Deepening the Role of the User: Neuro-Physiological Evidence as a Basis for Studying and Improving Search

Javed Mostafa
Info. Science & BRIC
UNC at Chapel Hill, NC, USA
jm@unc.edu

Jacek Gwizdka
School of Information
University of Texas at Austin, TX, USA
chiir2016@gwizdka.com

ABSTRACT
In this paper, the potential for expanding the set of scientific evidence and insights associated with the users' role during the search process is explored. As it is intended to be a position paper and not a systematic survey, a comprehensive review of literature is not presented here. However, the authors draw on some early stage research, in this emerging area, to describe and explain the generation of neuro-physiological evidence using three types of modalities. The modalities and the associated methods described here, presented in order of increasing complexity, include Eye-tracking, EEG, and fMRI. The paper concludes with a few critical observations regarding the promises and perils of using neuro-physiological approaches in studying search and search behavior.

General Terms
Experimentation, Human Factors, Measurement.

Author Keywords
Neuro-physiological methods; Eye-tracking; EEG; fMRI.

1. INTRODUCTION
With the emergence of the CHIIR as a new venue and an important milestone in the continued development of the IR field, it is appropriate to take a reflective view and investigate new potential avenues of research and scholarship that the CHIIR may catalyze and help advance. We argue that the CHIIR marks a deepening of interest in understanding the role and the influence of humans in the search process and therefore it behooves us to bring up some additional areas or topics that have not thus far found appropriate outlets in the mainline IR venues or even in the well-established HCI-oriented forums. One such area is neuro-physiological (NP) methods in explicating the search process, particularly interpreting human-responses to search tasks and the influence of search complexity on search outcomes. In this paper, we begin by broadening the conceptualization of the search process by using humans' psycho-physiological condition as a frame for understanding search. We then use three particular modalities, namely Eye-tracking, EEG, and fMRI, as examples of NP methods to discuss the utility of such methods in elucidating the search process. Then we present an example of application of NP methods. Following this, critical advantages and disadvantages of NP methods are described. Finally, the paper ends with a set of aims for advancing the area of HII&R by integrating NP methods into the field of IR.

CHIIR '16, March 13-17, 2016, Carrboro, NC, USA

DOI: http://dx.doi.org/10.1145/2854946.2854979

2. BACKGROUND
Search has largely been examined and studied in a manner which is agnostic to the human's psycho-physiological condition. That is in most IR studies that engage humans, the particular psychological or physiological condition of the humans are not directly measured and it is not the central focus of those studies. Yet, we must recognize that search is no longer an occasional activity that humans engage in. Humans no longer rely on search only to address specialized or complex information needs. Humans now conduct search frequently, under a variety of circumstances, and often as the first response when faced with an information need. It is therefore not a stretch to claim that search activity is now routine, highly personal, and interweaved with everyday activities of millions of users. A number of factors could be the cause for this change, but it is highly likely that wide-scale availability of computers and relatively affordable mobile devices promote more searching. It is well known that humans go through a wide variety of psycho-physiological conditions as they experience their world in the course of a single day, an hour, or even a few minutes. Hence, the psycho-physiological conditions of users are highly likely to trigger, shape, and influence humans' search behavior and performance. Imagine how a muted response from an otherwise jolly friend can quickly reveal something is off and can become a powerful signal indicating that the friend is experiencing a "down period". Now imagine how the same type of a psychological response, expressed in abnormal typing behavior or uncommon errors, detected by software could become a signal for tracking and understanding the human's condition. There is strong likelihood that specific NP conditions and tracking them may open up a new window to understand humans' behavior as they engage in searching. Given the fact that search has now become a significant activity of everyday behavior, it potentially is an untapped and underleverage source of insights and observation for humans' psycho-physiological condition.

2.1 NP Responses as a Two-way Street
It should be apparent from the discussion above that the psycho-physiological conditions of users both shape and in turn is shaped by their human-computer interaction (HCI) experience. To amplify our understanding of the role of the user in the interaction process, it is important that *the identification of specific NP responses is conducted in isolation from efforts that focus on shaping or changing the NP responses by manipulating the interaction*.

In other words to establish the actual roots and the patterns of NP responses, research on NP methods, must begin by establishing reliable "healthy" or "normal" baseline responses to common search tasks. It is important to start with baseline responses and a carefully selected set of search tasks so that the corresponding NP responses can be predicted accurately and consistently from the search tasks users perform (or typically perform). Such an investigation with the goal of establishing NP baselines may identify common regions in the brain that display neuronal activities as a response to fact-oriented searches (e.g., What is the capital of Colombia?). Or, in order to establish response to higher levels of cognitive load, such a line of NP research may purposefully and systematically increase the search task complexity and derive baseline EEG alpha wave forms and dynamics of pupil dilation from healthy and capable users that indicate the level of user engagement and deliberations necessary to execute complex search tasks.

At this point, we feel it is necessary to introduce a critical concept for explaining the importance of baseline NP patterns. The concept is known as "markers" in biological and clinical research. The concept refers to concrete and robust physiological human characteristics that indicate a specific disease or abnormality. For example, the existence of a mutated form of BRCA1 gene is considered a strong indicator of the likelihood of eventually leading to cancer and hence it is a now a well-established "bio" marker. Similarly, we believe it is possible, upon careful observation and analysis of NP data to establish behavioral correlates or markers that indicate normal or abnormal psycho-physiological conditions. For example, through careful scrutiny of neuronal activation patterns associated with simple fact- oriented searches we could establish a stronger basis for typical behavior and interaction responses of healthy and capable users. The baseline NP patterns would in other words assist in establishing the behavioral correlates (e.g., time to conduct such searches and error rates) as robust and consistent indicators or markers for a specific type of user group and a specific type of search tasks. Such behavioral markers in turn could be used to categorize search behavior more accurately, efficiently, perhaps even dynamically as search is taking place, without the need to rely on NP evidence. Hence, the larger goal of applying NP methods to establish baseline HCI patterns is to gain predictive accuracy and confidence in utilizing behavioral markers (**Figure 1**).

