Characterizing Information Seeking Processes with Multiple Physiological Signals

Kaixin Ji
RMIT University
Melbourne, Australia
kaixin.ji@student.rmit.edu.au

Danula Hettiachchi
RMIT University
Melbourne, Australia
danula.hettiachchi@rmit.edu.au

Flora D. Salim
The University of New South Wales
Sydney, Australia
flora.salim@unsw.edu.au

Falk Scholer
RMIT University
Melbourne, Australia
falk.scholer@rmit.edu.au

Damiano Spina
RMIT University
Melbourne, Australia
damiano.spina@rmit.edu.au

ABSTRACT

Information access systems are getting complex, and our understanding of user behavior during information seeking processes is mainly drawn from qualitative studies, such as observational studies or surveys. Leveraging the advances in sensing technologies, our study aims to characterize user behaviors with physiological signals, particularly in relation to cognitive load, affective arousal, and valence. We conduct a controlled lab study with 24 participants, and collect data including Electrodermal Activities, Photoplethysmogram, Electroencephalogram, and Pupillary Responses. This study examines informational search with four stages: the realization of Information Need (IN), Query Formulation (QF), Query Submission (QS), and Relevance Judgment (RJ). We also include different interaction modalities to represent modern systems, e.g., QS by text-typing or verbalizing, and RJ with text or audio information. We analyze the physiological signals across these stages and report outcomes of pairwise non-parametric repeated-measure statistical tests. The results show that participants experience significantly higher cognitive loads at IN with a subtle increase in alertness, while QF requires higher attention. QS involves demanding cognitive loads than QF. Affective responses are more pronounced at RJ than QS or IN, suggesting greater interest and engagement as knowledge gaps are resolved. To the best of our knowledge, this is the first study that explores user behaviors in a search process employing a more nuanced quantitative analysis of physiological signals. Our findings offer valuable insights into user behavior and emotional responses in information seeking processes. We believe our proposed methodology can inform the characterization of more complex processes such as conversational information seeking.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in ubiquitous and mobile computing**; • **Information systems → Users and interactive retrieval**.

KEYWORDS

information seeking; physiological signals; user studies

1 INTRODUCTION

One of the core concepts studied in Interactive information retrieval (IIR) is the continuous [65], problem-solving [8, 13, 35] process around information. Over the decades, theoretical models [8, 35, 42, 56, 66, 74] have attempted to characterize the interactions between users (searchers) and (search) systems from different perspectives. As outlined by Cole [13], common search system is "command-based" [74], which assumes that users already know what they are looking for ("known answer") and provide specific requests ("command") accordingly, rather than descriptive questions with "unknown answers". To come up with this search request, Taylor theorizes that information seeking is a process that transfers from the latter to the former [74]; in other words, digging deeper to uncover a more visceral level of need. This is similar to Kuhlthau's proposition that the cognitive state shifts from vague and ambiguous to clear and focused [35]. But both models convey the process as the interchange between affective and cognitive states, driven by the *feeling of uncertainty* and subsequently reacting with physical actions. Likewise, Nahl [56] narrates the exchanges among affect, cognition, and physical actions but emphasizes the role of appraisal as the drive. Overall, a search begins when users realize their inability or insufficiency of knowledge to solve a problem, prompting them to use a search engine. Each search session may contain multiple iterations of entering and executing queries, assessing search results, and evaluating the information quality. If users are unsatisfied with the collected information, they may reformulate the query and start another iteration [35, 42, 73].

Theoretical models have traditionally been formulated based on qualitative methods, such as observational studies and surveys, or facial expression analysis, e.g., see [4, 35, 39, 46]. By examining behavioral data and self-ratings, Gwizdka [21] reports that the

distribution of mental demand (*cognitive load*) varies across different search stages. These observational approaches have limited ability to capture the real situation at a detailed level [39]. Some affective activities happen but are not strong enough to be perceived by humans [68]. This might cause most experiments that rely on observations to find neutral affect as the most frequent during search interaction [39, 46]. The advancement of physiological sensors presents another opportunity to revisit and refine existing theoretical models [40]. In information searching or browsing, wearable sensors have been employed to detect user's interests [77], satisfaction [79], and engagement [16]. It has also been shown that sensor data can indicate affective appraisal (i.e., the continuous interplay between emotions and body perception of surroundings [14, 68]) in reading comprehension, for example, inferring a sense of preparedness, confidence, and activation of background knowledge when beginning reading [14].

This paper aims to validate and summarize human factors in theoretical information-seeking models and existing results. We revisit some of the phenomena observed in the literature by considering the use of physiological data. The physiological data are captured by wearable sensors, including Electrodermal Activity (EDA), Photoplethysmogram (PPG), Electroencephalogram (EEG), and Pupil Dilation (PD). Due to the complex nature of information activities and the sensitivity of physiological sensors [28], we conduct a highly controlled lab study to eliminate confounding variables as much as possible. We carefully scrutinize the study materials to minimize the influences of attitudes (relating to cognitive bias) and relevance. Our experimental design is inspired by the experiment by Moshfeghi and Pollick [54]. The novel hypotheses that we formalize and explore in this work are built upon the synthesis of established theoretical models and existing empirical results. This study focuses on four search stages in a single iteration: the realization of Information Need (IN), Query Formulation (QF), Query Submission (QS), and Relevance Judgment (RJ). Further, to account for diverse text- and voice-based systems, we include study conditions around different modalities of presenting and receiving information. In particular, a system receives queries or presents information in text or audio. Although QF and QS are usually consecutive stages in real-world scenarios, the literature suggests that their underlying activities diverge (discussed later in Section 2), especially when considering the impact of interaction modalities [29]. Hence, we treat them as separate stages in this study.

Overall, our results show that IN encounters higher cognitive loads and alertness, suggesting the update of knowledge gaps, than QF. And QF requires less cognitive demand but enhanced affective feelings than QS. Our study also observes more pronounced affective feelings at RJ. This reaction may be linked to the resolution of knowledge gaps, leading to increased interest and engagement. This study complements the understanding of cognitive activities and affective responses during information seeking by offering a detailed perspective with physiological signals. To the best of our knowledge, this is the first study in IIR to collect and analyze multimodal physiological data during interactive information search. The main contributions of our work are three-fold:

- Through a comprehensive analysis of literature in the areas of IIR, cognitive science, and affective and wearable computing, we formalize a novel set of hypotheses that allow us to study how search stages can be characterized with physiological signals.
- Our proposed controlled lab study design, allowed us to validate (either fully or partially) some of the hypotheses, while also obtaining insights into the rejected ones. This complements our existing knowledge of the role of cognitive and affective activities during the search stages of an information seeking process.
- Our study fills the gap of employing physiological wearable sensors in IIR. It can serve as a groundwork for future experiments using physiological sensors to characterize more complex search processes such as conversational information seeking with Large Language-based systems.

2 LITERATURE REVIEW & HYPOTHESES

Information-seeking models have been extensively studied in the field of information retrieval [8, 35, 42]. Although some work has aimed to understand the different search stages from a cognitive and affective point of view [3, 21, 23, 46, 54], little work has been done to characterize search processes with physiological signals captured from wearable devices [4, 71]. In this section, we draw attention to theories and findings that exist at the intersection of interactive information retrieval, cognitive science, and affective and wearable computing. Following the recommendation by Riedl et al. [64] to assure methodological rigor, we identify the hypotheses in terms of three low-level physiological constructs: *cognitive load*, *affective arousal*, and *affective valence*, and aim to validate them using the quantifiable physiological signals.

We start by summarizing how search stages are conceptualized by information-seeking models in Section 2.1. Section 2.2 details how the cognitive and affective activities in these stages have been studied in the literature and accordingly defines our hypotheses. Finally, Section 2.3 discusses how physiological signals and the derived indexes can be used to characterize cognitive load, affective arousal, and affective valence.

2.1 Information Seeking Models

Several information-seeking models have been proposed in the literature [8, 13, 35, 42]. Similarly to Moshfeghi and Pollick [54], we characterize the informational process with a sequence of search stages that reflect a consensus among these models: Realization of Information Need (IN), Query Formulation (QF), Query Submission (QS), and Relevance Judgment (RJ).[1] Note that this theoretical framework presented here is our adaptation of a handful of former models that, we view, were incomplete. Hence, it requires an amalgamation of former theories and unities, as shown in Figure 1.

