Computers in Human Behavior Reports 11 (2023) 100295

Contents lists available at ScienceDirect

Computers in Human Behavior Reports

journal homepage: www.sciencedirect.com/journal/computers-in-human-behavior-reports

Moderating effects of self-perceived knowledge in a relevance assessment task: An EEG study

Zuzana Pinkosova [a,*], William J. McGeown [b], Yashar Moshfeghi [a]

[a] NeuraSearch Laboratory, Computer & Information Sciences, University of Strathclyde, Glasgow, G1 1XH, UK
[b] Neuroanalycitcs Laboratory, The School of Psychological Sciences and Health, University of Strathclyde, Glasgow, G1 1XP, UK

ARTICLE INFO

Keywords:
Human-computer information retrieval (HCIR)
Relevance assessment
Binary relevance
EEG
ERPs
Self-perceived knowledge

ABSTRACT

Relevance assessment, a crucial Human-computer Information Retrieval (HCIR) aspect, denotes how well retrieved information meets the user's information need (IN). Recently, user-centred research benefited from the employment of brain imaging, which contributed to our understanding of relevance assessment and associated cognitive processes. However, the effect of contextual aspects, such as the searcher's self-perceived knowledge (SPK) on relevance assessment and its underlying neurocognitive processes, has not been studied. This work investigates the impact of users' SPK about a topic (i.e. 'knowledgeable' vs. 'not knowledgeable') on relevance assessments (i.e. 'relevant' vs. 'non-relevant'). To do so, using electroencephalography (EEG), we measured the neural activity of twenty-five participants while they provided relevance assessments during the Question and Answering (Q/A) Task. In the analysis, we considered the effects of SPK and specifically how it modulates the brain activity underpinning relevance judgements. Data-driven analysis revealed significant differences in cortical electrical activity modulated by searchers' SPK in the context of relevance assessment, suggesting that SPK affects cognitive processes associated with attention, semantic integration and categorisation, memory, and decision formation that underpin relevance assessment formation. Our findings are an important step toward a better understanding of the role users' SPK plays during relevance assessment.

1. Introduction

Despite the ongoing evolution of Human-computer Information Retrieval (HCIR), relevance assessment remains a fundamental construct and a major area of study in the field (Saracevic, 2007). Novel user-centred multidisciplinary research has significantly contributed to our understanding of relevance assessment through the investigation of users' behaviours and experiences within the information interaction context (Kelly, 2009). However, relevance assessments are complex, dynamic, multidimensional (Cool, Frieder, & Kantor, 1993; Cosijn & Ingwersen, 2000; Froehlich, 1994; Mizzaro, 1997, 1998; Schamber & Eisenberg, 1988; Schamber, Eisenberg, & Nilan, 1990) and often investigated in a context-independent manner (Jiang, He, Kelly, & Allan, 2017; Wang, 2010). Nonetheless, relevance assessment strongly depends on the users' cognitive states, perception, and knowledge (Barry, 1994; Ruthven, 2014), which provides psychological context determining the problem and situation at hand (Cosijn & Ingwersen, 2000; Ingwersen, 2006; Sanchiz, Chevalier, Fu, & Amadieu, 2017; Saracevic, 2007). This work, therefore, aims to better understand the role of users' self-perceived knowledge (SPK) within the relevance assessment context. Self-perceived knowledge refers to an individual's subjective assessment of their own knowledge and understanding of a particular topic (Park, Gardner, & Thukral, 1988). Subjective perception of insufficient or incomplete knowledge might drive the motivation to engage in information-seeking and searching behaviour in order to address the users' knowledge gaps (Kumar, 2013). Therefore, by considering users' subjective perceptions of their own knowledge, we can gain insight into the factors that influence their information-interaction behavior and subjective perception of relevance.

1.1. The SPK

Past HCIR studies have mainly focused on topical knowledge, referring to the relationship between one's prior knowledge and the conceptual aspects of the topic they engage in (Alexander, Schallert, & Hare, 1991). These studies have found that topical knowledge influences users' relevance criteria and information evaluation process (Fitzgerald, 2005; Ruthven, 2014; Vakkari & Hakala, 2000), as users rely on their

* Corresponding author. Computer & Information Sciences, University of Strathclyde, Glasgow, G1 1XH, UK.
E-mail address: zuzana.pinkosova@strath.ac.uk (Z. Pinkosova).

https://doi.org/10.1016/j.chbr.2023.100295
Received 14 December 2022; Received in revised form 16 March 2023; Accepted 30 April 2023
Available online 26 July 2023

knowledge to discriminate between relevant and non-relevant information. Furthermore, topical knowledge has also been shown to help users assess information credibility with higher accuracy (Jiang et al., 2017). Although topical knowledge plays an important role in information processing, users are often unaware of their knowledge anomalies (Versteeg & Steendijk, 2019) which significantly impact information search motivation, and decision-making (Radecki & Jaccard, 1995). From a neuroscientific perspective, information interaction is closely linked with reward circuity (Murayama, FitzGibbon, & Sakaki, 2019; Vellani, de Vries, Gaule, & Sharot, 2020) which might suggest that an increase in one's SPK state is related to a sense of pleasure or reward consequently modulating users' neurophysiological activity. This paper focuses on SPK, which refers to the self-assessment of knowledge that one believes they are holding (Park, 2001).

1.2. Relevance assessment in the context of SPK

Information science considers relevance assessment as a multidimensional (Cosijn & Ingwersen, 2000; Mizzaro, 1997; Saracevic, 2016), dynamic and complex process (Cool et al., 1993; Froehlich, 1994; Mizzaro, 1998; Schamber et al., 1990), which is difficult to quantify and which depends on the users' perception of information relating to the specific information need (IN) situation at a certain time point (Borlund, 2003; Levene, Bar-Ilan, & Zhitomirsky-Geffet, 2018; Mao et al., 2016; Saracevic, 2007). According to Ingwersen's cognitive theory of relevance (Ingwersen, 1992), relevance is underpinned by a range of cognitive processes, including perception, attention, memory, and reasoning. The user's information need is shaped by a complex set of contextual factors, including the user's task, goals, interests, and prior knowledge. Information is only considered relevant if it meets user's cognitive needs and interests. Therefore, relevance assessment, as a cognitive process, strongly depends on the user's internal context, such as the individual's knowledge and characteristics (Harter, 1992; Schamber & Eisenberg, 1988).

SPK plays an integral role in shaping cognition and influencing decision-making within information processing (Radecki & Jaccard, 1995) as it impacts perceived information importance value (Park et al., 1988). Despite the potential construct importance, SPK has not been investigated within the context of relevance assessment and past studies have predominantly focused on topical knowledge (Ruthven, 2014). Studies investigating topical knowledge during relevance assessment have found that topical knowledge can affect the interpretation of retrieved information (Ruthven, Baillie, & Elsweiler, 2007; Vakkari & Sormunen, 2004). However, the user's SPK is a better predictor of information-interaction behaviour than their topical knowledge (Radecki & Jaccard, 1995) as users are often unable to accurately assess their actual knowledge (Versteeg & Steendijk, 2019). The main aim of this work is to investigate the complex cognitive processes that underpin SPK within the relevance assessment context from a neuroscience perspective.

1.2.1. Capturing relevance assessment

The HCIR community has employed implicit and explicit feedback techniques to capture user's relevance assessment. Explicit feedback is easy to use but challenging to obtain due to the cognitive burden associated with it (Moshfeghi & Jose, 2013), as the user is required to explicitly state whether presented content is subjectively perceived as relevant or not (White, Ruthven, & Jose, 2002). Implicit feedback is an unobtrusive data collection method. Popular techniques used to measure implicit relevance assessment are mainly dwell time (document viewing time) (e.g. (Kelly & Belkin, 2004)), eye-tracking and pupillometry (Gwizdka, 2014), and/or the measurements of affective (Arapakis, Jose, & Gray, 2008), physiological and behavioural signals (Moshfeghi & Jose, 2013). Nevertheless, implicit feedback is often considered to be less accurate due to the noise associated with it (Allegretti et al., 2015). Currently, relevance assessment is receiving considerable attention from a neuroscientific point of view. The application of the neuroscientific approach was not only able to overcome conceptual ambiguities but also to introduce an innovative and effective method to capture relevance assessment in real-time.

1.2.2. Binary relevance assessment

Despite recent findings supporting the idea of categorical thinking (i.e. users divide retrieved results into 3–5 relevance categories) (Levene et al., 2018; Pinkosova, McGeown, & Moshfeghi, 2020; Zhitomirsky-Geffet, Bar-Ilan, & Levene, 2015), relevance assessment has been primarily considered in binary terms (i.e. 'relevant' vs 'non-relevant') (Saracevic, 2007). The binary division is considered to be a convenient approach, keeping the assessment cost low while maximising the number of relevant documents per topic, guaranteeing measure stability (Sormunen, 2002). Furthermore, binary relevance assessment can be accurately decoded based on physiological signals (Gwizdka, 2014; Gwizdka & Zhang, 2015).