The opposite side of the road in NP research is the potential for shaping and improving HCI as it takes place in searching. For example, as a derivative research, based on a good understanding of NP responses and behavioral markers (from the line of research described above), the interface and the interaction could be designed to be more intelligent and adaptive. If the behavioral markers indicate that the user is significantly off from "normal" responses, the response from the search software could be appropriately attenuated to gently "move" the user to more "normal" behavior or response. For example, the elements of the interface could be adjusted to facilitate improved interaction (e.g., font size or placement of content on the screen). The line of research involving generation of adaptive responses to robust behavioral markers requires a highly precise way to classify interaction signals and detect subtle changes dynamically. Studying such dynamic adaption of interaction and their influence on search performance could be driven by both prior and parallel research on NP patterns, which would increase the confidence in adjustments to be made to particular interface components and their influence on the user's response and performance.

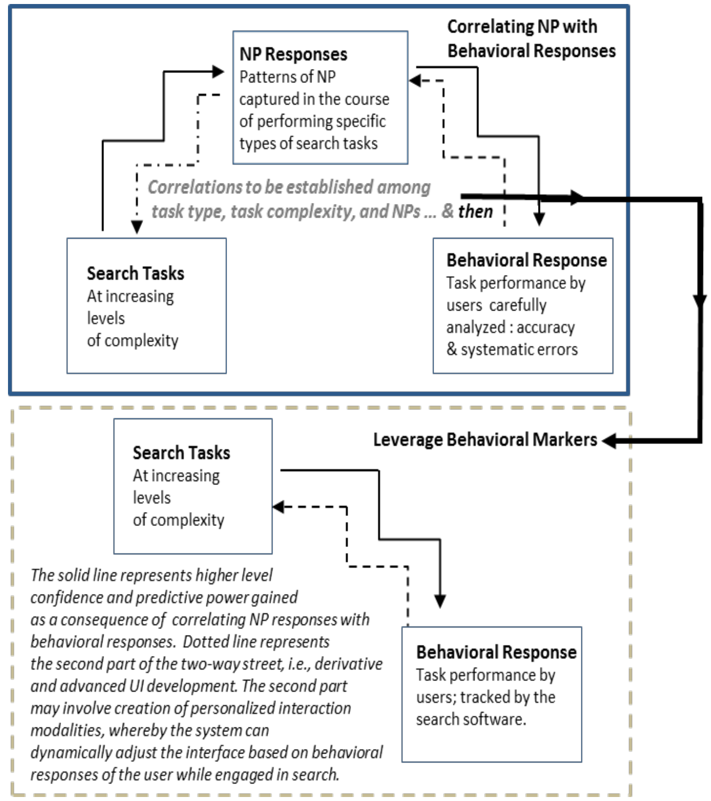

Figure 1. Two-way street of NP evidence: 1) First, correlating NP responses with specific behavioral responses in the context of search, and 2) Second, based on increased confidence in behavioral responses, leveraging them in adapting and improving interactions.

3. NP METHODS FOR STUDYING SEARCH

In this section, we discuss three NP modalities and methods. We chose Eye-tracking, EEG, and fMRI particularly as they have been used more often than other modalities and they represent a gradual increase in the level of complexity of NP methods. It is our intention to only offer an overview of these modalities in this section. We recommend the following [5, 10, 20, 27, 31, 35, 38–40, 47, 52] as relevant resources that scholars in the HII&R community may find useful.

Based on the overviews presented in this section, our goal is to point out and motivate the potential applications of the three modalities in the context of human-information interaction tasks and to deepen our knowledge of NPs and to discuss some of the critical limitations or the pitfalls.

3.1 Eye-tracking

Eye-tracking (ET) is the oldest (dating back to late XIX c., [10, 27]) and one of the most widely used NP modalities in IR research. The essential functioning of ET is based on eye-mind link hypothesis [29] which states that our attention is where our eyes are looking. While this statement needs to be qualified by considering covert attention, that is cases when our attention moves without our eyes moving, for the purpose of this paper, however, we will ignore this phenomenon, since, as explained below, it has little effect on textual IR.

At any point in time human eye can see with full acuity an area spanning less than 2° of visual field (roughly the size of a thumb at arm's length). Therefore to see surrounding environment sharply, we need to move our eyes constantly. To perceive an object (e.g., a word) our eyes remain focused on the object for a short period of time (between tens of milliseconds to seconds) – this is called a *fixation*. Our eyes then move very fast to the next fixation (e.g., the next word) – this rapid eye movement is called a *saccade*. Saccades are very short (30-80ms) and are the fastest movement human body is capable of (30-500° per second).

Eye-tracking hardware and software typically detects fixation events, while faster and more expensive eye-trackers are also able to detect saccade events. Fixations provide us with information where a user's attention is focused on screen, or, more generally, in the environment. This link to attention has been used in HCI and IR research for over two decades [17, 28]. In the context of search eye-tracking informs us about user interaction with a search environment at different stages of the search process (Figure 2).