Realization of Information Need (IN). See 'Stage 1' with blue-colored borders presented in Figure 1. Users start with a 'vague' idea of the problem and gradually gain clarity [35, 52]. Once information from external sources, such as visual or auditory channels, has been processed and understood [2, 54], the next step involves retrieving relevant information from long-term memory, e.g., past experiences, learned concepts, and memories [47, 55], to articulate any knowledge gaps or informational needs [8, 13, 68, 73] (Stage 1.2

---

[1]Satisfaction Judgment is not considered in this paper; to reduce complexity and possible confounding variables, we only use a single result item during the RJ process (rather than reading a SERP that presents a ranking of items, or a session).
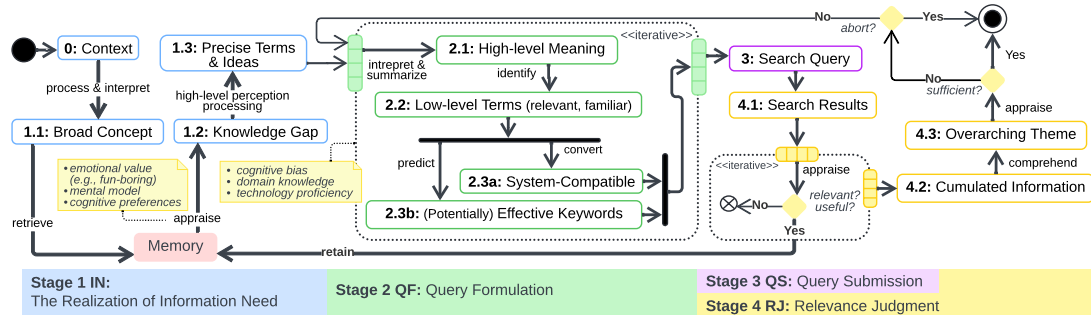
Figure 1: The flow chart presents how the information is transformed through search stages, 1) the realization of Information Need (IN), 2) Query Formulation (QF), 3) Query Submission (QS) and 4) Relevance Judgment (RJ), in information seeking process, based on the combination and unification of the previous models.

in Figure 1). Awareness is updated based on memory output [47, 55]. This is followed by high-level conceptualization [8, 55, 56, 73] to refine the broad concepts into more specific and detailed terms and ideas [13, 56, 73] (Stage 1.3).

The outcome is a comprehensive framework that connects the specific details of an information need to a more extensive network of knowledge. This network includes background and contextual information along with related concepts. Cole [13] refers to this framework as the "Information Need Frame" or "broad focus" as described by Moshfeghi and Pollick [55]. However, Cole [13] also claims that the information need developed so far only scratches the surface. The deeper level requires several iterations of collection and refinement. Mental models and cognitive preferences [13, 56, 73] might also steer the process, as in personalized understanding (e.g., filtering) and representation (e.g., organizing and structuring) of knowledge, and (emotional) value judgment[2] [68].

Query Formulation (QF). See 'Stage 2' with green-colored borders presented in Figure 1. Once the goal is clear, the initiative shifts from reactive (receiving information) to proactive (resolving uncertainty) [68]. The desired outcome here is a plan of action, specifically a strategy for obtaining useful information from the system.

To achieve that, the searcher progressively accumulates internal information and knowledge about the topic matter to enhance the understanding of their foreground information need [13]. (The background information need relates to distraction, see Jiang et al. [30].) Firstly, users interpret and create high-level meanings from the available information [35, 68], mainly from memory or prior experience (Stage 2.1). Next, they identify lower-level terms that are relevant and familiar [56] (Stage 2.2) and convert into a language that is compatible with the system [73] (Stage 2.3a). They also predict which keywords will effectively lead to the desired information [13, 35] (Stage 2.3b). Through multiple rounds of interpretation, identification, and prediction, the initial information need can connect with more specific and detailed needs [13, 73]. Here, users might also be influenced by their learned patterns of reasoning [56], cognitive bias, and technology proficiency [1], to plan their search effectively [68].

Query Submission (QS). See 'Stage 3' with purple-colored borders presented in Figure 1. When the search query is ready, the next step is to express the query to the system and execute it [54]. Modern systems offer various input modalities, such as typing via keyboard, speaking into a microphone, or more advanced approaches, such as, brain-computer interfaces [17, 80].

Relevance Judgment (RJ). See 'Stage 4' with yellow-colored borders presented in Figure 1. When search results are received, apart from comprehending and interpreting information like at IN, [2, 53, 54, 58, 60, 73], memory judgment [2] and inferential reasoning [58, 81] also apply to appraise the retrieved results [53]. The criteria include relevance, usefulness, and sufficiency [56, 73]. If the information is deemed relevant, it will be retained in long-term memory [60, 81] (from Stage 4.1 to 4.2). Moreover, Cole [13] envisages the user beginning to recognize a broader picture beyond just the facts or data, but also the societal aspects behind the information need, such as problem-goal or problem-solution frameworks, or task formulas. Moshfeghi et al. [53], Paisalnan et al. [58] identified the brain regions correspond to overarching theme comprehension activated at RJ (Stage 4.3). Lastly, the outcomes from appraisals influence the decision whether to continue the search [56], either by adopting the results or modifying the search query [73].

2.2 Cognitive & Affective Activities in Search

Cognitive load refers to the amount of cognitive resources in working memory exerted to complete a task. Working memory, an important cognitive system in informational processing, is responsible for processing sensory information, controlling and coordinating cognitive resources, as well as caching and processing recalled memory [21, 48]. Within its finite capacity, the more the working memory is used, the better task performance can be achieved [12, 36]. In terms of affective activities, Schubert's model [70] characterizes emotions with two main dimensions: (i) *affective arousal*, which refers to the intensity of a feeling and (ii) *affective valence*, which refers to the direction (positive or negative) of the feeling [68].

Given the search stages identified above, and using these physiological constructs, we formalize phenomena observed in the literature with a set of hypotheses – which we aim to test and validate

---

[2]These variables and activities are also important at the RJ stage, as discussed later.

with a laboratory user study. We denote the hypotheses of *cognitive load* as H_{cog}, *arousal* as H_{aro}, and *valence* as H_{val}.

Realization of Information Need (IN). This stage is about integrating information from the external context and internal memory. It requires demanding cognitive effort [68] allocated for three important components, Memory Retrieval, Information Flow Regulation, and Decision-Making [55].

In an experiment, participants are usually given some backstories which simulate a scenario and evoke the need of information to search [32]. A *feeling of uncertainty* is elicited because of a knowledge gap [8, 35, 52], and might lead to a combination of negative feelings, such as irritation, confusion, frustration, anxiety, and rage [68]. Even so, users still look forward to finding new information to solve their problems [35]. A neurological experiment of Moshfeghi and Pollick [54] encapsulates IN as a goal-setting process. Apart from the cognitive tasks for language processing, it also involves other tasks for which working memory is responsible, such as sustaining attention, planning, imagining, switching, maintaining instruction, and balancing and managing cognitive resources [47, 55, 58] (these are also involved in relevance judgment [2, 58]). Subsequently, goal-directed feelings appear which brings a sense of direction and temporary relief from negative feelings [35, 56]. These anticipatory feelings and previous negative feelings might balance out [68]. This explains the self-assessment results collected by Moshfeghi and Jose [52], that participants experienced uncertainty, but low anxiety and neutral emotions were predominant. It also shows that these affective activities only hover at the subconscious level in practice, compared to the theory [68].

Query Formulation (QF). Now that the goal has become clear and a plan has been set, the initial feeling of uncertainty gradually decreases while confidence and clarity increase [35, 52]. It is progressing from planning to action, the users are ready to begin to search [35]. Participants in Moshfeghi and Pollick [54] and Shovon et al. [71]'s experiments are also found to be prepared and ready to express at this stage. The cognitive activities here mainly involve term interpretation, identification, and prediction. We therefore expect the following relationships:

IN versus RJ :
$H_{cog}(1) :$ cognitive load$(IN) >$ cognitive load(QF) [47, 54, 55, 58]
$H_{aro}(1) :$ arousal$(IN) >$ arousal(QF) [35, 47, 55, 58]
$H_{val}(1) :$ valence$(IN) <$ valence(QF) [35, 52, 56, 58]

Query Submission (QS). It is observed by both Gwizdka [21] and Shovon et al. [71] that formulating and submitting queries requires more cognitive effort than passively receiving information (in relevance judgment). They reasoned that this is due to the simultaneous cognitive processes involved in recalling and producing terms being more demanding. The work of Moshfeghi and Pollick [54] differentiates these two, revealing that the brain activities at QF are primarily associated with semantic interpretation, keyword identifications and formulation, and prediction. At QS, they are centered around attention and motor processing for expressing (verbalizing) the query, and affective activities related to reward

processing. We therefore expect:

QF versus QS :
$H_{cog}(2) :$ cognitive load$(QF) <$ cognitive load(QS) [54]

QS versus RJ :
$H_{cog}(3) :$ cognitive load$(QS) >$ cognitive load(RJ) [2, 53, 54, 58]

For the affective activities, Savolainen [68] supposes that feelings are combined with positives related to brief elation and anticipation, and negatives such as confusion and sometimes anxiety, at QS. However, Lopatovska [39] found no significant varying of facial expressions during QS, yet collected insufficient self-rating emotion data. Explicitly, Moshfeghi and Pollick [54] found the brain regions responsible for affective appraisal activate at QS, confirming the occurrences of affective activities. This also implies that emotions are triggered by the expectation of the query's possible success and the inherent reward of finding the right information. Taken together, these results align with Savolainen's discussion [68], a balance between anticipatory emotions and overall emotional tone, so that most affective activities stay at a subconscious level. Furthermore, the difference between actively expressing at QS and passively receiving at QF might distinguish the arousal level. We expect:

