Engagement in Information Search

Ashlee Edwards and Diane Kelly

1 Introduction

Engagement is a multifaceted concept, and approaches to studying it vary across disciplines. In the education field, engagement is often understood to be the extent to which a student interacts with classroom material; research in this area often focuses on how teachers can present this material in more interesting and inviting ways [29]. In organizational psychology, the focus has been on how to create a workplace where employees stay invigorated and enthusiastic [40]. In cognitive psychology, engagement is studied from the perspective of understanding goal orientation, perceived ability, and motivation [31]. The common thread in all this work is the focus on creating positive subjective experiences for people so they stay motivated while performing activities. Engagement in the context of information search systems research has been anchored by similar goals: discovering what makes an interface (and search) engaging, creating search interfaces that promote engagement, and, to a lesser extent, understanding how to measure engagement. An underlying assumption of this work is that engagement creates a more positive search experience.

In this chapter, we focus on engagement in the context of interactive information search systems research, which "blends research from information retrieval (IR), information behavior, and human-computer interaction (HCI) to form a unique research specialty that is focused on enabling people to explore, resolve, and manage their information problems via interactions with information systems. Interactive information retrieval (IIR) research comprises studies of people's information search behaviors, their use of interfaces and search features, and their interactions with systems" [23, p. 745]. The main goal of an interactive information search

A. Edwards • D. Kelly (✉)
University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
e-mail: aedwards@unc.edu; dianek@email.unc.edu

system is to help people resolve their information needs by providing a mechanism for them to interact with a set of information objects (e.g., web pages, scholarly research articles, newspaper articles). Typically, such interaction is initiated with a query and continues until the searcher has resolved his or her information need. This might take place during a single search episode or across multiple search episodes. This basic mode of search is perhaps best illustrated by reference to Google, where a searcher submits a keyword query describing his or her information need and receives a set of search results. The searcher can then use these results as a way to access content. Most of the studies reviewed in this chapter focus on retrieval of textual information objects, although a few studies focusing on other types of objects are included when appropriate. Most of the studies also focus on retrieval in the context of Internet-based systems and general-purpose search services that are freely available on the web, as opposed to proprietary database systems, search services associated with a single website or digital library, or enterprise search systems. Some systems are studied in their natural states (e.g., studies of Google or Yahoo!), while others are experimental systems where the researchers either modified a commercial system or created a completely new system.

We first review conceptualizations and definitions of engagement. Next, we discuss the contexts in which researchers have examined engagement, which provide insight about the multifaceted nature of engagement and how many different aspects of the search system and search experience can impact engagement. Specifically, we review work that considers engagement in the context of search user interfaces, search tasks, content and architecture, and individual differences. Following this, we describe several large-scale search log studies of engagement. Next, we discuss several studies that have attempted to compare, contrast, and integrate different measures of engagement. Finally, we conclude by considering the future of engagement in information search research.

2   Defining and Measuring User Engagement in Search

Researchers have primarily examined engagement by focusing on signals that can be extracted from search logs such as clicks and dwell time. A search log is "a file of the communications (i.e., transactions) between a system and the users of that system" [21, p. 408]. Communications that are typically studied include queries people issue, clicks people make on hyperlinks, and the amount of time between subsequent communications (e.g., dwell time). These communications are typically referred to as search interaction data or search behavior data. The use of search log signals is the most common way researchers have measured engagement.

Generally, the underlying assumption is greater frequency of certain behaviors, such as clicks and dwell times, indicates more engaged users. In other words, users who are more engaged will communicate more with the search system and communicate with it for a longer period of time. This approach to defining engagement can be observed in both large-scale studies of search logs and small-scale laboratory

studies. Our examination of the literature also shows that the term *engagement* is increasingly being used to describe observed patterns and sets of search behaviors, even if it is not initially posited as a construct that drives research inquiry (cf. [15, 26]). Regardless of how it is defined, measured, or discussed, user engagement has become a key concern of information search systems researchers.

Search log signals can be useful since they provide information about a person's activities; however, using behavioral signals alone to define engagement is problematic because many of these same signals have been used to indicate other things about a person's search experience. For example, increases in behavioral signals have been used to indicate frustration, confusion about the task, uncertainty about where to find information, information relevance, and user satisfaction. Since these behavior-based measures do not capture the cognitive or affective parts of search engagement, important information about search context is missed. In addition, these signals alone do not wholly capture engagement as it has been conceptually defined in some of this research. For example, although Lehmann et al. [27] define engagement as the "quality of the user experience associated with a desire to use the web application" (p. 1), they only use behavioral signals to measure engagement, which do not adequately capture the quality of the experience or a person's desire to use an application. Solid measurement of latent constructs such as engagement relies on clear conceptual definitions of what is being measured as well as a clear mapping between conceptual and operational definitions (i.e., measures).

While much of the work has focused on using behavioral signals to measure engagement, O'Brien and colleagues [32–35] have used psychometric theory to create self-report measures to capture engagement. O'Brien and Toms [33] define engagement as a "category of user experience characterized by attributes of challenge, positive affect, endurability, aesthetic and sensory appeal, attention, feedback, variety/novelty, interactivity, and perceived user control" (p. 7). Building on this conceptual definition, O'Brien and Toms [34] created and evaluated a 31-item scale to measure engagement, called the user engagement scale (UES). Six attributes of engagement were identified using factor analysis: *perceived usability, aesthetics, focused attention, felt involvement, novelty*, and *endurability*, which capture the cognitive, affective, and usability-related attributes of user experience.

The UES is one of the first instruments to measure engagement in the context of information search. Importantly, it has undergone extensive validity and reliability testing [33, 34]. It is also one of the first measures of engagement designed to help researchers who are working more closely with users; while many of the behavioral-based measures are useful in the context of large-scale search log analysis, they do not characterize the entire user experience. Researchers often gather data directly from participants using questionnaires or interviews, and it is critical to have valid and reliable measures that allow researchers to obtain this type of feedback.

