HP-65; 1974), Smartphone (iPhone; 2007)*: Photo credit: modified from Wikimedia Commons

constant supervision, and adaptive systems that are capable of modifying their behavior in response to the changing demands of the environment and implicit user requirements. The design of computing systems have always been sensitive to the demands that they place on their users. A cursory search for *("mental workload" & "cognitive workload")* in the ACM digital library results in 72,660 articles, of which almost half were published in the last 7 years alone (see Figure 3). After all, computing machines are designed explicitly with the goal to assume part of the cognitive work that we would, otherwise, have to perform ourselves. For example, abaci, cameras, and programmable computers serve to reduce the cognitive effort that has to be expended in order to count, record memories, and simulate complex scenarios. The growing ubiquity and functional autonomy of computing machines elevates their statuses from being a servant to a collaborator (see Figure 2 for examples). If computing systems are no longer to be evaluated in terms of how much mental work they assume on our behalf but, instead, in terms of their ability to help us think more effectively, should we continue to design human-machine interactions with the objective of minimizing "mental workload"?

Mental workload is a diminishing concept

The concept of "mental workload" is best understood within the descriptive framework of capacity models [10, 13]. Capacity models are generally centered around the description that tasks places demands for (mental) resources, which the human (mind) attempts to supply in order to maintain a steady and acceptable level of performance. The total capacity of (mental) resources can be raised to meet the challenges of a more demanding task. However, it cannot do so indefinitely. Thus, demands can supercede supply and result in poor task performance. This characterization highlights an inconvenient truth, one that is often neglected—namely that task performance is not indicative of workload. A constant level of performance can be maintained in spite of diminishing spare capacity, namely the difference between one's total capacity and task demands.

Unfortunately, poorly designed machines can not only fail to ease our workload; they can increase our workload. There are at least two reasons for this. First, unintuitive machine interfaces can introduce unintended complexity into our interactions with the machine. This could pose a additional and non-negligible workload that might go unnoticed, so long as the job continues to get done and users do not express discomfort. For example, using an abacus (or beans) for counting might be preferable to using an Excel spreadsheet because it does not place a demand for abstract reasoning, which might be effortful for some. Second, user expectations could contradict the engineered purpose of a machine. This creates confounding results and more mental effort might be expended, than was anticipated, as users attempt to re-purpose a machine for a job that it was not designed for. The popularity of mobile computing devices (i.e., smartphones) might have arisen because it traversed both hurdles effectively. Touch devices with intuitive gesture interfaces do not require symbolic reasoning. Furthermore, a diverse application market has resulted in a software environment with functionalities that are driven by user choice and is, hence, more likely to be aligned with user expectations.

Thus, the "mental workload" experienced by users ought to be an important measure, at least of the extent to which computing machines relieve users of performing work. Unfortunately, there is no good measure for "mental workload". Subjective questionnaires, such as the popular NASA-TLX [9], require subjects to recall their experiences with interacting with a given system and broadly assumes that we are conscious of how much mental effort we expend. Physio-

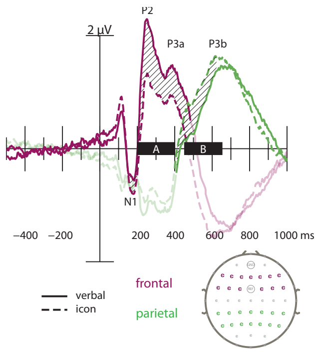

Figure 3: Speech notifications result in larger early neural responses, associated with detection processes, while auditory icons result in larger late responses, associated with context updating.

logical responses when applied without context can often be confounded with interesting but unrelated aspects of a user's state. For example, changes in pupil dilations can indicate a range of user states, from autonomic emotional arousal [2] to attention focus [12] to memory formation [8]. In this example, the claim that changes in pupil dilation relate to increases in "mental workload" is not entirely wrong, even after uninteresting influences are factored out e.g., ambient lighting [15]. However, we would contend that it is not particularly useful.