QF versus QS :
$H_{aro}(2) :$ arousal$(QF) <$ arousal(QS) [35]
$H_{val}(2) :$ valence$(QF) <$ valence(QS) [35]

Relevance Judgment (RJ). As depicted by Kuhlthau [35], when the search process nears completion, feelings generally shift to predominantly positive. The level of uncertainty decreases, and users feel more confident as they become better at finding relevant information. Interest also increases. In particular, users often experience satisfaction and a sense of direction when they come across useful information, as they can navigate through the information more effectively. Conversely, if the information is not useful, boredom can set in. This theory was later supported in experiments, such as self-reported perception from Moshfeghi and Jose [52], anticipatory electrodermal responses from Mooney et al. [51], and increasing sadness when search results fail to meet expectations from Arapakis et al. [3], Lopatovska [39]. In particular, self-assessment collects less neutral emotion [52], and the most frequent facial expression is surprise [3, 39, 46, 52]. The results captured by these approaches mean that feelings reach a conscious level, indicating the intensity of feelings is stronger at RJ. It is worth noting that high cognitive load is usually associated with high arousal [26], but not solely. Both cognitive and affective perspectives suggest RJ has the highest level of arousal. Therefore:

QS versus RJ :
$H_{aro}(3) :$ arousal$(QS) <$ arousal(RJ) [39, 54]
$H_{val}(3) :$ valence$(QS) <$ valence(RJ) [35]

Moreover, although IN and RJ both involve passively receiving information, the latter requires more demanding cognitive processes [58]. The efforts at RJ mainly are exerted to encode and maintain the task (e.g., relevance criteria), store and update information, and accumulate evidence during appraisal [2, 53, 58]. Meanwhile, negative feelings can still dominate at RJ, reflecting unsatisfying results, or greater mental effort or concentration when dealing with challenges like information overload, conflicts, or complex information

[35, 68]; the results of McDuff et al. [46] and Gwizdka [21] provide observational support. Accordingly, we expect:

RJ versus IN :
$H_{cog}(4)$:   cognitive load(RJ) > cognitive load(IN)   [58]
$H_{aro}(4)$:   arousal(RJ) > arousal(IN)   [51, 52]
$H_{val}(4)$:   valence(RJ) > valence(IN)   [2, 3, 39, 46, 52]

2.3 Physiological Indexes

Physiological indexes are measurable biological functions that provide insights into an individual's activities, such as their physical and emotional state, cognitive performance, and overall health.

Cognitive Load. The different intensities of signals generated by the human brain can indicate various cognitive activities. The frontal cortex plays a crucial role in attention, memory, and judgment. A common agreement that the frontal theta power (4–8 Hz) is a strong indicator for the change of cognitive load [12, 61], regardless of visual or auditory modalities [31] or cognitive or motor type of tasks [72]. Increased cognitive load is associated with enhanced frontal theta. Another brain wave, alpha power (8–12 Hz) is also frequently mentioned in relation to measuring cognitive load. Alpha power predominates when relaxing or inhibiting task-irrelevant activities [12, 61, 63]. Although there are some inconsistent results, Chikhi et al. [12] synthesizes the existing findings and reveals a prevalent negative correlation of cognitive load on alpha power in the parietal cortex – responsible for sensory processing. Combining these two, Theta-Alpha Ratio (TAR) has been validated by Raufi and Longo [63] as an index level of cognitive load.

Pupil dilation is also extensively used to measure cognitive load [23, 75], with increasing cognitive load being associated with increasing pupil dilation [61, 75]. Compared to the highly-sensitive nature of EEG with multiple channels, pupil data can provide a cleaner and simpler indication of cognitive load [61].

Both TAR and pupil dilation are typically positively correlated with *Cognitive Load*, but they contribute from different aspects. Pupil dilation is usually associated with the attentional aspect of *Cognitive Load* or general affective arousal [22, 23, 61, 75], whereas TAR is more specifically tied to the intensity of neural activity when engaged in cognitive or memory processing tasks [63, 67].

Arousal. Apart from theta and alpha, beta power (12-30 Hz) is also influenced by cognitive load. But it is caused by an associative relationship from emotional responses or other underlying mechanisms [12]; enhanced cognitive load might associate with enhanced affective arousal [26]. Beta power is associated with an alert or excited state of mind [62], while alpha power is associated with a relaxed state. They are often used as a robust index of *arousal*, computed as the Beta-Alpha Ratio (BAR) [45, 62]. When experiencing high arousal, the level of beta should be high while alpha should be low, resulting in a high BAR [45].

In addition, Electrodermal Activity (EDA) and Photoplethysmogram (PPG) are also robust indicators of arousal. Specifically, high arousal elicits presentation in higher Skin Conductance Level (SCL) [5, 19] and Heart Rate Variability (HRV) [9, 26, 59]. As mentioned above, Mooney et al. [51] has found increased EDA at RJ, indicating anticipatory feelings.
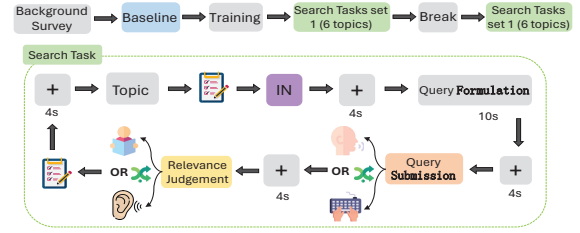

Figure 2: Experimental Procedure. IN: the realization of Information Need.

Valence. It is widely accepted in psychological studies that alpha power between the left and right frontal areas is associated with emotion [24, 25, 37]. In particular, enhanced left alpha is associated with negative emotion or withdrawal response, and vice versa. Regarding information activity, this withdrawal/approach response is represented as being open or conservative towards new information [35, 68]. Therefore, the level of asymmetry of alpha power in the frontal area, Frontal Alpha Asymmetry (FAA), is usually used to measure valence [25, 45, 62]. A negative FAA indicates relatively higher left alpha, thus negative emotion.

3 USER STUDY

3.1 Procedure

The experimental protocol is shown in Figure 2[3]. After calibration, the participants answer a background survey; information about handedness, sleep quality and caffeine intake are collected. Next, the participants complete a 15-second eyes-open (EO) and a 15-second eyes-closed (EC) section to collect the baseline data, followed by a training section containing the instruction and two practice tasks. Then they proceed to perform the search tasks (12 in total).

For the search task, participants start by looking at a fixed cross in the middle of a blank screen for 4 seconds. Next, a topic title is shown. Then, participants rate their interest, familiarity, and expected difficulty regarding the topic using a 5-level Likert item. Next, a backstory that evokes the information need (IN) is presented. Participants are then given 10 seconds to form a search query in their mind (QF), followed by submitting the query (QS) either written in text or via voice. Once the query is submitted, participants receive one relevant information snippet – either displayed as text on the screen, or played as an audio clip. Finally, they need to answer a binary factual judgment question (attention check) and rate their perceived relevance and difficulty in understanding the search result. In order to account for the delays on physiological responses a 4-second fixed cross gap is provided between search stages, i.e., IN, QF, QS, and RJ.

The sequences of topics and the interaction modalities (voice or text) are randomized. A mandatory 5-minute break is taken after 6 tasks. After completing the search tasks, participants verbally describe their experiences towards the experiment for quality purposes. Furthermore, to capture the activities precisely, we record

---

[3]Materials and code are available at https://github.com/kkkkk2017/IR-Physiological-Signals . The data can be requested by contacting the authors.

all the timestamps of page transactions, bottom press (to start/stop voice input), and first and last keystroke input.

3.2 Materials

Information Needs. We use the backstories in the Information-Needs dataset [6] created by Moffat et al. [49]. The dataset contains backstories that represent different information needs for 180 TREC topics. The information needs were categorized into three levels of cognitive complexity: Remember, Understand, and Analyze. We choose 12 topics from the middle level (Understand) to have enough room for unfamiliarity, but also to avoid risks of triggering emotions or cognitive bias. The Understand category involves searching and gathering relevant messages to construct meaning for the given topic. We randomly sample topics and remove those related to crises, wars, conspiracy, or politically sensitive topics. The original backstories have an average of 41 words ($SD = 6$). To ensure all selected backstories have a similar word count, we manually edited them, resulting in an average of 40 words ($SD = 1$).

Search Results. For each information need, participants receive one information snippet generated by combining relevant documents as follows. Although the backstories [49] were developed based on TREC topics, the qrels from the corresponding TREC test collection does not directly align with the Information Need. Therefore, given the TREC topic associated with the backstory, we manually select up to three documents judged as relevant in the qrels. Then, we use GPT-3.5[4] to generate a 150-word summary based on the provided documents[5] – as well as a binary factual judgment question that we used as attention check. To minimize the influences of word lengths or complexity, we further manually examine the generated summaries using the Flesch Reading Ease (FRE) score [18]. Overall, the summaries have an average word count of 148 ($SD = 3$) and an average FRE score of 11.9 ($SD = 0.9$).