When considering how engagement has been studied in research with self-report measures such as those administered via questionnaires, it is important to distinguish between measures such as the UES, which have undergone extensive testing, and sets of questions researchers ask that have not been tested. Researchers without training in the behavioral sciences often group all self-report measures together and

dismiss them because they are subjective. However, when constructed correctly, such measures can provide a valid and reliable signal. While many of the studies reviewed in this chapter elicit self-report data from research participants, in most cases, these items were created on an ad hoc basis, and there is no guarantee they adequately capture engagement. Such ad hoc collections of items also make it difficult to compare results across studies to generate a more thorough understanding of how and when engagement happens during information search and how it manifests itself in search behaviors.

Although the UES was initially evaluated in the context of e-commerce, it is increasingly being used to evaluate more general information search experiences [1, 4, 30]. Initial studies of its generalizability to the information search domain have been conducted [35] along with studies to understand its relationship to log data and other types of data such as physiological signals, eye tracking, and cursor movements [2, 32]. This work is discussed throughout this chapter. Ultimately, a variety of definitions and combinations of measures likely offer researchers the most robust understanding of search engagement and, depending on the type of study being conducted, different definitions and sets of measures might be more or less appropriate and feasible.

3   Search User Interfaces

The search user interface aids users "in the expression of their information needs, in the formulation of their queries, in the understanding of their search results, and in keeping track of the progress of their information-seeking efforts" [17, p. 1]. While researchers in information search have been interested in designing usable interfaces for quite some time, they have only recently moved beyond a focus on functional requirements and adopted the position that search interfaces should also be engaging and that search experiences should be pleasurable [6]. The work reviewed in this section has either used the UES to evaluate search interfaces or used terms like engagement when describing the goals and outcome of the work.

One of the first studies to use the UES to understand user experience in the context of search interfaces evaluated the display of vertical search results [4]. The study's authors examined differences between an interface that blended vertical results into web search engine results pages (SERPs) and an interface that displayed vertical results separately on individual SERPs that could be accessed via tabs. Arguello et al. [4] did not use the complete UES in their study and also made several modifications to the items they did use. This limited their ability to make strong claims about the validity of the modified set of items, which had been established in previous work [34]. However, previous testing of the UES was done in an e-commerce setting, and the authors argued that the changes were needed to make the scales more suitable for the evaluation of search interactions. Most of the changes consisted of replacing words like "shopping" with "searching." In addition, the researchers dropped the aesthetics subscale as the basic elements of

the interface remained constant throughout. Finally, the researchers indicated they deleted one item from each of the attention and endurability subscales after pilot participants reacted unexpectedly to them. Ultimately, [4] used the UES subscales focused attention, felt involvement, perceived usability, and endurability and added a subscale about search effectiveness. Reliability analysis of responses to these modified subscales demonstrated that these items had good reliability.

Arguello et al. [4] did not find any significant differences between responses to these items according to interface. They went on to compare participants' interface preferences with their post-task questionnaire ratings on these subscales and found that people who preferred one interface rated it higher along all aspects, specifically for attributes such as endurability and perceived usability. Participants who preferred one of the interfaces said they found it more visually appealing and felt the information was better organized and easier to understand, reinforcing the importance of usability. These findings are interesting because they show that engagement is related to a person's preferences, and without knowing this preference, aggregate engagement scores for two or more interfaces might appear similar even when they produce different user experiences.

Moshfeghi et al. [30] evaluated whether adding a timeline and a named-entity component to a news search system would improve engagement and whether engagement could be predicted based on interaction data. They created a search interface where a participant clicked on search results that were presented on a timeline in order to access content. In addition to the timeline, they added a list of entities for a given search result. For example, for a given entry such as (US) republican debates, the named entity list contained items such as "Newt Gingrich," "Herman Cain," and "Rick Perry."

Participants were recruited from Mechanical Turk and given explicit instructions about the assignment and how much time they would have to complete it (120 min). Engagement was measured using the UES. Similar to [4], Moshfeghi et al. [30] modified the UES by changing the wording of the items for a news context, and each question was structured to ensure forced choice instead of a range of values. They found that participants who used the enhanced interface rated felt involvement, endurability, novelty, and aesthetics (subscales of the UES) higher, which demonstrates the importance of moving beyond a purely functional assessment to more completely understand the user's search experience.

Bateman et al. [5] created an interface where participants interacted with their previous search data and were able to compare themselves to three archetypes: the typical participants, search experts, and topical experts. Search experts were defined as frequent users of search operators, and topical experts were defined as having visited ten search results within the category. One version of the interface allowed participants to compare themselves to these archetypes, and the other did not. Engagement was derived from interactions found in participants' log data, specifically attributes such as time spent examining search results, likelihood of returning to the dashboard, and an affective learning dimension. The researchers (without mentioning engagement directly) also referenced engagement when they discussed participants' interest in learning and insights when using the interface.

Participants were most interested in and felt they gained more insight about themselves from the data about characteristics of search engine use and data on special search engine features and advanced query operators they viewed. Participants rated the comparison interface much higher than one that did not allow comparison and were also more likely to report that the comparison interface would alter their search behavior later. Unlike the studies described above, this study focused on people's interactions with personalized content.

The structure and layout of a website and search interface is referred to as a representational context. Representational context includes the designer's decisions about how to represent actions that can be performed (e.g., search box, search button), the placement of elements and icons on a page, and even the icons themselves. Subsequently, representational *stability* refers to the extent to which this representational context is maintained over the course of the entire search experience. Representational stability can be examined both within a single system and also across systems that are used to perform a similar function. For example, most major search systems employ interfaces that use a single box for query entry and a rank ordering of search results. Duin and Archee [13] posited that representational context must remain stable in order for the participant to become engaged.

Webster and Ahuja's research [46] supports the relationship between engagement and representational stability by developing a model of disorientation and engagement in web systems. This model states that navigation systems affect perceived disorientation, which affects engagement, which affects both performance and future use. Engagement was operationalized as when a system "holds [a subject's] attention and they are attracted to it for their intrinsic rewards" [20, p. 58] and was measured with a seven-item questionnaire that contained items such as "the site kept me totally absorbed in the browsing" or "the site held my attention." To evaluate their model, Webster and Ahuja [46] tested a simple navigation system against a global navigation system and an enhanced navigation system. The simple navigation system contained only hyperlinks, and these hyperlinks disappeared while the participant scrolled. The global navigation system contained a site map, a search form, and nested navigation bars (i.e., a parent topic contained child topics), but the navigational features also disappeared while scrolling. The enhanced global navigation system had the same features as the global one but kept the features visible while scrolling. Their findings supported the model in that participants in both the global and enhanced global navigation conditions reported less disorientation, and participants in the enhanced global navigation system had the best performance. This group's high performance was also positively related to engagement, showing that navigational aspects of a search interface can affect engagement. Perceptions of navigation and orientation are shown to help maintain representational integrity, providing a link between engagement and usability.