1.2.3. Neurasearch research

NeuraSearch is an emerging interdisciplinary research field bridging Neuroscience and HCIR (Moshfeghi, 2021), which is able to bring new knowledge of IR phenomena as the field profits from direct access to neural signatures associated with user's mental processes (Kingphai and Moshfeghi, 2021a, 2022; Michalkova et al., 2022a, 2022b, 2022c) such as attention, cognitive workload and memory. Neuroscience has significantly contributed to improving and deepening the understanding of IN phenomenon (Moshfeghi, Triantafillou, & Pollick, 2016) (e.g. query formulation (Jacucci et al., 2019), search (Moshfeghi & Pollick, 2018; Zhang, Bao, & Xiao, 2019), relevance (e.g. (Kauppi et al., 2015; Pinkosova, McGeown, & Moshfeghi, 2022)) and search satisfaction (Paisalnan et al., 2022a, 2022b), leading to the potential development of novel information search models (Moshfeghi & Pollick, 2018) that can incorporate user's neurophysiological responses to the presented information. The most frequently used neuroimaging methods in the HCIR field have been functional magnetic resonance (fMRI) (Moshfeghi et al., 2013, 2016, 2019; Moshfeghi and Pollick, 2018, 2019, Paisalnan et al., 2021a, 2021b), magnetoencephalography (MEG) (Kauppi et al., 2015) and electroencephalography (EEG) (Allegretti et al., 2015; Barral, 2018; Golenia, Wenzel, Bogojeski, & Blankertz, 2018; Gwizdka, 2018; Gwizdka, Hosseini, Cole, & Wang, 2017; Jacucci et al., 2019; Kim & Kim, 2019; Kingphai & Moshfeghi, 2021b; Scharinger, Kammerer, & Gerjets, 2016; Slanzi, Balazs, & Velásquez, 2017; Wenzel, Bogojeski, & Blankertz, 2017). Within the context of relevance, the above-mentioned brain imaging techniques have been employed to investigate different neurological aspects of relevance, such as the brain's functional connectivity (Moshfeghi & Pollick, 2018), underlying cognitive processes and their timing (Allegretti et al., 2015). Past research has established that distinct relevance grades manifest themselves with specific brain activity on binary (e.g. Allegretti et al., 2015) and graded scale (Pinkosova et al., 2020).

1.2.4. NeuraSearch and relevance assessment

The neuroscientific approach might be categorised in two ways based on the experimental design used to investigate relevance assessment. The first line of brain-imaging research has considered users' IN and positioned relevance assessment within the HCIR task. Moshfeghi and colleagues (Moshfeghi et al., 2013) used fMRI to localise cortical activity differences during the processing of relevant vs non-relevant images that were related to visuospatial working memory (Moshfeghi et al., 2016; Moshfeghi & Pollick, 2018). Relevance assessment has also been studied within the context of the HCIR task for stimuli of different modalities, such as text (Jacucci et al., 2019), images (Allegretti et al., 2015) and videos (Kim & Kim, 2019). Allegretti et al. (2015) examined the processing of relevant vs non-relevant images, finding the most significant differences to occur between 500 and 800ms, reaching the peak in the central scalp areas. Kim & Kim (2019) explored the topical relevance of video skims and classified the neurological data based on specific patterns of electrical activity called event-related potentials (ERP), namely N400 and P600 components. The N400 and P600 ERP components are indicators of relevant and non-relevant assessments and are frequently associated with the processing of contextual information and decision-making. Apart from N400 and P600 ERP components, a recent relevance study has also observed significant differences in the processing of relevant vs. non-relevant content related to the P100 ERP component, suggesting a difference in processing effort when recognising relevant stimuli during relevance assessment (Pinkosova et al., 2022). Furthermore, recent findings have shown that relevance assessment can be automatically predicted using EEG data while the user engages with the HCIR task (Jacucci et al., 2019). The results of these studies suggest that human mental experience during the relevance assessment can be understood and accurately decoded.

Another approach employing EEG has placed relevance assessments in the word associations context (Eugster et al., 2014, 2016; Wenzel et al., 2017). Within the task, participants did not experience IN, but they judged associations between words and topics. The study findings have shown that neurological signals differ when subjects process relevant vs non-relevant words (Eugster et al., 2014). Later, Eugster et al. (2016) introduced a brain-relevance paradigm enabling information recommendation to users without any explicit user interaction, based on EEG signals alone evoked by users' text engagement.

Despite novel contributions and insight of previous neuroimaging studies, the complex nature of the user's SPK within the context of relevance assessment has not yet been investigated. It is not clear whether there are detectable neural signatures associated with SPK and if so, how do these signatures contribute to the formation of relevance assessments. Answering these research questions will play a key role in opening new doors to the design and implementation of novel HCIR techniques, which will be enabled to more accurately address and satisfy searchers' needs.

1.3. EEG recording and analysis preliminaries

An electroencephalogram (EEG) records, over time, the electrical potential of brain activity on the scalp surface that is generated by the activation of neurons. Sensors (also called electrodes), attached to specific scalp locations, enable the transfer of electrical activity from the scalp surface to the EEG input of the device. The EEG captures electric potential fluctuation changes between the channels and a reference point. The result is a wave representing the course of potential difference changes in time. Amplitude refers to the height of a waveform or the strength of the pattern in terms of microvolts (μV) of the EEG signal. The recorded signal consists of many waves with different characteristics. The rate at which the waveform data is sampled in order to convert it into a continuous digital format is known as the sampling rate (measured in Hz). The synchronisation between the behavioural responses of the participant and their brain signals is facilitated via the amplifier. Obtained neurological data may contain interfering elements at different frequencies with extracerebral origin (e.g. eye movements, muscle contractions or/and ambient electrical noise). Additionally, high-density EEG recordings are commonly associated with bad channels, which are common phenomena that arise due to various technical reasons, such as bad connection between the electrode and the scalp. To account for interfering elements, acquired (raw) data usually undergo a series of pre-processing steps (see Section 2.10) which aim to maximise the signal-to-noise ratio.

For the most accurate interpretation of brain activity, it is necessary to analyse the recorded neurological signal. A commonly used approach to analyse multichannel data between conditions is to quantify the difference of the topography in a given EEG segment or a time window of interest (i.e. epoch) and to test it for significance. This data-driven approach is not only applicable for the analysis of continuous EEG signal, but also for the analysis of ERPs. The ERP refers to scalp-recorded long latency electrical activity responses time-locked to an onset (start) of a specific event or stimulus. The ERP component represents a deflection from the baseline of EEG activity which correlates with cognitive processes. The names of ERP components begin with the letter P or N, which indicates the polarity of the component (i.e. positive vs negative). Next, the component is defined by a number that indicates its order or latency (from stimulus onset). In the data-driven ERP analysis, researchers often use statistical tests to identify regions-of-interest (ROIs), which refers to selected regions of neighbouring electrodes that jointly and significantly contribute toward neurophysiological phenomena of interest (Brooks, Zoumpoulaki, & Bowman, 2017). The data-driven identification compared to ERP component-driven analysis avoids the analytical biases introduced by apriori implication of known ERP components (Schmüser et al., 2014).

1.4. Capturing SPK states

We follow a common approach using post-trial assessment to evaluate participants' SPK states, which allows participants to be more cautious with the estimation of their SPK through the recognition of their anomalous knowledge states (Kruger & Dunning, 1999; Versteeg & Steendijk, 2019). However, both relevance assessment and SPK are dynamic, complex and subjective phenomena, which are difficult to quantify (Moshfeghi et al., 2013). Thus, the present study takes the neuroscience approach, which addresses the aforementioned challenges by offering the unique possibility of investigating these complex cognitive phenomena directly through the understanding of neurophysiological correlates of cognitive processes (Moshfeghi et al., 2013). Rather than focus our analyses on participants' general knowledge states, during the task, we capture users' SPK on a trial-by-trial basis. We aim to address two important questions:

- **RQ1:** "Are there clear and detectable neural manifestations associated with distinct users' SPK states during relevance assessment?"
- **RQ2:** "How do the neural mechanisms associated with different SPK states drive the cognitive processes underpinning the relevance assessment?"

This is the first study investigating user's SPK as a contextual aspect of relevance assessment during real-time information processing employing electrophysiological measurement. We capture the user's SPK, relevance assessments and associated brain activity in relation to the Q/A task. The data-driven approach employed in this study provides the benefit of avoiding potential analytical bias introduced by the restriction to distinct event-related potentials (ERPs) (Schmüser et al., 2014). Understanding brain activity associated with the user's cognitive states related to SPK could lead to innovative HCIR techniques improving retrieval performance and satisfying searchers' needs more effectively through the adaptation to individual differences.

2. Methodology

2.1. Participants

The study was carried out with a sample consisting of twenty-five individuals recruited using opportunistic sampling. Participants reported themselves to be neurologically and physically healthy with normal or corrected-to-normal vision. Seven participants were excluded from the final study analysis due to the high number of physiological artefacts present in the EEG data. The 18 remaining participants (11 females and 7 males) were between 19 and 39 years old and with a mean age of 24.5 and a standard deviation (SD) 4.91 years. Over half of the participants were students (55.56%), and the rest were either employed in skilled jobs (28.00%) or unemployed (16.67%). One participant reported being left-handed. Participants were either native English speakers (8) or had high English proficiency. On average, participants had an experience of 17.50 (SD = 3.88) years of formal education and indicated using search engines on average several times a day.

2.2. Experimental design

This user study followed a within-subject design in which participants engaged in the Q/A task. The aim of the study was to evaluate the differences in brain activity that were associated with participants' SPK states as they engaged in the relevance assessments during the Q/A task. The independent variables were user's SPK states (with two levels: "Knowledgeable" ('know'), "Not Knowledgeable" ('notknow')), and relevance assessments (with two levels: "Non-Relevant" ('nr') and "Relevant" ('rel')). The dependent variable was the EEG signal gathered during the Q/A task. We controlled the number of relevant and non-relevant answers presented to the participant, but we did not control the number of words each participant saw. This allowed us to simulate an information search and retrieval, as participants were not required to read through the whole answer. Instead, they were able to terminate the answer presentation once the relevance assessment was made.

2.3. Stimulus presentation

The stimuli were presented on a 22-inch colour Mitsubishi Diamond Pro 2040u NF CRT monitor (with a resolution of 2048 × 1536 and refresh rate of 75 Hz) using E-Prime 2.0. Participants were seated approximately 60 cm from the computer screen, and response keys were located on a QWERTY keyboard. All text events were presented in Arial font, size 16.