3.3 Equipment and Setup

Figure 3: Experiment setup (left), and the EEG electrode locations (right). The filled circles indicate the electrodes used to compute the indexes.

Four sensors are used in this study: a webcam camera for video recording, a Tobii Fusion eye-tracker[6] for pupillary responses (60Hz), an E4 wristband[7] for EDA (4Hz) and PPG (64Hz), and a 14-channel

Table 1: Data cleanup summary. Note that each baseline (in parentheses) corresponds to data from one participant.

Data cleanup step	Number of Trials (+ Baseline)	
	EEG & EDA & PPG	PUPIL
Original data	312 (+26)	276 (+23)
Bad data cleanup	300 (+25)	182 (+23)
Removal by self-ratings	177 (+25)	159 (+23)
Removal of 1 person with only 1 trial	176 (+24)	158 (+22)

Emotiv EPOC headset[8] for EEG data (128Hz). The experiment is conducted in an illuminated room. The participant sits in front of a desktop PC, which is mounted with an eye-tracker and a web camera. All participants use the computer mouse with their right hand, and wear the wristband on the left hand. We sanitize the electrodes and the participant's skin on the inner and outer wrist with alcohol wipes [5]. Then, the instructor helps the participants to wear the headset, and adjusts the positions of the electrodes. The experiment material is deployed using the Qualtrics[9] platform.

3.4 Participants

The study received human research ethics approval from RMIT University, and participants provided written informed consent prior to the experiment. To ensure a minimum of additional effort involved for language, we recruit participants with at least a professional working proficiency level in English. A total of 29 participants are recruited. The data collected for 3 of these participants are discarded due to environmental disturbances. Due to software errors, the eye-tracking data from 3 participants could not be obtained. For results concerning EEG, EDA, and PPG, we use valid data from 26 participants (15M, 11F). There are 77% of the participants with full professional proficiency or are native English speakers. For results related to eye data, we use valid 23 participants (13M,10F).

4 DATA CLEAN-UP & ANALYSIS

First, the data obtained from all sensors are synchronized by timestamp. Each recording is then denoised, explained in further detail below, and divided into 13 trials corresponding to 1 baseline (EYEOPEN/CLOSE) and 12 search tasks. As per our experimental methodology, each trial starts with a 4-second fixation, and contains 4 Events of Interest (EOI), i.e., IN, QF, QS, RJ. To deal with time inconsistency, we only analyze the first 10 seconds of each EOI, selected by the lower quartile [72].

4.1 Data Processing

EEG data is processed using the MNE Python library.[10] The break section is excluded. Following similar procedures to Martínez-Santiago et al. [44] and Gwizdka et al. [23], each EEG recording is first denoised with a Butterworth filter (1–50Hz, 5th), removed the signal mean, and re-referenced with the common average. Next, the data is further cleaned and interpolated with the Autoreject [27] package. Lastly, to remove the artifacts (e.g., blinking), we use the

Independent Component Analysis (ICA) combined with ICLabel [38]. One recording is removed because of bad quality of EEG data. The power spectral density of each EEG channel is then calculated using Welch's method and hamming window and normalized [33, 37, 72]. The indexes are then computed from the set of EEG electrodes as follows. **Theta-Alpha Ratio (TAR)** is computed by $avg(\theta(AF3, AF4, F3, F4, F7, F8))/avg(\alpha(P7, P8))$ [63]. **Beta-Alpha Ratio (BAR)** and **Frontal Alpha Asymmetry (FAA)** are computed $BAR = \beta(AF3 + AF4 + F3 + F4)/\alpha(AF3 + AF4 + F3 + F4)$, $FAA = \log(\alpha(F4)/\beta(F4)) - \log(\alpha(F3)/\beta(F3))$ [25, 45, 62].

Pupil data are cleaned following the procedure described by Kret and Sjak-Shie [34] and Gwizdka et al. [23]. The left and right pupils are first processed separately. Samples with dilation speed above the median absolute deviation or the gap between two data points above (75 ms) are removed. This is done twice for each side to remove the edge values. Then, the cleaned data of both sides are combined by taking the arithmetic mean, and linear interpolation is applied to fill in the blink gaps. Finally, a zero-phase Butterworth filter (4Hz, 3rd) is applied to remove outliers [43]. As our experiment includes sub-tasks that do not require on-screen visuals (i.e., QF via voice and RJ via audio), some sub-tasks are significantly lacking in pupil data. The EOIs with > 20% missing data are excluded for analysis [23], and the trials that do not include all 4 EOIs are subsequently excluded. **Relative Pupil Dilation (RPD)** calculates the relative changes of current pupil diameter compared to a baseline value [23]: $RPD_t^i = (P_t - P_{baseline}^i)/P_{baseline}^i$, where t is time, i is participant, and baseline is the average pupil diameter across all tasks.

EDA and PPG signals obtained from the wristband are processed using the NeuroKit2 [41] Python library, following a similar procedure as by Bota et al. [10], Braithwaite et al. [11], Di Lascio et al. [15]. For EDA, a low-pass (0.5Hz) Butterworth filter followed by a rolling median with a 3-second window [5] and min-max normalization are applied. The convex optimization cvxEDA method [20] is then applied to decompose the tonic value, i.e., the **Skin Conductance Level (SCL)**. The raw PPG data is cleaned with the default approach in NeuroKit2. Then, the time between consecutive heartbeats is computed, representing **Heart Rate Variability (HRV)** in milliseconds.

4.2 Assumptions & Trial Selection

When forming the hypotheses, it is worth noting that the following assumptions are made when considering possible factors that might interfere with physiological responses, such as information complexity [44], relevance [2, 7, 17, 57, 80], and interest [77, 78].

In this experiment, the participants report average scores of 3.5 ($SD = 1.1$) interest, 2.5 ($SD = 1.1$) difficulty, 2.6 ($SD = 1.3$) familiarity towards the topics, and 4.0 ($SD = 1.1$) relevance, 2.0 ($SD = 1.1$) difficulty to the search results. To meet the assumption before conducting any analysis, we first select from the trials based on self-ratings, using the following thresholds:

- Users are fairly interested in the topics (1 < topic_interest < 5).
- The search results are relevant to the submitted queries (info_relevance ≥ 3).
- Search results are not difficult to understand (info_difficulty ≤ 3).
- Participants are engaged in the tasks.

Table 2: Summary of hypothesis validation. Pairs with significant differences that confirm the hypothesis (✓), a significant but opposite relationship (✗) or no significant difference (–) . $p < .001^{***}$, $p < .01^{**}$, $p < .05^{*}$.

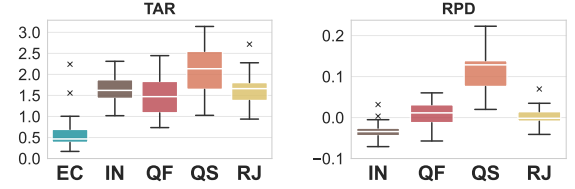| H_{cog}	TAR	RPD	H_{aro}	BAR	SCL	HRV	H_{val}	FAA
IN > QF	✓***	✗***	IN > QF	✓*	–	–	IN < QF	–
QF < QS	✓***	✓***	QF < QS	–	–	✗***	QF < QS	–
QS > RJ	✓***	✓***	QS < RJ	–	✓*	✓***	QS < RJ	–
RJ > IN	–	✓***	RJ > IN	–	–	✓**	RJ > IN	–

Figure 4: Distribution of indexes for measuring Cognitive Load across all participants. The values of one participant are aggregated into one data point.

The summary of data cleanup is presented in Table 1. It is noteworthy that QF and QS are usually consecutive phases. Although our experimental design attempts to separate them, it cannot guarantee the complete removal of automatic progression. It might also involve QF-related cognitive activities at QS, such as recalling and re-evaluating the terms.

4.3 Statistical Analysis

The proposed hypotheses are tested in a within-subject setting. As the data is not normally distributed, we conduct the non-parametric Wilcoxon signed-rank tests for each physiological index between pairs of EOIs: IN and QF, QF and QS, QS and RJ, RJ and IN. The Bonferroni correction for multiple comparisons is applied to adjust p-values before they are compared to the α significance thresholds.

5 RESULTS

5.1 Baseline

EYECLOSE (EC) represents a relaxed state of participants, potentially indicating a minimum level of cognitive effort, arousal, and valence. However, EC is excluded when comparing RPD as pupil data is unavailable. TAR and BAR have significant differences at EC compared to all EOIs, but FAA or SCL does not. Nevertheless, SCL is lower than all EOIs (refer to Figure 5B). HRV has significant differences at EC compared to QS or RJ.