More recent applications of electroencephalography (EEG) as objective and direct measurements of "mental workload" suffer from similar inferential limitations. Decreased power in alpha oscillations (i.e., 8-12 Hz) in the EEG signal is an established marker for "mental workload" [3, 16, 1]. However, investigations that use functional magnetic resonance imaging to associate brain regions with this broad EEG measure have revealed that alpha oscillations are likely to signal a neural baseline for "inattention". In this light, it is unsurprising that alpha power decreases when the mind is occupied. While it is correct to say that alpha power is a metric for "mental workload", it is similiarly appropriate to question whether it is a useful metric for evaluating human cognition when interacting with novel machine interfaces.

Mental workload may no longer be a useful concept, especially if we are no longer designing computing systems to alleviate workload but rather to facilitate our cognitive processes instead. The increasing sophistication of artificial intelligence motivates this objective of fluent human-machine collaborations. Arguably, computing systems are often designed to perform targeted computational tasks instead of relieving *mental workload* in general. Thus, it could be preferable to determine, in the first place, whether a machine is assisting the user in performing a given task in the first place rather than to ask whether or not it results in less *mental workload*.

Designing for cognitive processes

Here we provide two examples where we first address the cognitive process that a given interface is supposed to target, before we evaluate it for whether or not it does with the use of EEG methods. By identifying and targeting specific cognitive processes that are well-described in psychology ((i.e., discrimination, context-updating, working memory load) and are the aspects that an interface is designed for, we are able to focus the questions that we ask concerning the utility of the system that we are evaluating.

Our first example is drawn from auditory notification displays, specifically in the domain of in-vehicle notifications [7, 4]. There is no clear consensus on what constitutes an optimal auditory notification for an in-vehicle environment. Although meaningful guidelines and recommendations exist [14], it is unclear whether speech commands or auditory icons might be preferable for any given purpose. In a recent project, we evaluated auditory notifications that were explicitly designed for in-vehicle task management [11]. While a previous evaluation suggested that users responded faster to speech commands, which led to the recommendation of their use [5], we found a more subtle distinction between the two sounds. Using EEG/ERP techniques, we found that speed commands evoked larger early brain responses at around 236–304 ms than auditory icons after they were played, while auditory icons evoked larger late brain responses at around 352–468 ms subsequently. This suggests that the different notifications have preferential access to different cognitive mechanisms, as opposed to the simplistic assertion that one notification is more readily detected than another. Indeed, the earlier ERP component (i.e., P2) is typically associated with target discrimination while the later ERP component (i.e., P3b) is typically associated with context-updating [17]. In other words, it is more appropriate to describe speech commands as being more discriminable notifications and auditory icons as being more representative or vivid notifications. Indeed, it was for these reasons that motivated their consideration as suitable notification candidates in the first place.

In a separate example, we describe a recent evaluation that we performed using EEG to validate our belief that *in situ* displays result in less "mental workload" because they reduce the number of items that users have to hold in memory while performing a task [6]. Until this evaluation was performed, *in situ* projections that presented the next assembly piece in concert with the user's activity consistently resulted in faster assembly performance and higher subjective preference, compared to conventional methods (e.g., paper instructions). While we theorized that such systems resulted in less "mental workload" because it provided an interface that assumed the function of working memory, this could be directly verified; Subjective reports of "mental workload" were either inconsistent or simply not attainable given that such systems were developed to assist the cognitively impaired. To verify our design beliefs, we first identified the EEG frequency bandwidth for which gradual attenuations were apparent specifically for when they performed a visuospatial working memory task; this bandwidth varied across individuals about the 10 Hz region. Subsequently, we confirmed that EEG power in the same bandwidth was significantly less attenuated when participants relied the *in situ* display compared to when they did not. This provided converging evidence that an *in situ* display reduced working memory load for visuo-spatial objects.