2.4. Questionnaires

Throughout the experiment, participants were asked to fill in the Entry, Post-Task and Exit Questionnaires. The Entry questionnaire was administered at the beginning of the experiment to gather demographic information about the participants (i.e. age, gender, handedness, occupation) and to determine their inclusion in the experiment. Inclusion criteria included individuals between 18 and 55 years of age, without any pre-existing neurological or psychiatric condition, and not under influence of drugs or medication that might impact the EEG signal recordings. There were no selection criteria based on handedness. After completing the task, participants filled in the Post-Task questionnaire, which assessed their perception of the task by presenting items covering difficulty, enjoyment, interest, stress, readability, relevance and understanding. At the end of the experiment, participants completed the Exit Questionnaire, designed to examine the participants' perception of their overall performance and to gather feedback regarding their experience to identify factors that may have influenced their performance (i.e. tiredness, clarity of task instruction, subjective performance satisfaction, simplicity of the procedure, monitor luminance, font size, perceived pressure, stimuli presentation speed, perceived time pressure).

2.5. Question-answering data sets

The data set employed in the study was developed and used by Moshfeghi, Triantafillou, and Pollick (Moshfeghi et al., 2016). We have chosen this data set as it has been proven effective in investigating HCIR phenomena from a neuroscience standpoint (Pinkosova et al., 2020, 2022). The data set was further adapted to address our research questions and expanded through the additional question and answer selection from TREC-8 and TREC-2001. These two Tracks were selected as they (i) cover a wide discipline range, (ii) they are independent of one another, and (iii) they provide a correct question answer. We ensured that the selected questions were accurate and not time-dependent or ambiguous. The data set contained 128 questions, answers, and relevance assessments in total. To reduce the fatigue, 64 questions were carefully selected for each participant to be balanced based on relevance (i.e. 50% relevant and 50% non-relevant), length (i.e. 50% long vs 50% short) and difficulty (i.e. 50% difficult vs 50% easy).[1] This was done to reduce any possible bias that could result from the focus on one form of Q/A. An example of an easy question presented to the participants was "What is epilepsy?", which was followed by the short, relevant answer "Epilepsy is a brain disorder characterised by seizures". The order of the presented questions was randomised also for each participant.

2.6. EEG recordings

Brain signals were acquired using the 128-channel geodesic sensor net (Electrical Geodesic Inc) and recorded within the standard EGI package Net Station 4.3.1. A Net Amps 200 amplifier was used for the recording and to facilitate the synchronisation between the behavioural response of the participant and their brain signals. To set the system for recording, we followed Electrical Geodesic Inc guidelines. We aimed to keep the electrode impedances below 50 kΩ, according to the recommended system value. Raw EEG data were recorded at a sampling rate of 1000 Hz and referenced to the vertex electrode (Cz). Before fitting the sensor net over the scalp, the electrodes were soaked in KCl electrolyte solution to facilitate conductivity between the skin and electrodes. The participant's head was measured to select the correct net size and determine the vertex point, which ensured the accurate placement of the net. The vertex point was determined using standardised procedures, measuring the halfway between inion and nasion and halfway between both bilateral preauricular points.

2.7. Procedure

The experiment was approved and carried out in accordance with the University of Strathclyde Ethics Committee guidelines. To ensure participants' informed consent, the researchers provided them with an information sheet explaining the purpose of the experiment before requesting that they sign a consent form indicating their willingness to participate voluntarily. Participants were explicitly informed of their right to withdraw from the experiment at any time, without needing to provide a reason, to ensure that they had complete control over their involvement in the study. After that, they filled in an Entry Questionnaire. Prior to the main experimental trials, participants underwent a number of training trials, which resembled the main experimental task. Participants were able to repeat the training until they confirmed to have a good understanding of the procedure. In total, every participant completed 64 trials. To avoid fatigue, the trails were split into two equally long blocks separated by a break. On average, participants were presented with 810.06 words (±134.77) and the main experimental task lasted approximately 53.69 min (±9.74). After completing the main experimental task, participants were instructed to fill out the Post-Task and Exit Questionnaires. A debriefing sheet was provided at the end of the experimental session.

2.8. Task

The study has embraced an artificial task, which is often preferable in EEG research as it can be more tightly controlled and allow for the manipulation of specific variables to isolate cognitive processes of interest. This can be particularly useful for investigating the neural mechanisms underlying complex cognitive processes such as relevance assessment. Complex real-world search tasks can also introduce more noise into the EEG signals. This is because these tasks typically require more cognitive effort and physical engagement from the participant, which can lead to more variability in the EEG data as well as the

[1] To assess the difficulty level, two annotators separately judged question difficulty (i.e. difficult vs easy). The overall inter-annotator agreement was reasonably high (Cohen's kappa, $\kappa = 0.72$).

presence of muscle and ocular artefacts contaminating the signal.

The schematic task representation is depicted in Fig. 1. At the beginning of the task, participants were presented with the instructions. Next, they viewed a question from the data set presented in a randomised order. Once the participant read and fully understood the question, they pressed a button to start. Next, a fixation cross was presented for 950ms, which indicated the location of the answer presentation. To control free-viewing and minimise the presence of any confounding artefacts (i.e. saccades), the answer was presented in the middle of the screen word by word. Each word was presented for 950ms, which has been deemed to be a sufficient duration to model fluent reading and to avoid the overlapping effect of two consecutive words on the ERPs (Eugster et al., 2016). The ERP components were, therefore, time-locked to the word presentation. This approach has been commonly applied to examine neurological signatures of reading in the ERP studies (e.g. Dien, Michelson, & Franklin, 2010). Participants were instructed to carefully read individual words that would form either relevant or non-relevant answers and to assess their relevance on a binary scale (i.e. relevant vs non-relevant). Once participants gathered enough information and submitted their relevance judgements, they had the option to terminate the word presentation sequence (and to continue to the next step), or to view the sequence in full. As brain activity was recorded during the reading, to avoid the possibility of confounding hemispheric effects (due to motor planning or execution), counter-balancing was used, and participants were instructed to interact with the keyboard using either their left or right hand. Participants were then asked whether they already knew the answer to the presented question (i.e. SPK assessment - ‘know’ vs ‘notknow’). The SPK evaluation was performed after completing each trial (after seeing the answer to the question). This allowed participants to make a more informed judgement about their knowledge state, as opposed to asking participants about their knowledge state prior to seeing the answer. This is because participants may not be completely aware of whether they know the answer (e.g., there are difficulties in distinguishing whether someone actually knows something or is instead simply familiar with it, whether they can recall or only recognise information they believe to have knowledge of, and additionally levels of confidence and criterion levels for judgements of this nature can vary across participants) (Kruger & Dunning, 1999; Versteeg & Steendijk, 2019). In other words, asking the Q after the participant sees the answer, can make participants aware of any anomalies in their knowledge (Versteeg & Steendijk, 2019). The interpretation of relevance assessment categories depended on each participant’s subjectively perceived information accumulation process, which enabled capturing the subjective nature of relevance assessment (Saracevic, 2007).

2.9. Pilot studies

Before commencing the main user study, we performed a pilot study with 4 participants whose data were not included in the final analysis. Based on the participants’ experience and feedback, we adjusted the study design and presentation. After the pilot study, it was determined that the participants were able to complete the user study without problems, including having adequate time to comfortably read and respond to presented stimuli, and that the system was correctly logging participants’ behavioural responses and neurological signals.

2.10. Pre-processing steps

The brain activity was recorded from participants as they engaged with relevant and non-relevant content, up to the point where the participant stopped the answer presentation. To prepare data for analysis, an automated pre-processing pipeline was built through the implementation of the EEGLAB tools. The EEG data pre-processing steps were based on Makoto’s Pre-processing Pipeline .[2] All collected neurological data were first visually inspected. Then a low-pass filter of 30Hz was applied. We down-sampled the data from 1000Hz to 250Hz. Downsampling, a commonly applied procedure, is used to reduce file size for easier data manipulation. Then a high pass filter of 0.3Hz was applied. Filtering is another common procedure used to attenuate frequencies associated with noise rather than a signal of interest. We then automatically rejected bad channels (EEG sensors that were not functioning properly during the data acquisition and that were high in noise throughout the task). On average, we removed 13.94 bad channels (±7.67). The re-referencing to average (across all electrodes) was subsequently performed (to provide an approximation of zero μV for the reference at each timepoint). The CleanLine EEGLAB plugin was used to filter line noise. All epochs (the time windows of interest) were then extracted from 200ms before stimulus presentation to 950ms afterwards. To detect and remove components associated with ocular, cardiac and muscular artefacts based on their power spectrum and time-course, we performed Independent Component Analysis and rejected artefacts using ADJUST (Mognon, Jovicich, Bruzzone, & Buiatti, 2011). A mean number of 18.17 (±9.17) components were removed. Bad channels were interpolated using a spherical interpolation method. The spherical interpolation method refers to a common data-repair method that computes interpolated data in the bad channels based on the good channel values (Perrin, Pernier, Bertrand, & Echallier, 1989). Next, we removed the two outermost belts of electrodes of the sensor net [3] which are prone to show muscular artefacts, following the approaches of Bian et al. (2014) and Calbi et al. (2019). Epochs were then extracted again from 100ms before stimulus presentation to 950ms afterwards based on the stimulus labels for every condition of interest (i.e. ‘know’, ‘notknow’, ‘know_rel’, ‘notknow_rel’, ‘know_nr’ and ‘notknow_nr’). We used automatic epoch rejection based on thresholding (i.e. rejecting epochs by detecting outlier values greater than ±100 μV). The mean number and SD of accepted and rejected epochs are displayed in Table 1. All epochs were baseline corrected. After pre-processing the data, epochs of interest were grand averaged.