5.2 Cognitive Load (H_{cog})

$H_{cog}(1)$: *IN versus QF.* Both TAR and RPD show significant differences between IN and QF ($W_{TAR} = 28, p < .001, W_{RPD} = 20, p < .001$). But interestingly, they present opposite trends, as shown in Figure 4. TAR is higher at IN than QF, whereas RPD is lower. These opposite results can be presumably explained by the different cognitive demands required at these EOIs. As discussed in
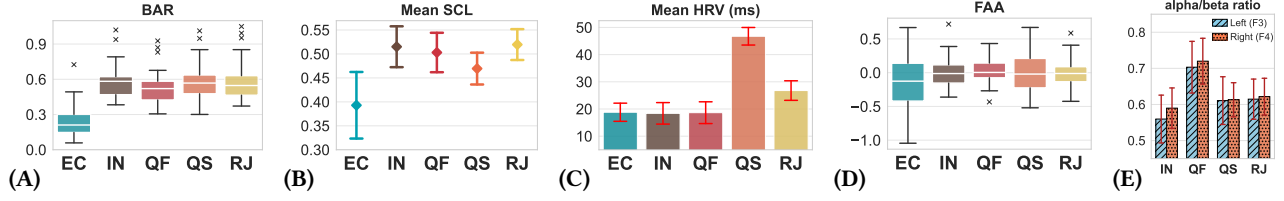
Figure 5: Distribution of indexes for measuring *affective arousal* (A, B, C) and *valence* (D, E). Error bars indicate standard error.

Section 2, the primary cognitive activities at IN involve information processing and memory retrieval; thus, higher TAR. In contrast, those at QF primarily entail problem-solving to generate an effective search query, attentional resources are dominant; thus, higher RPD. Overall, $H_{cog}(1)$ is partially supported.

$H_{cog}(2)$: QF versus QS, and $H_{cog}(3)$: QS versus RJ. Both TAR and RPD are significantly lower at QF than QS ($W_{TAR} = 1, p < .001$, $W_{RPD} = 0, p < .001$), which supports $H_{cog}(2)$. They are also lower at RJ than QS ($W_{TAR} = 3, p < .001, W_{RPD} = 1, p < .001$), which supports $H_{cog}(3)$. These results suggest the demands for either cognitive processing or attention are lower at QF or RJ when compared to QS. The difference between QS and RJ is consistent with the results by Gwizdka [21] and Shovon et al. [71].

The difference between QF and QS further distinguishes these two EOIs. The high TAR and RPD at QS are potentially due to the simultaneous cognitive activities for recalling information, and forming and expressing the query. If the task involves typing, it may also require additional effort to coordinate hand movements. However, a large deviation for QF in Figure 4 might indicate disengagement. Some participants disclosed that they do not always think about the query during given seconds. This was also a potential issue reported in Moshfeghi and Pollick [54]'s experiment.

$H_{cog}(4)$: RJ versus IN. No significant difference is found for TAR ($W_{TAR} = 135$), but RPD is significantly lower at IN ($W_{RPD} = 15, p < .001$) when compared to RJ. The TAR result may arise because both EOIs involve language comprehension, memory retrieval, and appraisal, and the cognitive demands for these tasks are not substantially different. On the other hand, even though the cognitive processes involved in both scenarios are similar, participants may exhibit more interest at RJ where they get search results that complete their knowledge gap. This heightened interest leads to greater engagement, and consequently, more directing of attention resources. As a result, RPD at RJ is higher than at IN. We further discuss this finding with the results of $H_{aro}(4)$ in the following section. As the causation might not primarily be attributed to cognitive load, $H_{cog}(4)$ is partially supported.

5.3 Arousal (H_{aro})

$H_{aro}(1)$: IN versus QF. No significant difference is observed in SCL or HRV, while BAR is significantly different ($W_{BAR} = 52, p < .05$). A higher BAR at IN compared to QF may imply that the participants are becoming aware of their knowledge gap [47] and consequently more alert in addressing it. The lack of significant difference in SCL

or HRV suggests that alertness might have been mostly subconscious and not strong enough to reach a level of affective arousal. $H_{aro}(1)$ is partially supported.

$H_{aro}(2)$: QF versus QS. As shown in Figure 5B and 5C, HRV is significantly lower at QF as compared to QS ($W_{HRV} = 1, p < .001$). SCL is lower at QS than at QF, although the difference is not significant. When a person experiences a high level of arousal, SCL will increase while HRV will decrease [50, 59]. These findings suggest arousal is relatively higher at QF than QS. Thus, $H_{aro}(2)$ is rejected.

$H_{aro}(3)$: QS versus RJ. Significant differences are observed for both SCL ($W_{SCL} = 58, p < .05$) and HRV ($W_{HRV} = 3, p < .001$). In addition to the lower SCL and higher HRV at QS than RJ, as demonstrated in Figure 5B and 5C, indicating arousal is lower at QS than RJ, thereby supporting $H_{aro}(3)$.

$H_{aro}(4)$: RJ versus IN. HRV is significantly longer at RJ than IN ($W_{HRV} = 34, p < .01$), suggesting a higher level of arousal at RJ. This finding can be linked with the observation for $H_{cog}(4)$ in Section 5.2, where a significantly larger RPD at RJ than IN. This elevated RPD is likely due to enhanced interest and engagement, which in turn translates to heightened arousal. Therefore, the result in HRV further rejects $H_{cog}(4)$ and supports $H_{aro}(4)$.

5.4 Valence (H_{val})

Our conclusions around valence mostly rely on FAA. The statistical results of FAA fail to reject the null hypothesis of any pair. This suggests valence between EOIs is not substantially different; H_{val} s are rejected. These results may add up to a balance of conflicting feelings, e.g. expectation and anxiety [68]. Refer to Figure 5D, all EOIs have FAA averages and medians nearly 0, implying neutral valence levels. A closer look at its components (Figure 5E) reveals that the right alpha power is relatively higher than the left at IN. Higher right alpha usually represents withdrawal motivation, associated with negative valence [45, 62]. Right alpha is still slightly higher than left alpha at QF. At QS and RJ, the alpha levels are balanced. It might suggest valence tends to shift from negative to positive when participants figure out search queries and find relevant information.

6 DISCUSSION & LIMITATIONS

This study aims to characterize the information seeking process from 3 aspects, in relation to 3 physiological constructs: *cognitive load*, *affective arousal*, and *valence*. Physiological signals are captured to compute the indexes infer these constructs. Our hypotheses are primarily built upon the findings of prior work in neurophysiology [54] and behavioral analysis [21] (see Section 2).

Measured through TAR and RPD, *Cognitive Load* shows statistically significant differences across search stages, supporting hypotheses $H_{cog}(1–3)$. A noteworthy reversal between TAR and RPD at IN and QF offers insights into Moshfeghi and Pollick's findings [54]. At IN, goal-directed brain functions are more activated, aligning with the elevated TAR, while attention-directing functions are more activated at QF, reflected in the higher RPD. Then, our results further distinguish between QF and QS, that QS is more demanding. However, this is likely to be influenced by disengagement caused at the 10-seconds QF. Moreover, QS requires higher cognitive loads than RJ– observed in both TAR and RPD – which aligns with Gwizdka's findings [21]. Lastly, between RJ and IN, heightened RPD at RJ may be attributed to increased interest, along with engagement and arousal. This aligns with Paisalnan et al.'s findings [58] of similar but more demanding cognitive processes at RJ compared to IN.

Three physiological indexes, BAR, SCL, and HRV, are used to measure *affective arousal*. The hypotheses H_{aro} are primarily validated with SCL and HRV. For *affective valence*, the results from FAA fail to validate any hypotheses H_{val}, with no significant differences observed. But some variations are seen when examining the components of FAA, i.e., left and right frontal alpha. At the beginning of a search (IN), the elevated TAR and BAR suggest a knowledge gap is updated to the awareness. Additionally, right frontal alpha dominance implies a relatively negative valence. These physiological signs might infer that the *feeling of uncertainty* primarily stays in a cognitive state, without the corresponding emotional reactions. Thus, emotional responses as a sign of need-to-search might not be effective. Arousal decreases and valence tends to be neutral at QF, suggesting the participants are planning the action. Then, at QS, arousal is lower compared to QF. This suggests that the pleasantness of being able to act, the expectation of success, and increased confidence in finding relevant results. Or it might because the participants were disengaged at QF then back to the task at QS. However, with no significant difference found in FAA, we cannot infer whether the feeling is positive or negative. Consequently, when relevant search results are presented at RJ, arousal further increases than QS. These findings are related to the reward-seeking feelings discussed in previous work [39, 54].

Although IN and RJ may share similarities, our findings of significant differences in HRV and RPD, and observations of higher FAA, further highlight the difference between these two stages. The affective difference could be attributed to different appraising criteria [56], such as having sufficient knowledge to solve the problem or accepting the found answers. Yet, the result of FAA is insufficient to conclude a positive emotion at RJ, so levels of satisfaction cannot be inferred from FAA. The results might relate to interest and curiosity. The difference in RPD can be transformed as the web-logging behaviors, such as longer dwell time on the relevant results.