Feild et al. [15] also wanted to support orientation and engagement during the transition from the SERP to the content page by adding clickable snippets on the SERP. These snippets contained text from the document that matched the participant's query, and clicking on these snippets took participants directly to where

that text was located in the document. Feild et al. [15] measured engagement with the system by calculating differences in views on the landing page, path length, gaze fixations, time until fixation on the answer passage, and scroll distance. They found that participants had lower time until fixation on the passage with the answer, lower fixations, and lower scroll distance when using the system with the clickable snippets with gradual transition. This indicates that for most of Feild et al.'s metrics, participants were engaged with and performed better using systems with clickable snippets. These results demonstrate that interventions can improve engagement and search performance, but they challenge prior notions that familiarity and comfort (i.e., a stable representational context) with a system are a necessary but not sufficient condition for engagement. More work is needed to understand what kinds of interventions improve engagement while not overwhelming the user.

This notion of stability is also supported by work on the effect of different user interface interaction modalities on engagement [42]. Interaction modality refers to mouse-based interaction patterns, specifically zoom, drag, slide, mouseover, cover flow, and click to download. Sundar et al. investigated these modalities on six artificial websites. Layout, page content, and color were kept constant between interaction modalities. These modalities allowed participants to access "hotspots" or links to information embedded in the website. Sundar et al. [42] defined engagement as a combination of participant attitudes, actions, skill, and behavior toward the content. They hypothesized that different types of interaction modalities would lead to different levels of perceptual bandwidth or the "range of sensory and preliminary attentional resources available to individuals" (p. 1478), referring to the resources a person has for understanding and perceiving interactivity in an interface; Sundar et al. defined this conceptually as "users memory for interface content" (p. 1478). Reeves and Nass [38] stated that perceptual bandwidth is increased by perceptual interfaces, which offer people "more and different sensory channels" (p. 65) than traditional interfaces. This suggests that perceptual interfaces or increases in perceptual bandwidth can change interest in the content of an interface.

Perceptual bandwidth was measured in terms of recall and recognition, perceived interactivity, actions, behavioral intention toward content and the website, and attitudes toward content. Sundar et al. [42] found significant differences between modalities; specifically, the slide modality showed higher recall than the zoom in/out modality. Participants who used the cover flow and mouseover actions performed more actions overall than the other modality types. Some participants preferred modality types that gave them more control over their content, while others preferred modality types that allowed them to perform more actions. Sundar et al. remind us that interaction modalities can make content more absorbing and generate positive feelings, which are closely related to the interest and cognitive absorption that occurs during engagement. The distinct preference for modality among participants indicates that users want to maintain representational stability, though representational stability may be subject to variation across individuals.

Sundar et al. [42] collected attitude data and found that certain actions such as the mouseover led to more positive attitudes than cover flow, which led to

more negative attitudes. This also shows that some interaction types are generally more preferable than others. Some users, referred to as "power users" (who were identified based on a questionnaire containing items about liking, skill, and dependence on technology) preferred modality types that gave them more control over their content, while other users who were not "power users" preferred modality types that allowed them to perform more actions, demonstrating the importance of individual differences. Other research has suggested that control is important in engagement [46], and this work showed that control might be more critical to engaging some users than others.

Teevan et al.'s findings [43] challenge the notion of representational stability as necessary for engagement. In this study, Teevan et al. studied one important structural element of search systems: latency. Latency refers to the interval between an action and the response. High latency can be thought of as disruptive to representational stability because it disrupts a person's ability to maintain representational context. The purpose of this study was to examine how participants interacted with a search system that prioritized high-quality results over speed. Specifically, Teevan et al. looked at querying behavior with navigational queries (those that "targeted specific web pages" (p. 2)) and informational query types (those that are "intended to find information about a topic" (p. 2)). The researchers also examined two post-query behaviors: abandonment rate and time to first click. Engagement was examined and was defined as engagement with the search results in the form of more search interaction behaviors. Teevan et al. found that click frequency decreased as page load times increased, which the authors claimed signaled a loss of interest. However, the results showed no increase in search abandonment (which is also posited as evidence of disengagement) as load times increased. They explain this by stating there is a point beyond which load times can increase without causing higher search abandonment rates. It is also possible that the clicking was more deliberate, as participants anticipated the page load times and wanted to be sure they clicked on the most fruitful result. Participants were asked how long they would be willing to wait if they knew search engines would give them the best response, versus an acceptable response, and most said they were willing to wait much longer for the best response. This indicated that participants may be able to tolerate shifts in their representational context and adapt to them if they receive some benefit.

Arapakis et al. [1] investigated the impact of response latency on the click behavior of participants and the point at which response latency becomes noticeable in two studies. The first study looked at participants' sensitivity to latency and used two manipulations: response latency and site speed. Response latency refers to the time between a user's action and the perception of the response. Site speed was operationalized as either slow (a search site with a slow response) or fast (a search site with a fast response). They found that participants were more likely to notice the response latency if it climbed above 1000 ms. In the second study, they measured the effect of response latency on user engagement using the *focused attention* subscale of the UES (modified for a search context), satisfaction, and click behavior. They found a small effect for focused attention in participants in the fast condition, which suggests that these participants felt more deeply involved

in the search task. They also found that though there were no significant differences in frustration, participants' positive search engine bias (the belief that the search system was helpful) was correlated with focused attention and perceived usability in both speed conditions. This suggests that search engine bias affects the way that participants interpret system response. Lastly, they found that participants were more likely to click on a result from a SERP that had been returned with low latency. This paper showed that conditions we may see as unfavorable to engagement (such as low latency) could encourage positive behaviors such as more examination of search results.