2.11. Statistical analysis of EEG data

Participants’ brain activity was recorded for self-perceived known vs not known (i.e. ‘know’ vs ‘notknow’) information within the relevance assessment task. After data pre-processing, 49.54% of accepted epochs were marked as ‘know’ and 50.46% as ‘notknow’. To test for statistically significant differences in the neurological processing associated with the assessment of ‘know’ vs ‘notknow’ information, we employed a data-driven approach, which is particularly effective in whole-brain analysis of complex mental phenomena as it minimises the upfront assumptions and allows for the contribution of many distinct areas at different time points (Schmüser et al., 2014). To identify significant cortical differences, we compared the values for 109 electrode pairs at every time point (every 4ms, 237-time points in total) over the 100–950ms time window. The initial time interval (0–100ms) was excluded from the main analysis as we were not interested in the initial sensory processing of stimulus features (Liu et al., 2020). The data-driven approach applied a non-parametric permutation-based paired *t*-test (1000 permutations) using the *statcond* function implemented in the EEGLAB (Delorme & Makeig, 2004). Differences were considered significant at a threshold of $p < 0.05$.

[2] https://sccn.ucsd.edu/wiki/Makoto's_preprocessing_pipeline

[3] We removed 38 peripheral channels: E1, E8, E14, E17, E21, E25, E32, E38, E43, E44, E48, E49, E56, E57, E63, E64, E68, E69, E73, E74, E81, E82, E88, E89, E94, E95, E99, E100, E107, E113, E114, E119, E120, E121, E125, E126, E127, E128.

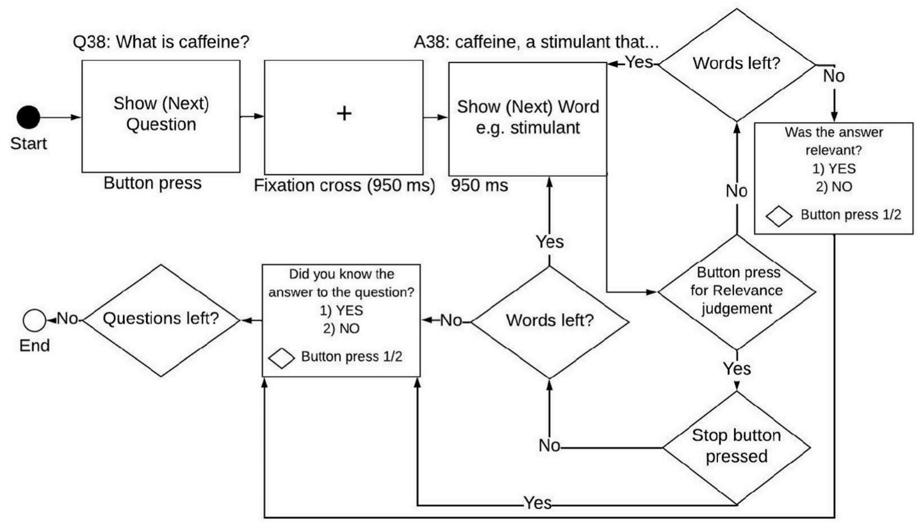

Fig. 1. The figure illustrates the task structure from the left (Start) to finish (End). The process is repeated for all 64 questions.

Table 1
The Mean number and SD of rejected epochs for every condition of interest.

Condition		Rejected Epochs		Accepted Epochs	
		Mean	SD	Mean	SD
know		52.78	65.42	261.28	114.35
notknow		88.61	104.60	407.39	142.12
know	rel	26.78	34.90	141.61	66.45
	nr	26.00	31.92	119.67	59.60
notknow	rel	42.61	57.83	188.11	76.07
	nr	46.00	49.91	219.28	94.27

2.12. ROIs

As the present study utilises a data-driven approach, for optimal detection of effects, the ROIs were determined based on statistically significant differences between compared conditions of interest. Therefore, we used the features of the data under analysis to position the ROIs. We were not interested in isolated electrodes where a test statistic might happen to be large. Instead, we applied the method utilised by Laganaro and colleagues (Laganaro & Perret, 2011). To identify potential ROIs, we only considered clusters with at least five electrodes next to each other extending over at least 20 ms and retained with an alpha criterion of 0.05 (Laganaro & Perret, 2011).

3. Results

3.1. Questionnaire Results

Prior to the main experimental result analysis, we examined participants' task perception using the Post-task Questionnaire. Additionally, we analysed the Exit Questionnaire results to understand participants' subjectively perceived performance and their task-related impressions. Both questionnaires used a 7-point Likert Scale (answers: 1: "Strongly Disagree", 2: "Disagree", 3: "Somewhat Disagree", 4: "Neither Agree nor Disagree", 5: "Somewhat Agree", 6: "Agree", 7: "Strongly Agree").

The results of the Post-task Questionnaire shown in Fig. 2 indicate that participants found the task (M = 5.94, SD = 1.35), questions (M = 6.06, SD = 0.73) and selected question topics (M = 5.94, SD = 1.11) somewhat interesting. Perceived difficulty of the task (M = 4.39, SD = 1.85), questions (M = 4.17, SD = 1.69) and selected question topics (M = 4.17, SD = 1.69) were rated as moderate. Presented questions (M = 5.78, SD = 1.40) and task in general (M = 5.72, SD = 1.23) were overall considered readable. Additionally, both, questions (M = 5.61, SD = 1.58) and task (M = 5.56, SD = 1.20) were also considered understandable. Overall, participants indicated that they somewhat enjoyed the task (M = 5.39, SD = 1.42). On average, participants felt moderate physical comfort (M = 5.22, SD = 1.56) and task was not rated as too stressful (M = 3.44, SD = 1.92). Questions selected for the experiment were perceived by participants as moderately familiar (M = 4.95, SD = 1.26) and relevant to them (M = 5.28, SD = 1.53). In general, the results of the Post-Task Questionnaire indicate that participants did not perceive any difficulties with the experimental design that might have caused them discomfort and impacted their engagement.

The Exit Questionnaire results, displayed in Fig. 3, suggest that despite participants feeling under moderate pressure (M = 4.22, SD = 1.17) and experiencing some degree of tiredness (M = 4.78, SD = 1.35), they were overall satisfied with their performance (M = 5.72, SD = 1.07) and found the procedure to be easy (M = 6.06, SD = 0.87). They found task instructions to be clear (M = 6.67, SD = 0.49), and font size (M = 6.61, SD = 0.78), monitor luminance (M = 5.89, SD = 1.18) and presentation speed (M = 6.56, SD = 0.62) were perceived as appropriate. Furthermore, participants felt that they had enough time to submit the response by pressing the button (M = 6.22, SD = 0.94). In general, the results of the Exit Questionnaire indicate that participants did not perceive any difficulties with the experimental design that might have made caused them discomfort and impacted their engagement.

3.2. Relevance assessment

The data-driven approach was used to investigate the brain activity differences associated with 'rel' vs 'nr' assessments. Our findings are consistent with the previous studies, suggesting that the processing of content perceived as 'rel' is associated with higher P300 and P600/Late Positive Component (LPC) amplitudes [4] (Allegretti et al., 2015; Eugster et al., 2014, 2016). On the other hand, processing of 'nr' content is associated with larger N400 deflections, which is consistent with other studies (e.g. Eugster et al., 2016; Kim & Kim, 2019). However, we do not expand on these findings as it is not the main scope of this work.

[4] The labels LPC and P600 are frequently used interchangeably. Past studies have frequently associated relevance assessment with the P600 ERP component (e.g. Allegretti et al., 2015; Eugster et al., 2014; Eugster et al., 2016). However, the P600 component is associated with 'syntactic re-analyses' in language studies. Therefore, the label LPC might be more appropriate to use while focusing on relevance assessment, as the LPC has been linked to memory and recognition.

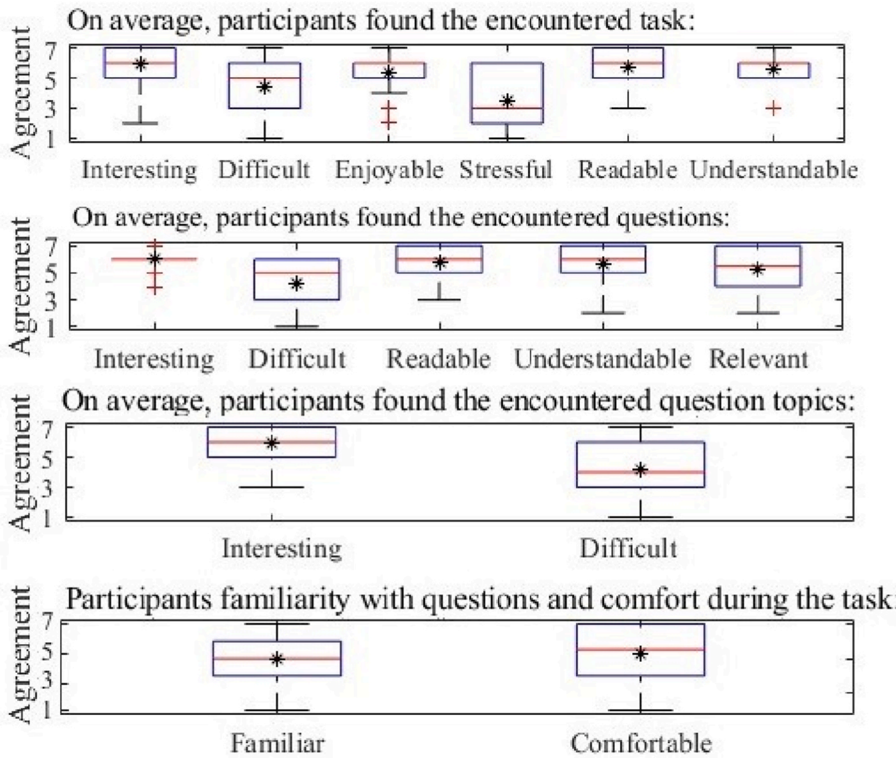

Fig. 2. Post-task questionnaire.