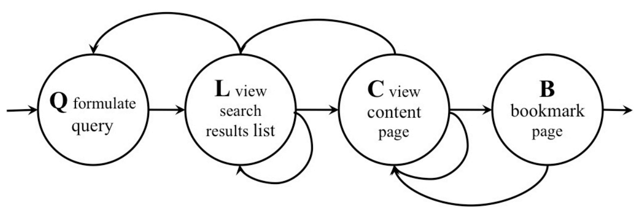

Figure 2. A simplified four-stage search process [21].

At the query formulation and re-formulation stage (Q), we can learn from eye-tracking, for example, whether users pay attention to query suggestions or auto-completion [25, 37]. We can also learn the likely sources of user entered query terms [12]. We can do this by considering whether the entered words were fixated on in query suggestions or in documents. At the search results list review stage (L), we learn which search result surrogates users pay attention to, in which order, and how much time they spend on each of them [9, 17, 19]. At the document viewing stage (C), we learn which elements of a document are being paid attention to and which are not, how carefully a document is read and how much it is just skimmed.

Eye-tracking can also help in understanding the search process complexity. In particular, in reading, which is relatively well understood in traditional settings [46], several indicators of cognitive effort were identified. The effort indicators include: fixation duration, the existence and number of regression fixations in the reading sequence (moving back in the sequence), the spacing of fixations in the sequence, the reading speed. Plausibly higher cognitive effort is taking place when users are acquiring new information, or are integrating that information with existing concepts. In this way, the eye-tracking based cognitive effort indicators may also identify when and how much learning is taking place during the search process.

While examination of areas of visual attention based on fixation duration and their timing provides useful information, it is only one type of information eye-tracking delivers. More advanced uses of eye-tracking data include considering eye movement patterns and pupil dilation.

In basic and applied psychology research, eye movements were used, for example, as a diagnostic tool in detecting dyslexia [16, 43], mental workload [18] and even in biometric identification of a person [26, 32]. More interestingly for research in IR, distinct eye movements were found to be associated with memory recall [6], thus indicating a possibility of distinguishing internal (memory) and external (query suggestions, document text) sources of query terms.

In interactive information retrieval, eye movement patterns have been used in investigating reading patterns on relevant vs. irrelevant documents/web pages [1, 2, 7, 20, 49]. We learn that different patterns are associated with different levels of document relevance (Figure 3). One should note that these patterns may depend on the text layout, which could limit their use as NP responses. More generally, investigation into reading eye movement patterns in the search process has a potential to move beyond the settings traditionally studied by psychology and to bring new knowledge not only to IR but also to cognitive and educational psychology.

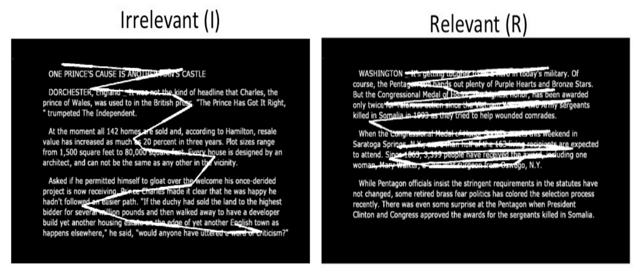

Figure 3. Eye movement reading patterns on irrelevant and relevant document [20].

Pupil dilation is arguably the most intriguing and most under-utilized aspect of the eye-tracking method in application to IR. Pupil diameter is controlled by the Autonomic Nervous System [42]. Under constant illumination, it has been associated with a number of cognitive functions, including mental effort [30], interest [34] surprise [45], and making a decision. The sources of variation in pupil's size are related to attention [24, 53] and, thus, plausibly, it should reflect some aspects of cognitive processing of documents. Indeed, recent results obtained in two experiments (N=24 and N=32) by one of the paper authors demonstrated pupil dilation on relevant and pupil contraction on irrelevant text documents and web pages [20, 23]. This effect is particularly apparent for changes in pupil size in the last couple of seconds before a participant's relevance judgment. Other research have shown similar differences on image stimuli [41]. Pupil dilation on visits to relevant stimuli indicates, in part, a higher mental effort and an increased level of attention paid to relevant information. Pupil dilation analysis is independent of the layout, and thus it presents a good candidate for an NP response signal.

3.2 EEG

Electroencephalography gained popularity in clinical use, particularly to detect abnormal electrical signals that are highly correlated with brain disorders, drug effects, and lesions or brain damages. It is well-known that the brain is an extremely dense structure, consisting of billions of neurons, which interchanges signals at the rate of once every 5 milliseconds (thousandth of a sec). Generally, in order to map the physiological responses of a brain the objective is to pinpoint the location of the signal as well as the strength of the signal. EEG is temporally precise in detecting signal instantiations, and based on modern signal converters the signals can be quickly characterized (or "filtered") into specific wave patterns. The patterns are dependent on number of waves per second and the amplitude or height of each wave. The amplitudes correspond to the strengths of the electric signals and modern

EEGs can detect signals in the range of 1/1 millionth of volt (or micro volt). On the other hand, EEG is far less precise in terms of establishing specific brain locations where signals originate from, as it is placed on the surface of the brain and can detect signals close to the skull level and only in the range of centimeters[1].