Like all experimental studies, our investigation is subject to limitations. Firstly, although we attempted to eliminate confounding variables, other factors, such as users' search skills and prior domain knowledge, may have influenced the results. Secondly, the indexes derived from physiological signals are insufficient to disclose all of the raw information they capture. The intricacies between cognitive load and emotions make it difficult to entirely separate their effects in physiological signals. In the next phase of this work, we plan to explore pattern analysis with machine learning models [28],

incorporating all features extracted from the signals. Despite this limitation, the specific indexes we used are based on empirical findings in the literature, which can provide a more robust description of subconscious user behavior. We note that the temporal relationship between search stages may influence physiological signals. However, for the sake of brevity, this paper compresses the entire signal into a single value, omitting the temporal relationship. Alternative aggregation methods will be explored, such as dividing into three equal-length segments – beginning, middle, and end – as applied by Gwizdka et al. [23], or performing sophisticated temporal analysis of the entire signal, as shown by van Rij et al. [76].

7 CONCLUSION

This study aimed to characterize user behaviors during an information seeking process using physiological signals. Our experiment focused on a scenario of searching for unknown knowledge and understanding a topic, i.e., searching to resolve a knowledge gap. A lab study was conducted to collect physiological signals through four search stages involving: the realization of Information Need (IN), Query Formulation (QF), Query Submission (QS), and Relevance Judgment (RJ). The *cognitive load*, *affective arousal* and *valence* were analyzed using well-established indexes derived from the signals.

Our results indicate that cognitive demands are higher, but attentional resources are lower at IN compared to QF. At IN, a little rise in alertness might capture the recognition of the knowledge gap. But this response does not elicit any negative affective feelings, at least not to the extent that peripheral signals were able to detect in our experiment. Next, cognitive load is more intense at QS than at the previous (QF) or subsequent stage (RJ), which supplements the findings by Gwizdka [21] and Shovon et al. [71]. This further indicates that simultaneous cognitive processes are highly demanding at QS, potentially explaining higher affective feelings than at QF. Finally, our results indicate that affective feelings are more active at RJ. Compared to IN, the incremental feelings and attentions at RJ suggest greater interest, engagement, and curiosity as the results resolve the searcher's knowledge gap.

This study extended the existing understanding of how users engage in information seeking processes by complementing existing theories and observational studies with the characterization of search stages using physiological signals. Our findings serve as a baseline for future experiments investigating affective and cognitive feedback, as well as physiological signals, for search interactions. There is a growing interest in employing wearable physiological sensors in search systems due to their mobility, decreasing cost, and information-rich advantages [69]. By better understanding the intrinsic states of searchers in a continuous process, our proposed methodology can contribute to improving the overall search experience and devising real-time solutions. We believe our experimental setting – validated in the context of known information seeking models – can help characterizing cognitive load and affective arousal in less established IIR settings, including Large Language Model-based conversational search.

ACKNOWLEDGMENTS

REFERENCES

[1] Marwah Alaofi, Luke Gallagher, Dana Mckay, Lauren L. Saling, Mark Sanderson, Falk Scholer, Damiano Spina, and Ryen W. White. 2022. Where Do Queries Come From?. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Madrid, Spain) *(SIGIR '22)*. Association for Computing Machinery, New York, NY, USA, 2850–2862. https://doi.org/10.1145/3477495.3531711

[2] Marco Allegretti, Yashar Moshfeghi, Maria Hadjigeorgieva, Frank E. Pollick, Joemon M. Jose, and Gabriella Pasi. 2015. When Relevance Judgement is Happening? An EEG-based Study. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Santiago, Chile) *(SIGIR '15)*. Association for Computing Machinery, New York, NY, USA, 719–722. https://doi.org/10.1145/2766462.2767811

[3] Ioannis Arapakis, Joemon M. Jose, and Philip D. Gray. 2008. Affective feedback: an investigation into the role of emotions in the information seeking process. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Singapore, Singapore) *(SIGIR '08)*. Association for Computing Machinery, New York, NY, USA, 395–402. https://doi.org/10.1145/1390334.1390403

[4] Ioannis Arapakis, Ioannis Konstas, and Joemon M. Jose. 2009. Using facial expressions and peripheral physiological signals as implicit indicators of topical relevance. In *Proceedings of the 17th ACM International Conference on Multimedia* (Beijing, China) *(MM '09)*. Association for Computing Machinery, New York, NY, USA, 461–470. https://doi.org/10.1145/1631272.1631336

[5] Ebrahim Babaei, Benjamin Tag, Tilman Dingler, and Eduardo Velloso. 2021. A Critique of Electrodermal Activity Practices at CHI. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 177, 14 pages. https://doi.org/10.1145/3411764.3445370

[6] Peter Bailey, Alistair Moffat, Falk Scholer, and Paul Thomas. 2014. Information Needs for TREC 2002-4 (2014). v2. CSIRO. Data Collection. https://doi.org/10.4225/08/55D0B6A098248

[7] Oswald Barral, Manuel J.A. Eugster, Tuukka Ruotsalo, Michiel M. Spapé, Ilkka Kosunen, Niklas Ravaja, Samuel Kaski, and Giulio Jacucci. 2015. Exploring Peripheral Physiology as a Predictor of Perceived Relevance in Information Retrieval. In *Proceedings of the 20th International Conference on Intelligent User Interfaces* (Atlanta, Georgia, USA) *(IUI '15)*. Association for Computing Machinery, New York, NY, USA, 389–399. https://doi.org/10.1145/2678025.2701389

[8] Nicholas J. Belkin. 1980. Anomalous States of Knowledge as a Basis for Information Retrieval. *Canadian Journal of Information and Library Science* 5, 1 (1980), 133–143.

[9] Nattapat Boonprakong, Xiuge Chen, Catherine Davey, Benjamin Tag, and Tilman Dingler. 2023. Bias-Aware Systems: Exploring Indicators for the Occurrences of Cognitive Biases When Facing Different Opinions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 27, 19 pages. https://doi.org/10.1145/3544548.3580917

[10] Patricia J. Bota, Chen Wang, Ana L.N. Fred, and Hugo Placido Da Silva. 2019. A Review, Current Challenges, and Future Possibilities on Emotion Recognition Using Machine Learning and Physiological Signals. *IEEE Access* 7 (2019), 140990–141020. https://doi.org/10.1109/ACCESS.2019.2944001

[11] Jason J. Braithwaite, Derrick G. Watson, Robert Jones, and Mickey Rowe. 2015. *A Guide for Analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for Psychological Experiments*. Technical Report 2. Behavioural Brain Sciences Centre, University of Birmingham.

[12] Samy Chikhi, Nadine Matton, and Sophie Blanchet. 2022. EEG power spectral measures of cognitive workload: A meta-analysis. *Psychophysiology* 59, 6 (2022). https://doi.org/10.1111/psyp.14009

[13] Charles Cole. 2011. A theory of information need for information retrieval that connects information to knowledge. *Journal of the American Society for Information Science and Technology* 62, 7 (2011), 1216–1231. https://doi.org/10.1002/asi.21541

[14] Samantha G Daley, John B Willett, and Kurt W Fischer. 2014. Emotional responses during reading: Physiological responses predict real-time reading comprehension. *Journal of Educational Psychology* 106, 1 (2014), 132–143. https://doi.org/10.1037/a0033408

[15] Elena Di Lascio, Shkurta Gashi, and Silvia Santini. 2018. Unobtrusive Assessment of Students' Emotional Engagement during Lectures Using Electrodermal Activity Sensors. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 3, Article 103 (sep 2018), 21 pages. https://doi.org/10.1145/3264913

[16] Ashlee Edwards and Diane Kelly. 2017. Engaged or Frustrated? Disambiguating Emotional State in Search. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Shinjuku, Tokyo, Japan) *(SIGIR '17)*. Association for Computing Machinery, New York, NY, USA, 125–134. https://doi.org/10.1145/3077136.3080818

[17] Manuel JA Eugster, Tuukka Ruotsalo, Michiel M Spapé, Oswald Barral, Niklas Ravaja, Giulio Jacucci, and Samuel Kaski. 2016. Natural brain-information interfaces: Recommending information by relevance inferred from human brain

signals. *Scientific reports* 6, 1 (2016), 38580. https://doi.org/10.1038/srep38580

[18] Rudolph Flesch. 1948. A New Readability Yardstick. *Journal of Applied Psychology* 32, 3 (1948), 221–233. https://doi.org/10.1037/h0057532

[19] Alberto Greco, Gaetano Valenza, Luca Citi, and Enzo Pasquale Scilingo. 2017. Arousal and Valence Recognition of Affective Sounds Based on Electrodermal Activity. *IEEE Sensors Journal* 17, 3 (2017), 716–725. https://doi.org/10.1109/JSEN.2016.2623677

[20] Alberto Greco, Gaetano Valenza, Antonio Lanata, Enzo Pasquale Scilingo, and Luca Citi. 2016. cvxEDA: A Convex Optimization Approach to Electrodermal Activity Processing. *IEEE Transactions on Biomedical Engineering* 63, 4 (2016), 797–804. https://doi.org/10.1109/TBME.2015.2474131

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

[22] Jacek Gwizdka. 2018. Inferring Web Page Relevance Using Pupillometry and Single Channel EEG. In *Information Systems and Neuroscience*. Springer International Publishing, Cham, 175–183. https://doi.org/10.1007/978-3-319-67431-5_20

[23] Jacek Gwizdka, Rahilsadat Hosseini, Michael Cole, and Shouyi Wang. 2017. Temporal Dynamics of Eye-Tracking and EEG during Reading and Relevance Decisions. *J. Assoc. Inf. Sci. Technol.* 68, 10 (oct 2017), 2299–2312.