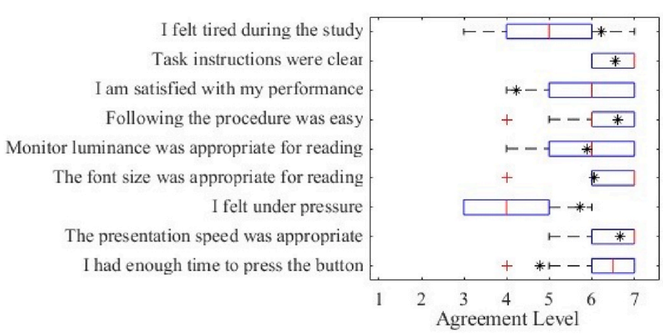

Fig. 3. Exit questionnaire.

3.3. *Effects of SPK*

The data-driven comparison of 'know' and 'notknow' conditions (irrespective of relevance assessment) revealed no statistically significant differences in brain activity. On the other hand, the comparisons of 'know_rel' vs 'notknow_rel' and 'know_nr' vs 'notknow_nr' conditions were associated with significant brain signal differences within multiple time intervals and with wide scalp distributions. The main findings of significant pairwise comparison of the conditions of interest are displayed in Table 2. The Time Window column presents the specific time intervals when ERP components were observed, the ERP column indicates associated ERP components, the electrode cluster column lists the significant electrode clusters and the cortical region indicates the cortical location where the corresponding ERP demonstrates statistical significance. SPK, therefore, has an effect on relevance assessment and can modulate this process at the neural level.

3.3.1. *100–350ms*

3.3.1.1. Non-relevant assessments. The earliest neural activity differences for information assessed as not relevant emerged in the 100–300ms interval for the comparison of 'know_nr' vs 'notknow_nr' conditions. The 'know_nr' condition was associated with a significantly greater right centro-parietal positivity compared to the 'notknow_nr' condition. Significant electrode clusters, time intervals and ERP waveforms, as well as topographic plots, are displayed in Fig. 4, row I. Given the topographies and waveform peaks at around 300ms post-stimulus, the differences are likely to reflect variability in the P300/Centro-parietal positivity (CPP) (similar distributions are reported, e.g., by Tagliabue et al., 2019; Twomey, Murphy, Kelly, & O'Connell, 2015). The higher amplitude observed when people indicated to have SPK might suggest processing ease associated with reduced cognitive load (see e.g. Polich, 2007).

Table 2
Significant differences in ERP components and related electrode clusters in the pairwise comparison of conditions of interest.

Condition comparison	Time window	ERP	Electrode Cluster	Cortical Region
Non-Relevant Assessments: know vs. notknow	100–300ms	P300/CPP	E7 E31 E78 E79 E80 E85 E86 E87 E93 E105 E106 E111 E112	Right Centro-parietal
	400–500ms	N400	E85 E91 E92 E93 E97 E98 E103	Right Centro-parieto-temporal
Relevant Assessments: know vs. notknow	250–300ms	P300/CPP	E78 E80 E86 E87 E104 E105	Right Centro-parietal
	300–350ms	N400	E29 E30 E40 E35 E36 E42 E47	Left Central
	350–400ms	N400	E3 E10 E11 E16 E18	Bilateral Frontal
	600–700ms	LPC	E53 E54 E55 E60 E61 E67 E72 E75 E76 E77 E78 E86	Bilateral Centro-parietal

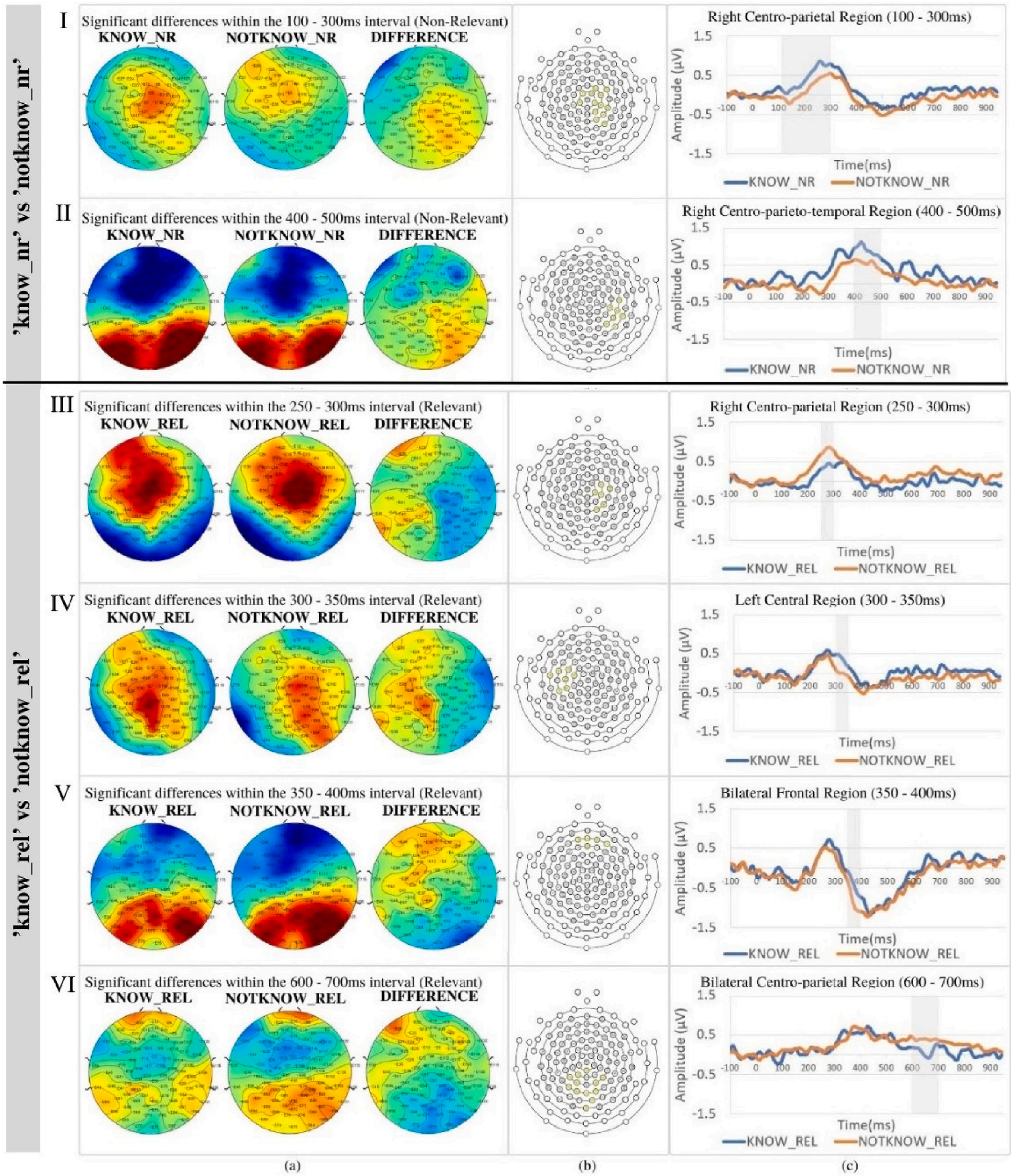

Fig. 4. (a) Topographic plots for conditions of interest including a mean difference plot for each significant time-window. Reddish scalp topography colours indicate positive ERP values, whereas bluish colours indicate negative ERP values. (b) The 128-channel net graph with highlighted statistically significant electrode sites for each significant time interval (see Table 2). (c) The comparison of grand averaged ERP waveforms for every condition of interest. Grand averages were calculated for each group of electrodes that showed significant difference within the time period of interest. Significant time intervals are highlighted in grey. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

3.3.1.2. Relevant assessments. The comparison of 'know_rel' and 'notknow_rel' conditions was associated with statistically significant differences in the right-centro-parietal and left central regions. The differences in the right-centro-parietal region were significant within the 250–300ms time interval and reflected greater positivity in the no knowledge condition, whereas the differences in the left central region were significant within the 300–350ms time interval and reflected greater positivity when SPK was reported. Significant electrode clusters, time intervals and ERP waveforms, as well as topographic plots, are displayed in Fig. 4, row III and IV. The results demonstrate that for information judged as relevant, a lack of SPK is linked to an earlier P300 rise and a higher peak over the right centro-parietal region, whereas SPK is linked to a more sustained P300 over the left central region. Interpretation of these seemingly contradictory effects should be made cautiously as P300/CPP amplitudes are known to be modulated by a variety of factors. As the reason is unlikely to be processing ease (which might be expected when people report having SPK), higher amplitude P300/CPP components in response to relevant information when people do not report having SPK might instead reflect differences in memory access and selective attention allocation underlying the decision-making process (van Vugt, Beulen, & Taatgen, 2019).

3.3.2. 350–500ms

3.3.2.1. Non-relevant assessments. The comparison of 'know_nr' and 'notknow_nr' conditions revealed significant differences in the right centro-parieto-temporal cluster within the 400–500ms time interval, as displayed in Fig. 4, row II. The significant differences were driven by the higher centro-parieto-temporal positivity associated with 'know_nr' compared to the 'notknow_nr' condition. Posterior positivity of this nature has been shown to co-occur with the N400 (Savostyanov et al., 2020). If interpreting this difference in the N400 context, the greater positivity may indicate that SPK attenuates semantic incongruity (e.g., perhaps through a process where SPK informs the participant that the information is not relevant, and they, therefore, do not focus as intently on the relationship and the incongruence between the answer and question, as someone who does not have SPK).