The standard or common EEG wave patterns include alpha, beta, theta, and gamma waves. The common wave patterns represent specific frequency and amplitude patterns. There are now several relatively inexpensive EEG tools in the market, specifically aimed at usability and human factors researchers that come with built-in hardware and software capabilities to capture, amplify, and represent all the common EEG wave patterns. The relatively inexpensive EEG systems are wired and have few electrodes. While at the higher end are those that have numerous electrodes (thus offering better spatial resolution), have built-in amplifiers in each electrode, are wireless, and have customized classifiers and visualization software to support easier interpretation of data. Based on the alpha, beta, theta, and gamma waves it is possible to capture and determine brain signals that are un-induced (i.e., no special stimuli or tasks are used) and by comparing these wave patterns to well-known "normative" or "abnormal" patterns one can draw general conclusions regarding the status of the subject's functioning brain.

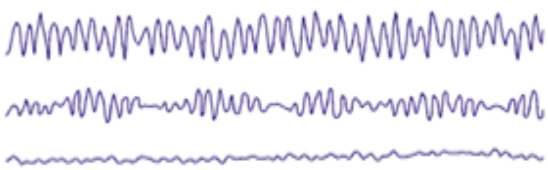

Figure 4. Variations observed in Alpha waves: 1) Top-most pattern representing the characteristic, typical wave form during relaxation, 2) The pattern in the middle is an intermediate stage when the subject is beginning to concentrate, and 3) The last wave, with the almost flat pattern, represents full attenuation while the subject is completely engaged in an activity with open eyes [11].

In contrast to understanding the current status of a functioning brain, a researcher may wish to investigate the response of the brain to specific stimuli or tasks. When a specific NP response is elicited as a consequence of presenting a stimuli or a task, such an NP is called an induced or evoked response. It is possible to rely on the standard alpha, beta, theta, and gamma waves to discern and clarify their association with certain characteristic NP responses and what these responses imply. For example, it is well-known that when a subject is in a relaxed mood and/or has their eyes closed alpha waves are particularly prominent and have high amplitudes, and in contrast when the same subject opens her eyes and concentrates on a demanding task there is attenuation (i.e., flattening) of the alpha waves (see Figure 4). Similarly, it has been shown that numerous occurrences of beta waves in diverse frequencies and with low amplitudes indicate active concentration and anxious states.

There are more precise characteristic response waves that have been demonstrated to strongly correlate with the onset of certain specific stimuli and these characteristic NP responses have been utilized in a wide variety of psycho-social studies. These responses are termed event-related potentials (ERPs) [36]. For example, the occurrence of P300 wave patterns[2] has been linked to categorization and evaluation tasks, and particularly strongly associated in execution of odd-ball tasks involving detection of low-probability occurrences of stimuli. Similarly, the N400 wave patterns have shown to be prominently associated with analysis and processing of meaningful stimuli such as words, icons, faces, etc. Given the potential for generating human-responses such as concentration, categorization, and processing of symbolic and semantic stimuli such as words, it should be apparent to the reader by now that there are obvious opportunities to apply the EEG modality exclusively or complementarily in establishing a clearer understanding of the search process. We find examples of early applications of the EEG modality in IR in recently published papers [3, 13]. However, there remain certain critical limitations associated with the EEG modality that the human-information interaction researcher must be aware of and we will discuss some of these limitations in our concluding section.

3.3 fMRI

During brain's engagement in support of executing tasks, blood flow to neurons in the specialized areas of the brain relevant to those tasks increases and correspondingly the oxygenation level in those specialized areas also increases. The powerful magnets in the MR machines are capable of detecting the subtle changes in the blood oxygenation levels. Thus, the measurement of neurological response based on the differentiated oxygenation levels is called blood-oxygenation level dependent measure or BOLD. A synonymous term also used to refer to changes in the blood oxygenation levels is known as the hemodynamic response (HR). In contrast to the EEG, the mapping of active sites or locations based on BOLD is highly precise, and it is typically captured as voxels (volumetric pixels) at the range of millimeter (one tenth of a centimeter). In contrast, however, the temporal resolution measured using BOLD is rather coarse – typically in the range of few seconds.

Similar to the process of measuring the impact of evoked responses in EEG, while inside an MR machine a subject is presented with a stimuli or asked to conduct a simple task. In order to enhance the effect of the task and gain clear evidence of the impact, subjects are typically requested to conduct the same task or the same type of task repeatedly while inside the MR machine. The tasks are usually interspersed with short durations of "non-activities" or "resting" periods. An experimental session which utilizes repeated exposure to similar tasks with boundaries defined by resting periods is called an event-based session (or event-based design). To avoid the potential impact of the subject gradually learning from the repeated exposure to the same tasks, the session usually involves conducting a fixed of set of different types of tasks that are randomly interspersed in the even-based design.

In a recent pilot experimental session, conducted by one of the authors of this paper, an event-based design was employed which involved two types of search result presentations: high precision (i.e., relevant search results were placed in high rank order) and low-precision. Subjects, while inside the MR machine, were shown a search question, for example, What is the capital of Japan? It was followed by a display one of two types of result-screens. Subjects were then requested to identify one relevant result from the screen by pressing a button representing the rank order of the relevant

[1] For a comparison, consider the fact that a square millimeter (one tenth of a centimeter) contains approximately 100,000 neurons.