Work on engagement and search interfaces has shown that the interface can be crucial in fostering and maintaining engagement throughout the search session and that altering the traditional search interface to include elements that allow users to reflect on their own behaviors, and compare them to others, can potentially improve user engagement. This body of research also shows that representational stability, while important to engagement, may be one facet where individual differences are important. The literature reviewed here shows that users can tolerate shifts in their representational contexts and that users can express preferences for different kinds of interaction.

4   Search Tasks

When people decide to use an information search system, they often do so because they have an information need. In much of the search system research, a user's information need is encapsulated in a search task. This is especially true in laboratory studies where researchers assign search tasks to users so they have some (controlled and prescribed) reason to use a search system. Search tasks have been defined as "goal-directed activities carried out using search systems" [44, p. 1134] and as "a task that users need to accomplish through effective interaction with information systems" [28, p. 1823]. The impact of different types of search tasks on information search behavior and the user experience has been of great interest to information search researchers during the last 10 years [44]. While in the past researchers have measured task properties, such as difficulty and complexity [47], recently researchers have begun to evaluate the relationship between search tasks, search task properties (e.g., difficulty), and engagement.

O'Brien and Toms [35] evaluated the generalizability of the UES to exploratory search tasks by asking people to complete three complex, situated tasks that required them to make a decision. Using results from 381 participants, they found that the UES factor loadings differed from those observed in the initial evaluation of the UES, which was evaluated in the context of an e-commerce setting [34]. While the perceived usability, aesthetics, and focused attention factors remained distinct, the novelty, felt involvement, and endurability were indistinguishable. They also found a negative correlation between focused attention and perceived usability, which differed from the original study [33]. O'Brien and Toms explain this by stating

that the laboratory setting may have impacted the relationship between flow and usability and that in naturalistic settings, flow and usability may be more closely correlated. The assigned tasks may have inhibited the participant's ability to achieve a flow state. Focused attention scores were also lower than scores for other factors, which O'Brien and Toms [33] speculate was a result of participants' focus on task completion rather than on the content of the task.

Jiang et al. [22] illustrated the importance of task when they measured how different kinds of tasks affected search behavior, relevance judgments, and interest. While Jiang et al. did not conceptually define engagement, they used the term interest when characterizing participants' search behaviors. Participants were given tasks defined by a goal (specific or amorphous) and the required information behavior (either factual or intellectual). These two dimensions created four sets of tasks: known item (factual and well defined), known subject (factual and amorphous), interpretive (intellectual and well defined), and exploratory (intellectual and amorphous). Jiang et al. measured interest in the search results by unique clicks per query, unique fixations per query, and SERP views per query. All of these behavioral measures dropped significantly over the course of the search session, and Jiang et al. present this as evidence that a person's interest in the search task decreased during the course of the search session. When considered alongside O'Brien and Toms' findings [35], this finding is likely related to the search context, that is, a laboratory study with assigned search tasks and a task time limit. While these findings might have limited applicability to real-world search, especially when task time is unlimited, they suggest what researchers might expect when assigning search tasks with a fixed search time to research participants; that is, participants might begin to disengage as they approach the task time limit. This is also consistent with conceptual model of engagement in [33], which depicted points of disengagement arising from external forces and constraints.

It is also useful to consider the work of Borlund et al. [8] and others [9, 37] who have investigated differences in user experience and interest between assigned search tasks and genuine search tasks, or search tasks created by users, as this work demonstrates how the content of the task can potentially impact engagement. This is consistent with earlier conceptualizations of flow, where interest in one's task was found to be central to experiences of flow [12] and Borlund's recommendations [7] that simulated work tasks be those to which participants relate and find topically interesting. Borlund et al. [8] found that 76 % of participants attributed time spent searching on their genuine task to it being interesting versus 38 % for a simulated task. They also found that participants spent more time searching during genuine tasks and generally found them more difficult. This finding is interesting since it suggests that difficulty is, in part, related to interest in the task, and in an unexpected way. It might be the case that when a person is more interested in a task, they have more emotional investment, which translates into greater perceptions of task difficulty. Poddar and Ruthven [37] found that participants had greater positive emotions and made more use of various search strategies when completing their own search tasks versus assigned search tasks, so the source of the task can impact user experience and effort expended.

While the studies described above focused on understanding how search task properties impact engagement, at least one study has examined the relationship between search task type and engagement in the context of creating reusable tasks for laboratory search studies [24]. Kelly et al. [24] created and evaluated a set of search tasks that were proposed to vary in terms of cognitive complexity. The researchers examined participants' behaviors as they completed tasks of different levels of cognitive complexity as well as their ratings of these tasks along a number of dimensions, including difficulty and engagement. The hope was that more cognitively complex tasks would be rated as more engaging because of the increasing amounts of cognition they required. Results showed that the two most cognitively complex tasks were rated as significantly more engaging than the least cognitively complex task. Participants also exhibited significantly more effort (e.g., queries, clicks) completing more cognitively complex tasks, which is aligned with work that posits increased search behavior is related to increased engagement. The difficulty, of course, is untangling temporal order to show cause and effect; that is, do more engaging tasks cause a person to exhibit more search effort, or does more search effort cause a person to become more engaged?

Interestingly, in Kelly et al.'s study [24], when participants were asked to rank tasks according to level of engagement, the signal was not as clear (except for the least cognitively complex tasks which were mostly rated as the least engaging). This result suggests that the content of the task likely played a role in engagement and, more specifically, the user's interest in the content. This implies that researchers who are constructing assigned search tasks for laboratory use should not only consider the structure of the tasks but also the topic of the tasks if they wish to study searchers who are engaged. Of course, discovering what interests an individual participant before a study becomes a challenge as well as maintaining some parity among the potentially large number of topical areas that are likely to interest participants. Thus, an important future research direction is to understand how search tasks can be developed to foster or inhibit engagement in experimental settings.