3.3.2.2. Relevant assessments. The comparison of 'know_rel' and 'notknow_rel' conditions revealed significant differences in the bilateral frontal region within the 350–400ms time interval (see Fig. 4, row V). Greater frontal negativity was observed in the 'notknow_rel' compared to the 'know_rel' condition. The negativity reflects the N400 component, which has been previously described (Savostyanov et al., 2020; Song et al., 2020; Spironelli & Angrilli, 2021). The decreased N400 deflection, when judging information to be relevant and aligned with the question, appears to indicate that SPK helps to decrease semantic incongruity and to integrate the words into context (Dien et al., 2010).

3.3.3. 600–700ms

3.3.3.1. Non-relevant assessments. There were no significant differences between the 'know' vs 'notknow' conditions for non-relevant information in the 600–700ms time-window.

3.3.3.2. Relevant assessments. Significant differences between 'know_rel' vs 'notknow_rel' conditions were observed within the 600–700ms time-window over the centro-parietal region (see Fig. 4, row VI). The differences were associated with higher positive-going ERP amplitudes associated with the processing of 'notknow_rel' compared to 'know_rel' information. The topographic distribution with a characteristic posterior positivity can be attributed to the LPC component. The LPC component is a positive-going deflection, emerging around 600ms post-stimulus usually largest over the medial posterior brain areas (Curran & Dien, 2003; Friedman & Johnson, 2000). Greater LPC amplitudes have been associated with information accumulation and decision-making processes (Mueller, White, & Kuchinke, 2017) and reflect effort invested in working memory maintenance. Additionally, the LPC reflects the information learning process (Wachinger et al., 2018) through codification and strengthening of episodic memory (Bermúdez-Margaretto, Beltrán, Cuetos, & Domínguez, 2019). Greater posterior positivity across 'notknow_rel' compared to the 'know_rel' condition might therefore reflect the enhanced episodic memory activation, enabling the lexico-semantic facilitation of learning novel information.

4. Discussion

The current experiment was carried out to investigate the role of SPK when making relevance assessments. The main finding, which addresses **RQ1**, is that there are significant differences in neural activity associated with the user's SPK when they perceive information as relevant or as non-relevant. Data-driven analyses revealed distinct significant time intervals and cortical differences driven by the self-perceived level of knowledge the user had about the question during relevance assessment. The differences in neural activity suggest that a user's SPK affects a variety of cognitive processes, which underpin relevance assessment formation, such as attentional engagement, perception of semantic relatedness and working memory engagement (addressing **RQ2**).

4.1. SPK & non-relevance assessment

When judging information to be non-relevant, SPK was associated with greater P300/CPP amplitudes, which may reflect greater processing ease in terms of memory access and retrieval. This processing ease might be associated with one's perceived ability to retrieve knowledge stored in memory (Radecki & Jaccard, 1995) which might, in turn, guide the assessment of information relevance from text sections (Park, 1993).

Another key difference emerged within the 400–500ms time interval in relation to right centro-parieto-temporal positivity (during a time period corresponding with the N400). SPK might facilitate the cognitive expectancy process and potentially help with information integration. If the positivity is taken to reflect the same processes as the N400 (given the bipolar representation across the scalp), then the greater amplitude in relation to SPK might reflect a greater degree of perceived semantic congruency (e.g., the answer is not relevant and the participant is aware of that). Users with SPK might experience reduced uncertainty levels and make more accurate information relevance predictions (Jiang et al., 2017).

4.2. SPK & relevance assessment

The results suggest that SPK within relevant information processing is associated with P300/CPP in two ways. During an earlier significant time interval (and with a right centro-parietal focus), the P300/CPP peak amplitudes may be influenced by the amount of cognitive control (Liu et al., 2020), referring to high-level executive functions such as attention, salience detection, working memory and task management. Furthermore, the P300/CPP is modulated by motivational factors, such as the level of engagement and interest of the participant. Early processing of relevant information for which one has no SPK might therefore involve complex attentional selection, stimulus evaluation, and evidence accumulation processes. This initial period associated with the P300/CPP component could also be linked to intrinsic motivation based on how participants subjectively perceive their own knowledge gaps, supporting the study of Kumar et al. (Kumar, 2013). However, in the later time interval (relating to the left central site), the P300/CPP may be related to a process such as recognition of previously encountered information (Meixner & Rosenfeld, 2014).

A reduction of the N400 within the time interval of 350–400ms, more prominent for subjectively-perceived known relevant information, might be related to semantic information retrieval (Dien et al., 2010). The N400 amplitude positively correlates with the ease of semantic processing (Voss & Federmeier, 2011), and information recognition during the presentation of self-important information (Savostyanov et al., 2020). This may suggest that SPK decreases cognitive effort when processing information within a subjectively relevant context (Debruille, 2007). It is possible that the P300/CPP component and N400 deflections associated with the processing of subjectively-perceived known relevant information are interdependently modulating relevance assessments (Arbel, Spencer, & Donchin, 2011), as both of these components have been frequently linked to relevance processing (e.g. Pinkosova et al., 2020). However, further research is required to provide clarification. Another important difference was seen in the LPC, which is commonly reported to follow the N400 (Stróżak, Bird, Corby, Frishkoff, & Curran, 2016) and it is a key component that relates to memory-based decisions (Ratcliff, Sederberg, Smith, & Childers, 2016). No SPK conditions might, therefore, require higher memory effort during decision-making tasks that require relevance considerations (Yang et al., 2019). Also, past studies have reported that learning is correlated with an increase in LPC amplitude (Wachinger et al., 2018) which supports Ingwersen's Cognitive Theory, suggesting that HCIR facilitates information transfer into knowledge and novel cognition (Ingwersen, 1994). Furthermore, the LPC amplitude might be associated with reward and positive emotional valence (Yan, Liu, Li, Zhang, & Cui, 2017), playing a role in the motivational modulation of attention during the presentation of stimuli that are relevant to users' goals or INs (Murayama et al., 2019; Vellani et al., 2020). This further highlights the importance of affective features in the relevance assessment reported in previous studies (Arapakis, Konstas, & Jose, 2009; Moshfeghi & Jose, 2013). The absence of differences in the LPC component during the non-relevant content processing might suggest that there are no differences in memory effort invested in maintaining task-relevant working memory representations (Gunseli, Meeter, & Olivers, 2014).

A possible study limitation might be related to the introduction of IN as an external and artificial factor through Q/A task. To address this issue, future studies should consider incorporating more naturalistic representations of participants' true INs, such as allowing them to submit their own queries that are better representations of their actual INs. This research primarily focused on one aspect of IR and, therefore, examining relevance while taking into account all interactive IR components and incorporating participants' actual INs could further enhance the validity of the study. It is important, however, to consider the cognitive demands as well as the physical demands placed on the participants resulting from their interactions with the system while submitting queries and searching through the results. Such a complex system interaction might introduce noise and variability into the EEG data, making it more difficult to accurately interpret the neural activity that is related to information processing. Another general limitation of the study is that participants were presented with questions and answers word-by-word instead of continuous text. While this method is commonly used in many EEG studies that examine textual processing to minimise eye-movement-related artefacts (Dimigen, Sommer, Hohlfeld, Jacobs, & Kliegl, 2011), presenting participants with continuous text would be a more appropriate way to simulate naturalistic information interaction. Furthermore, we did not select the participants on the basis of their demographics or characteristics. Therefore, a future avenue of research investigation could involve examining how these demographic factors might influence participant perceptions or behaviours, their neurological signatures, and how these effects may vary depending on the context or experimental conditions during IR task.

5. Conclusion

To conclude, our findings demonstrate the potential of exploiting differences in SPK during relevance assessments to gain a better understanding of the cognitive and neural processes underlying relevance assessments. Our findings indicate that there are significant variations in neural activity among users who report having SPK versus those who do not, suggesting that SPK is a crucial contextual factor that influences how users evaluate relevance. This modulation of users' underlying cognitive processes is in line with Ingwersen's cognitive relevance theory (Ingwersen, 1999). In particular, we observed significant ERP differences in P300/CPP, N400, LPC, suggesting the ease of cognitive processing (attention, semantic integration and categorisation, memory, and decision formation) when participants had indicated to have self-perceived knowledge of the answer to the question. Our results strengthen the theoretical basis of IR and provide a foundation for future research to further explore how SPK interacts with other factors in relevance assessment, such as difficulty level or graded relevance assessment. By operationalising the concept of SPK for the HCIR community, our study paves the way for improved information retrieval systems that can take into account the user's SPK and cognitive states when presenting search results. Overall, this explorative study highlights the potential of leveraging SPK in relevance assessments to gain insights into the complex cognitive and neural mechanisms involved in this process. By investigating the cognitive and contextual factors, such as SPK, that influence relevance assessment, we can better understand how individuals process and evaluate information to design better information retrieval systems.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

This work was supported by the Engineering and Physical Sciences Research Council [grant number EP/R513349/1].

References

Alexander, P. A., Schallert, D. L., & Hare, V. C. (1991). Coming to terms: How researchers in learning and literacy talk about knowledge. *Review of Educational Research, 61*(3), 315–343.

Allegretti, M., Moshfeghi, Y., Hadjigeorgieva, M., Pollick, F. E., Jose, J. M., & Pasi, G. (2015). When relevance judgement is happening? An eeg-based study. In *Sigir '15* (pp. 719–722). NY, USA: ACM.

Arapakis, I., Jose, J. M., & Gray, P. D. (2008). Affective feedback: An investigation into the role of emotions in the information seeking process. In *Sigir '08* (pp. 395–402). NY, USA: ACM.

Arapakis, I., Konstas, I., & Jose, J. M. (2009). Using facial expressions and peripheral physiological signals as implicit indicators of topical relevance. In *Proceedings of the 17th ACM international conference on Multimedia* (pp. 461–470).

Arbel, Y., Spencer, K. M., & Donchin, E. (2011). The n400 and the p300 are not all that independent. *Psychophysiology, 48*(6), 861–875.

Barral, O. (2018). *Implicit interaction with textual information using physiological signals*. Helsingin yliopisto.