[2] The numeric designator 300 represents a latency factor in milliseconds, post introduction of stimuli.

item in the result-screen, where each result item was labeled with the corresponding rank order e.g., 1, 2, 3. Each subject was requested to perform 80 scan tasks, divided into blocks of 16, with 5 resting boundary intervals separating each block of tasks. During the resting intervals subjects were presented with a screen containing only a "+" sign. Every task took about half a minute and the intervals were about 1 second long. To minimize the impact of the learning-effect, the scan tasks were randomly interspersed in the event-based session. The hemodynamic responses or BOLD measures were collected from all the subjects (n=12), and they were pooled, averaged, and mapped to the surface area of the brain in order to facilitate interpretation of the resulting neurological response. As can be seen in the Figure 5, the neurological impact of scanning for relevant items is clearly differentiated from the resting modes (also known as fixation modes).

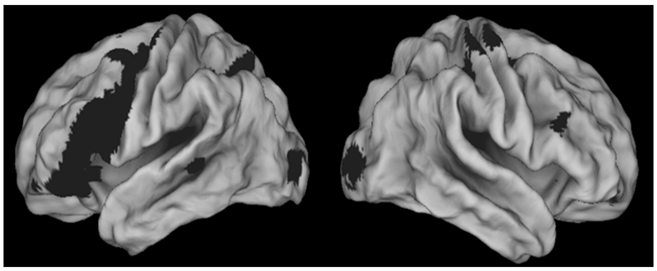

Figure 5. Brain-surface area mapping of fMRI BOLD signals collected from 12 subjects contrasting fixation with the search result scanning. The left and right lateral views of the brain surfaces are shown. The large dark patches on the brain surfaces represent those areas that were active during scanning modes but NOT active during the resting or fixation modes.

A fulsome description of the fMRI pilot study is out-of-scope for this position paper. Additional details on the experimental design, results, and analysis can be found in a recently published paper [40]. Here, we only provide a small glimpse of the fMRI results to point to the potential application of fMRI for generating NP evidence for investigating the impact of search related activities on the brain. As can be seen in Figure 5, a highly active area while the scanning tasks were being completed is the prefrontal cortex (PFC). It is known that the lateral PFC is generally associated with working memory and decision making; particularly, in the PFC the ventral regions are associated with working memory and the dorsal regions are associated with decision making process. Given that the scanning task is strongly verbal in nature, which involves reading and interpreting the search question and the search results, it is not a surprise to find the dominant role of various PFC regions while subjects execute the scanning tasks.

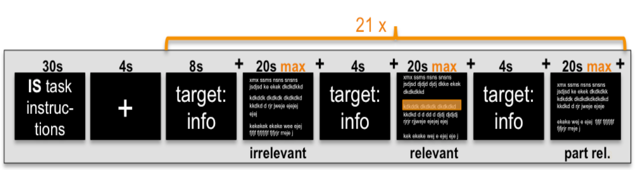

Figure 6. Experimental Design of the second pilot study.

In another pilot fMRI experiment conducted by one of the paper authors, a difference in brain activations between reading relevant and not-relevant short text documents was investigated [22]. The fMRI environment and the essentials of this experimental design bear many similarities with the first study described above. The experimental design was also event-based, but in this case it consisted of 21 trials that contained irrelevant, partially relevant and topical text documents in randomly changing order (Figure 6).

The findings from this experiment indicated that reading and deciding on relevance of relevant vs. not-relevant documents was characterized by increased activity in middle front gyrus, an area of frontal lobe that has been generally associated with executive mechanisms and decision-making. This finding also comes with little surprise, since judging relevance certainly involves areas of brain responsible for decision-making. While this early results shed some light on cognitive processes engaged in relevance judgments, we are just at the beginning of this new research territory.

In order to establish more precisely the contributions of specific brain regions to execute search tasks, it would be necessary to examine full search cycles: starting with a question, to query formulation, to execution, to revision, and finally to interpretation stages. As far as we know, no one has examined full search cycles inside fMRI machines. Additionally, of course, the association of search types (i.e., tasks) and search difficulty with specific brain regions also remain unexamined. To be sure, we must note here, the challenges associated with applying fMRI as a tool for generating NP evidence are immense and some are potentially impossible to overcome. We will discuss some of the key challenges associated fMRI and the other two modalities in the next section.

4. PROMISES AND PERILS OF NP EVIDENCE

In this section, we will discuss some fundamental advantages and disadvantages associated with NP-based methods in investigating search interactions.

4.1 Balancing the Good with the Bad

When researchers study search behavior, either retroactively by scrutinizing search log data or proactively by collecting data from search performance, the analysis is always strongly dependent on the context of the search (e.g., task scenario, time allocated, and even location of experiments). It is also well-known that search experiments that involve collecting self-report data face the risk that the nature of interactions or relationship between the experimenter and the searcher may impact the quality of the experimental data generated (i.e., the risk of response bias). Thus, any insights to be gained from such behavior-oriented experiments regarding critical cognitive and psychological factors, causal or associated, must always be achieved second-hand, based on assumptions or stipulations regarding what the data represents about the user's actual mental condition or physical status.

Neuro-physiological signals as indicators of human engagement and responses to search tasks comes with the great promise of "lifting the lid" of the black-box that is the searcher's state-of-mind or his/her physical condition. If NP data are collected carefully and with appropriate level of control, while subjects engage in search tasks, it may reveal a wide array of additional clues regarding factors such as level of cognitive effort (e.g., load), type of effort (e.g., verbal vs. motor planning), anxiety, stress, curiosity, arousal, and even pleasure. We contend that a particularly powerful approach to amplify and clarify the clues regarding searching is to collect and correlate behavioral data with NP data. Such an approach may lead to more precise prediction of search activities and the quality of search performance, based only on monitoring behavioral responses during the search process[3]. Last but not least, NP approaches open up the possibility of gathering evidence that are not directly expressed (or even expressible) by the searcher. Thus, it may be possible via an NP-driven approach to unearth deeper and more nuanced insights regarding a subject's response to a search task and arguably such evidence is immune from being tainted by response bias.