5   Content and Architecture

In the previous section, we discussed how the content of a search task potentially impacts a person's experiences of engagement. Research has also shown that the content of the information sources with which a user interacts plays an important role in engagement. Arapakis et al. [3] used the *focused attention* subscale of the UES in conjunction with other measures to observe what attributes of news articles and comments were important to participants. They examined several attributes: genre, sentimentality of the article (the richness of the emotional tone of the article), polarity (positivity or negativity), and time of publication. Articles were then selected from three categories: crime, entertainment, and science. Participants indicated their interest before and after the task. Arapakis et al. found that participants who read articles they labeled as interesting exhibited higher levels

of focused attention. They also found that interest in the article and enjoyment experienced from reading it were higher when the topic of the article had a strong sentiment and negative connotations.

Linking content focus and attention, Rokhlenko et al. [39] looked at how interest in peripheral content, such as advertisements, varied based on interest in the primary content on the page. Participants (Mechanical Turk workers) were asked to read news articles until they felt they had discovered the purpose of an article and then were instructed to answer questions based on the text. Results showed most participants missed the ads entirely; only a quarter of participants paid any attention to the ad image surrogates. Rokhlenko et al. [39] found that participants who spent a lot of time reading the content on a web page had higher recall for the advertisement images than participants who read less. If interest can serve as an indicator of engagement, then this study showed that engagement with content could lead to higher recall for peripheral images. This study also helps confirm that when participants are engaged, they tend to display deeper information processing behaviors such as reading and absorbing more content. If engaged participants are able to recall many different types of information, then it is possible that engagement could lend itself to expanding attentional resources.

Song et al. [41] examined whether degraded search relevance had an effect on engagement. The researchers defined engagement both in terms of frequency of search engine reuse and behavioral signals. Participants in this study were given a search algorithm that provided low-quality search results or received the normal search engine algorithm. Song et al. analyzed the session data of search logs from 2.2 million users. Query attributes such as queries issued per session, length, success, click-through rate, type, and session length as well as frequency of search engine usage were used to measure engagement. Song et al. found that though engagement decreased overall, there was some indication that participants might have been engaged. Participants in the treatment group issued more queries overall, issued more navigational queries, reformulated their queries more, and clicked on more results. They surmise that this search behavior could reflect increased effort, a consequence of struggling to complete the search with poor search results. This means that, for the engagement metrics defined in this study, engagement was initially negatively correlated with relevance. Song et al. then tried to predict engagement using search behaviors and found that the number of clicks was the highest correlated feature with engagement. This study established a link between behavioral signals and engagement as induced by effort. In particular, effort invokes the factors of felt involvement and focused attention, which, as this study showed, can be induced by negative influences rather than positive ones.

Perhaps the most revealing studies are those that combine changes in both content and navigational structure. Chen et al. [10] examined the effect of disorientation on engagement with a website given the breadth, familiarity, and media richness of the site. Two websites were created with different structures: the "broad" structure contained two levels, while the "deep" structure contained four levels. Familiar sites contained stationery products, while unfamiliar sites contained industrial products. Media richness was also manipulated; "media rich" sites contained images and

videos, while "lean media" sites contained only text. Chen et al. [10] found that participants preferred websites that had a deeper structure and were more engaged with a site that had unfamiliar structure and lean media richness in addition to deeper structure. Higher disorientation was linked to less engagement and lower intentions to use the website in future. This study shows that engagement does not always occur when a participant is completely comfortable and familiar with a web interface. Rather, a combination of novelty and familiarity can foster engagement.

Colbert and Boodoo [11] examined the effect of web content noncompliance on engagement. Some attributes of noncompliance were minor, such as grammatical errors, but also included direct barriers to information-seeking such as a lack of same-page links and obscure heading levels. Participants were subject to an advertising campaign on both sites, and there were 11 advertisements, with between 7 and 43 keywords and phrases per advertisement. Four attributes of engagement were defined in this study: time spent on site, pages per visit, ratio of revisits to first visit, and bounce rate, or whether the participant spends time on a single page only versus multiple pages. They found the compliant website more engaging across all metrics and in particular for return visits. Colbert and Boodoo believed that web standard compliance in the form of fewer well-placed words increases engagement. This study suggests that engagement can occur at the micro-level; if the structure and content of a site does not facilitate information-seeking, then a person will not be engaged and may leave the site prematurely. It is encouraging as it suggests that by following web standards, website designers can increase the engagement of their sites. These studies demonstrate that the content of a search system is just as important as the representational context in keeping users engaged with a website, suggesting that engagement is highly sensitive to both major and minor changes in a search interface.

6   Individual Differences

So far, we have discussed how interface features, search tasks, and information content and design relate to engagement. Individual differences have also been shown to have an effect on engagement. Heinstrom [18] looked at the relationship among individual differences, information-seeking behaviors, and engagement in the context of a naturalistic study. The individual differences investigated included personality traits, learning approach preferences, and disciplinary differences. Master's students writing theses were chosen to participate, and three questionnaires were used: the NEO Five-Factor Inventory, the Approaches and Study Skills Inventory for Students, and a questionnaire about information-seeking behaviors. Three information-seeking patterns were discovered: fast surfing, broad scanning, and deep diving. Fast surfing was a search pattern characterized by minimal effort in both information-seeking behaviors and content analysis, while broad scanning was an exploratory search pattern characterized by wide searches and many information searches. Deep diving, however, was characterized by expending considerable effort

on the search as well as looking for high-quality documents. The behavior most closely related to engagement was deep diving. Heinstrom [18] noted that they seemed "focused and structured" (p. 1446) in their searches and searched to gain a thorough understanding of the topic rather than just scanning for information. There was also an interaction among information-seeking pattern, engagement, and content: broad scanners were more engaged with documents that gave them new information, while fast surfers were more interested in documents that were easy to read and were less academically challenging. Topical engagement was more likely to occur in relaxed settings presumably because of the absence of time pressure.

This work supports the notion that engagement is highly context and topic dependent. Since the students in this study were completing master's theses, there was an inherent interest in the topic and task that likely lent natural motivation to searches. These results also suggest that differences in personality and information-seeking styles will impact what experiences a person finds engaging. This finding is similar to the one described in [4], where participants' engagement ratings were tied to their interface preferences.