Barry, C. L. (1994). User-defined relevance criteria: An exploratory study. *JASIST, 45*(3), 149–159.

Bermúdez-Margaretto, B., Beltrán, D., Cuetos, F., & Domínguez, A. (2019). Novel word learning: Event-related brain potentials reflect pure lexical and task-related effects. *Frontiers in Human Neuroscience, 13*, 347.

Bian, Z., Li, Q., Wang, L., Lu, C., Yin, S., & Li, X. (2014). Relative power and coherence of eeg series are related to amnestic mild cognitive impairment in diabetes. *Frontiers in Aging Neuroscience, 6*, 11.

Borlund, P. (2003). The concept of relevance in ir. *JASIST, 54*(10), 913–925.

Brooks, J. L., Zoumpoulaki, A., & Bowman, H. (2017). Data-driven region-of-interest selection without inflating type i error rate. *Psychophysiology, 54*(1), 100–113.

Calbi, M., Siri, F., Heimann, K., Barratt, D., Gallese, V., Kolesnikov, A., et al. (2019). How context influences the interpretation of facial expressions: A source localization high-density eeg study on the "kuleshov effect". *Scientific Reports, 9*(1), 1–16.

Cool, C., Frieder, O., & Kantor, P. (1993). Characteristics of text affecting relevance judgments. *Proceedings of the 14th National Online Meeting, 14*.

Cosijn, E., & Ingwersen, P. (2000). Dimensions of relevance. *IP&M, 36*(4), 533–550.

Curran, T., & Dien, J. (2003). Differentiating amodal familiarity from modality-specific memory processes: An erp study. *Psychophysiology, 40*(6), 979–988.

Debruille, J. B. (2007). The n400 potential could index a semantic inhibition. *Brain Research Reviews, 56*(2), 472–477.

Delorme, A., & Makeig, S. (2004). Eeglab: An open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9–21.

Dien, J., Michelson, C. A., & Franklin, M. S. (2010). Separating the visual sentence n400 effect from the p400 sequential expectancy effect: Cognitive and neuroanatomical implications. *Brain Research, 1355*, 126–140.

Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A. M., & Kliegl, R. (2011). Coregistration of eye movements and eeg in natural reading: Analyses and review. *Journal of Experimental Psychology: General, 140*(4), 552.

Eugster, M. J., Ruotsalo, T., Spapé, M. M., Barral, O., Ravaja, N., Jacucci, G., et al. (2016). Natural brain-information interfaces: Recommending information by relevance inferred from human brain signals. *Scientific Reports, 6*, Article 38580.

Eugster, M. J., Ruotsalo, T., Spapé, M. M., Kosunen, I., Barral, O., Ravaja, N., et al. (2014). Predicting term-relevance from brain signals. In *Sigir '14* (pp. 425–434). NY, USA: ACM.

Fitzgerald, M. A. (2005). *Skills for evaluating web-based information*.

Friedman, D., & Johnson, R., Jr. (2000). Event-related potential (erp) studies of memory encoding and retrieval: A selective review. *Microscopy Research and Technique, 51*(1), 6–28.

Froehlich, T. J. (1994). Relevance reconsidered—towards an agenda for the 21st century: Introduction to special topic issue on relevance research. *JASIST, 45*(3), 124–134.
Golenia, J.-E., Wenzel, M. A., Bogojeski, M., & Blankertz, B. (2018). Implicit relevance feedback from electroencephalography and eye tracking in image search. *Journal of Neural Engineering, 15*(2), Article 026002.
Gunseli, E., Meeter, M., & Olivers, C. N. (2014). Is a search template an ordinary working memory? Comparing electrophysiological markers of working memory maintenance for visual search and recognition. *Neuropsychologia, 60*, 29–38.
Gwizdka, J. (2014). Characterizing relevance with eye-tracking measures. In *IIiX '14* (pp. 58–67). NY, USA: ACM.
Gwizdka, J. (2018). *Inferring web page relevance using pupillometry and single channel eeg*. Springer.
Gwizdka, J., Hosseini, R., Cole, M., & Wang, S. (2017). Temporal dynamics of eye-tracking and eeg during reading and relevance decisions. *JASIST, 68*(10), 2299–2312.
Gwizdka, J., & Zhang, Y. (2015). Differences in eye-tracking measures between visits and revisits to relevant and irrelevant web pages. In *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval* (pp. 811–814).
Harter, S. P. (1992). Psychological relevance and information science. *JASIST, 43*(9), 602–615.
Ingwersen, P. (1992). *Information retrieval interaction, 246*. Taylor Graham London.
Ingwersen, P. (1994). Polyrepresentation of information needs and semantic entities elements of a cognitive theory for information retrieval interaction. In *SIGIR'94* (pp. 101–110). Springer.
Ingwersen, P. (1999). Cognitive information retrieval. *Annual Review of Information Science & Technology, 34*, 3–52.
Ingwersen, P. (2006). Context in information interaction–revisited. In *Proceedings of the fourth biennial DISSAnet conference* (pp. 13–23). Pretoria: University of Pretoria, 2006.
Jacucci, G., Barral, O., Daee, P., Wenzel, M., Serim, B., Ruotsalo, T., et al. (2019). Integrating neurophysiologic relevance feedback in intent modeling for information retrieval. *JASIST, 70*, 917–930.
Jiang, J., He, D., Kelly, D., & Allan, J. (2017). Understanding ephemeral state of relevance. In *Chiir '17* (pp. 137–146). NY, USA: ACM.
Kauppi, J.-P., Kandemir, M., Saarinen, V.-M., Hirvenkari, L., Parkkonen, L., Klami, A., et al. (2015). Towards brain-activity-controlled information retrieval: Decoding image relevance from meg signals. *NeuroImage, 112*, 288–298.
Kelly, D. (2009). Methods for evaluating interactive information retrieval systems with users. *Found. Trends Inf. Retr., 3*, 1–224.
Kelly, D., & Belkin, N. J. (2004). Display time as implicit feedback: Understanding task effects. In *Sigir '04* (pp. 377–384). NY, USA: ACM.
Kim, H. H., & Kim, Y. H. (2019). Erp/mmr algorithm for classifying topic-relevant and topic-irrelevant visual shots of documentary videos. *JASIST, 70*(9), 931–941.
Kingphai, K., & Moshfeghi, Y. (2021a). *Mental workload prediction level from eeg signals using deep learning models*.
Kingphai, K., & Moshfeghi, Y. (2021b). On eeg preprocessing role in deep learning effectiveness for mental workload classification. In *International symposium on human mental workload: Models and applications* (pp. 81–98). Springer.
Kingphai, K., & Moshfeghi, Y. (2022). On time series cross-validation for deep learning classification model of mental workload levels based on eeg signals. In *Advanced online & onsite course & symposium on artificial intelligence & neuroscience*.
Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121.
Kumar, A. (2013). Assessing the information need and information seeking behavior of research scholars of mbpg college: A case study. *International Journal of Digital Library Systems, 3*(3), 1–12.
Laganaro, M., & Perret, C. (2011). Comparing electrophysiological correlates of word production in immediate and delayed naming through the analysis of word age of acquisition effects. *Brain Topography, 24*(1), 19–29.
Levene, M., Bar-Ilan, J., & Zhitomirsky-Geffet, M. (2018). *Categorical relevance judgment*. Journal of the Association for Information Science and Technology.
Liu, X., Zhou, H., Jiang, C., Xue, Y., Zhou, Z., & Wang, J. (2020). Cognitive control deficits in alcohol dependence are a trait-and state-dependent biomarker: An erp study. *Frontiers in Psychiatry, 11*, 1389.
Mao, J., Liu, Y., Zhou, K., Nie, J.-Y., Song, J., Zhang, M., et al. (2016). When does relevance mean usefulness and user satisfaction in web search?. In *Sigir '16* (pp. 463–472). NY, USA: ACM.
Meixner, J. B., & Rosenfeld, J. P. (2014). Detecting knowledge of incidentally acquired, real-world memories using a p300-based concealed-information test. *Psychological Science, 25*(11), 1994–2005.
Michalkova, D., Parra Rodriguez, M., & Moshfeghi, Y. (2022b). Confidence perceptions as part of searcher's cognitive context. In *Advanced online & onsite course & symposium on artificial intelligence & neuroscience*.
Michalkova, D., Parra Rodriguez, M., & Moshfeghi, Y. (2022c). Drivers of information needs: A behavioural study–exploring searcher's feeling-of-knowing. In *Proceedings of the 2022 ACM SIGIR international conference on theory of information retrieval* (pp. 171–181).
Michalkova, D., Rodriguez, M. P., & Moshfeghi, Y. (2022a). Information need awareness: An eeg study. In *Special interest group on information retrieval* (p. 2022). SIGIR).
Mizzaro, S. (1997). Relevance: The whole history. *JASIST, 48*(9), 810–832.
Mizzaro, S. (1998). How many relevances in information retrieval? *Interacting with Computers, 10*(3), 303–320.
Mognon, A., Jovicich, J., Bruzzone, L., & Buiatti, M. (2011). Adjust: An automatic eeg artifact detector based on the joint use of spatial and temporal features. *Psychophysiology, 48*(2), 229–240.
Moshfeghi, Y. (2021). Neurasearch: Neuroscience and information retrieval. *CEUR Workshop Proceedings, 2950*, 193–194.
Moshfeghi, Y., & Jose, J. M. (2013). An effective implicit relevance feedback technique using affective, physiological and behavioural features. In *Sigir '13* (pp. 133–142). NY, USA: ACM.
Moshfeghi, Y., Pinto, L. R., Pollick, F. E., & Jose, J. M. (2013). Understanding relevance: An fmri study. In *ECIR'13* (pp. 14–25). Berlin, Heidelberg: Springer-Verlag.
Moshfeghi, Y., & Pollick, F. E. (2018). Search process as transitions between neural states. In *WWW '18, international world wide web conferences steering committee, republic and canton of geneva, CHE* (pp. 1683–1692).
Moshfeghi, Y., & Pollick, F. E. (2019). Neuropsychological model of the realization of information need. *JASIST, 70*(9), 954–967.
Moshfeghi, Y., Triantafillou, P., & Pollick, F. E. (2016). Understanding information need: An fmri study. In *Sigir '16* (pp. 335–344). NY, USA: ACM.
Moshfeghi, Y., Triantafillou, P., & Pollick, F. (2019). *Towards predicting a realisation of an information need based on brain signals*. ACM, NY, USA: WWW '19.
Mueller, C. J., White, C. N., & Kuchinke, L. (2017). Electrophysiological correlates of the drift diffusion model in visual word recognition. *Human Brain Mapping, 38*(11), 5616–5627.
Murayama, K., FitzGibbon, L., & Sakaki, M. (2019). Process account of curiosity and interest: A reward-learning perspective. *Educational Psychology Review, 31*, 875–895.
Paisalnan, S., Moshfeghi, Y., & Pollick, F. (2021). Neural correlates of realisation of satisfaction in a successful search process. *Proceedings of the Association for Information Science and Technology, 58*(1), 282–291.
Paisalnan, S., Moshfeghi, Y., & Pollick, F. E. (2022). Neural correlates of satisfaction of an information need. In *Advanced online & onsite course & symposium on artificial intelligence & neuroscience*.
Paisalnan, S., Pollick, F., & Moshfeghi, Y. (2021). Towards understanding neuroscience of realisation of information need in light of relevance and satisfaction judgement. In *International conference on machine learning, optimization, and data science* (pp. 41–56). Springer.
Paisalnan, S., Pollick, F., & Moshfeghi, Y. (2022). 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* (pp. 41–56). Grasmere, UK: Springer. October 4–8, 2021, Revised Selected Papers, Part I.
Park, T. K. (1993). The nature of relevance in information retrieval: An empirical study. *The Library Quarterly, 63*(3), 318–351.
Park, C. Y. (2001). News media exposure and self-perceived knowledge: The illusion of knowing. *International Journal of Public Opinion Research, 13(4): 419–42*.
Park, C. W., Gardner, M. P., & Thukral, V. K. (1988). Self-perceived knowledge: Some effects on information processing for a choice task. *American Journal of Psychology*, 401–424.
Perrin, F., Pernier, J., Bertrand, O., & Echallier, J. F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology, 72*(2), 184–187.
Pinkosova, Z., McGeown, W. J., & Moshfeghi, Y. (2020). The cortical activity of graded relevance. In *Sigir '20* (pp. 299–308). NY, USA: ACM.
Pinkosova, Z., McGeown, W., & Moshfeghi, Y. (2022). Revisiting neurological aspects of relevance: An eeg study. In *Advanced online & onsite course & symposium on artificial intelligence & neuroscience*.
Polich, J. (2007). Updating p300: An integrative theory of p3a and p3b. *Clinical Neurophysiology, 118*(10), 2128–2148.
Radecki, C. M., & Jaccard, J. (1995). Perceptions of knowledge, actual knowledge, and information search behavior. *Journal of Experimental Social Psychology, 31*(2), 107–138.
Ratcliff, R., Sederberg, P. B., Smith, T. A., & Childers, R. (2016). A single trial analysis of eeg in recognition memory: Tracking the neural correlates of memory strength. *Neuropsychologia, 93*, 128–141.
Ruthven, I. (2014). Relevance behaviour in trec. *Journal of Documentation, 70*(6), 1098–1117.
Ruthven, I., Baillie, M., & Elsweiler, D. (2007). The relative effects of knowledge, interest and confidence in assessing relevance. *Journal of Documentation, 63*(4), 482–504.
Sanchiz, M., Chevalier, A., Fu, W.-T., & Amadieu, F. (2017). Relationships between age, domain knowledge and prior knowledge pre-activation on information searching. In *Chiir '17* (pp. 289–292). NY, USA: ACM.
Saracevic, T. (2007). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part iii: Behavior and effects of relevance. *JASIST, 58*(13), 2126–2144.
Saracevic, T. (2016). The notion of relevance in information science: Everybody knows what relevance is. but, what is it really? *Synthesis Lectures on Information Concepts, Retrieval, and Services, 8*(3), i–109.
Savostyanov, A., Bocharov, A., Astakhova, T., Tamozhnikov, S., Saprygin, A., & Knyazev, G. (2020). The behavioral and erp responses to self-and other-referenced adjectives. *Brain Sciences, 10*(11), 782.
Schamber, L., & Eisenberg, M. B. (1988). *Relevance: The search for a definition*.
Schamber, L., Eisenberg, M. B., & Nilan, M. S. (1990). A re-examination of relevance: Toward a dynamic, situational definition. *IP&M, 26*(6), 755–776.
Scharinger, C., Kammerer, Y., & Gerjets, P. (2016). Fixation-related eeg frequency band power analysis: A promising neuro-cognitive methodology to evaluate the matching-quality of web search results? In C. Stephanidis (Ed.), *HCI international 2016 – posters' extended abstracts* (pp. 245–250). Cham: Springer International Publishing.