We would, however, like to end this paper with a strong note of caution regarding problems associated with NP methods. First and foremost, it should be amply clear by now that NP approaches may be extremely intrusive, and while they may reduce or eliminate one type of problem (e.g., contamination of data due to response bias), at the same time they may introduce new type of problems. How a researcher should set up an experiment to elicit useful NP evidence while preserving realism of the search context remains a great barrier. Applying NP approaches may involve transforming the search context to such a degree that it becomes an artificial or unrealistic condition, thus compromising external validity and generalizability of the findings. Second, a big obstacle is the wide variety of standards and metrics that NP tools employ in categorizing and reporting the critical interaction data. For example, we know fixation durations and pupil dilations are important parameters that are employed in measuring eye movement activities during interactions, however, much more evidence is needed as to what they actually represent in terms of cognitive processing and mental load associated with specific types of search tasks. The lack of clarity regarding these metrics remains, in large part, due to the opacity and obstacles researchers face when they seek information from NP tool vendors. Third, although one may be able to pinpoint a specific a physiological location associated with an NP signal, it may not actually translate to a clear understanding of what that signal means. The challenge in associating a specific function to a location is particularly relevant to the brain. Researchers must be cautious to avoid reverse inferencing [44]; that is, after already being aware of functions that are associated with specific brain locations, based on evidence collected by past studies, the inclination is to attribute the causality of, or the association with specific triggering factors in the new experimental condition (i.e., specific stages or activities of a search task) to those brain regions. Given that a brain region is actually capable of playing different roles for different human activities, additional research is needed to identify the functions those regions dominantly or consistently represent and how they contribute to search tasks. It has to be emphasized here that functional mapping of the brain is still an emerging area, and, particularly, the knowledge of connectivity among brain regions and the information exchange among specific regions is just beginning to be established. Therefore, establishing contributions of specific brain regions to a high-level cognitive activity such as search needs to be approached extremely cautiously. Fourth and the final point we would like to make is regarding the parameters and settings associated with instruments, both software and hardware, used during NP experiments and the lack of a tradition in sharing key details regarding these parameters. Utilizing each NP instrument and the software demands manipulation and appropriate initialization of a large number of parameters. The omission of details along with published results deters reproducibility, makes validation difficult, and may delay further advances in the field.

[3] We referred to such behavioral indicators as "markers" earlier in this article.

5. AN ILLUSTRATIVE EXAMPLE

One example will serve as an illustration of how one might move from considering information retrieval phenomena to the space of NP concepts and then to NP markers. Historically, dwell time has been suggested as a good indicator of user interest in a document and thus of its relevance [8]. Subsequent research showed that the length of time a user spends viewing a document does not correlate significantly with the user's explicit relevance assessments [33]. Ambiguity of dwell time comes from its association with user interest in information as well as with problems encountered by users in processing this information. Total dwell time on a document includes periods of actual reading and processing a document and periods of not attending to the document. We translate interest into attention given to a document and possible problems associated with reading difficulty or lack of attention (e.g., mind wandering) and model dwell time on a document as composed of several critical components,

$$t_d = t_{tr} + t_{dr} - t_{mw} + \varepsilon$$

where t_d is dwell time on a document; t_{tr} is typical reading time; t_{dr} is additional time required for reading of difficult text; t_{mw} is time spent on mind wandering; and, ε is error term.

At a simple level, attention can be assessed by considering time spent on reading. This metric can be obtained from eye-tracking as total duration of eye fixations on text. Prior research showed that using eye-tracking measures improves document relevance classification as compared to using dwell time alone (Gwizdka 2014). Let us now consider text difficulty (readability) and word frequency. These metrics can help to account for individual differences in reading time between users as well as for differences between documents. We can further improve measurement of dwell time components by detecting and subtracting periods of mind wandering ("daydreaming"). Mind wandering is defined as decoupling attention from perception and from immediate task. During reading, when the mind wanders to unrelated feelings and thoughts, the eyes continue to scan the words but without close attention to their meaning. The periods of mindless reading that result are characterized by longer fixations and reduced sensitivity to lexical features [48, 51]. Application of these measures requires building a model of text, making it somewhat difficult to apply the measures. However, a few other NP markers, can be applied without considering text features. Larger pupil dilation (PD) was found as a NP marker of mind-wandering while reading [14]. Other NP markers of potential interest include spectral features of EEG signal and ERPs. For example, mind wandering was shown to generally reduce brain signal responses to perceptual events (within 100ms) and to task related events (the aforementioned P300 wave patterns) [50]. Furthermore, mind wandering was found to be characterized by a power amplitude increase in the theta and delta frequency bands and an amplitude decrease in the alpha and beta frequency bands. The alpha-peak frequency increase was found to be a likely marker of the attentional switch between mind wandering and the focused task [4]. These NP markers may be potentially helpful in developing a "cleaner" measure of dwell time on documents. IR researchers who plan to take advantage of NP metrics presented in the example here should take into consideration some key factors associated with them. While some NP markers were obtained based on search activities (e.g., PD during reading) and are thus directly applicable, others such as the EEG metrics were obtained in different contexts (e.g., based on simple visual and auditory tasks). This points out the difficulty of translating research findings from cognitive neuroscience to IR and the need for further studies that exclusively concentrate on NP metrics associated with search tasks .