7 Large-Scale Analysis of Commercial Search Logs

Lehmann and colleagues have conducted a number of studies that provide a good illustration of how engagement has been studied in the context of large-scale search logs [25–27]. In one of their first studies, Lehmann et al. [25] proposed and evaluated three interaction-based models of engagement: a general model, a time-based model, and a user-based model. Using search log data from millions of people, three measures of engagement were defined and examined in the context of each model: popularity, activity, and loyalty. *Popularity* was defined as the number of users that visit a site (including number of clicks). *Loyalty* was defined as the frequency with which a person returns to a site and how often they dwell on the site. *Activity* was defined as total dwell time on the site and number of page views per visit. Lehmann et al.'s general model of engagement [25] focused primarily on popularity and clicks on a site, the time-based model was more focused on loyalty, and the user-based model was more focused on an individual user's behavioral patterns.

Lehmann et al. [26] continued this work by proposing the concept of *networked user engagement*, which refers to engagement within a network of websites. This work focused on user clicks among different websites and posited users with high network engagement would make clicks among the websites within the network. They found that users performed more goal-oriented behaviors on a weekday (Wednesday), while they performed more browsing activities on the weekend. They also found that some users who were more active with regard to search behavior (referred to as VIP users) navigated more frequently between sites and had higher rates of return to previously visited sites than users who were less active. This

conceptualization differed from the previous one in that it focused on activity within a collection of websites as opposed to activity at an individual website.

Lehmann et al. [27] furthered their work on engagement by focusing on user engagement with many tasks simultaneously and analyzed online multitasking and engagement using two behavioral signals: dwell time and page views. Transforming these signals into metrics like attention shift, attention range, cumulative actions, visits, and sessions, Lehmann et al. grouped different kinds of sites based on levels of engagement and proposed a model in which dwell time and page views were conceptualized as tree-streams, or paths through which participants click at the session level. Shopping and mail sites were found to have high activity per visit and also short times between visits, indicating that participants progressively became more focused on their tasks. Search sites, front pages, and auction sites had lower dwell time overall but higher dwell time per session and had high cumulative activity numbers, indicating that participants spent more time completing more activities. The most engaging set of sites had high ranges of attention shift and attention range, indicating that when participants did return to the site, they spent more time than before.

Dupret and Lalmas [14] investigated the usefulness of absence time, or the time between two user visits, as a metric of user engagement. This was based on the assumption that users who are engaged will return to a site sooner, meaning that their absence times will be shorter. Most unique to this work was that the researchers used survival analysis. Survival analysis is based on a "death and resuscitation" model, whereby users "survive" past a given time, and "hazard rate," which refers to the probability a user will die at a given time; thus, a higher hazard rate implies a lower survival rate. In this example, a high hazard rate is associated with a lower absence time (meaning a user is returning more frequently). They found that faster time to click was associated with higher hazard rate, suggesting that if users click quickly, they are likely more engaged. They suggest that click three, which contributed weakly to the hazard rate of five clicks, may be associated with greater user engagement because it suggests more perusal of search results and thus more cognitive investment. Lastly, they found more views than distinct queries are associated with longer absence times.

Finally, Ortiz-Cordova and Jansen [36] investigated whether behavioral signals could be used to identify "high-revenue" participants, or those who were more engaged with site content and advertisements. They defined engagement in terms of new visits, number of pages visited during the duration of the session, time spent on a site, click-through rate, ads clicked, ad impressions, and rate of return to the site. The researchers classified participants into three clusters, low, medium, and highly engaged, and identified different kinds of revenue streams generated by each cluster. Participants in the highly engaged cluster spent the most time on the site, visited the most content, and clicked on the most ads, while those in the low engagement cluster typically visited few pages and clicked on little content. Generally, revenue streams were higher if participants clicked on the ads and had a higher number of page visits.

8 Multiple Measures of Engagement

We started this chapter by discussing how engagement has been defined and measured in information search research. We turn our attention back to this topic by examining studies that have attempted to combine different measures including behavioral measures, self-report measures, and eye-tracking and physiological data. While many of the studies discussed above used both objective and subjective measures of engagement, researchers have only recently started investigating how these measures are related to one another, along with other types of measures. O'Brien and Lebow [32] used the UES in conjunction with the Cognitive Absorption and System Usability Scales to examine which attributes were important during information-seeking experiences within an online news context. They combined these self-report measures with physiological signals in order to get a better understanding of engagement. Participants completed one task, with a time limit of 20 min, followed by the psychometric scales and an interview. O'Brien and Lebow [32] found that participants who rated their level of interest in an article higher were also more engaged. They also found that participants who were less engaged spent more time browsing and visited more web pages but had increased physiological signals. Participants who reported the highest levels of engagement spent the least amount of time browsing, visited the least amount of web pages, and spent the least time reading but had lower physiological signals. This study also found low, negative correlations between physiological signals and the psychometric scales used.

Arapakis et al. [2] investigated the usefulness of mouse gestures and gaze behavior as possible indicators of engagement with search results. Participants were given one interesting and one uninteresting news-related task (and a corresponding corpus) and had their cursor movements and eye fixations recorded. In addition to this, affect was measured via the PANAS and focused attention subscale of the UES, modified to reflect a news context. Arapakis et al. found a correlation between gaze behavior and engagement, specifically that participants had more fixations when reading an interesting article, looked more at the content of the article, and had longer visits. When the article was not interesting to participants, they fixated on other content on the page. Arapakis et al. also found that negative emotions had a greater influence on mouse movements than positive ones, suggesting that lack of engagement may be more detectable through cursor movement than engagement.

In our review, we found two studies by Grafsgaard and colleagues that have used part of the UES to evaluate intelligent tutoring systems, along with facial expression analysis and skin conductance [16, 45]. Grafsgaard et al. [16] were interested in investigating the usefulness of facial expression analysis in understanding the affective states of engagement and frustration. Sixty-five participants interacted with a programming tutor through a web interface. Their facial expressions and skin conductance were recorded, though the researchers did not report the skin conductance results in their paper. Students were given the endurability subscale of the UES, modified for a learning context, as well as questions about temporal

demand, performance, and frustration from the NASA-TLX. They found that endurability was predicted by rises in participants' inner eyebrows, while temporal demand was predicted by rises in participants' outer eyebrows. Performance was predicted by mouth dimpling, and frustration was predicted by brow lowering. The major contribution of this paper was its linkages of facial expressions to measures of engagement and frustration.