Schmüser, L., Sebastian, A., Mobascher, A., Lieb, K., Tüscher, O., & Feige, B. (2014). Data-driven analysis of simultaneous eeg/fmri using an ica approach. *Frontiers in Neuroscience, 8*, 175.
Slanzi, G., Balazs, J. A., & Velásquez, J. D. (2017). Combining eye tracking, pupil dilation and eeg analysis for predicting web users click intention. *Information Fusion, 35*, 51–57.
Song, Z., Liu, C., Shi, R., Zhang, M., Wang, H., & Mei, Y. (2020). Neural activities during the evaluation of luxury goods-to-service brand extension: An event-related potentials (erps) study. *Journal of Neuroscience, Psychology, and Economics, 13*(3), 127.
Sormunen, E. (2002). Liberal relevance criteria of trec-: Counting on negligible documents?. In *Sigir '02* (pp. 324–330). ACM.
Spironelli, C., & Angrilli, A. (2021). Complex time-dependent erp hemispheric asymmetries during word matching in phonological, semantic and orthographical matching judgment tasks. *Symmetry, 13*(1), 74.
Stróżak, P., Bird, C. W., Corby, K., Frishkoff, G., & Curran, T. (2016). Fn400 and lpc memory effects for concrete and abstract words. *Psychophysiology, 53*(11), 1669–1678.
Tagliabue, C. F., Veniero, D., Benwell, C. S., Cecere, R., Savazzi, S., & Thut, G. (2019). The eeg signature of sensory evidence accumulation during decision formation closely tracks subjective perceptual experience. *Scientific Reports, 9*(1), 1–12.
Twomey, D. M., Murphy, P. R., Kelly, S. P., & O'Connell, R. G. (2015). The classic p300 encodes a build-to-threshold decision variable. *European Journal of Neuroscience, 42*(1), 1636–1643.
Vakkari, P., & Hakala, N. (2000). Changes in relevance criteria and problem stages in task performance. *Journal of Documentation*, 295–310.
Vakkari, P., & Sormunen, E. (2004). The influence of relevance levels on the effectiveness of interactive information retrieval. *JASIST, 55*(11), 963–969.
Vellani, V., de Vries, L. P., Gaule, A., & Sharot, T. (2020). A selective effect of dopamine on information-seeking. *Elife, 9*, Article e59152.
Versteeg, M., & Steendijk, P. (2019). Putting post-decision wagering to the test: A measure of self-perceived knowledge in basic sciences? *Perspectives on Medical Education, 8*(1), 9–16.
Voss, J. L., & Federmeier, K. D. (2011). Fn400 potentials are functionally identical to n400 potentials and reflect semantic processing during recognition testing. *Psychophysiology, 48*(4), 532–546.
van Vugt, M. K., Beulen, M. A., & Taatgen, N. A. (2019). Relation between centro-parietal positivity and diffusion model parameters in both perceptual and memory-based decision making. *Brain Research, 1715*, 1–12.
Wachinger, C., Volkmer, S., Bublath, K., Bruder, J., Bartling, J., & Schulte-Körne, G. (2018). Does the late positive component reflect successful reading acquisition? A longitudinal erp study. *NeuroImage: Clinica, 17*, 232–240.
Wang, P. (2010). Contextualizing user relevance criteria: A meta-ethnographic approach to user-centered relevance studies. In *IIiX '10* (pp. 293–298). NY, USA: ACM.
Wenzel, M. A., Bogojeski, M., & Blankertz, B. (2017). Real-time inference of word relevance from electroencephalogram and eye gaze. *Journal of Neural Engineering, 14*(5), Article 056007.
White, R. W., Ruthven, I., & Jose, J. M. (2002). The use of implicit evidence for relevance feedback in web retrieval. In F. Crestani, M. Girolami, & C. J. van Rijsbergen (Eds.), *Advances in information retrieval* (pp. 93–109). Berlin, Heidelberg: Springer Berlin Heidelberg.
Yang, H., Laforge, G., Stojanoski, B., Nichols, E. S., McRae, K., & Köhler, S. (2019). Late positive complex in event-related potentials tracks memory signals when they are decision relevant. *Scientific Reports, 9*(1), 1–15.
Yan, C., Liu, F., Li, Y., Zhang, Q., & Cui, L. (2017). Mutual influence of reward anticipation and emotion on brain activity during memory retrieval. *Frontiers in Psychology, 8*, 1873.
Zhang, T., Bao, C., & Xiao, C. (2019). Promoting effects of color-text congruence in banner advertising. *Color Research & Application, 44*(1), 125–131.
Zhitomirsky-Geffet, M., Bar-Ilan, J., & Levene, M. (2015). How and why do users change their assessment of search results over time? *ASIST, 52*(1), 1–4.