6. CONCLUSION

Despite the challenges, NP approaches hold immense promise in advancing human-information interaction scholarship. Beyond IR, NP approaches are finding wider usage in domains such as learning, entertainment, and health care. Some of the press releases and video announcements describing the amplification of human capacities can be quite seductive (e.g., children playing computer games by wearing EEG skull caps). Generally, we do hope that human-information interaction researchers would look beyond IR and would investigate and learn from a wide variety of interaction modalities in related domains. Such a broad approach in studying interaction, however, should be conducted with some caution, as the contexts of user engagement, their roles, and the goals can vary greatly.

In this paper, we focused exclusively on investigating the prospects of pursing NP approaches as they pertain to information retrieval. We end this position paper with a list of goals as a way to integrate NP approaches into mainstream research and methods in HII that focus on search. First, researchers need to concentrate on establishing standardized search tasks. These search tasks need to be carefully established in terms of individual steps and realistic contexts. Second, there is a strong need to evaluate the standardized search tasks and generate baseline NP data. The latter line of research could be driven with the goal of creating a solid foundation of behavioral markers and their NP correlates. Third, whenever possible, the NP outcomes and their interpretations (e.g., cognitive load or stress) should be cross-validated with more than one NP modalities. For example, EEG can be combined with Eye-tracking while the subjects are engaged in the same set of tasks [15]. Fourth and the final recommendation is to conduct NP-oriented research as transparently and as openly as possible by sharing details regarding the environmental contexts such as software and device settings and other critical details regarding the experimental conditions.

7. REFERENCES

[1] Ajanki, A. et al. 2009. Can eyes reveal interest? Implicit queries from gaze patterns. *User Modeling and User-Adapted Interaction*. 19, 4 (2009), 307–339.

[2] Ajanki, A. 2013. Inference of relevance for proactive information retrieval. (2013).

[3] Allegretti, M. et al. 2015. When Relevance Judgement is Happening?: An EEG-based Study. *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval* (New York, NY, USA, 2015), 719–722.

[4] Braboszcz, C. and Delorme, A. 2011. Lost in thoughts: Neural markers of low alertness during mind wandering. *NeuroImage*. 54, 4 (Feb. 2011), 3040–3047.

[5] Bui, A.A.T. and Taira, R.K. 2009. *Medical Imaging Informatics*. Springer Science & Business Media.

[6] Bulling, A. and Roggen, D. 2011. Recognition of visual memory recall processes using eye movement analysis. *Proceedings of the 13th international conference on Ubiquitous computing* (New York, NY, USA, 2011), 455–464.

[7] Buscher, G. et al. 2012. Attentive documents: Eye tracking as implicit feedback for information retrieval and beyond. *ACM Trans. Interact. Intell. Syst.* 1, 2 (2012), 9:1–9:30.

[8] Claypool, M. et al. 2001. Implicit Interest Indicators. *Proceedings of the 6th International Conference on Intelligent User Interfaces* (New York, NY, USA, 2001), 33–40.

[9] Cutrell, E. and Guan, Z. 2007. What are you looking for?: an eye-tracking study of information usage in web search. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2007), 407–416.

[10] Duchowski, A.T. 2007. *Eye Tracking Methodology: Theory & Practice*. Springer-Verlag.

[11] EEG: Instructions (by Carl Rosenkilde): *http://www.rosenkilde.com/EEG--Instructions.html*. Accessed: 2015-09-05.

[12] Eickhoff, C. et al. 2015. An Eye-Tracking Study of Query Reformulation. *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval* (New York, NY, USA, 2015), 13–22.

[13] Eugster, M.J.A. et al. 2014. Predicting Term-relevance from Brain Signals. *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval* (New York, NY, USA, 2014), 425–434.

[14] Franklin, M.S. et al. 2013. Window to the wandering mind: Pupillometry of spontaneous thought while reading. *The Quarterly Journal of Experimental Psychology*. 66, 12 (Dec. 2013), 2289–2294.

[15] Frey, A. et al. 2013. Decision-making in information seeking on texts: an Eye-Fixation-Related Potentials investigation. *Frontiers in Systems Neuroscience*. 7, 39 (2013).

[16] Gomathi, P.M. and Nasira, G.M. 2014. Detection Of Dyslexia From Eye Movements Using Anfis & Bbwpe Feature Extraction Methods. *International Review on Computers and Software (IRECOS)*. 9, 3 (Mar. 2014), 526–532.

[17] Granka, L.A. et al. 2004. Eye-tracking analysis of user behavior in WWW search. *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval* (New York, NY, USA, 2004), 478–479.

[18] Greef, T. et al. 2009. Eye Movement as Indicators of Mental Workload to Trigger Adaptive Automation. *Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience*. D.D. Schmorrow et al., eds. Springer Berlin Heidelberg. 219–228.

[19] Guan, Z. and Cutrell, E. 2007. An eye tracking study of the effect of target rank on web search. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2007), 417–420.

[20] Gwizdka, J. 2014. Characterizing Relevance with Eye-tracking Measures. *Proceedings of the 5th Information Interaction in Context Symposium* (New York, NY, USA, 2014), 58–67.

[21] Gwizdka, J. 2010. Distribution of cognitive load in Web search. *Journal of the American Society for Information Science and Technology*. 61, 11 (2010), 2167–2187.

[22] Gwizdka, J. 2013. Looking for Information Relevance In the Brain. *Gmunden Retreat on NeuroIS 2013* (Gmunden, Austria, Jun. 2013), 14.