[24] Eddie Harmon-Jones. 2003. Clarifying the emotive functions of asymmetrical frontal cortical activity. *Psychophysiology* 40, 6 (2003), 838–848. https://doi.org/10.1111/1469-8986.00121

[25] Eddie Harmon-Jones and Philip A Gable. 2018. On the role of asymmetric frontal cortical activity in approach and withdrawal motivation: An updated review of the evidence. *Psychophysiology* 55, 1 (2018). https://doi.org/10.1111/psyp.12879

[26] Maarten A Hogervorst, Anne-Marie Brouwer, and Jan BF Van Erp. 2014. Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload. *Frontiers in neuroscience* 8 (2014), 322. https://doi.org/10.3389/fnins.2014.00322

[27] Mainak Jas, Denis A. Engemann, Yousra Bekhti, Federico Raimondo, and Alexandre Gramfort. 2017. Autoreject: Automated artifact rejection for MEG and EEG data. *NeuroImage* 159 (2017), 417–429. https://doi.org/10.1016/j.neuroimage.2017.06.030

[28] Kaixin Ji, Damiano Spina, Danula Hettiachchi, Flora Dilys Salim, and Falk Scholer. 2023. Examining the Impact of Uncontrolled Variables on Physiological Signals in User Studies for Information Processing Activities. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval* (<conf-loc>, <city>Taipei</city>, <country>Taiwan</country>, </conf-loc>) *(SIGIR '23)*. Association for Computing Machinery, New York, NY, USA, 1971–1975. https://doi.org/10.1145/3539618.3591981

[29] Kaixin Ji, Damiano Spina, Danula Hettiachchi, Flora Dylis Salim, and Falk Scholer. 2023. Towards Detecting Tonic Information Processing Activities with Physiological Data. In *Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing* (Cancun, Quintana Roo, Mexico) *(UbiComp/ISWC '23 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 5 pages. https://doi.org/10.1145/3594739.3610679

[30] Tingting Jiang, Shiting Fu, Sanda Erdelez, and Qian Guo. 2022. Understanding the seeking-encountering tension: Roles of foreground and background task urgency. *Information Processing & Management* 59, 3 (2022), 102910. https://doi.org/10.1016/j.ipm.2022.102910

[31] Maciej Kaminski, Aneta Brzezicka, Jan Kaminski, and Katarzyna J Blinowska. 2016. Information Transfer During Auditory Working Memory Task. In *XIV Mediterranean Conference on Medical and Biological Engineering and Computing*. Springer, Cham, 19–24. https://doi.org/10.1007/978-3-319-32703-7_4

[32] Diane Kelly. 2009. Methods for Evaluating Interactive Information Retrieval Systems with Users. *Foundations and Trends® in Information Retrieval* 3, 1–2 (2009), 1–224. https://doi.org/10.1561/1500000012

[33] Vladimir Kosonogov, Danila Shelepenkov, and Nikita Rudenkiy. 2023. EEG and peripheral markers of viewer ratings: a study of short films. *Frontiers in Neuroscience* 17 (2023), 1148205. https://doi.org/10.3389/fnins.2023.1148205

[34] Mariska E Kret and Elio E Sjak-Shie. 2019. Preprocessing pupil size data: Guidelines and code. *Behavior Research Methods* 51 (2019), 1336–1342. https://doi.org/10.3758/s13428-018-1075-y

[35] Carol Collier Kuhlthau. 2005. Information Search Process. *CITE Seminar: Information Literacy and Pre-service Programs, Hong Kong, China* 7 (2005), 226.

[36] Naveen Kumar and Jyoti Kumar. 2016. Measurement of Cognitive Load in HCI Systems Using EEG Power Spectrum: An Experimental Study. *Procedia Computer Science* 84 (2016), 70–78. https://doi.org/10.1016/j.procs.2016.04.068 Proceeding of the Seventh International Conference on Intelligent Human Computer Interaction (IHCI 2015).

[37] Minji Lee, Gi-Hwan Shin, and Seong-Whan Lee. 2020. Frontal EEG Asymmetry of Emotion for the Same Auditory Stimulus. *IEEE Access* 8 (2020), 107200–107213. https://doi.org/10.1109/ACCESS.2020.3000788

[38] Adam Li, Jacob Feitelberg, Anand Prakash Saini, Richard Höchenberger, and Mathieu Scheltienne. 2022. MNE-ICALabel: Automatically annotating ICA components with ICLabel in Python. *Journal of Open Source Software* 7, 76 (2022), 4484. https://doi.org/10.21105/joss.04484

[39] Irene Lopatovska. 2014. Toward a model of emotions and mood in the online information search process. *Journal of the Association for Information Science and Technology* 65, 9 (2014), 1775–1793.

[40] Irene Lopatovska and Ioannis Arapakis. 2011. Theories, methods and current research on emotions in library and information science, information retrieval and human-computer interaction. *Inf. Process. Manage.* 47, 4 (jul 2011), 575–592. https://doi.org/10.1016/j.ipm.2010.09.001

[41] Dominique Makowski, Tam Pham, Zen J. Lau, Jan C. Brammer, François Lespinasse, Hung Pham, Christopher Schölzel, and S. H. Annabel Chen. 2021. Neurokit2: A Python Toolbox for Neurophysiological Signal Processing. *Behavior Research Methods* 53, 4 (Aug. 2021), 1689–1696. https://doi.org/10.3758/s13428-020-01516-y

[42] Gray Marchionini. 1995. Information-seeking perspective and framework. In *Information-Seeking in Electronic Environments*. Cambridge University Press, 27–60.

[43] Joel T Martin, Joana Pinto, Daniel P Bulte, and Manuel Spitschan. 2021. PyPlr: A versatile, integrated system of hardware and software for researching the human pupillary light reflex. *Behavior Research Methods* 54 (2021), 2720–-2739. https://doi.org/10.3758/s13428-021-01759-3

[44] Fernando Martínez-Santiago, Alejandro A Torres-García, Arturo Montejo-Ráez, and Nicolás Gutiérrez-Palma. 2023. The impact of reading fluency level on interactive information retrieval. *Universal Access in the Information Society* 22, 1 (2023), 51–67.

[45] Tomáš Matlovič. 2016. Emotion Detection using EPOC EEG device. *IIT. SRC* (2016), 1–6.

[46] Daniel McDuff, Paul Thomas, Nick Craswell, Kael Rowan, and Mary Czerwinski. 2021. Do Affective Cues Validate Behavioural Metrics for Search?. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Virtual Event, Canada) *(SIGIR '21)*. Association for Computing Machinery, New York, NY, USA, 1544–1553. https://doi.org/10.1145/3404835.3462894

[47] Dominika Michalkova, Mario Parra-Rodriguez, and Yashar Moshfeghi. 2022. Information Need Awareness: An EEG Study. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Madrid, Spain) *(SIGIR '22)*. Association for Computing Machinery, New York, NY, USA, 610–621. https://doi.org/10.1145/3477495.3531999

[48] Randall K. Minas, Robert F. Potter, Alan R. Dennis, Valerie Bartelt, and Soyoung Bae. 2014. Putting on the Thinking Cap: Using NeuroIS to Understand Information Processing Biases in Virtual Teams. *Journal of Management Information Systems* 30, 4 (2014), 49–82. https://doi.org/10.2753/MIS0742-1222300403

[49] Alistair Moffat, Peter Bailey, Falk Scholer, and Paul Thomas. 2014. Assessing the Cognitive Complexity of Information Needs. In *Proceedings of the 2014 Australasian Document Computing Symposium* (Melbourne, VIC, Australia) *(ADCS '14)*. ACM, New York, NY, USA, Article 97, 4 pages. https://doi.org/10.1145/2682862.2682874

[50] Mahtab Mohammadpoor Faskhodi, Mireya Fernández Chimeno, and Miguel Ángel García González. 2023. Arousal detection by using ultra-short-term heart rate variability (HRV) analysis. *Frontiers in Medical Engineering* 1, article 1209252 (2023). https://doi.org/10.3389/fmede.2023.1209252

[51] Colum Mooney, Micheál Scully, Gareth JF Jones, and Alan F Smeaton. 2006. Investigating Biometric Response for Information Retrieval Applications. In *Advances in Information Retrieval: 28th European Conference on IR Research*. Springer, Berlin, Heidelberg, 570–574. https://doi.org/10.1007/11735106_67

[52] Yashar Moshfeghi and Joemon M. Jose. 2013. On cognition, emotion, and interaction aspects of search tasks with different search intentions. In *Proceedings of the 22nd International Conference on World Wide Web* (Rio de Janeiro, Brazil) *(WWW '13)*. Association for Computing Machinery, New York, NY, USA, 931–942. https://doi.org/10.1145/2488388.2488469