In a follow-up study, Vail et al. [45] reviewed the utility of one of the Big Five personality traits (extraversion and its opposite introversion) in conjunction with facial and postural gestures as predictors of engagement and frustration. Seventy-seven participants had their personality traits measured and facial and postural gestures recorded during a web-based tutoring session. Engagement was measured via the focused attention, felt involvement, and endurability subscales of the UES, modified for a learning context. They found that feedback from the tutor was a feature of the predictive model for extraverts and that engagement and learning gains were positively and negatively affected by feedback from the tutor. Frustration was more often correlated with changes in posture and seat movement for extraverts. For introverts, engagement was correlated with forward postural movements, while frustration was correlated with backward postural movements, indicating that introverts express their feelings behaviorally rather than with dialogue. This study reinforces the idea that individual differences, and specifically personality differences, can play a role in how a person experiences and expresses engagement.

9   Conclusions

It has been observed that engagement is integral to system success [19], and the work reviewed in this chapter supports this idea. System success is a complex mix of attributes such as system response time, content and results quality, the user interface, and subjective experience. While in the past research in the area of information search has emphasized the functional aspects of search systems such as performance and usability, there has been growing interest in creating engaging search experiences for searchers and understanding more about searchers' emotional experiences during search. This chapter reviewed some of the research within the information search research specialty that has focused on engagement and related constructs like interest.

This review illustrated the challenges of studying engagement. The extent to which someone is engaged depends on a variety of factors including the structure and ease of use of the system, performance of the system, content within it, complexity and difficulty of search tasks, how searchers perceive all of these variables, and whatever individual differences they bring to the search situation. Our role as researchers is to meet the needs of searchers by understanding how these aspects contribute to experiences of engagement and subsequently applying this understanding to the design of search tools that foster engaging experiences. In recent years, information search research has been dominated by studies of

searchers' interactions with SERPs; an interesting consequence of focusing on engagement is that now a wider view including both interactions with SERPs and the information objects themselves is required.

One way we can address some of the challenges of studying engagement is to examine both the behavior of the searcher and their subjective experiences. While it is easy to discount self-report data as flawed and unreliable, searchers are really the only ones who can tell us if they are engaged. It is also easy to discount behavioral-based measures because they can be ambiguous and only represent the potential manifestation of engagement. However, this physical manifestation is an important part of engagement and provides a useful and unobtrusive way to operationalize engagement. In many environments and contexts, especially at scale, it is not possible to ask people about their experiences, so refining the use of these signals as standalone measures is important. The examination of physiological signals, facial expressions, and eye-tracking data, along with behavioral and self-report measures, are likely productive ways to start refining theoretical models of engagement and methods for measuring engagement in different contexts.

References

1. Arapakis, I., Bai, X., Cambazoglu, B.B.: Impact of response latency on user behavior in web search. In: Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval, pp. 103–112 (2014)
2. Arapakis, I., Lalmas, M., Valkanas, G.: Understanding within-content engagement through pattern analysis of mouse gestures. In: Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, pp. 1439–1448 (2014)
3. Arapakis, I., Lalmas, M., Cambazoglu, B.B., Marcos, M.C., Jose, J.M.: User engagement in online news: under the scope of sentiment, interest, affect, and gaze. J. Assoc. Inf. Sci. Technol. **65**, 1988–2005 (2014)
4. Arguello, J., Wu, W.C., Kelly, D., Edwards, A.: Task complexity, vertical display and user interaction in aggregated search. In: Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 435–444 (2012)
5. Bateman, S., Teevan, J., White, R.W.: The search dashboard: how reflection and comparison impact search behavior. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1785–1794. ACM Press, New York (2012)
6. Belkin, N.J.: Some(what) grand challenges for information retrieval. SIGIR Forum **42**, 47–54 (2008)
7. Borlund, P.: The IIR evaluation model: a framework for evaluation of interactive information retrieval systems. Inf. Res. **8**(3), 1–5 (2003)
8. Borlund, P., Dreier, S., Byström, K.: What does time spent on searching indicate? In: Proceedings of the 4th Information Interaction in Context Symposium, pp. 184–193 (2012)
9. Capra, R., Sams, B., Seligson, P.: Self-generated versus imposed tasks in collaborative search. In: Collaborative Information Seeking: Bridging the Gap Between Theory and Practice (CIS). Workshop at the Meeting of the American Society for Information Science and Technology (2011)
10. Chen, J.V., Lin, C., Yen, D.C., Linn, K.P.: The interaction effects of familiarity, breadth and media usage on web browsing experience. Comput. Hum. Behav. **27**, 2141–2152 (2011)