[23] Gwizdka, J. and Zhang, Y. 2015. Towards Inferring Web Page Relevance – An Eye-Tracking Study. *Proceedings of iConference'2015* (Mar. 2015), 5.

[24] Hoeks, B. and Levelt, W.J.M. 1993. Pupillary dilation as a measure of attention: a quantitative system analysis. *Behavior Research Methods, Instruments, & Computers*. 25, 1 (Mar. 1993), 16–26.

[25] Hofmann, K. et al. 2014. An Eye-tracking Study of User Interactions with Query Auto Completion. *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management* (New York, NY, USA, 2014), 549–558.

[26] Holland, C. and Komogortsev, O.V. 2011. Biometric identification via eye movement scanpaths in reading. *2011 International Joint Conference on Biometrics (IJCB)* (Oct. 2011), 1 –8.

[27] Holmqvist, K. et al. 2011. *Eye Tracking: A comprehensive guide to methods and measures*. Oxford University Press.

[28] Jacob, R.J.K. 1991. The use of eye movements in human-computer interaction techniques: what you look at is what you get. *ACM Trans. Inf. Syst.* 9, 2 (Apr. 1991), 152–169.

[29] Just, M.A. and Carpenter, P.A. 1980. A theory of reading: From eye fixations to comprehension. *Psychological Review*. 87, 4 (1980), 329–354.

[30] Kahneman, D. and Beatty, J. 1966. Pupil Diameter and Load on Memory. *Science*. 154, 3756 (Dec. 1966), 1583–1585.

[31] Kandel, E. 2013. *Principles of Neural Science, Fifth Edition*. McGraw Hill Professional.

[32] Kasprowski, P. and Ober, J. 2004. Eye Movements in Biometrics. *Biometric Authentication*. D. Maltoni and A. Jain, eds. Springer Berlin / Heidelberg. 248–258.

[33] Kelly, D. and Belkin, N.J. 2004. Display Time As Implicit Feedback: Understanding Task Effects. *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (New York, NY, USA, 2004), 377–384.

[34] Krugman, H.E. 1964. Some applications of pupil measurement. *JMR, Journal of Marketing Research (pre-1986)*. 1, 000004 (Nov. 1964), 15.

[35] Luck, S. 2005. *An introduction to the event-related potential technique*. MIT Press.

[36] Makeig, S. 2009. Electrophysiology: EEG and ERP Analysis. *Encyclopedia of Neuroscience*. Academic Press. 879–882.

[37] Mitra, B. et al. 2014. On User Interactions with Query Auto-completion. *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval* (New York, NY, USA, 2014), 1055–1058.

[38] Moshfeghi, Y. et al. 2013. Understanding Relevance: An fMRI Study. *Advances in Information Retrieval*. P. Serdyukov et al., eds. Springer Berlin Heidelberg. 14–25.

[39] Moshfeghi, Y. and Jose, J.M. 2013. An effective implicit relevance feedback technique using affective, physiological and behavioural features. *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval* (New York, NY, USA, 2013), 133–142.

[40] Mostafa, J. et al. 2015. Identifying Neurological Patterns Associated with Information Seeking: A Pilot fMRI Study. *Information Systems and Neuroscience*. F.D. Davis et al., eds. Springer International Publishing. 167–173.

[41] Oliveira, F.T.P. et al. 2009. Discriminating the relevance of web search results with measures of pupil size. *Proceedings of the 27th international conference on Human factors in computing systems* (Boston, MA, USA, 2009), 2209–2212.

[42] Onorati, F. et al. 2013. Characterization of affective states by pupillary dynamics and autonomic correlates. *Frontiers in Neuroengineering*. 6, (2013), 9.

[43] Pavlidis, G.T. 1985. Eye Movements in Dyslexia Their Diagnostic Significance. *Journal of Learning Disabilities*. 18, 1 (Jan. 1985), 42–50.

[44] Poldrack, R.A. 2006. Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*. 10, 2 (Feb. 2006), 59–63.

[45] Preuschoff, K. et al. 2011. Pupil dilation signals surprise: evidence for noradrenaline's role in decision making. *Frontiers in Decision Neuroscience*. 5, (2011), 115.

[46] Rayner, K. et al. 2011. *Psychology of Reading*. Psychology Press.

[47] Regan, D. 1989. *Human brain electrophysiology: evoked potentials and evoked magnetic fields in science and medicine*. Elsevier.

[48] Reichle, E.D. et al. 2010. Eye Movements During Mindless Reading. *Psychological Science*. 21, 9 (2010), 1300 –1310.

[49] Salojärvi, J. et al. 2005. Implicit Relevance Feedback from Eye Movements. *Artificial Neural Networks: Biological Inspirations – ICANN 2005*. W. Duch et al., eds. Springer Berlin Heidelberg. 513–518.

[50] Smallwood, J. et al. 2007. Going AWOL in the Brain: Mind Wandering Reduces Cortical Analysis of External Events. *Journal of Cognitive Neuroscience*. 20, 3 (Nov. 2007), 458–469.

[51] Smallwood, J. 2011. Mind-wandering While Reading: Attentional Decoupling, Mindless Reading and the Cascade Model of Inattention. *Language and Linguistics Compass*. 5, 2 (2011), 63–77.

[52] Ward, J. 2015. *The Student's Guide to Cognitive Neuroscience*. Psychology Press.

[53] Wierda, S.M. et al. 2012. Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution. *Proceedings of the National Academy of Sciences*. 109, 22 (May 2012), 8456–8460.