[53] Yashar Moshfeghi, Luisa R Pinto, Frank E Pollick, and Joemon M Jose. 2013. Understanding relevance: An fMRI study. In *Advances in Information Retrieval: 35th European Conference on IR Research, ECIR 2013*. Springer, Berlin, Heidelberg, 14–25. https://doi.org/10.1007/978-3-642-36973-5_2

[54] Yashar Moshfeghi and Frank E. Pollick. 2018. Search Process as Transitions Between Neural States. In *Proceedings of the 2018 World Wide Web Conference* (Lyon, France) *(WWW '18)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 1683–1692. https://doi.org/10.1145/3178876.3186080

[55] Yashar Moshfeghi and Frank E. Pollick. 2019. Neuropsychological Model of the Realization of Information Need. *Journal of the Association for Information Science and Technology* 70, 9 (2019), 954–967. https://doi.org/10.1002/asi.24242

[56] Diane Nahl. 2007. Social–Biological Information Technology: An IntegratedConceptual Framework. *Journal of the American Society for Information Science and Technology* 58, 13 (2007), 2021–2046. https://doi.org/10.1002/asi.20690

[57] Flavio T.P. Oliveira, Anne Aula, and Daniel M. Russell. 2009. Discriminating the relevance of web search results with measures of pupil size. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 2209–2212.

https://doi.org/10.1145/1518701.1519038

[58] Sakrapee Paisalnan, Frank Pollick, and Yashar Moshfeghi. 2021. Towards Understanding Neuroscience of Realisation of Information Need in Light of Relevance and Satisfaction Judgement. In *Machine Learning, Optimization, and Data Science: 7th International Conference, LOD 2021, Grasmere, UK, October 4–8, 2021, Revised Selected Papers, Part I* (Grasmere, United Kingdom). Springer-Verlag, Berlin, Heidelberg, 41–56. https://doi.org/10.1007/978-3-030-95467-3_3

[59] Tam Pham, Zen Juen Lau, SH Annabel Chen, and Dominique Makowski. 2021. Heart Rate Variability in Psychology: A Review of HRV Indices and an Analysis Tutorial. *Sensors (Basel)* 21, 12 (2021), 3998. https://doi.org/10.3390/s21123998

[60] Zuzana Pinkosova, William J. McGeown, and Yashar Moshfeghi. 2023. Revisiting Neurological Aspects of Relevance: An EEG Study. In *Machine Learning, Optimization, and Data Science: 8th International Conference, LOD 2022, Certosa Di Pontignano, Italy, September 18–22, 2022, Revised Selected Papers, Part II* (Certosa di Pontignano, Italy). Springer-Verlag, Berlin, Heidelberg, 549–563. https://doi.org/10.1007/978-3-031-25891-6_41

[61] Sébastien Puma, Nadine Matton, Pierre-V Paubel, Éric Raufaste, and Radouane El-Yagoubi. 2018. Using theta and alpha band power to assess cognitive workload in multitasking environments. *International Journal of Psychophysiology* 123 (2018), 111–120.

[62] Rafael Ramirez and Zacharias Vamvakousis. 2012. Detecting Emotion from EEG Signals Using the Emotive Epoc Device. In *International Conference on Brain Informatics (Lecture Notes in Computer Science)*. Springer, Berlin, Heidelberg, 175–184. https://doi.org/10.1007/978-3-642-35139-6_17

[63] Bujar Raufi and Luca Longo. 2022. An Evaluation of the EEG Alpha-to-Theta and Theta-to-Alpha Band Ratios as Indexes of Mental Workload. *Frontiers in Neuroinformatics* 16 (2022), 44. https://doi.org/10.3389/fninf.2022.861967

[64] René Riedl, Fred Davis, and Alan Hevner. 2014. Towards a NeuroIS Research Methodology: Intensifying the Discussion on Methods, Tools, and Measurement. *Journal of the Association for Information Systems* 15, 10 (2014), 4. https://doi.org/10.17705/1jais.00377

[65] Ian Ruthven and Diane Kelly. 2011. *Interactive Information Seeking, Behaviour and Retrieval*. Facet. https://doi.org/10.29085/9781856049740

[66] Tefko Saracevic and Paul B Kantor. 1997. Studying the value of library and information services. Part I. Establishing a theoretical framework. *Journal of the American Society for Information Science* 48, 6 (1997), 527–542.

[67] Paul Sauseng, Wolfgang Klimesch, W Gruber, Michael Doppelmayr, Waltraud Stadler, and Manuel Schabus. 2002. The interplay between theta and alpha oscillations in the human electroencephalogram reflects the transfer of information between memory systems. *Neuroscience Letters* 324, 2 (2002), 121–124. https://doi.org/10.1016/S0304-3940(02)00225-2

[68] Reijo Savolainen. 2015. The interplay of affective and cognitive factors in information seeking and use: Comparing Kuhlthau's and Nahl's models. *Journal of Documentation* 1 (2015).

[69] Christina Schneegass, Max L Wilson, Horia A. Maior, Francesco Chiossi, Anna L Cox, and Jason Wiese. 2023. The Future of Cognitive Personal Informatics. In *Proceedings of the 25th International Conference on Mobile Human-Computer Interaction* (Athens, Greece) *(MobileHCI '23 Companion)*. Association for Computing Machinery, New York, NY, USA, Article 35, 5 pages. https://doi.org/10.1145/3565066.3609790

[70] Emery Schubert. 1999. Measuring emotion continuously: Validity and reliability of the two-dimensional emotion-space. *Australian Journal of Psychology* 51, 3 (1999), 154–165. https://doi.org/10.1080/00049539908255353

[71] Md. Hedayetul Islam Shovon, D (Nanda) Nandagopal, Jia Tina Du, Ramasamy Vijayalakshmi, and Bernadine Cocks. 2015. Cognitive Activity during Web Search. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Santiago, Chile) *(SIGIR '15)*. Association for Computing Machinery, New York, NY, USA, 967–970. https://doi.org/10.1145/2766462.2767784

[72] Winnie K. Y. So, Savio W. H. Wong, Joseph N. Mak, and Rosa H. M. Chan. 2017. An evaluation of mental workload with frontal EEG. *PLOS ONE* 12, 4 (04 2017), 1–17. https://doi.org/10.1371/journal.pone.0174949

[73] Alistair Sutcliffe and Mark Ennis. 1998. Towards a cognitive theory of information retrieval. *Interacting with Computers* 10, 3 (1998), 321–351. https://doi.org/10.1016/S0953-5438(98)00013-7 HCI and Information Retrieval.

[74] Robert S. Taylor. 1968. Question-Negotiation and Information Seeking in Libraries. *College and Research Libraries* 29, 3 (1968), 178–194.

[75] Pauline van der Wel and Henk Van Steenbergen. 2018. Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychon Bull & Review* 25 (2018), 2005–2015. https://doi.org/10.3758/s13423-018-1432-y

[76] Jacolien van Rij, Petra Hendriks, Hedderik van Rijn, R Harald Baayen, and Simon N Wood. 2019. Analyzing the Time Course of Pupillometric Data. *Trends in Hearing* 23 (2019), 2331216519832483. https://doi.org/10.1177/2331216519832483

[77] Ryen W. White and Ryan Ma. 2017. Improving Search Engines via Large-Scale Physiological Sensing. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Shinjuku, Tokyo, Japan) *(SIGIR '17)*. Association for Computing Machinery, New York, NY, USA, 881–884. https://doi.org/10.1145/3077136.3080669

[78] Kevin Wise, Hyo Jung Kim, and Jeesum Kim. 2009. The Effect of Searching Versus Surfing on Cognitive and Emotional Responses to Online News. *Journal of Media Psychology: Theories, Methods, and Applications* 21 (2009), 49–59. Issue 2. https://doi.org/10.1027/1864-1105.21.2.49

[79] Yingying Wu, Yiqun Liu, Ning Su, Shaoping Ma, and Wenwu Ou. 2017. Predicting Online Shopping Search Satisfaction and User Behaviors with Electrodermal Activity. In *Proceedings of the 26th International Conference on World Wide Web Companion* (Perth, Australia) *(WWW '17 Companion)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE,

855–856. https://doi.org/10.1145/3041021.3054226

[80] Ziyi Ye, Xiaohui Xie, Qingyao Ai, Yiqun Liu, Zhihong Wang, Weihang Su, and Min Zhang. 2024. Relevance Feedback with Brain Signals. *ACM Trans. Inf. Syst.* 42, 4, Article 93 (feb 2024), 37 pages. https://doi.org/10.1145/3637874

[81] Ziyi Ye, Xiaohui Xie, Yiqun Liu, Zhihong Wang, Xuesong Chen, Min Zhang, and Shaoping Ma. 2022. Towards a Better Understanding of Human Reading Comprehension with Brain Signals. In *Proceedings of the ACM Web Conference 2022* (Virtual Event, Lyon, France) *(WWW '22)*. Association for Computing Machinery, New York, NY, USA, 380–391. https://doi.org/10.1145/3485447.3511966