11. Colbert, M., Boodoo, A.: Does' letting go of the words' increase engagement: a traffic study. In: ACM CHI 2011 Extended Abstracts on Human Factors in Computing Systems, pp. 655–667. ACM Press, New York (2011)
12. Csikszentmihalyi, M.: Flow: The Psychology of Optimal Experience. Harper-Perennial, New York (1991)
13. Duin, A.H., Archee, R.: Distance learning via the world wide web: information, engagement, and community. In: Selber, S. (ed.) Computers and Technical Communication: Pedagogical and Programmatic Perspectives, pp. 149–169. Erlbaum, Hillsdale, NJ (1997)
14. Dupret, G., Lalmas, M.: Absence time and user engagement: evaluating ranking functions. In: Proceedings of the 6th ACM International Conference on Web Search and Data Mining, pp. 173–182. ACM Press, New York (2013)
15. Feild, H., White, R.W., Fu, X.: Supporting orientation during search result examination. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2999–3008. ACM Press, New York (2013)
16. Grafsgaard, J.F., Wiggins, J.B., Boyer, K.E., Wiebe, E.N., Lester, J.C.: Automatically recognizing facial expression: predicting engagement and frustration. In: Proceedings of the 6th International Conference on Educational Data Mining, pp. 43–50 (2013)
17. Hearst, M.: Search User Interfaces. Cambridge University Press, Cambridge (2009)
18. Heinström, J.: Broad exploration or precise specificity: two basic information seeking patterns among students. J. Am. Soc. Inf. Sci. Technol. **57**, 1440–1450 (2006)
19. Hwang, M.I., Thorn, R.G.: The effect of user engagement on system success: a meta-analytical integration of research findings. Inf. Manag. **35**(4), 229–336 (1999)
20. Jacques, R., Precce, J, Carey, J.T.: Engagement as a design concept for hypermedia. Can. J. Educ. Commun. **24**, 49–59 (1995)
21. Jansen, B.J.: Search log analysis: what it is, what's been done, how to do it. Libr. Inf. Sci. Res. **28**, 407–432 (2006)
22. Jiang, J., He, D., Allan, J.: Searching, browsing, and clicking in a search session: changes in user behavior by task and over time. In: Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval, pp. 607–616. ACM Press, New York (2014)
23. Kelly, D., Sugimoto, C.R.: A systematic review of interactive information retrieval evaluation studies, 1967–2006. J. Am. Soc. Inf. Sci. Technol. **64**(4), 745–770 (2013)
24. Kelly, D., Arguello, J., Edwards, A., Wu, W.C.: Development and evaluation of search tasks for IIR experiments using a cognitive complexity framework. In: Proceedings of SIGIR International Conference on the Theory of Information Retrieval, pp. 7–10 (2015)
25. Lehmann, J., Lalmas, M., Yom-Tov, E., Dupret, G.: Models of user engagement. In: User Modeling, Adaptation, and Personalization, pp. 164–175 (2012)
26. Lehmann, J., Lalmas, M., Baeza-Yates, R., Yom-Tov, E.: Networked user engagement. In: Proceedings of the 1st Workshop on User Engagement Optimization, pp. 7–10 (2013)
27. Lehmann, J., Lalmas, M., Dupret, G., Baeza-Yates, R.: Online multitasking and user engagement. In: Proceedings of the 22nd ACM International Conference on Information and Knowledge Management, pp. 519–528 (2013)
28. Li, Y., Belkin, N.J.: A faceted approach to conceptualizing tasks in information seeking. Inf. Process. Manag. **44**, 1822–1837 (2008)
29. Linnenbrink, E.A., Pintrich, P.R.: The role of self-efficacy beliefs in student engagement and learning in the classroom. Read. Writ. Q. **19**(2), 119–137 (2003)
30. Moshfeghi, Y., Matthews, M., Blanco, R., Jose, J.M.: Influence of timeline and named-entity components on user engagement. In: Advances in Information Retrieval, pp. 305–317 (2013)
31. Nes, L.S., Segerstrom, S.C., Sephton, S.E.: Engagement and arousal: optimism's effects during a brief stressor. Personal. Soc. Psychol. Bull. **31**(1), 111–120 (2005)
32. O'Brien, H., Lebow, M.: Mixed-methods approach to measuring user experience in online news interactions. J. Am. Soc. Inf. Sci. Technol. **64**, 1543–1556 (2013)
33. O'Brien, H., Toms, E.G.: What is user engagement? A conceptual framework for defining user engagement with technology. J. Am. Soc. Inf. Sci. Technol. **59**, 938–955 (2008)

34. O'Brien, H., Toms, E.G.: The development and evaluation of a survey to measure user engagement. J. Am. Soc. Inf. Sci. Technol. **61**, 50–69 (2010)
35. O'Brien, H., Toms, E.G.: Examining the generalizability of the user engagement scale (UES) in exploratory search. Inf. Process. Manag. **49**, 1092–1107 (2013)
36. Ortiz-Cordova, A., Jansen, B.J.: Classifying web search queries to identify high revenue generating customers. J. Am. Soc. Inf. Sci. Technol. **63**, 1426–1441 (2012)
37. Poddar, A., Ruthven, I.: The emotional impact of search tasks. In: Proceedings of the 3rd Symposium on Information Interaction in Context, pp. 35–44 (2010)
38. Reeves, B., Nass, C.: The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. Cambridge University Press, New York, NY (1996)
39. Rokhlenko, O., Golbandi, N., Lempel, R., Leibovich, L.: Engagement-based user attention distribution on web article pages. In: Proceedings of the 24th ACM Conference on Hypertext and Social Media, pp. 196–201 (2013)
40. Schaufeli, W.B., Salanova, M.: Enhancing work engagement through the management of human resources. In: Naswall, K., Hellgren, J., Sverke, M. (eds.) The Individual in the Changing Working Life, pp. 380–402. Cambridge University Press, Cambridge (2008)
41. Song, Y., Shi, X., Fu, X.: Evaluating and predicting user engagement change with degraded search relevance. In: Proceedings of the 22nd International Conference on World Wide Web, pp. 1213–1224 (2013)
42. Sundar, S.S., Xu, Q., Bellur, S., Oh, J., Jia, H.: Beyond pointing and clicking: how do newer interaction modalities affect user engagement? In: CHI'11 Extended Abstracts on Human Factors in Computing Systems. ACM Press, New York (2011)
43. Teevan, J., Collins-Thompson, K., White, R.W., Dumais, S.T., Kim, Y.: Slow search: information retrieval without time constraints. In: Proceedings of the Symposium on Human-Computer Interaction and Information Retrieval, pp. 1–10 (2013)
44. Toms, E.: Task-based information searching and retrieval. In: Ruthven, I., Kelly, D. (eds.) Task-Based Information Searching and Retrieval, pp. 43–59 (2011)
45. Vail, A.K., Grafsgaard, J.F., Wiggins, J.B., Lester, J.C., Boyer, K.E.: Predicting learning and engagement in tutorial dialogue: a personality-based model. In: Proceedings of the 16th International Conference on Multimodal Interaction, pp. 255–262 (2014)
46. Webster, J., Ahuja, J.S.: Enhancing the design of web navigation systems: the influence of user disorientation on engagement and performance. MIS Q. **30**, 661–678 (2006)
47. Wildemuth, B.W., Freund, L., Toms, E.G.: Untangling search task complexity and difficulty in the context of interactive information retrieval studies. J. Doc. **70**, 1118–1140 (2014)