Conclusion and Outlook

In this position paper, we argue that there is a growing need to address the specific cognitive role of humans with regards to their interactions with computing systems that will be designed to facilitate human work instead of assume it. In view of this trend, we argue that the unitary concept of "mental workload" is unlikely to offer a level of insight that will be effective in guiding interface design or characterizing how we cognitively respond to novel interfaces. To understand how a given interface modifies or supports our cognitive processes it is necessary to have an appreciation for the design motivation of the interfaces themselves, the functional resolution of the methods used to infer cognitive activity, and, above all, a good working definition of the cognitive processes that are being targeted. Two examples were provided from our research to illustrate our methods that relied on EEG. EEG is a useful tool not because it offers a direct measurement for cognitive processes but because it allows us to distinguish between potentially separable cognitive processes, which are suggested by psychological models.

REFERENCES

1. H. Berger. 1930. Über das elektrenkephalogramm des menschen. II. *Journal for Psychology and Neurology (Leipzig)* 40 (1930), 160–179.

2. M.M. Bradley, L. Miccoli, M.A. Escrig, and P.J. Lang. 2008. The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology* 45, 4 (2008), 602–607.

3. A-M. Brouwer, M.A. Hogervorst, J.B.F. Van Erp, T. Heffelaar, P.H. Zimmerman, and R. Oostenveld. 2012. Estimating workload using EEG spectral power and ERPs in the n-back task. *Journal of neural engineering* 9, 4 (2012), 045008.

4. L.L. Chuang, C. Glatz, and S. Krupenia. 2017. Using EEG to understand why behavior to auditory in-vehicle notifications differs across test environments. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, 123–133.
5. J. Fagerlönn, S. Lindberg, and A. Sirkka. 2015. Combined Auditory Warnings For Driving-Related Information. In *Proceedings of the Audio Mostly 2015 on Interaction With Sound (AM '15)*. ACM, New York, NY, USA, Article 11, 5 pages. DOI: http://dx.doi.org/10.1145/2814895.2814924
6. Markus Funk and Albrecht Schmidt. 2015. Cognitive assistance in the workplace. *IEEE Pervasive Computing* 14, 3 (2015), 53–55.
7. C. Glatz, S. Krupenia, H.H. Bülthoff, and L.L. Chuang. 2018. Use the Right Sound for the Right Job: Verbal Commands and Auditory Icons for a Task-Management System Favor Different Information Processes in the Brain. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, xx–xx.
8. S.D. Goldinger and M.H. Papesh. 2012. Pupil dilation reflects the creation and retrieval of memories. *Current Directions in Psychological Science* 21, 2 (2012), 90–95.
9. S.G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 50. 904–908. DOI: http://dx.doi.org/10.1177/154193120605000909
10. Daniel Kahneman. 1973. *Attention and effort*. Vol. 1063. Prentice-Hall Englewood Cliffs, NJ.
11. S. Krupenia, A. Selmarker, J. Fagerlönn, K. Delsing, A. Jansson, B. Sandblad, and C. Grane. 2014. The Methods for Designing Future Autonomous SystemsâĂŹ (MODAS) project: Developing the cab for a highly autonomous truck. In *Proceedings of the 5th International Conference on Applied Human Factors and Ergonomics (AHFE2014)* (Krakow, Poland). 19–23.
12. S. Mathôt and S. Van der Stigchel. 2015. New light on the mindâĂŹs eye: The pupillary light response as active vision. *Current directions in psychological science* 24, 5 (2015), 374–378.
13. Neville Moray. 1979. *Mental workload: Its theory and measurement*. Vol. 8. Springer Science & Business Media.
14. M.A. Nees and B.N. Walker. 2011. Auditory displays for in-vehicle technologies. *Reviews of human factors and ergonomics* 7, 1 (2011), 58–99.
15. B. Pfleging, D.K. Fekety, A. Schmidt, and A.L. Kun. 2016. A model relating pupil diameter to mental workload and lighting conditions. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 5776–5788.
16. G. Pfurtscheller, A. Stancak Jr, and C. Neuper. 1996. Event-related synchronization (ERS) in the alpha bandâĂŤan electrophysiological correlate of cortical idling: a review. *International journal of psychophysiology* 24, 1-2 (1996), 39–46.
17. Terence W Picton. 2010. *Human auditory evoked potentials*. Plural Publishing.